Performance of Top Quark and W Boson Tagging in Run 2 with ATLAS

The ATLAS Collaboration

The performance of hadronically-decaying top-quark and W-boson taggers in pp collisions at $\sqrt{s} = 13$ TeV recorded by the ATLAS experiment at the Large Hadron Collider is presented. A set of techniques, including some new to the data recorded in 2015 and 2016, are studied to determine a set of optimal cut-based taggers for use in physics analyses. A further extension is made to study the utility of combinations of substructure observables as a multivariate tagger using boosted decision trees and deep neural networks in comparison with taggers based on two-variable combinations. The performance of these taggers is studied with the data collected during 2015 and 2016 in $t\bar{t}$, dijet and $\gamma +$ jet event topologies.
1 Introduction

With the increase to 13 TeV center-of-mass energy in Run 2 of the Large Hadron Collider (LHC) [1], it is increasingly important for searches for physics beyond the Standard Model to probe processes involving highly boosted massive particles, such as $W$ bosons and top quarks, with recent examples from ATLAS in Refs. [2–5].

To fully exploit these final states, reconstructing the hadronic decay modes of these massive particles is of great importance. The identification of the origin of a hadronic jet can be used as an effective tool to reject events produced by background processes and improve the sensitivity in searches for physics beyond the Standard Model.

For $W$-boson and top-quark jet identification, physically motivated observables exploiting the radiation pattern within the jet have been used to effectively tag large-radius jets [6–10]. This work is intended to expand upon those studies and provide a more comprehensive investigation of tagging techniques applicable to the identification of both $W$ bosons and top quarks.

The contents of this note are organized as follows: Section 2 briefly describes the ATLAS detector, followed by Section 3 with a description of the Monte Carlo and data samples used in the analysis. The set of substructure techniques investigated here is described in Section 4 where a signal definition based on the parton decay products is also introduced. The optimization procedure used to determine the optimal tagger for use in searches following this signal definition and new inputs is described in Section 5. Additionally, comparisons are made of $W$-boson and top-quark identification techniques using simulated data at $\sqrt{s} = 13$ TeV. In Section 6, the data recorded in 2015 and 2016 is used to investigate the performance of these tagging techniques by measuring signal and background efficiencies using boosted lepton+jet $t\bar{t}$, dijet and $\gamma +$ jet topologies. Concluding remarks are given in Section 7.
2 ATLAS detector

The ATLAS detector [11] at the LHC covers nearly the entire solid angle around the collision point. It consists of an inner tracking detector (ID) surrounded by a thin superconducting solenoid, electromagnetic and hadronic calorimetry, and a muon spectrometer composed of three large superconducting toroid magnets. For this study, important subsystems are the calorimeters, which cover the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel and endcap high-granularity lead/liquid-argon (LAr) electromagnetic calorimeters, with an additional thin LAr presampler covering $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters. Hadronic calorimetry is provided by a steel/scintillating-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadronic endcap calorimeters. The forward region $3.2 < |\eta| < 4.9$ is instrumented with copper/LAr and tungsten/LAr calorimeter modules. Inside the calorimeters, there is a 2 T solenoid magnet that surrounds the inner tracking detector which measures charged-particle trajectories covering a pseudorapidity range $|\eta| < 2.5$ with pixel and silicon microstrip detectors (SCT), and additionally covering the region $|\eta| < 2.0$ with a straw-tube transition radiation tracker (TRT).

The muon spectrometer (MS) comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by superconducting air-core toroids. The precision chamber system covers the region $|\eta| < 2.7$ with three layers of monitored drift tubes, complemented by cathode strip chambers in the forward region where the background is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive plate chambers in the barrel and thin gap chambers in the endcap regions. A two-level trigger system is used to select events for offline analysis [12]. The first pass, named the level-1 trigger, is implemented in hardware and uses a subset of detector information to reduce the event rate to 100 kHz. This is followed by a software-based high-level trigger which reduces the final event rate to less than 1 kHz.

---

1 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, $\phi$ being the azimuth's angle around the z-axis. The pseudo-rapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta \eta = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$. 

---

3
3 Samples

The taggers described in this note are studied using three Monte Carlo (MC) samples for signal and background processes. The dijet process is used to simulate jets from gluons and other (non-top) quarks and this process is modelled using the Pythia8 (v8.186) [13] generator with the NNPDF2.3LO [14] parton distribution function (PDF) set and the A14 tune [15]. Events are generated in slices of jet $p_T$ to sufficiently populate the kinematic region of interest. Event-by-event weights are applied to correct for this generation and to produce the expected smoothly-falling $p_T$ distribution expected for multijet background.

The signal samples containing either high-$p_T$ top-quark or $W$-boson jets are obtained from two physics processes modelling phenomena beyond the Standard Model. For the $W$-boson sample, the high-mass $W' \rightarrow WZ \rightarrow qqqq$ is used. For the top-quark sample, high-mass $Z' \rightarrow t\bar{t}$ events are used as a source of signal jets. Both the $W$ bosons and top quarks are required to decay hadronically. The two signal processes are simulated using the Pythia8 generator with the NNPDF2.3LO PDF set and A14 tune for multiple values of the resonance ($W'$ or $Z'$ boson) mass between 400 and 5000 GeV in order to populate the entire $p_T$ range$^2$ from 200 to 2500 GeV and to ensure that the calculated signal efficiencies have small statistical uncertainties.

For the study of $W$-boson and top-quark jets in data, described in Section 6, a number of Monte Carlo samples are needed to model both the $t\bar{t}$ signal and backgrounds. The Powheg-Box v2 generator [16] is used to simulate $t\bar{t}$ and single-top-quark production in the $Wt$- and $s$-channels, while for the single-top-quark $t$-channel process, Powheg-Box v1 generator is used. The matrix-element generation in Powheg-Box is interfaced to the CT10 [17] NLO PDF set. For all processes involving top quarks, the parton shower, fragmentation, and the underlying event are simulated using Pythia6 (v6.428) [18] with the CTEQ6L1 [19] PDF set and the corresponding Perugia 2012 tune (P2012) [20]. The top-quark mass is set to 172.5 GeV. The $h_{\text{damp}}$ parameter, which controls the $p_T$ of the first additional emission beyond the Born configuration, is set to the mass of the top quark. The $t\bar{t}$ process is normalized to the cross-sections predicted to next-to-next-to-leading-order (NNLO) in $\alpha_s$ and next-to-next-to-leading logarithm in soft-gluon terms while the single-top-quark processes are normalized to the NNLO cross-section predictions.

Several additional variations of the $t\bar{t}$ generator are used for the estimation of modeling uncertainties. Estimates for the parton-showering and hadronization modelling uncertainty are derived by comparing results with the Powheg-Box v2 generator in tandem with Herwig++ (v2.7.1) [21] instead of Pythia6. To estimate the hard-scattering-modelling uncertainty, the MadGraph5_aMC@NLO (v2.2.1) generator [22] (hereafter referred to as MC@NLO) with Pythia6 is used. To estimate the uncertainty on modelling of additional radiation, the Powheg-Box v2 generator with Pythia6 is used with modified renormalization and factorization scales (both $\times 2$ or $\times 0.5$) and simultaneously modified $h_{\text{damp}}$ parameter ($h_{\text{damp}} = m_{\text{top}}$ or $h_{\text{damp}} = 2 \times m_{\text{top}}$).

Samples of $W/Z$+jets and Standard Model diboson ($WW/WZ/ZZ$) production are generated with final states that include either one or two charged leptons. The Sherpa [23] generator version 2.1.1 and version 2.2.1 are used with the CT10 PDF set to simulate the diboson production and $W/Z$+jets processes, respectively. The $W/Z$+jets events are normalised to the NNLO cross sections.

$^2$ As the combination of these signal samples with different generated heavy resonance masses results in irregular top-quark and $W$-boson $p_T$ distributions, the events are reweighted on the generator level to either a flat or a QCD background-like falling $p_T$ distribution in the following.
For the study of $\gamma +$ jets in data, events containing a photon with associated jets are simulated using the Sherpa 2.1.1 [23] generator, requiring a photon transverse momentum above 140 GeV. Matrix elements are calculated with up to 4 partons at LO and merged with the Sherpa parton shower [24] using the ME+PS@LO prescription [25]. The CT10 PDF set is used in conjunction with dedicated parton shower tuning developed by the Sherpa authors.

The Monte Carlo samples are processed through the full ATLAS detector simulation [26] based on Geant4 [27]. Additional simulated proton–proton collisions generated using Pythia8 (v8.186) with the A2 [28] tune and MSTW2008LO PDF set [29] are overlaid to simulate the effects of additional collisions from the same and nearby bunch crossings (pile-up), with a mean number of 24 collisions per bunch crossing. All simulated events are then processed using the same reconstruction algorithms and analysis chain as is used for real data.

Data are collected in three broad categories to study the signal and the background. For the signal, a set of observed top-quark and $W$-boson candidates is obtained from a sample of $t\bar{t}$ candidates in which one top quark decays semi-leptonically and the other decays hadronically, the so-called lepton plus jets decay signature. The background is studied using data samples of dijet events and $\gamma +$ jet events. In addition to covering different $p_T$ regions, the dijet and $\gamma +$ jet samples differ in what partons initiated the jets under study. In the $\gamma +$ jet topology the jets are dominantly initiated by quarks over the full $p_T$ range studied, while for the dijet topology the fraction of quarks initiating the jets is slightly smaller than the gluon fraction at low $p_T$ and becomes large at high $p_T$. The data for the $t\bar{t}$ and $\gamma+$ jets studies were collected during normal operations of the detector and correspond to an integrated luminosity of $36.1 \text{ fb}^{-1}$. For the dijet analysis additional data where the toroid magnet was turned off can be used, resulting in $36.7 \text{ fb}^{-1}$ in total.

The signal lepton plus jets events are collected with a set of single-electron and single-muon triggers that become fully efficient for $p_T$ of the reconstructed lepton greater than 28 GeV. The dijet events are triggered by a single large-$R$ jet trigger that becomes fully efficient for an offline-jet $p_T$ of approximately 450 GeV. Additionally for the dijet studies, small contributions from signal processes are accounted for with the use of the simulated Standard Model $W/Z$ plus jets and all-hadronic $t\bar{t}$ samples. The $\gamma +$ jet events are triggered by a single-photon trigger that becomes fully efficient for an offline-photon $p_T$ of approximately 155 GeV. Additionally, for the $\gamma +$ jet studies the small contributions from signal processes are accounted for with the use of the simulated $W/Z + \gamma$ and $t\bar{t} + \gamma$ samples.
4 Jet substructure techniques

The reconstruction and identification of hadronic decays of boosted $W$ bosons and top quarks can broadly be divided into two stages, grooming and tagging. A number of techniques and observables pertaining to these two categories have been described and investigated extensively in previous work [8, 9] with only a short summary of the relevant techniques presented here.

4.1 Jet reconstruction and grooming

Jets within ATLAS are reconstructed from noise-suppressed topological clusters [30] that are individually calibrated to correct for effects such as the non-compensation of the calorimeter response and inactive material [31]. The clusters are set to be massless. They form the basis for the set of constituents from which large-$R$ jets are reconstructed using the anti-$k_t$ algorithm [32] with a radius of $R = 1.0$ and further trimmed to remove the effects of pile-up and underlying event. Trimming [33] is a grooming technique in which the original constituents of the jets are reclustered using the $k_t$ algorithm [34] with a distance parameter $R_{\text{sub}}$ in order to produce a collection of subjets. These subjets are then discarded if they carry less than a specific fraction ($f_{\text{cut}}$) of the $p_T$ of the original jet. The trimming parameters used here are $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 5\%$. These jets are then calibrated in a two-step procedure that first corrects the jet energy scale and then the jet mass scale [31, 35].

The resultant set of constituents forms the basis from which further observables are calculated. The notable exceptions are the inputs for the shower deconstruction [36] and the HEPTopTagger [37, 38] algorithms, described later in more detail, that make use of the Cambridge/Aachen (C/A) jet algorithm [39, 40].

For the purpose of identifying the flavor of the jet at truth level in Monte Carlo events, a second set of jets is formed from truth particles with lifetimes greater than 10 picoseconds, except for muons and neutrinos which are not included. These jets are reconstructed with the anti-$k_t$ algorithm with a distance parameter of $R = 1.0$ but are not modified with the trimming algorithm. These jets are referred to as truth jets and the related quantities such as the $p_T$ of the jets are referred to as $p_T^{\text{truth}}$.

In the MC-simulation-based study presented in Section 5, the event selection isolates ensembles of jets which are representative of those originating from either $W$ bosons or top quarks (signal) and gluon or other quarks\(^3\) (background). Initially, events which contain a reconstructed primary vertex with at least two tracks are selected. In each event the two highest-$p_T$ truth jets are retained if they satisfy $|\eta| < 2$ and have a $p_T$ greater than 200 GeV in the case of the $W$-boson or background quark and gluon jets, and greater than 350 GeV for top-quark jets. The retained truth jets and the reconstructed jets that are truth-matched to those as described below are used.

It is important to note that the tagging techniques used to identify signal $W$-boson or top-quark jets are typically designed under the hypothesis of a signal model for the jets. In the study presented in Section 5, signal jets are defined as hadronically-decaying $W$ bosons or top quarks when all partonic decay products are fully contained within the region of interest of the reconstructed jet in a three-step process. First, reconstructed jets are matched to truth jets. Next, those truth jets are matched to truth $W$ bosons and top quarks ($W, t$). Lastly their partonic decay products (two light quarks for hadronically-decaying $W$ bosons and an additional $b$ quark for top quarks decaying into a $W$ boson and a $b$ quark) are matched to the initial reconstructed jet. All stages of this matching procedure are performed with simple matching in ($\eta, \phi$)

\(^3\) This includes quark flavours other than top quarks.
with $\Delta R < 0.75^4$. In particular, the containment depends strongly on the $p_T$ of the particle, as shown in Figure 1.

![Figure 1: Containment of the W boson and top quark decay products in a single truth level anti-$k_t$, $R = 1.0$ jet as a function of the particle’s transverse momentum.](image)

### 4.2 Jet mass

The jet mass provides the most powerful single discriminant and is typically constructed purely from the topocluster constituents of the jet. However, at extremely high $p_T$, it becomes advantageous to use the spatial granularity of tracks reconstructed in the inner detector to construct an observable called the track-assisted jet mass ($m_{TA}$), defined as

$$m_{TA} = m_{\text{track}} \times \frac{p_{T \text{calo}}}{p_{T \text{track}}}$$

in which $m_{\text{track}}$ and $p_{T \text{track}}$ are the invariant mass and $p_T$ calculated from tracks associated with the large-$R$ trimmed calorimeter jet and $p_{T \text{calo}}$ is the $p_T$ of the original trimmed large-$R$ jet.

In order to take full advantage of this new observable and the calorimeter-based jet mass, a new combined mass ($m_{\text{comb}}$) definition uses a linear combination of the two. The variables combined in a weighted average. The assigned weight $w_{TA}$ for the track-assisted mass is defined as

$$w_{TA} = \frac{\sigma_{\text{calo}}^{-2}}{\sigma_{\text{calo}}^{-2} + \sigma_{TA}^{-2}}$$

where $\sigma_{TA}$ and $\sigma_{\text{calo}}$ are the $m_{TA}$ and calorimeter mass resolutions, respectively. The calorimeter weight, $w_{\text{calo}} = 1 - w_{TA}$, is defined so that the sum of both weights equals unity. An illustration of the $m_{\text{comb}}$ distribution for signal and background events is shown in Figure 2. Further details can be found in Ref. [35].

---

4 For $R = 1.5$ C/A jets the containment cuts are increased to $0.75 \cdot 1.5 = 1.125$. 
4.3 Tagging techniques

In addition to the jet mass, a number of observables and techniques can be used to further identify a jet as originating from a $W$ boson or a top quark.

4.3.1 Jet moments

The first broad class of observables are those referred to as jet moments or simply jet substructure variables that are analytically defined and that quantify the nature of the radiation pattern via a single quantity calculated from the set of constituents of the trimmed jet. A brief summary of all observables tested here is provided in Table 1 and a more complete description of the observables under study can be found in Refs. [8, 9].

In general, these observables are constructed with the intent to quantify how clustered or uniform the constituents are. This can be done by explicitly using a set of axes (e.g. N-subjettiness, $\tau_{21}$ and $\tau_{32}$), declustering the jet (e.g. splitting measures, $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$), or using all jet constituents to quantify the dispersion of the jet constituents in an axis independent way (e.g. energy correlation ratios). In previous ATLAS studies [6–9], it was found that for $W$-boson tagging, that energy correlation variables, in particular $D_2$, were the best performing tagging observable while for top-quark tagging the N-subjettiness ratio, $\tau_{32}$, was found to be optimal. This has recently been understood from an analytical point of view and attributed to additional wide-angle radiation present in parton jets originating from $W$-boson decays, which is more fully exploited in the energy correlation functions than in the N-subjettiness moments [41].
### Observables

<table>
<thead>
<tr>
<th>Observable</th>
<th>Variable</th>
<th>Used For</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jet mass</td>
<td>$m_{\text{comb}}$</td>
<td>top, W</td>
<td>[35]</td>
</tr>
<tr>
<td>Energy Correlation Ratios</td>
<td>$ECF_1$, $ECF_2$, $ECF_3$</td>
<td>top, W</td>
<td>[41, 42]</td>
</tr>
<tr>
<td></td>
<td>$C_2, D_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-subjettiness</td>
<td>$\tau_1, \tau_2, \tau_3$</td>
<td>top, W</td>
<td>[43, 44]</td>
</tr>
<tr>
<td></td>
<td>$\tau_{21}, \tau_{32}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Center of Mass Observables</td>
<td>Fox Wolfram ($R^{FW}_2$)</td>
<td>$W$</td>
<td>[45]</td>
</tr>
<tr>
<td>Splitting Measures</td>
<td>$Z_{\text{cut}}$, $\sqrt{d_{12}}, \sqrt{d_{23}}$</td>
<td>top, W</td>
<td>[46, 47]</td>
</tr>
<tr>
<td>Planar Flow</td>
<td>$P$</td>
<td>$W$</td>
<td>[48]</td>
</tr>
<tr>
<td>Angularity</td>
<td>$a_3$</td>
<td>$W$</td>
<td>[49]</td>
</tr>
<tr>
<td>Aplanarity</td>
<td>$A$</td>
<td>$W$</td>
<td>[50]</td>
</tr>
<tr>
<td>KtDR</td>
<td>$KtDR$</td>
<td>$W$</td>
<td>[51]</td>
</tr>
<tr>
<td>Qw</td>
<td>$Q_w$</td>
<td>top</td>
<td>[46]</td>
</tr>
</tbody>
</table>

Table 1: Summary of tagging techniques and resultant variables that have been studied. In the case of the energy correlation observables, the angular exponent $\beta$ is set to 1.0 and for the N-subjettiness observables, the winner-take-all [52] configuration is used.

#### 4.3.2 Shower deconstruction

The shower deconstruction (SD) algorithm [36] is a different approach to tagging that involves deconstructing the jet into its constituents and that uses those constituents as inputs into an algorithm that tests different parton shower hypotheses. During Run 1, extensive work was done to test this technique in the context of top-quark tagging [53], and in this work the technique is further studied as a top tagger. The goal of the shower deconstruction algorithm is to assign a variable to a jet which can be used to discriminate between predefined signal and background models. In order to calculate this variable, defined as $\chi$, a simplified parton shower model is used.

The constituents of an input jet are reclustered into subjets and these are used as the inputs to the algorithm. For a set of input subjets ($\{p_N\}$), a set of potential shower histories is constructed for the signal and background models. Each shower history represents a possible way that the chosen model could have resulted in the given input subjet configuration. A probability is assigned to each shower history based on the parton shower model and the $\chi$ variable is defined as the likelihood ratio

$$\chi(\{p\}_N) = \frac{\sum_{\text{histories}} P(\{p\}_N | S)}{\sum_{\text{histories}} P(\{p\}_N | B)}.$$  \hspace{1cm} (3)

Then log $\chi$ is used as a tagging discriminant. In addition, the SD algorithm can only define log $\chi$ when the subjets are kinematically compatible with a hadronic top-quark decay.

This leads to the following requirements: the large-$R$ jet has at least three subjets; two or more subjets must have a mass in a window centered around the $W$ boson mass ($\Delta M_W$); and at least one more subjet can be added to obtain a total mass in a window centered around top mass ($\Delta M_{\text{top}}$). Values of $\Delta M_W$ and $\Delta M_{\text{top}}$ equal to 20 GeV and 40 GeV were found to give the best results in terms of signal efficiency and background rejection. Moreover, the computation time needed for the calculation of log $\chi$ grows exponentially with the subjet multiplicity. This effect was also studied by restricting the number of subjets. The optimal
Table 2: The HEPTopTagger parameter settings used in this study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{\text{cut}}$</td>
<td>50 GeV</td>
</tr>
<tr>
<td>$R_{\text{hit}}^{\text{max}}$</td>
<td>0.25</td>
</tr>
<tr>
<td>$N_{\text{filt}}$</td>
<td>5</td>
</tr>
<tr>
<td>$f_W$</td>
<td>15%</td>
</tr>
</tbody>
</table>

maximum number of subjets was found to be 6. This limits computation time without loss of rejection power.

In previous ATLAS studies, the subjets were defined by running the C/A jet algorithm [39, 40] with $R = 0.2$ over the large-$R$ jet constituents [53]. This definition of the subjets was found to have a low signal efficiency for very high transverse momentum. This is due to the fact that the distance between the decay products of the top ($\Delta R_{t,\text{const}}$) is expected to be lower than 0.2 due to the kinematic boost. This leads to a number of reconstructed subjets lower than three. This is improved in this study by defining subjets using exclusive jets [54]. First, the $k_t$ algorithm with $R = 1.0$ is run over the large-$R$ constituents and then the $k_t$ reclustering is stopped if splitting scales bigger than 15 GeV are found. At that stage, the reclustered protojets are used as subjets. Since splitting scales are less dependent on the large-$R$ jet $p_T$ than the $\Delta R_{t,\text{const}}$, an improvement in the signal efficiency at high transverse momentum is expected.

4.3.3 HEPTopTagger

An alternative approach to top-quark tagging is the HEPTopTagger (HTT) algorithm [37, 38]. Unlike the previous observables which are calculated from the constituents of the large-$R$ trimmed jets, this technique relies on reconstructing jets using the C/A algorithm with $R = 1.5$ to allow the tagging of fully contained boosted tops to reach lower values of $p_T$ and to take advantage of the C/A clustering sequence. To define jet kinematics at reconstruction level, the C/A $R = 1.5$ jets are groomed to mitigate pile-up effects. Various configuration of trimming and filtering techniques are tested. Trimming with subject radius of $R_{\text{sub}} = 0.2$ and momentum fraction $f_{\text{cut}} = 0.05$ is found to produce jets that are independent of the average number of interactions per bunch crossing.

The C/A $R = 1.5$ jets are analysed with the HEPTopTagger algorithm, which identifies the hard jet substructure and tests it for compatibility with the 3-prong pattern of hadronic top-quark decays. This tagger was developed to find top quarks with $p_T > 200$ GeV and to achieve a high rejection of background, with the latter being largest for low-$p_T$ large-$R$ jets. The HEPTopTagger studied in this paper is the original algorithm, not the extended HEPTopTagger2 algorithm [55]. The settings used here are given in Table 2. The large-$R$ jet is considered to be tagged if the top-quark-candidate mass is between 140 and 210 GeV and the top-quark-candidate $p_T$ is larger than 200 GeV.
5 Tagger optimization

As described in Section 4, a wide variety of techniques for identifying W-boson and top-quark jets exist. To assist in guiding the efficient use of these taggers and exploit their full potential in searches and measurements, a comparison is performed between the different techniques in two broad ways. First, a simple approach to W-boson and top-quark tagging is pursued, where one-dimensional selections on two jet substructure tagging observable are combined, as described in Section 5.1. Second, the usage of deep neural networks and boosted decision trees that use substructure observables as inputs is studied to attempt to fully exploit the information content of multiple observables at once, as described in Section 5.2.

The simulated signal samples described in Sections 3 and 4 are combined and weighted (separately for W bosons and top quarks) such that the truth $p_T$ distribution of the ensemble of signal jets matches that of the dijet background to remove any bias on the tagging performance due to the difference in the $p_T$ spectrum of the signal and background jet samples. To evaluate the performance of each tagger, the two primary quantitative figures of merit, the signal efficiency and background rejection, are used and are defined as

\[
\text{Signal efficiency} = \epsilon_{\text{sig}} = \frac{N_{\text{tagged}}^{\text{sig}}}{N_{\text{total}}^{\text{sig}}} \quad (4)
\]

\[
\text{Background rejection} = \frac{1}{\epsilon_{\text{bkg}}} = \frac{N_{\text{total}}^{\text{bkg}}}{N_{\text{tagged}}^{\text{bkg}}} \quad (5)
\]

For each tagger, the performance is quantified in terms of the background rejection as a function of jet $p_T$. Working points are established at 50% and 80% efficiency as a function of $p_T$ for W-boson and top-quark tagging. For these studies, the jet $p_T$ that is used to parameterize the performance is that of the associated anti-$k_T R = 1.0$ truth jet ($p_T^{\text{truth}}$), thereby allowing comparisons of taggers employing different jet clustering algorithms.

5.1 Cut based optimization

A straightforward cut-based tagger combines a selection of jet mass and a second observable, and these have been studied by ATLAS in 2015 [6, 7]. The primary goal of these taggers is to provide a set of selections that vary based on the $p_T$ of the jet and provide an approximately constant signal efficiency. The selections are parameterized as a function of the $p_T$ of the associated anti-$k_T R = 1.0$ truth jet to make the choice of the optimal set of variables only.

For this study the optimization is unified for both W-boson and top-quark tagging. For fixed signal efficiencies the background rejection of two-variable combinations is scanned and the combination of cuts leading to the largest background rejection is considered optimal.

In case of the substructure pattern of hadronic top-quark decays, cuts which can be either a single-sided upper or lower cut on a substructure variable are found to discriminate best against multi-jet patterns. In the case of W-boson tagging a two-sided cut on the jet mass variable is used, alongside a single-sided cut on another substructure variable. The two-sided mass cut was found to improve the background rejection due to the expected Gaussian-like distributed shape of the mass. To provide a smooth cut function, the optimised cut values as a function of jet $p_T$ are fitted using two different kinds of functions. All single-sided
cuts are fitted with a polynomial function to flexibly describe features which occur due to correlation of the combined-tagger variables. For the $W$-boson tagging jet mass optimization, an empirical cut function is used to fit the two-sided mass cuts, $\sqrt{(A/p_T + B)^2 + (C \cdot p_T + D)^2}$, where $A$, $B$, $C$ and $D$ are fit to the optimised cut values in each $p_T$ bin.

Different combinations of 2 variables are tested and their performance is compared at signal efficiencies of 50% and 80% with a parametrization yielding flat signal efficiencies as a function the jet $p_T$. The resulting background rejection versus $p_T$ distributions are shown in Figure 3 for the best performing combinations found.

The two-variable combinations that provided the best overall performance for $W$-boson and top-quark tagging were chosen:

- combined jet mass and $D_2$ for $W$-boson tagging and
- $\sqrt{D_{23}}$ and $\tau_{32}$ for top-quark tagging.

For the chosen taggers the working points are re-defined as a function of the reconstructed jet $p_T$ and they are parametrized so that they yield flat signal efficiencies versus $p_T$. In addition, only one efficiency working point parametrized in $p_T$ will be presented in the following: 50% for $W$-boson tagging and 80% for top-quark tagging, as those are commonly used in ATLAS analyses.

![Figure 3: $W$-boson (left) and top-quark tagging (right) background rejection distributions as a function of $p_T$ for the best performing two-variable combinations at 50% and 80% signal efficiency.](image)

### 5.2 Deep neural network and boosted decision tree based taggers

Some of the observables described in Section 4 contain complementary information. It has been shown that the combination of these observables to create a multivariate $W$-boson or top-quark classifier provides higher discrimination, albeit to differing degrees [56–58]. In this work boosted decision tree (BDT) and deep neural network (DNN) algorithms are applied to explore such techniques following a similar procedure as Ref. [58] and the ability of these algorithms to discriminate $W$-boson and top-quark jets
from the gluon and light-quark jet background are studied in parallel. Study of these two algorithms in parallel is motivated to explore if one of the architectures is better suited to exploit input correlations among high-level variables, and to study the performance of both algorithms when applied to data.

The BDT and DNN used here are similar to the ones described in detail in Ref. [58]. The description of the training criteria, of the training and testing sets, and of the relevant event weights is not repeated here. A similar hyperparameter scan was performed yielding similar results. While the procedure has only changed minimally in comparison to the one in Ref. [58], the choice of input observables are slightly adapted and they are shown in Table 3. As the jet mass is now used as an input, the BDT and DNN taggers in this work do not use an additional mass cut except the very low mass threshold that is part of the training criteria. Although the BDT and DNN optimization studies are carried out using only the jets that satisfy the training criteria, the discrimination power obtained by $m_{\text{comb}} > 40$ GeV is included as part of the BDT and DNN taggers in Section 5.3 in order to compare taggers using jets with similar kinematics as inputs.

Additionally, the optimization studies are carried out in a wide $p_T$ truth bin and the relative performance gain is evaluated with flat $p_T$ truth spectra. However, the comparison of taggers in Section 5.3 is made with $p_T$ truth distributions for signal jets weighted to match that of the dijet background samples.

To find the optimal set of BDT input variables, single input variables that give the largest increase in relative performance are sequentially added to the network. This procedure is initiated with the input variables listed in Table 3. The relative performance is evaluated using jets from the testing sample which pass the training criteria and with the training weights described in Ref. [58] in the kinematic range $200 < p_T^{\text{truth}} < 2000$ GeV. During the input and hyperparameter optimization of the multivariate techniques, the jets which do not pass the jet mass ($m_{\text{comb}} > 40$ GeV) and number of constituents ($N_{\text{const}} > 2$) training criteria are not included in the relative signal efficiency or background rejection evaluation. At each step, the variable which gives the greatest increase in relative background rejection at a fixed relative signal efficiency of 50% ($W$-boson tagging) and 80% (top-quark tagging), when added to the existing set of variables, is retained. The minimum set of variables which reaches the highest relative background rejection within statistical uncertainties are selected. The minimal number of variables is 12

<table>
<thead>
<tr>
<th>Tagging Type</th>
<th>Observable Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$-boson tagging</td>
<td>$m_{\text{comb}}$, $p_T$</td>
</tr>
<tr>
<td></td>
<td>$e_3$, $C_2$, $D_2$</td>
</tr>
<tr>
<td></td>
<td>$\tau_1$, $\tau_2$, $\tau_21$</td>
</tr>
<tr>
<td></td>
<td>$R_{FW}^{\tau}$, $p_T$, $a_3$, $A$</td>
</tr>
<tr>
<td></td>
<td>$Z_{\text{cut}}$, $\sqrt{d_{12}}$, $\text{KtDR}$</td>
</tr>
<tr>
<td>Top-Quark Tagging</td>
<td>$m_{\text{comb}}$, $p_T$</td>
</tr>
<tr>
<td></td>
<td>$e_3$, $C_2$, $D_2$</td>
</tr>
<tr>
<td></td>
<td>$\tau_1$, $\tau_2$, $\tau_3$, $\tau_32$, $\tau_31$</td>
</tr>
<tr>
<td></td>
<td>$\sqrt{d_{12}}$, $\sqrt{d_{23}}$, $Q_W$</td>
</tr>
</tbody>
</table>

Table 3: Summary of variables used in the DNN and BDT taggers studies for $W$-boson and top-quark tagging. Here $e_3$ is defined as $e_3 = ECF_3 / ECF_1^3$.  

5 To ensure that all used jet substructure features are well defined for the training jets, two additional selection criteria are applied on the jet mass ($m_{\text{comb}} > 40$ GeV) and number of constituents ($N_{\text{const}} > 2$). The jets which fail the training criteria are not used in the training.

6 Jets that fail the mass criteria are tagged as background jets. Jets that pass the mass requirement but fail the number of constituents requirement are tagged as signal jets. The fraction of jets found to be categorized in this way is less than 1%.  

13
for $W$-boson tagging and 10 for top-quark tagging. The relative background rejections achieved are shown in Figure 4.

![Figure 4: BDT relative background rejection (blue) for different sets of variables with successively adding more variables at the 50% ($W$-boson tagging) and 80% (top-quark tagging) relative signal efficiency working point for $W$-boson (left) and top-quark tagging (right). Only jets which pass the training criteria are considered while calculating the relative signal efficiency and relative background rejection. The performance is evaluated with flat $p_T^{\text{truth}}$ spectra. Uncertainties are not presented.](image)

Similar to the BDT training, the DNN is trained on different sets of input variables in order to find the optimal set of input variables with the training weights described in Ref. [58]. The relative performance is evaluated using the jets in the testing set that pass the training criteria. Unlike the BDT, sets of input variables are not defined by successively adding variables but are defined by grouping the inputs variables related to the corresponding signal. The grouping is chosen by selecting variables based on their dependence on the momentum scale of the jet’s substructure objects, on what features of the substructure they describe and on their dependence on other substructure variables. A summary of all the variables tested for the DNN is shown in Tables 4 and 5. The relative background rejections achieved inclusively in jet $p_T$ are presented in Figure 5. As observed in Ref. [58], the performance of the DNN tagger depends both on the number of variables and the information content in the group. The chosen groups of inputs for $W$-boson tagging and top-quark tagging are listed in Table 6. Within the statistical uncertainties the number of variables necessary for maximum rejection at the relative fixed signal efficiency (50% for $W$-boson, 80% for top-quark) is found to be 12 variables for $W$-boson tagging (Group 8 in Table 4) and 13 variables for top-quark tagging (Group 9 in Table 5).
For top-quark tagging, six taggers are compared:

- BDT top-quark and DNN top-quark taggers are both composed of a requirement on the relevant discriminant described in Section 5.2, optimized for top-quark tagging, and a fixed $m^{\text{comb}}$ requirement of $m^{\text{comb}} > 40$ GeV. The requirement on the BDT or DNN discriminant is varied to obtain the desired signal efficiency.

- Simple $m^{\text{comb}} + D_2$ is composed of a pre-defined fixed mass requirement of $60 < m^{\text{comb}} < 120$ GeV and a varying $D_2$ requirement chosen to obtain the desired signal efficiency. This simple tagger is included for comparison purposes.

- The two-variable optimized 50% W-boson tagger is composed of a varying $m^{\text{comb}}$ and $D_2$ requirement as described in Section 5.1. As this tagger is designed and optimized for the specific fixed signal efficiency working points, it is represented by a point in Figure 6 that gives the desired signal efficiency chosen for comparison.

For top-quark tagging, six taggers are compared:

- BDT top-quark and DNN top-quark taggers are both composed of a requirement on the relevant discriminant described in Section 5.2, optimized for top-quark tagging, and a fixed $m^{\text{comb}}$ require-

### Table 4: W-boson tagging inputs groups for DNN as in Figure 5.

<table>
<thead>
<tr>
<th>Group</th>
<th>Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\tau_1, \tau_2, e_3, m^{\text{comb}}, p_T$</td>
</tr>
<tr>
<td>2</td>
<td>$\tau_1, \tau_2, e_3, m^{\text{comb}}, p_T, \sqrt{d_{12}}, \text{KtDR}$</td>
</tr>
<tr>
<td>3</td>
<td>$\tau_1, C_2, D_2, R^F_W, P, a_3, A, Z_{\text{cut}}$</td>
</tr>
<tr>
<td>4</td>
<td>$\tau_1, C_2, D_2, R^F_W, P, a_3, A, Z_{\text{cut}}, m^{\text{comb}}$</td>
</tr>
<tr>
<td>5</td>
<td>$\tau_1, C_2, D_2, R^F_W, P, a_3, A, Z_{\text{cut}}, m^{\text{comb}}, p_T$</td>
</tr>
<tr>
<td>6</td>
<td>$\tau_1, \tau_2, e_3, m^{\text{comb}}, p_T, R^F_W, \sqrt{d_{12}}, \text{KtDR}, a_3, A$</td>
</tr>
<tr>
<td>7</td>
<td>$\tau_1, C_2, D_2, R^F_W, P, a_3, A, Z_{\text{cut}}, m^{\text{comb}}, \sqrt{d_{12}}, \text{KtDR}$</td>
</tr>
<tr>
<td>8</td>
<td>$\tau_1, C_2, D_2, R^F_W, P, a_3, A, Z_{\text{cut}}, m^{\text{comb}}, p_T, \sqrt{d_{12}}, \text{KtDR}$</td>
</tr>
<tr>
<td>9</td>
<td>$\tau_1, \tau_2, \tau_{21}, \sqrt{d_{12}}, C_2, D_2, e_3, m^{\text{comb}}, p_T, R^F_W, P, a_3, A, Z_{\text{cut}}, \text{KtDR}$</td>
</tr>
</tbody>
</table>

### Table 5: Top-quark tagging inputs groups for DNN as in Figure 5.

<table>
<thead>
<tr>
<th>Group</th>
<th>Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$C_2, D_2, \tau_{21}, \tau_{32}$</td>
</tr>
<tr>
<td>2</td>
<td>$C_2, D_2, \tau_{21}, \tau_{32}, m^{\text{comb}}$</td>
</tr>
<tr>
<td>3</td>
<td>$C_2, D_2, \tau_{21}, \tau_{32}, m^{\text{comb}}, p_T$</td>
</tr>
<tr>
<td>4</td>
<td>$\tau_1, \tau_2, \tau_3, e_3, m^{\text{comb}}, p_T$</td>
</tr>
<tr>
<td>5</td>
<td>$C_2, D_2, \tau_{21}, \tau_{32}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$</td>
</tr>
<tr>
<td>6</td>
<td>$C_2, D_2, \tau_{21}, \tau_{32}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W, m^{\text{comb}}$</td>
</tr>
<tr>
<td>7</td>
<td>$\tau_1, \tau_2, \tau_3, e_3, m^{\text{comb}}, p_T, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$</td>
</tr>
<tr>
<td>8</td>
<td>$C_2, D_2, \tau_{21}, \tau_{32}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W, m^{\text{comb}}, p_T$</td>
</tr>
<tr>
<td>9</td>
<td>$\tau_1, \tau_2, \tau_3, \tau_{21}, \tau_{32}, \sqrt{d_{13}}, \sqrt{d_{23}}, Q_W, C_2, D_2, e_3, m^{\text{comb}}, p_T$</td>
</tr>
</tbody>
</table>

### 5.3 Summary of tagger performance studies in simulation

The discrimination power of the taggers studied is compared in this section in two ways. First, for various taggers the background rejection versus the signal efficiency is shown in receiver operating characteristic (ROC) curves. In addition, the background rejection versus the $p_T^{\text{truth}}$ is presented for fixed signal efficiencies. For W-boson tagging, four taggers are compared:

- BDT W-boson and DNN W-boson taggers are both composed of a requirement on the relevant discriminant described in Section 5.2, optimized for W-boson tagging, and a fixed $m^{\text{comb}}$ requirement of $m^{\text{comb}} > 40$ GeV. The requirement on the BDT or DNN discriminant is varied to obtain the desired signal efficiency.

- Simple $m^{\text{comb}} + D_2$ is composed of a pre-defined fixed mass requirement of $60 < m^{\text{comb}} < 120$ GeV and a varying $D_2$ requirement chosen to obtain the desired signal efficiency. This simple tagger is included for comparison purposes.

- The two-variable optimized 50% W-boson tagger is composed of a varying $m^{\text{comb}}$ and $D_2$ requirement as described in Section 5.1. As this tagger is designed and optimized for the specific fixed signal efficiency working points, it is represented by a point in Figure 6 that gives the desired signal efficiency chosen for comparison.
Figure 5: Distributions showing the training with different set of variables and relative improvement in performance for the DNN W-boson and top-quark taggers at the 50% and 80% relative signal efficiency working point, respectively. Only jets which pass the training criteria are considered while calculating the relative signal efficiency and relative background rejection. The performance is evaluated with flat $p_T^{\text{truth}}$ spectra. Uncertainties are not presented.

<table>
<thead>
<tr>
<th>Observable</th>
<th>W-Boson Tagging</th>
<th>Top-Quark Tagging</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{\text{comb}}$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$p_T$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$\eta$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$C_2$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$D_2$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$\tau_3$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$\tau_{21}$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$\tau_{32}$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$R_2^W$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$\vec{p}$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$a_3$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$A$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$Z_{\text{cut}}$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$\sqrt{d_{12}}$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$\sqrt{d_{23}}$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$KtDR$</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>$Q_w$</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

Table 6: Summary of the set of observables that are chosen with respect to Figures 4 and 5.
ment of $m^{\text{comb}} > 40$ GeV. The requirement on the BDT or DNN discriminant is varied to obtain the desired signal efficiency.

- The shower deconstruction tagger is composed of a requirement on the log $\chi$ variable described in Section 4.3.2 and a fixed $m^{\text{comb}}$ requirement of $m^{\text{comb}} > 60$ GeV. Similar to the BDT and DNN taggers, the requirement on log $\chi$ is varied to obtain the desired signal efficiency.

- The HEPTopTagger selection consists of identifying a top-quark candidate followed by a mass requirement of $140 < m^{\text{comb}} < 210$ GeV on this candidate. This tagger is represented by a point in Figure 7 as it provides a working point in each $p_T$ bin.

- A tagger using the $m^{\text{comb}}$ and $\tau_{32}$ variables requires a fixed $m^{\text{comb}} > 60$ GeV and a varying maximum $\tau_{32}$ cut to obtain the desired signal efficiency. This simple tagger is included for comparison purposes.

- The two-variable optimized 80% top-quark tagger is composed of a varying $\sqrt{d_{23}}$ and $\tau_{32}$ requirement as described in Section 5.1. As this tagger is optimized for the specific fixed signal efficiency working points, it is represented by a point in Figure 7 that gives the desired signal efficiency chosen for comparison.

The performance of the taggers is evaluated in two wide jet $p_T$ bins in terms of ROC curves, as shown in Figures 6 and 7 for the $W$-boson and top-quark taggers, respectively. Furthermore, the background rejection at the fixed 50% and 80% signal efficiency working point as a function of $p_T$ are presented for $W$-boson and top-quark tagging, respectively, in Figure 8.

The dependence of the taggers on the pile-up conditions is illustrated in Figures 9 and 10 using the two-variable $W$-boson and top-quark taggers optimized for the full dataset (independent of the average number of interactions in the event) as examples and showing the signal efficiency and background rejection obtained for different average number of interactions $\mu$. While the top-quark tagger shows hardly any dependence ($< 1\%$) on the pile-up conditions, there is some dependence for the $W$-boson tagger due to the use of the $D_2$ variable that is more sensitive to pile-up. At the 50% working point, the variation in the $W$-boson tagger efficiency and background rejection in different $\mu$ ranges is observed to be up to 5% and 6%, respectively.

For $W$-boson tagging the best performing taggers are the DNN and BDT taggers over the whole range in $p_T$ studied. They perform very similarly and outperform the two-variable tagger, especially at low $p_T$. The optimization of the two-variable taggers yields a better performance than the cut and scan combination of the same variables.

For top-quark tagging the DNN and BDT taggers also yield the best performance over the full $p_T$ range. They show very similar performance overall, with the BDT being a bit better at low signal efficiencies and $p_T$. Shower deconstruction algorithm yields the best non multivariate result, with the much simpler HEPTopTagger (in its original version) being close in performance. The HEPTopTagger is the only tagger able to tag top quarks down to a $p_T$ of 200 GeV.

In the next sections, the taggers that are compared in this section are applied to data and compared to MC samples. For this purpose working points are defined for the BDT, DNN, shower deconstruction taggers and they are parametrized so that they yield flat signal efficiencies versus reconstructed jet $p_T$ like the two-variable taggers.

---

Footnote: HEPTopTagger is not included in this comparison as it provides a single working point in each $p_T$ bin.
Figure 6: The performance comparison of the $W$-boson taggers in a low-$p_T^{\text{truth}}$ (left) and high-$p_T^{\text{truth}}$ (right) bin. The performance is evaluated with the $p_T^{\text{truth}}$ distribution of the signal jets weighted to match that of the dijet background samples.

Figure 7: The performance comparison of the top-quark taggers in a low-$p_T^{\text{truth}}$ (left) and high-$p_T^{\text{truth}}$ (right) bin. The performance is evaluated with the $p_T^{\text{truth}}$ distribution of the signal jets weighted to match that of the dijet background samples.
Figure 8: The background rejection comparison of \( W \)-boson taggers at fixed 50\% signal efficiency working point (left) and top-quark taggers at fixed 80\% signal efficiency working point (right). The performance is evaluated with the \( p_T^{\text{true}} \) distribution of the signal jets weighted to match that of the dijet background samples. Statistical uncertainties on the background rejection are presented.

Figure 9: The comparison of the signal efficiency (left) and background rejection (right) of the 50\% signal efficiency two-variable \( W \)-boson tagger for different pile-up conditions by using \( \mu \) bins. The performance is evaluated with the \( p_T^{\text{true}} \) distribution of the signal jets weighted to match that of the dijet background samples. Statistical uncertainties are presented.
Figure 10: The comparison of the signal efficiency (left) and background rejection (right) of the 80% signal efficiency two-variable top-quark tagger for different pile-up conditions by using $\mu$ bins. The performance is evaluated with the $p_T^{\text{truth}}$ distribution of the signal jets weighted to match that of the dijet background samples. Statistical uncertainties are presented.
6 Performance in data

The taggers studied here are validated by studies of signal and background-enriched data samples. The data used were taken in 2015 and 2016 at a center of mass energy of $\sqrt{s} = 13$ TeV and correspond to integrated luminosities between 36.1 and 36.7 fb$^{-1}$. Only data are used in which all subsystems of the detector, as well as the trigger system, were fully functional. Events are required to have at least one reconstructed primary vertex with at least five associated ID tracks which is consistent with the LHC beam spot.

From this dataset, three different samples of events are analyzed to study the performance of $W$-boson and top-quark tagging algorithms in data:

- a $t\bar{t}$ signal sample, enriched in hadronically decaying top quarks,
- a $\gamma +$ jet background sample, covering the $p_T$ range from 200 GeV to about 2 TeV,
- a multijet background sample, covering the $p_T$ range from 500 GeV to about 3.5 TeV.

6.1 Signal efficiency in boosted $t\bar{t}$ events

To study the signal efficiency of $W$-boson and top-quark tagging, samples enriched in hadronically decaying $W$ bosons and top quarks are obtained by selecting $t\bar{t}$ events where one top quark decays hadronically and the other semi-leptonically. Events are selected in both the electron and the muon decay channels, using criteria similar to the ones used in Refs. [8, 9] and the $b$ quarks from the top-quark decay are used to select $W$-boson and top-quark enriched samples.

6.1.1 Analysis and selection

The events are required to pass either an inclusive electron or muon trigger, where the thresholds varied between the 2015 and 2016 data set due to increases in instantaneous luminosity. In the electron channel, events from the 2015 data-taking period are required to pass at least one of three triggers. The first is designed for isolated electrons with $p_T > 24$ GeV, while the second trigger requires electrons with $p_T > 60$ GeV without the isolation requirement and the third trigger requires electrons with $p_T > 120$ GeV without the isolation and less strict identification criteria. In the 2016 data-taking period, the $p_T$ thresholds of these electron triggers were increased to $p_T > 26$ GeV, $p_T > 60$ GeV and $p_T > 140$ GeV , respectively. In the muon channel, events from the 2015 data-taking period are required to pass at least one of two muon triggers. One selects isolated muons with $p_T > 20$ GeV and the other selects muons with $p_T > 50$ GeV without the isolation requirement. In the 2016 data-taking period, the $p_T$ thresholds of these triggers were increased to $p_T > 26$ GeV and $p_T > 50$ GeV, respectively.

Events are then required to contain exactly one electron or muon candidate with $p_T > 30$ GeV. Events with additional electron or muon candidates with $p_T > 20$ GeV are rejected. In this analysis, electron candidates are identified as ID tracks that are matched to a cluster of energy in the electromagnetic calorimeter. Electron candidates must satisfy a likelihood-based identification criterion [59, 60] based on shower shape and track selection criteria, and are selected using the “tight” working point. Lastly, they are required to be within $|\eta| < 2.47$, excluding the calorimeter transition region from $1.37 < |\eta| < 1.52$. In the case of muons, tracks found in the ID are matched to either full tracks or track segments reconstructed
in the muon spectrometer and are required to satisfy the “medium” muon identification quality criteria defined in Ref. [61]. They are required to have $|\eta| < 2.5$.

In the following, all distributions show the combined electron and muon channel events.

In addition to the identification of leptons, small-radius calorimeter jets are used to reconstruct the missing transverse momentum ($E_T^{\text{miss}}$) and identify the signal topology. These jets are reconstructed from topoclusters calibrated to the electromagnetic scale using the anti-$k_t$ algorithm with $R = 0.4$ radius parameter. The energy of these jets is corrected for the effects of pile-up by using a technique based on jet area [62] and the jet energy is further corrected using a Monte Carlo and data-based jet energy scale calibration [63]. To ensure that the reconstructed jets are well measured, they are required to have $p_T > 20$ GeV and to pass “loose” quality criteria to prevent mismeasurements due to calorimeter noise spikes and non-collision backgrounds. For jets with $p_T < 60$ GeV, a requirement that the jets arise from the primary vertex using the ID tracks associated with the jet is made to suppress pile-up jets [64].

For identification of $b$-quark candidate jets, jets reconstructed from ID tracks with anti-$k_t$ algorithm with radius parameter $R = 0.2$ are used, which are $b$-tagged using a multivariate discriminant based on impact parameter and secondary vertex information [65]. The 70% signal efficiency point selection is used.

The missing transverse momentum is reconstructed as the vectorial sum of the momenta of all reconstructed physics objects, which in this case consists of the single identified lepton and the full set of reconstructed and calibrated small-$R$ jets, and a contribution from unassociated energy deposits as measured via ID tracks. In the following the magnitude of the missing transverse momentum vector, $E_T^{\text{miss}}$, is used.

To select events with a leptonically-decaying $W$ boson, the following requirements are imposed. Events are required to have missing transverse momentum $E_T^{\text{miss}} > 20$ GeV and the scalar sum of $E_T^{\text{miss}}$ and the transverse mass of the leptonic $W$-boson candidate\(^8\) must satisfy $E_T^{\text{miss}} + M_W^T > 60$ GeV. In addition, requirements are made to ensure a topology consistent with a lepton plus jets final state originating from a $t\bar{t}$ event. At least one small-$R$ jet is required to have a transverse momentum of $p_T > 25$ GeV and to be close to the lepton ($\Delta R(\text{lepton, jet}) < 1.5$) to be representative of the semi-leptonically decaying top quark. To study $W$-boson and top-quark tagging, at least one large-$R$ jet, either a trimmed jet or a CA jet in the case of HEPTopTagger, with $p_T > 200$ GeV and $|\eta| < 2.0$ is required. At least one large-$R$ jet is required to be well-separated from the semi-leptonic top-quark decay by requiring it to be $\Delta R > 1.5$ away from the small-$R$ jet that is closest to the lepton. Additionally, the angular separation in the transverse plane between the lepton and the large-$R$ jet is required to be $\Delta \phi > 2.3$. Lastly, at least one $b$-tagged track jet is required.

The $t\bar{t}$ MC sample is divided into multiple components based on the truth partons from the hadronic top quark decay matched to the jet. For matched “$t\bar{t}$(top)” large-$R$ jets, two light quarks from top-quark decay and a $b$ quark must be within $\Delta R < R_{\text{match}}$ of the large-$R$ jet. For matched “$t\bar{t}$(W)” the light quarks from the $W$ boson decay must be within $\Delta R < R_{\text{match}}$ and no $b$ quark arising from a top-quark decay can be within $\Delta R < R_{\text{match}}$. All large-$R$ jets from $t\bar{t}$ events that are not matched to a true top-quark or a $W$-boson are put into the “$t\bar{t}$(other)” category.

The value of $R_{\text{match}}$ is taken to be $0.75 \cdot R_{\text{jet}}$, where $R_{\text{jet}}$ is the large-$R$ jet radius. Therefore, $R_{\text{match}} = 0.75$ for the large-$R$ anti-$k_t$ jet collection and $R_{\text{match}} = 1.125$ for the large-$R$ CA jets.

---

\(^8\) $M_W^T = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos \Delta \phi)}$ is calculated from the transverse momentum of the lepton, $p_T^\ell$ and $E_T^{\text{miss}}$ in the event. $\Delta \phi$ is the azimuthal angle between the lepton momentum and the $E_T^{\text{miss}}$ direction.
To illustrate the effectiveness of the $t\bar{t}$ selection in the data, the $p_T$ and mass distributions of the trimmed jet and the trimmed C/A jets for HEPTopTagger are shown in Figures 11 and 12. A breakdown of the systematic uncertainties shown in these and following figures is given in Section 6.3. For these distributions the total event yield in the simulation is normalized to match the one in data. These distributions show that the jet mass variable is reasonably well-modelled for the full range of masses. However, the $p_T$ of the large-$R$ jet is not reproduced by the MC models, which is a feature observed in differential cross-section measurements of top-quark production [66]. The disagreement in the $p_T$ spectrum is not relevant for the studies in this note, given the strategy to optimize the taggers to fixed working points as a function of top-quark $p_T$.

Figure 11: A comparison of the observed data and predicted MC distributions of the $p_T$ (left) and mass (right) of the $p_T$-leading large-$R$ jet in the event, where the $t\bar{t}$ MC sample is decomposed using generator level information.
Figure 12: A comparison of the observed data and predicted MC distributions of the $p_T$ (left) and mass (right) of the $p_T$-leading C/A $R = 1.5$ in the event, where the $t\bar{t}$ MC sample is decomposed using generator level information.
Before applying a tagger to the large-\(R\) jets, the purities of the samples of \(W\)-boson and top-quark signal jets are increased by splitting the events into two samples. A sample enriched in top quarks ("top-quark selection") is defined by requiring a \(b\)-tagged track jet to be within \(\Delta R(\text{\(b\)-jet, large-\(R\) jet}) < 1.0\) of the large-\(R\) jet, while the inverted requirement applies for the "\(W\)-boson selection". In addition, a requirement of \(p_T > 350\) GeV for the large-\(R\) jet is required to enhance the fraction of fully contained top quarks. This divides the inclusive \(t\bar{t}\) sample into two subsamples of relatively pure \(W\)-boson and top-quark jets as shown in Figure 13, where the event yield in the simulation is again fixed to match that in the data. This normalization is kept constant in the following. Analogous enrichment of top-quark signal jets is performed for C/A jets, where a \(b\)-tagged track jet is required to satisfy \(\Delta R(\text{\(b\)-jet, large-\(R\) C/A jet}) < 1.5\) and a \(p_T > 200\) GeV cut for the large-\(R\) jet is required. Commonly used jet substructure moments for \(W\)-boson tagging and top-quark tagging are presented for those topologies in Figures 14 and 15, respectively.

![Figure 13](image-url)

**Figure 13:** A comparison of the observed data and predicted MC distributions of the mass of the \(p_T\)-leading large-\(R\) jet in the event for the \(W\)-boson (left) and top-quark (right) selected samples, where the \(t\bar{t}\) MC sample is decomposed using generator level information. For the top-quark selection, an additional \(p_T > 350\) GeV cut on the large-\(R\) jet is required.
Figure 14: A comparison of the observed data and predicted MC distributions of the large-$R$ jet $D_2$ for the $W$-boson selected samples, where the $t\bar{t}$ MC sample is decomposed using generator level information.

Figure 15: A comparison of the observed data and predicted MC distributions of the large-$R$ jet $r_{32}$ (left) and $\sqrt{d_{23}}$ (right) for the top-quark selected samples, where the $t\bar{t}$ MC sample is decomposed using generator level information.
The distributions of the discriminant outputs of the BDT and the DNN W-boson and top-quark taggers are shown in Figure 16.

Figure 16: A comparison of the observed data and predicted MC distributions of the BDT (left) and DNN (right) discriminants for W-boson (top) and top-quark (bottom) tagging for the W-boson (left) and top-quark (right) selected samples. The $t\bar{t}$ MC sample is decomposed using generator level information. For the top-quark selection, an additional $p_T > 350$ GeV cut on the large-$R$ jet is required.

The effect of various taggers is illustrated by showing mass distributions for events that pass and fail the tagging. The combined mass distributions after application of the two-variable W-boson and top-quark taggers are shown in Figure 17. The combined mass distributions for the events after application of the multivariate W-boson and top-quark taggers are shown in Figures 18 and 19. The combined mass distributions after application of the shower deconstruction top-quark tagger are shown in Figure 20. The groomed C/A jet mass distributions after application of the HEPTopTagger tagger is shown in Figure 21. For all of the distributions after tagging has been applied, the MC prediction is normalized based on the ratio of data to predicted yields before applying a tag, in order to reveal potential normalisation
discrepancies between data and MC arising from the mismodelling of the tagging.

Figure 17: Combined jet mass distributions for events that pass (left) and fail (right) the two-variable tagging requirements are shown. The observed data are compared with the Monte Carlo prediction after application of the $D_2$ selection (i.e. no mass cut) of the two-variable $W$-boson tagger (top) and the two-variable top-quark tagger (bottom). For the top-quark selection, an additional $p_T > 350$ GeV cut on the large-$R$ jet is required.
Figure 18: Combined jet mass distributions for events that pass the BDT (left) and DNN (right) $W$-boson tagger (top) and top-quark tagger (bottom) taggers. The observed data are compared with the Monte Carlo prediction. For the top-quark selection, an additional $p_T > 350$ GeV cut on the large-$R$ jet is required. The systematics uncertainties for these distributions do not include large-$R$ jet uncertainties.

6.2 Background rejection from dijets and $\gamma +$ jet events

The efficiency of the $W$-boson and top-quark taggers tagging light quark and gluon jets is measured using a dijet sample and $\gamma +$ jet sample.

Dijet events are required to have at least one trimmed anti-$k_T$ jet with $p_T > 450$ GeV matching the large-$R$ jet that the event was triggered on. This results in a large sample of light quark and gluon jets, which are compared to the Pythia dijet sample described in Section 3.
by performing the procedure multiple times, varying the signal normalization by ±1%. Uncertainty in background efficiency measurements due to this normalization procedure can be estimated from the reconstructed photon (Δφ(jet, γ) > π/2) is required.

An identical procedure is used to normalize the simulated prediction to the data in both the dijet and γ + jet selections, taking into account the small contribution from W-boson, Z-boson and top-quark events. First the predicted contribution from signal processes is subtracted from data to get signal-subtracted data. The Monte Carlo background samples are then normalized to the integral of this distribution. The uncertainty in background efficiency measurements due to this normalization procedure can be estimated by performing the procedure multiple times, varying the signal normalization by ±25%. Variations in the
Figure 20: Combined jet mass distributions for events that pass (left) and fail (right) shower deconstruction top-quark tagger are shown. The observed data are compared with the Monte Carlo prediction. For the top-quark selection, an additional $p_T > 350$ GeV cut on the large-$R$ jet is required.

Figure 21: C/A jet mass distributions for events that pass (left) and fail (right) HEPTopTagger top-quark tagger are shown. The observed data are compared with the Monte Carlo prediction. For the top-quark selection, an additional $p_T > 350$ GeV cut on the large-$R$ jet is required.

resulting background rejection measurement from this signal normalization uncertainty are observed to be $< 1\%$ in all cases and thus neglected.

In addition to covering different $p_T$ regions, the dijet and $\gamma + $ jet samples differ in what partons initiated the jets under study. In the $\gamma + $ jet topology the jets are dominantly initiated by quarks over the full $p_T$ range studied, while for the dijet topology the fraction of quarks initiating the jets is slightly smaller than the gluon fraction at low $p_T$ and becomes large at high $p_T$. 


Figure 22 shows a comparison of the distributions of the leading jet mass and $p_T$ in the dijet and $\gamma + \text{jet}$ selections. Additionally, commonly used jet substructure variables for $W$-boson tagging and top-quark tagging are also presented in Figure 23 for the $\gamma + \text{jet}$ selection.

Figure 22: Distributions of combined jet mass (top) and $p_T$ in the dijet and $\gamma + \text{jet}$ selections. Compared are dijet (left) and $\gamma + \text{jet}$ (right) events in data and MC simulation. The dijet data are compared to two different dijet MC samples, PYTHIA8 (red) and HERWIG++ (blue), therefore two lines are shown in the data/MC ratios.
Figure 23: Distributions of the jet substructure variables used by the two-variable taggers ($D_2$, $\tau_{22}$ and $\sqrt{d_{23}}$). The observed data are compared to the MC prediction in the $\gamma$ + jet selection.
The discriminant output distributions of the BDT and the DNN W-boson and top-quark taggers are shown in Figure 24 in the dijet selection and in Figure 25 in the γ + jet selection.

Figure 24: Distributions of the BDT (left) and DNN (right) discriminants in the dijet selection for W-boson tagging (top row) and top-quark tagging (bottom row). Only statistical uncertainties are shown. The dijet data are compared to two different dijet MC samples, PYTHIA8 (red) and HERWIG++ (blue), therefore two lines are shown in the data/MC ratios.
Figure 25: A comparison of the observed data and predicted MC distributions of the BDT (left) and DNN (right) discriminants in the $\gamma$ + jet selection for $W$-boson tagging (top row) and top-quark tagging (bottom row). Only statistical uncertainties are shown.
6.3 Systematic uncertainties

A number of systematic uncertainties are considered in the individual topologies for the studies of tagger performance using data, and are summarized in Table 7. Individual systematic uncertainties are symmetrized and summed in quadrature, unless explicitly specified otherwise.

In the signal efficiency measurement, in general the \( t\bar{t} \) modelling and large-\( R \) jet scale uncertainties are the dominant sources of uncertainties. A more detailed summary of the systematic uncertainties on the

<table>
<thead>
<tr>
<th>Source</th>
<th>Affected topologies</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event generator choice</td>
<td>( t\bar{t} l+\text{jets} )</td>
<td>Hard-scattering modelling uncertainty estimated as difference between Powheg+Herwig and MC@NLO+Herwig sample.</td>
</tr>
<tr>
<td>Showering choice</td>
<td>( t\bar{t} l+\text{jets} )</td>
<td>Parton shower and hadronization modelling uncertainty estimated as difference between Powheg+Pythia6 and Powheg+Herwig sample.</td>
</tr>
<tr>
<td>Modelling of extra QCD radiation</td>
<td>( t\bar{t} l+\text{jets} )</td>
<td>Variation of renormalization and factorisation scales and the ( h_{\text{damp}} ) parameter of the Powheg+Pythia6 sample.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Affected topologies</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trimmed ( k_t ) ( R = 1.0 ) jet energy scale and jet substructure scale</td>
<td>( t\bar{t} l+\text{jets, dijets, } \gamma+\text{jets} )</td>
<td>Uncertainties on ( p_T ), mass, ( \tau_{32} ), ( D_2 ), ( \sqrt{d_{32}} ) and ( Q_w ) scales derived from comparing calorimeter response to response of tracks matched to jets [8]. These are referred to as “large-( R ) jet uncertainties”.</td>
</tr>
<tr>
<td>Trimmed C/A ( R = 1.5 ) subjet jet energy scale</td>
<td>( t\bar{t} l+\text{jets, dijets, } \gamma+\text{jets} )</td>
<td>A flat 3% uncertainty is assigned as an upper limit given the Run I studies [9].</td>
</tr>
<tr>
<td>Anti-( k_t ) ( R = 0.4 ) jet energy scale and resolution</td>
<td>( t\bar{t} l+\text{jets} )</td>
<td>Uncertainties derived from simulation and in-situ calibration measurements based on momentum balance between jets and reference objects (photon, Z boson) [31].</td>
</tr>
<tr>
<td>Anti-( k_t ) ( R = 0.2 ) track jet ( b )-tagging</td>
<td>( t\bar{t} l+\text{jets} )</td>
<td>Uncertainties on scale factors correcting the mis-modelling of tagging efficiencies of ( b ) and mis-tagging efficiencies of ( c ) and light jets [65, 67].</td>
</tr>
<tr>
<td>Uncertainties related to leptons</td>
<td>( t\bar{t} l+\text{jets} )</td>
<td>Uncertainties on scale factors correcting the mis-modelling of electron and muon trigger and identification efficiencies and uncertainties on energy scale and resolution [68–70].</td>
</tr>
<tr>
<td>QCD fake leptons normalisation</td>
<td>( t\bar{t} l+\text{jets} )</td>
<td>A 50% uncertainty on the data-driven prediction yield is assigned, based on the estimate in Ref. [2].</td>
</tr>
</tbody>
</table>

Table 7: Summary of systematic uncertainties considered in the performance measurements in data.
yield in the $t\bar{t}$ signal efficiency control plots is given below.

- The leading-$p_T$ large-$R$ jet spectra (Figures 11 and 12) uncertainties are dominated by $t\bar{t}$ modelling uncertainties, in particular radiation together with hard-scattering modelling in the high-$p_T$ region and showering/hadronization together with hard-scattering modelling in the low-$p_T$ region.

- In the combined mass spectra (Figures 11-13), the $t\bar{t}$ modelling uncertainties are in general a dominant source of uncertainties with the exception of certain specific regions. In the $W$-boson and top-quark mass peaks the large-$R$ jet scale uncertainties dominate, with the showering/hadronization modelling uncertainties being the second largest source of uncertainty.

- After the application of a higher $p_T$ cut and the $\Delta R(b, large-R)$ jet) < 1.0 (for C/A jets, the $\Delta R$ cut is 1.5) in the top-quark selection (Figure 13), in the low-mass region, the $b$-tagging uncertainties dominate. In the regions of the (now suppressed) $W$-boson peak and top-quark peak, the showering/hadronization uncertainties are the dominant source of uncertainty. The large-$R$ jet scale uncertainties are of similar size in the top mass peak compared to $t\bar{t}$ modelling uncertainties.

- The substructure distribution (Figures 14-15) uncertainties are dominated by $t\bar{t}$ modelling. The low-tail regions of $\tau_{32}$ and $D_2$ are in particular dominated by modelling of extra QCD radiation. High-tail regions of all three substructure observables are dominated by large-$R$ scale uncertainties.

- The BDT and DNN discriminant distributions' (Figure 16) uncertainties are dominated by $t\bar{t}$ modelling, with similar contributions in magnitude from all generator uncertainty sources. The low-discriminant region mostly populated by 1-prong jets is dominated by uncertainty on additional QCD radiation modelling in $t\bar{t}$.

- Post-tag mass distributions (Figures 17-20) exhibit similar relative uncertainty contributions as pre-tag distributions. However the $t\bar{t}$ modelling uncertainties (showering/hadronization in particular) in tagged top-quark and $W$-boson mass regions are the same magnitude as the large-$R$ jet uncertainties.

- For the HepTopTagger distributions (Figures 12 and 21), the uncertainty contributions are similar to those for anti-$k_t$ top-tagging related distributions. The notable difference is that the subject energy scale uncertainties are a conservative estimate based on Run I studies, and these uncertainties dominate in the top-quark mass peak.

In the background rejection studies, the uncertainties in dijet and $\gamma$+jets topologies are dominated by large-$R$ jet uncertainties.

### 6.4 Summary of performance measurements in data

#### 6.4.1 Signal efficiencies

Given the relatively high purity samples of $W$-boson and top-quark jets that result from the selection described in Section 6.1.1, the tagging efficiency can be measured in data and the modelling can be compared to that predicted by MC simulation.

The following quantities are extracted:

$$\epsilon_{MC} = \frac{N_{\text{tagged truth top}}/W}{N_{\text{total truth top}}/W}$$

(6)
\[ \epsilon_{\text{data}} = \frac{N_{\text{tagged data}}}{N_{\text{total data}}} - \frac{N_{\text{tagged truth non-top/W}}}{N_{\text{total truth non-top/W}}} \]

(7)

where \( N_{\text{total}} \) are all events in simulation that are top-quark matched before top-quark tagging and \( N_{\text{truth top/W}} \) are all events that are top-quark matched. In data, \( N_{\text{tagged data}} \) are all events that are tagged and \( N_{\text{total data}} \) are all events in data before tagging. The backgrounds for the data efficiency measurement correspond to the sample of truth-unmatched events from \( t\bar{t} \) MC and other non-\( t\bar{t} \) background samples, where \( N_{\text{tagged truth non-top/W}} \) are events after top-quark tagging and \( N_{\text{total truth non-top/W}} \) are all events before tagging. A normalization of prediction to data is performed, such that \( N_{\text{total truth top/W}} + N_{\text{total truth non-top/W}} = N_{\text{total data}} \), i.e. the total prediction before tagging is normalised to data. This normalization of prediction is performed in individual large-\( R \) jet \( p_T \) bins separately.

The signal efficiency as a function of the \( p_T \) of the large-\( R \) jet is shown for the two-variable \( W \)-boson tagger and for the two-variable top-quark tagger in Figure 26 in data and simulation. In Figures 27 and 28 the signal efficiencies for the SD and HEPTopTagger top-quark taggers and the BDT and DNN \( W \)-boson taggers are shown.

![Figure 26: The signal efficiency for the two-variable \( W \) (left) and top (right) taggers as a function of the large-\( R \) jet \( p_T \) in data and simulation. The top panel shows the statistical uncertainties on the signal efficiency measurement in data and MC. In the bottom panel, the ratio Data/MC is shown with its statistical uncertainty (black dots and associated error bars). The green bands are centered around one, and represent the same statistical uncertainty (dark green) and the quadratic sum of statistical and systematic uncertainties (light green) on the ratio. The systematic uncertainties do not include the large-\( R \) jet uncertainties (see Table 7).](image)

### 6.4.2 Background rejections

The rejection as a function of the \( p_T \) of the large-\( R \) jet is shown for the two-variable \( W \)-tagger and for the two-variable top-quark tagger in Figures 29 and 30, respectively, in data and simulation for the dijet and \( \gamma + \text{jet} \) selections. In Figures 31 to 34 the background rejection for the SD and HEPTopTagger taggers
Figure 27: The signal efficiency for the shower deconstruction top-quark tagger (left) and the HEPTopTagger algorithm (right) as a function of the large-$R$ jet $p_T$ in data and simulation. The top panel shows the statistical uncertainties on the signal efficiency measurement in data and MC. In the bottom panel, the ratio Data/MC is shown with its statistical uncertainty (black dots and associated error bars). The green bands are centered around one, and represent the same statistical uncertainty (dark green) and the quadratic sum of statistical and systematic uncertainties (light green) on the ratio. The systematic uncertainties do not include the large-$R$ jet uncertainties (see Table 7).

and the BDT and DNN $W$-boson and top-quark taggers are shown. For the dijet topology, the PYTHIA8 prediction for the background rejection describes the observed rejection, while the HERWIG++ prediction is higher than the rejection in data, especially for lower values of jet $p_T$.

Although the rejections for the two topologies are similar, there are relatively large uncertainties at higher jet $p_T$. Differences in the observed rejection may also arise from different gluon and light quark compositions in the samples.
Figure 28: The signal efficiency for the BDT (left) and DNN (right) $W$-boson (top) and top-quark (bottom) taggers as a function of the large-$R$ jet $p_T$ in data and simulation. The top panel shows the statistical uncertainties on the signal efficiency measurement in data and MC. In the bottom panel, the ratio Data/MC is shown with its statistical uncertainty (black dots and associated error bars). The green bands are centered around one, and represent the same statistical uncertainty (dark green) and the quadratic sum of statistical and systematic uncertainties (light green) on the ratio. The systematic uncertainties do not include the large-$R$ jet uncertainties (see Table 7).
Background rejection (1/\(bkg\))

Figure 29: QCD background rejection as a function of the leading jet \(p_T\) for the two-variable \(W\)-boson tagger. The left (right) plot show a Data/MC comparison for the dijet (\(\gamma + \text{jet}\)) selections. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection. For the MC points, systematic uncertainties (light green shaded area) are due to the large-\(R\) jet systematic uncertainties.

Figure 30: QCD background rejection as a function of the leading jet \(p_T\) for the two-variable top-quark tagger. The left (right) plot show a Data/MC comparison for the dijet (\(\gamma + \text{jet}\)) selections. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection. For the MC points, systematic uncertainties (light green shaded area) are due to the large-\(R\) jet systematic uncertainties.
Figure 31: QCD background rejection as a function of the leading jet $p_T$ for the SD top-quark tagger. Shown is a Data/MC comparison for the dijet selection. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection.

Figure 32: QCD background rejection as a function of the leading jet $p_T$ for the HEPTopTagger top-quark tagger. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection. For the MC points, systematic uncertainties (light green shaded area) are obtained from the C/A subjet energy scale uncertainties that are applied in both the trimming and the HEPTopTagger algorithm.
Figure 33: QCD background rejection as a function of the leading jet $p_T$ for the BDT (top) and DNN (bottom) $W$-boson taggers. The left (right) plot show a Data/MC comparison for the dijet ($\gamma + \text{jet}$) selections. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection.
Figure 34: QCD background rejection as a function of the leading jet $p_T$ for the BDT (top) and DNN (bottom) top-quark taggers. The left (right) plot show a Data/MC comparison for the dijet ($\gamma +\text{jet}$) selections. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection.
6.4.3 Pile-up dependence of the BDT and DNN taggers in data

The pile-up dependence observed in the data and the MC simulations are compared for the dijet, $\gamma +$ jet and $t\bar{t}$ signal events. For $t\bar{t}$ signal events the dependence is shown for fully contained top-quark/$W$-boson jets in $t\bar{t}$ MC, and for background-subtracted data. For the background subtraction, the yield of the total MC prediction in each bin of the average number of interactions, $\mu$, is normalised to the yield in data and then all of the non-$t\bar{t}$ backgrounds as well as the partially contained top-quark/$W$-boson jets from the $t\bar{t}$ MC sample are subtracted from the data. The resulting two-dimensional distributions of BDT or DNN discriminant versus mean number of interactions per bunch crossing $\mu$ are smoothed and from these the means of BDT or DNN discriminant in $\mu$ are obtained. The dependences in the examined topologies are presented and compared in Figures 35 and 36. Only statistical uncertainties are shown. Most of the missing systematic uncertainties would likely be correlated across $\mu$ bins.

For both the BDT and the DNN $W$-boson taggers a negative slope of the discriminants versus $\mu$ is seen and the $t\bar{t}$ simulation describes the data for the $W$-boson signal. For the background samples there are offsets between data and simulation, but the overall negative slopes are similar in data and simulation. A similar offset in the mean discriminant value is observed for the backgrounds for the BDT and DNN top-quark taggers, but no slope is seen in either data or simulation. While the DNN discriminant for the top quark signal from $t\bar{t}$ events seems to be compatible to data in simulation, there is a clear difference visible in the high $\mu$ data compared to MC simulation for the BDT discriminant that is not compatible within statistical uncertainties.

Figure 35: The mean of the BDT (left) and DNN (right) $W$-boson tagger discriminants in data and MC as a function of $\mu$. Both the dijet and $\gamma +$ jet background and $t\bar{t}$ signal events are presented to observe the separation. The large-$R$ jets from $t\bar{t}$ MC contain a fully contained truth-matched hadronically-decaying top quark. For the data points in $t\bar{t}$ topology, the non-$t\bar{t}$ backgrounds and fraction of $t\bar{t}$ not containing fully contained truth-matched hadronically-decaying top quark within a large-$R$ jet are subtracted from data. For the dijets MC prediction, the nominal sample generated by PYTHIA8 is used.
Figure 36: The mean of the BDT (left) and DNN (right) top-quark tagger discriminants in data and MC as a function of $\mu$. Both the dijet and $\gamma$+jets background and $t\bar{t}$ signal events are presented to observe the separation. The large-$R$ jets from $t\bar{t}$ MC contain a fully contained truth-matched hadronically-decaying top quark. For the data points in $t\bar{t}$ topology, the non-$t\bar{t}$ backgrounds and fraction of $t\bar{t}$ not containing fully contained truth-matched hadronically-decaying top quark within a large-$R$ jet are subtracted from data. For the dijets MC prediction, the nominal sample generated by PYTHIA8 is used.
7 Conclusion

Various techniques to tag boosted, hadronically-decaying $W$ bosons and top quarks are studied in data and simulation.

Tagger performance is evaluated using MC simulation for jets in the $p_T$ range from 500 to 2000 GeV. In these studies two-variable, BDT and DNN taggers are optimized. The performance of these optimised taggers is evaluated, and comparisons are made among each other as well as to the shower deconstruction and HEPTopTagger algorithms. In simulation the BDT and DNN taggers perform similarly and outperform all other taggers for both $W$-boson and top-quark tagging. The study is limited by an incomplete assessment of the systematic uncertainties that could have an effect on these conclusions.

The performance of the taggers is studied using the data collected in 2015 and 2016. The signal and background enriched event topologies are studied using $t\bar{t}$ lepton+jets, dijet and $\gamma$ + jet samples. The $\gamma$ + jet sample is used in a tagging performance study for the first time in ATLAS and has a different combination of light-quark- and gluon-initiated jets than the dijet sample. The observed large-$R$ jet kinematic and substructure variables, which are the inputs to the taggers, are well described by the MC simulations within uncertainties. Signal efficiencies and background rejections for all taggers evaluated are found to be also well described.

The performance of the BDT and DNN $W$-boson and top-quark taggers are studied in data for the first time in ATLAS. The BDT and DNN discriminants in the data is adequately described by the MC simulation within the considered uncertainties. However, these comparisons do not include systematic uncertainties arising from the large-$R$ jets.

This analysis demonstrates that the inputs to and the performance of the studied top-quark and $W$-boson taggers currently in use in physics analyses are well modeled by simulations.
Appendix

A Combined Background Topologies

Figure 37: Background rejection as a function of leading jet $p_T$ for the BDT and DNN $W$-boson taggers (top) and top-quark taggers (bottom). In each plot, $\gamma + \text{jet}$ data (simulation) are shown as solid blue (hollow black) triangles and dijet data (simulation) are shown as red (hollow black) circles. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection. Only the statistical uncertainties are shown.
Figure 38: Background rejection as a function of leading jet $p_T$ for the two-variable $W$-boson tagger (left) and top-quark tagger (right). In each plot, $\gamma +$ jet data (simulation) are shown as solid blue (hollow black) triangles and dijet data (simulation) are shown as red (hollow black) circles. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection. Only the statistical uncertainties are shown.

Figure 39: Background rejection as a function of leading jet $p_T$ for the HEPTopTagger top-quark tagger. In each plot, $\gamma +$ jet data (simulation) are shown as solid blue (hollow black) triangles and dijet data (simulation) are shown as red (hollow black) circles. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection. Only the statistical uncertainties are shown.
References


[3] ATLAS Collaboration, *A Search for Resonances Decaying to a W or Z Boson and a Higgs Boson in the $q\bar{q}(c\bar{c})b\bar{b}$ Final State*, ATLAS-CONF-2016-083, 2016, url: https://cds.cern.ch/record/2206276.


[51] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, 
*Longitudinally invariant $K_t$ clustering algorithms for hadron hadron collisions*, 

[52] A. J. Larkoski, D. Neill and J. Thaler, 
*Jet Shapes with the Broadening Axis*, 

[53] ATLAS Collaboration, 
*Performance of shower deconstruction in ATLAS*, 

[54] M. Cacciari, G. P. Salam and G. Soyez, 
*FastJet User Manual*, 

*Resonance Searches with an Updated Top Tagger*, 

[56] CMS Collaboration, 
*Identification techniques for highly boosted $W$ bosons that decay into hadrons*, 

[57] CMS Collaboration, 

[58] ATLAS Collaboration, 
*Identification techniques for highly boosted $W$ bosons and top quarks using high-level features as input to boosted decision trees and deep neural networks in ATLAS at $\sqrt{s} = 13$ TeV*, 

[59] ATLAS Collaboration, 
*Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton–proton collision data*, 

[60] ATLAS Collaboration, 
*Electron identification measurements in ATLAS using $\sqrt{s} = 13$ TeV data with 50 ns bunch spacing*, 

[61] ATLAS Collaboration, 
*Muon reconstruction performance in early $\sqrt{s} = 13$ TeV data*, 

[62] M. Cacciari, G. P. Salam and G. Soyez, 
*The Catchment Area of Jets*, 

[63] ATLAS Collaboration, 
*Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, 

[64] ATLAS Collaboration, 
*Tagging and suppression of pileup jets with the ATLAS detector*, 

[65] ATLAS Collaboration, 
*Optimisation of the ATLAS $b$-tagging performance for the 2016 LHC Run*, 

[66] ATLAS Collaboration, 
*Comparison of Monte Carlo generator predictions to ATLAS measurements of top pair production at 7 TeV*, 

