Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector


DOI
10.1007/JHEP09(2013)076

Publication date
2013

Document Version
Final published version

Published in
The Journal of High Energy Physics

Citation for published version (APA):
Performance of jet substructure techniques for large-$R$ jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: This paper presents the application of a variety of techniques to study jet substructure. The performance of various modified jet algorithms, or jet grooming techniques, for several jet types and event topologies is investigated for jets with transverse momentum larger than 300 GeV. Properties of jets subjected to the mass-drop filtering, trimming, and pruning algorithms are found to have a reduced sensitivity to multiple proton-proton interactions, are more stable at high luminosity and improve the physics potential of searches for heavy boosted objects. Studies of the expected discrimination power of jet mass and jet substructure observables in searches for new physics are also presented. Event samples enriched in boosted $W$ and $Z$ bosons and top-quark pairs are used to study both the individual jet invariant mass scales and the efficacy of algorithms to tag boosted hadronic objects. The analyses presented use the full 2011 ATLAS dataset, corresponding to an integrated luminosity of $4.7\pm0.1$ fb$^{-1}$ from proton-proton collisions produced by the Large Hadron Collider at a centre-of-mass energy of $\sqrt{s} = 7$ TeV.

KEYWORDS: Hadron-Hadron Scattering

ArXiv ePrint: 1306.4945
## Contents

1 Introduction  
1.1 Motivation  
1.2 Definitions  
1.2.1 Jet algorithms  
1.2.2 Jet properties and substructure observables  
1.2.3 Jet grooming algorithms  
1.2.4 HEPTopTagger  

2 The ATLAS detector and data samples  
2.1 The ATLAS detector  
2.2 Data and Monte Carlo samples  

3 Jet reconstruction and calibration  
3.1 Inputs to jet reconstruction  
3.2 Jet quality criteria and selection  
3.3 Jet calibration and systematic uncertainties  
3.3.1 Monte Carlo based calibration  
3.3.2 Jet mass scale validation in inclusive jet events using track-jets  
3.3.3 Jet mass scale validation using hadronic $W$ decays in $t\bar{t}$ events  
3.4 In situ validation of the subjet energy scale  
3.5 Calibration of subjets  

4 Jet substructure and grooming in the presence of pile-up  
4.1 Impact of pile-up on the jet energy scale and the jet mass scale  
4.2 Impact of pile-up on jet substructure properties  
4.3 Impact of pile-up on signal and background in simulation  

5 Jet substructure and grooming with boosted objects in data and simulation  
5.1 Expected performance of jet substructure and grooming in simulation  
5.1.1 Jet mass resolution for background  
5.1.2 Jet mass resolution for simulated signal events  
5.1.3 Signal and background comparisons with and without grooming  
5.2 Inclusive jet data compared to simulation with and without grooming  
5.3 Performance of jet grooming in boosted top-quark events  
5.3.1 Semi-leptonic $t\bar{t}$ selection  
5.3.2 Performance of trimming in $t\bar{t}$ events  
5.3.3 Application of the HEPTopTagger in $t\bar{t}$ events  

6 Conclusions  
The ATLAS collaboration
1 Introduction

The dominant feature of high-energy proton-proton ($pp$) collisions at the Large Hadron Collider (LHC) is the production of highly collimated sprays of energetic hadrons, called jets, that originate from the quarks and gluons in the primary collisions. The large centre-of-mass energy at the LHC enables the production of Lorentz-boosted heavy particles, whose decay products can be reconstructed as one large-area jet. The study of the internal structure of jets goes beyond the four-momentum description of a single parton and yields new approaches for testing Quantum Chromodynamics (QCD) and for searching for new physics in hadronic final states. However, many of the new tools developed for the study of jet substructure at the LHC have only recently been validated with data in a hadron-hadron collider environment. For example, the effect of multiple $pp$ interactions on large-area jet measurements has not been extensively studied experimentally.

This paper presents a comprehensive set of studies designed to establish the efficacy, accuracy, and precision of several of the tools available for determining and analysing the internal structure of jets at the LHC. New jet algorithms and strategies, referred to as jet grooming, that refine the definition of a jet in a high-luminosity environment, are studied using data taken at a centre-of-mass energy of $\sqrt{s} = 7$ TeV during 2011. A variety of techniques and tagging algorithms intended to improve the mass resolution in the reconstruction of boosted objects that decay hadronically are studied in the data both in inclusive jet samples and in samples enriched in events containing boosted $W/Z$ bosons and top quarks. Evaluations of the systematic uncertainties for jet mass measurements are presented for a variety of jet algorithms. Comparisons of the discrimination between signal and background provided by various observables are also evaluated for a selection of models of new physics containing boosted hadronic particle decays.

The organization of the paper is as follows. In this section, motivation for the use of new jet reconstruction techniques for Lorentz-boosted particles is given and the jet algorithms and jet substructure variables that are used in the analyses presented here are defined. Section 2 provides descriptions of the ATLAS detector and the Monte Carlo simulations, and section 3 defines the jet reconstruction and calibration procedures that are used throughout. The latter section includes a discussion of the jet mass scale and the subjet energy scale, which are important ingredients in the jet grooming algorithms. Section 4 describes studies of the effect of jet grooming on jet properties in the presence of pile-up, which represents a major experimental challenge at the present and future LHC machine.

Studies of the performance of the various jet algorithms are conducted with three classes of event samples in both data and Monte Carlo simulation in section 5: inclusive jet events, which are dominated by light-quark or gluon jets whose properties are defined primarily by soft gluon emission; boosted hadronically decaying $W$ and $Z$ bosons, which form jets that are dominated by two high-$p_T$ components; and top-quark decays, where the $W$ boson decays hadronically, which form jets that have three prominent components (due to the $b$-quark in addition to the $W$). In section 5.1, the effect on jet resolution of the various jet grooming algorithms is compared in simulated events separately for signal
(W, Z, and top jets) and background from light-quark and gluon jets. The discrimination between background and signal is then studied using a number of grooming configurations by comparing jet properties for the different types of events before and after grooming. This is followed in section 5.2 by a direct comparison of multiple Monte Carlo predictions and inclusive jet data. Lastly, section 5.3 presents jet grooming studies on boosted top-quark events. Finally, conclusions are drawn in section 6.

1.1 Motivation

The centre-of-mass energy of the LHC has opened new kinematic regimes to experimental study. The new phase space available for the production of Standard Model (SM) particles with significant Lorentz boosts, or even new massive particles that decay to highly boosted SM particles, necessitates new techniques to conduct measurements in novel final states. For example, when sufficiently boosted, the decay products of \( W \) bosons \([1–5]\), top quarks \([6–8]\), and Higgs bosons \([9]\) can become collimated to the point that standard reconstruction techniques begin to fail. When the separation of the quarks in these boosted topologies becomes smaller than the radius parameter of the jets, they often fail to be individually resolved by standard jet algorithms and configurations. Moreover, the high-luminosity conditions at the LHC can further degrade even the most complex procedures for reconstructing decays of boosted hadronic objects. Multiple \( pp \) interactions per bunch crossing (pile-up) produce soft particles unrelated to the hard scattering that can contaminate jets in the detector considerably more than at previous hadron-hadron colliders. In events where boosted particle decays are fully contained within individual large-radius jets, a diminished mass resolution due to pile-up may dramatically weaken sensitivity to new physics processes. It is crucial that the above issues be addressed together, as the efficacy of a given technique for boosted object reconstruction may depend critically on its vulnerability to experimental conditions.

One example of a new physics process that may produce heavy objects with a significant Lorentz boost is the decay of a new heavy gauge boson, the \( Z' \), to top-quark pairs. Figure 1 shows the angular separation between the \( W \) and \( b \) decay products of a top quark in simulated \( Z' \rightarrow t\bar{t} \) \((m_{Z'} = 1.6 \text{ TeV})\) events, as well as the separation between the light quarks of the subsequent hadronic decay of the \( W \) boson. In each case, the angular separation of the decay products is approximately

\[
\Delta R \approx \frac{2m}{p_T},
\]

where \( \Delta R = \sqrt{\Delta y^2 + (\Delta \phi)^2} \), and \( p_T \) and \( m \) are the transverse momentum and the mass, respectively, of the decaying particle.\(^1\) For \( p_T^W > 200 \text{ GeV} \), the ability to resolve the

\(^1\)The ATLAS coordinate system is a right-handed system with the \( x \)-axis pointing to the centre of the LHC ring and the \( y \)-axis pointing upwards. The polar angle \( \theta \) is measured with respect to the LHC beam-line. The azimuthal angle \( \phi \) is measured with respect to the \( x \)-axis. The rapidity is defined as \( y = 0.5 \times \ln[(E + p_z)/(E - p_z)] \), where \( E \) denotes the energy and \( p_z \) is the component of the momentum along the beam direction. The pseudorapidity \( \eta \) is an approximation for rapidity \( y \) in the high-energy limit, and it is related to the polar angle \( \theta \) by \( \eta = -\ln \tan(\theta/2) \). Transverse momentum and energy are defined as \( p_T = p \times \sin \theta \) and \( E_T = E \times \sin \theta \), respectively.
Figure 1. (a) The angular separation between the $W$ boson and $b$-quark in top decays, $t \to Wb$, as a function of the top-quark transverse momentum ($p_T^t$) in simulated PYTHIA [10] $Z' \to t\bar{t}$ ($m_{Z'} = 1.6$ TeV) events. (b) The angular distance between the light quark and anti-quark from $t \to Wb$ decays as a function of the $p_T$ of the $W$ boson ($p_T^W$). Both distributions are at the generator level and do not include effects due to initial and final-state radiation, or the underlying event.

individual hadronic decay products using standard narrow-radius jet algorithms begins to degrade, and when $p_T^t$ is greater than 300 GeV, the decay products of the top quark tend to have a separation $\Delta R < 1.0$. Techniques designed to recover sensitivity in such cases focus on large-$R$ jets in order to maximize efficiency. In this paper, large-$R$ refers to jets with a radius parameter $R \geq 1.0$. At $\sqrt{s} = 7$ TeV, nearly one thousand SM $t\bar{t}$ events per fb$^{-1}$ are expected with $p_T^W$ greater than 300 GeV. New physics may appear in this region of phase space, the study of which was limited by integrated luminosity and available energy at previous colliders.

A single jet that contains all of the decay products of a massive particle has significantly different properties than a jet of the same $p_T$ originating from a light quark. The characteristic two-body or three-body decays of a high $p_T$ vector boson or top quark result in a hard substructure that is absent from typical high $p_T$ jets formed from gluons and light quarks. These subtle differences in substructure can be resolved more clearly by removing soft QCD radiation from jets. Such adaptive modification of the jet algorithm or selective removal of soft radiation during the process of iterative recombination in jet reconstruction is generally referred to as jet grooming [9, 11, 12].

Recently many jet grooming algorithms have been designed to remove contributions to a given jet that are irrelevant or detrimental to resolving the hard decay products from a boosted object (for recent reviews and comparisons of these techniques, see for example refs. [13, 14]). The structural differences between jets formed from gluons or light quarks and individual jets originating from the decay of a boosted hadronic particle form the basis for these tools. The former are characterized primarily by a single dense core of energy surrounded by soft radiation from the parton shower, hadronization, and underlying event (UE) remnants [15–17]. Jets containing the decay products of single massive particles,
Figure 2. Single-jet invariant mass distribution for Cambridge–Aachen (C/A) $R = 1.2$ jets in simulated events containing highly boosted hadronically decaying $Z$ bosons before and after the application of a grooming procedure referred to as mass-drop filtering. The technical details of this figure are explained in section 1.2. The normalization of the groomed distribution includes the efficiency of mass-drop filtering with respect to the ungroomed large-$R$ jets for comparison. The local cluster weighting (LCW) calibration scheme is described in section 3.3.1.

on the other hand, can be distinguished by hard, wide-angle components representative of the individual decay products that result in a large reconstructed jet mass, as well as typical kinematic relationships among the hard components of the jet [4, 7, 9, 18–23]. Grooming algorithms are designed to retain the characteristic substructure within such a jet while reducing the impact of the fluctuations of the parton shower and the UE, thereby improving the mass resolution and mitigating the influence of pile-up. These features have only recently begun to be studied experimentally [24–31] and have been exploited heavily in recent studies of the phenomenological implications of such tools in searches for new physics [9, 32–42]. A groomed jet can also be a powerful tool to discriminate between the often dominant multi-jet background and the heavy-particle decay, which increases signal sensitivity. Figure 2 demonstrates this by comparing the invariant mass distribution of single jets in events containing highly boosted hadronically decaying $Z$ bosons before and after the application of a grooming procedure referred to as mass-drop filtering. In this simulated $Z \rightarrow q\bar{q}$ sample described in section 2.2, pile-up events are also included. Prior to the application of this procedure, no distinct features are present in the jet mass distribution, whereas afterwards, a clear mass peak that corresponds to the $Z$ boson is evident.

1.2 Definitions

1.2.1 Jet algorithms

In this paper, three jet algorithms are studied: the anti-$k_t$ algorithm [43], the Cambridge–Aachen (C/A) algorithm [44, 45], and the $k_t$ algorithm [46–48]. These are implemented within the framework of the FastJet software [49, 50]. They represent the most widely
used infrared and collinear-safe jet algorithms available for hadron-hadron collider physics today. Furthermore, in the case of the $k_t$ and C/A algorithms, the clustering history of the algorithm — that is, the ordering and structure of the pair-wise subjet recombinations made during jet reconstruction — provides spatial and kinematic information about the substructure of that jet. The anti-$k_t$ algorithm provides jets that are defined primarily by the highest-$p_T$ constituent, yielding stable, circular jets. The compromise is that the structure of the jet as defined by the successive recombinations carried out by the anti-$k_t$ algorithm carries little or no information about the $p_T$ ordering of the shower or wide angular-scale structure. It is, however, possible to exploit the stability of the anti-$k_t$ algorithm and recover meaningful information about the jet substructure: anti-$k_t$ jets are selected for analysis based on their kinematics ($\eta$ and $p_T$), and then the jet constituents are reclustered with the $k_t$ algorithm to enable use of the $k_t$-ordered splitting scales described in section 1.2.2. The four-momentum recombination scheme is used in all cases and the jet finding is performed in rapidity-azimuthal angle ($y-\phi$) coordinates. Jet selections and corrections are made in pseudorapidity-azimuthal angle ($\eta-\phi$) coordinates.

1.2.2 Jet properties and substructure observables

Three observables are used throughout these studies to characterize jet substructure and distinguish massive boosted objects from gluons or light quarks: mass, $k_t$ splitting scales, and $N$-subjettiness.

**Jet mass:** the jet mass is defined as the mass deduced from the four-momentum sum of all jet constituents. Depending on the input to the jet algorithm (see section 3.1), the constituents may be considered as either massive or massless four-momenta.

**$k_t$ splitting scales:** the $k_t$ splitting scales [2] are defined by reclustering the constituents of a jet with the $k_t$ recombination algorithm, which tends to combine the harder constituents last. At the final step of the jet recombination procedure, the $k_t$ distance measure, $d_{ij}$, for the two remaining proto-jets (intermediate jet-like objects at each stage of clustering), referred to as subjets in this case, can be used to define a splitting scale variable as:

$$\sqrt{d_{ij}} = \min(p_{T_i}, p_{T_j}) \times \Delta R_{ij},$$

where $\Delta R_{ij}$ is the distance between the two subjets in $\eta-\phi$ space. With this definition, the subjets identified at the last step of the reclustering in the $k_t$ algorithm provide the $\sqrt{d_{12}}$ observable. Similarly, $\sqrt{d_{23}}$ characterizes the splitting scale in the second-to-last step of the reclustering. The parameters $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ can be used to distinguish heavy-particle decays, which tend to be reasonably symmetric when the decay is to like-mass particles, from the largely asymmetric splittings that originate from QCD radiation in light-quark or gluon jets. The expected value for a two-body heavy-particle decay is approximately $\sqrt{d_{12}} \approx m_{\text{particle}}/2$, whereas jets from the parton shower of gluons and light quarks tend to have smaller values of the splitting scales and to exhibit a steeply falling spectrum for both $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ (see for example figure 30).
**N-subjettiness:** the $N$-subjettiness variables $\tau_N$ \cite{18, 51} are observables related to the jet multiplicity. The $\tau_N$ variable is calculated by clustering the constituents of the jet with the $k_t$ algorithm and requiring that exactly $N$ subjets be found. This is done using the exclusive version of the $k_t$ algorithm \cite{48} and is based on reconstructing clusters of particles in the jet using all of the jet constituents. These $N$ final subjets define axes within the jet. The variables $\tau_N$ are then defined in eq. (1.3) as the sum over all constituents $k$ of the jet, such that

$$\tau_N = \frac{1}{d_0} \sum_k p_{T_k} \times \min(\delta R_{1k}, \delta R_{2k}, \ldots, \delta R_{Nk}), \quad \text{with} \quad d_0 \equiv \sum_k p_{T_k} \times R$$

where $R$ is the jet radius parameter in the jet algorithm, $p_{T_k}$ is the $p_T$ of constituent $k$ and $\delta R_{ik}$ is the distance from subjet $i$ to constituent $k$. Using this definition, $\tau_N$ describes how well jets can be described as containing $N$ or fewer $k_t$ subjets by assessing the degree to which constituents are aligned with the axes of these subjets for a given hypothesis $N$. The ratios $\tau_2/\tau_1$ and $\tau_3/\tau_2$ can be used to provide discrimination between jets formed from the parton shower of initial-state gluons or light-quarks and jets formed from two hadronic decay products (from $Z$-bosons, for example) or three hadronic decay products from boosted top quarks. These ratios are herein referred to as $\tau_{21}$ and $\tau_{32}$ respectively. For example, $\tau_{21} \simeq 1$ corresponds to a jet that is very well described by a single subjet whereas a lower value implies a jet that is much better described by two subjets than one.

### 1.2.3 Jet grooming algorithms

Three jet grooming procedures are studied in this paper. Mass-drop filtering, trimming, and pruning are described, and performance measures related to each are defined. The different configurations of the grooming algorithms described in this section are summarized in table 1. Additionally, a technique to tag boosted top quarks using the mass-drop filtering method is introduced. Unless otherwise specified, the jet $p_T$ reported for a groomed jet is that which is calculated after the grooming algorithm is applied to the original jet.

**Mass-drop filtering:** the mass-drop filtering procedure\(^2\) seeks to isolate concentrations of energy within a jet by identifying relatively symmetric subjets, each with a significantly smaller mass than that of the original jet. This technique was developed and optimized using C/A jets in the search for a Higgs boson decaying to two $b$-quarks: $H \rightarrow bb$ \cite{9}. The C/A algorithm is used because it provides an angular-ordered shower history that begins with the widest combinations when reversing the cluster sequence. This provides useful information regarding the presence of potentially large splittings within a jet (see section 4 and section 5). Although the mass-drop criterion and subsequent filtering procedure are not based specifically on soft-$p_T$ or wide-angle selection criteria, the algorithm does retain the hard components of the jet through the requirements placed on its internal structure. The first measurements of the jet mass of these filtered jets was performed using 35 pb\(^{-1}\) of data collected in 2010 by the ATLAS experiment \cite{25}. The mass-drop filtering procedure has two stages:

\(^2\)In this paper, mass-drop filtering is often shortened to only filtering in figures and captions.
• Mass-drop and symmetry. The last stage of the C/A clustering is undone. The jet “splits” into two subjets, \(j_1\) and \(j_2\), ordered such that the mass of \(j_1\) is larger: \(m_{j_1} > m_{j_2}\). The mass-drop criterion requires that there be a significant difference between the original jet mass \(m_{\text{jet}}\) and \(m_{j_1}\) after the splitting:

\[
m_{j_1}/m_{\text{jet}} < \mu_{\text{frac}},
\]

where \(\mu_{\text{frac}}\) is a parameter of the algorithm. The splitting is also required to be relatively symmetric, which is approximated by the requirement that

\[
\min[(p_T^{j_1})^2, (p_T^{j_2})^2] \times \Delta R_{j_1,j_2}^2 > y_{\text{cut}},
\]

where \(\Delta R_{j_1,j_2}\) is a measure of the opening angle between \(j_2\) and \(j_1\), and \(y_{\text{cut}}\) defines the energy sharing between the two subjets in the original jet. For the analyses presented here, \(y_{\text{cut}}\) is set to 0.09, the optimal value for identifying two-body decays, obtained in previous studies \cite{9}. To give a sense of the kinematic requirements that this places on a given decay, consider a hadronically decaying W boson with \(p_T^W \approx 200\) GeV. According to the approximation given by eq. (1.1), the average angular separation of the two daughter quarks is \(\Delta R_{j_1,j_2} \sim 0.8\). The symmetry requirement determined by \(y_{\text{cut}}\) in eq. (1.5) thereby implies that the transverse momentum of the softer (in \(p_T\)) of the two subjets is greater than approximately 30 GeV. Generally, this requirement entails a minimum \(p_T\) of the softer subjet of \(p_T^{\text{subjet}} / p_T^{\text{jet}} > 0.15\), thus forcing both subjets to carry some significant fraction of the momentum of the original jet. This procedure is illustrated in figure 3(a). If the mass-drop and symmetry criteria are not satisfied, the jet is discarded.

• Filtering. The constituents of \(j_1\) and \(j_2\) are reclustered using the C/A algorithm with radius parameter \(R_{\text{filt}} = \min[0.3, \Delta R_{j_1,j_2}/2]\), where \(R_{\text{filt}} < \Delta R_{j_1,j_2}\). The jet is then filtered; all constituents outside the three hardest subjets are discarded. The choice of three allows one additional radiation from a two-body decay to be captured. In isolating \(j_1\) and \(j_2\) with the C/A algorithm, the angular scale of any potential massive particle decay is known. By dynamically reclustering the jet at an appropriate angular scale able to resolve that structure, the sensitivity to highly collimated decays is maximized. This is illustrated in figure 3(b).

In this paper, three values of the mass-drop parameter \(\mu_{\text{frac}}\) are studied, as summarized in table 1. The values chosen for \(\mu_{\text{frac}}\) are based on a previous study \cite{9} which has shown that \(\mu_{\text{frac}} = 0.67\) is optimal in discriminating \(H \to b\bar{b}\) from background. A subsequent study regarding the factorization properties of several groomed jet algorithms \cite{52} found that smaller values of \(\mu_{\text{frac}}\) (0.20 and 0.33) are similarly effective at reducing backgrounds, and yet they remain factorizable within the soft collinear effective theory studied in that analysis.
Trimming: the trimming algorithm \cite{12} takes advantage of the fact that contamination from pile-up, multiple parton interactions (MPI) and initial-state radiation (ISR) in the reconstructed jet is often much softer than the outgoing partons associated with the hard-scatter and their final-state radiation (FSR). The ratio of the $p_T$ of the constituents to that of the jet is used as a selection criterion. Although there is some spatial overlap, removing the softer components from the final jet preferentially removes radiation from pile-up, MPI, and ISR while discarding only a small part of the hard-scatter decay products and FSR. Since the primary effect of pile-up in the detector is additional low-energy deposits in clusters of calorimeter cells, as opposed to additional energy being added to already existing clusters produced by particles originating from the hard scattering process, this allows a relatively simple jet energy offset correction for smaller radius jets ($R = 0.4, 0.6$) as a function of the number of primary reconstructed vertices \cite{53}.

The trimming procedure uses a $k_t$ algorithm to create subjets of size $R_{sub}$ from the constituents of a jet. Any subjets with $p_{Ti}/p_{T}\text{jet}^\text{cut}$ are removed, where $p_{Ti}$ is the transverse momentum of the $i$th subjet, and $f_{\text{cut}}$ is a parameter of the method, which is typically a few percent. The remaining constituents form the trimmed jet. This procedure is illustrated in figure 4. Low-mass jets ($m_{\text{jet}} < 100 \text{ GeV}$) from a light-quark or gluon lose typically 30–50% of their mass in the trimming procedure, while jets containing the decay products of a boosted object lose less of their mass, with most of the reduction due to the removal of pile-up or UE (see, for example, figures 29 and 32). The fraction removed increases with the number of $pp$ interactions in the event.
Six configurations of trimmed jets are studied here, arising from combinations of $f_{\text{cut}}$ and $R_{\text{sub}}$, given in table 1. They are based on the optimized parameters in ref. [12] ($f_{\text{cut}} = 0.03, R_{\text{sub}} = 0.2$) and variations suggested by the authors of the algorithm. This set represents a wide range of phase space for trimming and is somewhat broader than considered in ref. [12].

**Pruning:** the pruning algorithm [3, 11] is similar to trimming in that it removes constituents with a small relative $p_T$, but it additionally applies a veto on wide-angle radiation. The pruning procedure is invoked at each successive recombination step of the jet algorithm (either C/A or $k_t$). It is based on a decision at each step of the jet reconstruction whether or not to add the constituent being considered. As such, it does not require the reconstruction of subjets. For all studies performed for this paper, the $k_t$ algorithm is used in the pruning procedure. This results in definitions of the terms *wide-angle* or *soft* that are not directly related to the original jet but rather to the proto-jets formed in the process of rebuilding the pruned jet.

The procedure is as follows:

- The C/A or $k_t$ recombination jet algorithm is run on the constituents, which were found by any jet finding algorithm.

- At each recombination step of constituents $j_1$ and $j_2$ (where $p_T^{j_1} > p_T^{j_2}$), either $p_T^{j_2}/p_T^{j_1+j_2} > z_{\text{cut}}$ or $\Delta R_{j_1,j_2} < R_{\text{cut}} \times (2m_{\text{jet}}/p_T^{\text{jet}})$ must be satisfied. Here, $z_{\text{cut}}$ and $R_{\text{cut}}$ are parameters of the algorithm which are studied in this paper.
Jet finding algorithms used | Grooming algorithm | Configurations considered
--- | --- | ---
C/A | Mass-Drop Filtering | $\mu_{\text{frac}} = 0.20, 0.33, 0.67$
Anti-$k_t$ and C/A | Trimming | $f_{\text{cut}} = 0.01, 0.03, 0.05$
 |  | $R_{\text{sub}} = 0.2, 0.3$
Anti-$k_t$ and C/A | Pruning | $R_{\text{cut}} = 0.1, 0.2, 0.3$
C/A | HEPTopTagger | $z_{\text{cut}} = 0.05, 0.1$

Table 1. Summary of the grooming configurations considered in this study. Values in boldface are optimized configurations reported in ref. [9] and ref. [12] for filtering and trimming, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Default</th>
<th>Tight</th>
<th>Loose</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>1.5 or 1.8</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>$m_{\text{cut}}$ [GeV]</td>
<td>30</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>$R_{\text{jet}}$</td>
<td>0.3</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>$N_{\text{subjet}}$</td>
<td>5</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>$f_{W}[%]$</td>
<td>15</td>
<td>10</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 2. The settings used for studying the performance of the HEPTopTagger.

- $j_2$ with $j_1$ are merged if one or both of the above criteria are met, otherwise, $j_2$ is discarded and the algorithm continues.

The pruning procedure is illustrated in figure 5. Six configurations, given in table 1, based on combinations of $z_{\text{cut}}$ and $R_{\text{cut}}$ are studied here. This set of parameters also represents a relatively wide range of possible configurations.

### 1.2.4 HEPTopTagger

The HEPTopTagger algorithm [32, 54] is designed to identify a top quark with a hadronically decaying $W$ boson daughter over a large multi-jet background. The method uses the C/A jet algorithm and a variant of the mass-drop filtering technique described in section 1.2.3 in order to exploit information about the recombination history of the jet. This information is used to search for evidence within the jet of the presence of $W$ decay products as well as an additional energy deposition — the $b$-quark — that are consistent with the $W$ and top masses and the expected angular distribution of the final-state quarks. The HEPTopTagger algorithm is optimized for top-quark transverse momentum as low as 200 GeV and therefore uses a correspondingly large jet radius parameter. The algorithm proceeds as follows and is illustrated in figure 6.

**Decomposition into substructure objects:** The mass-drop criterion defined in eq. (1.4) is applied to a large-$R$ C/A jet, where $j_1$ and $j_2$ are the two subjets from the last
stage of clustering, with \( m^{j_1} > m^{j_2} \). If the criterion is satisfied, the same prescription is followed to split both \( j_1 \) and \( j_2 \) further. The iterative splitting continues until the subjets either have masses \( m_i \) less than a tunable parameter \( m_{\text{cut}} \), or represent individual constituents, such as calorimeter energy deposits, tracks, or generator-level particles (i.e. no clustering history); see section 3.1 for definitions. This procedure results in \( N_i \) subjets. If at any stage \( m^{j_1} > (m^{\text{jet}} \times \mu_{\text{frac}}) \), the mass-drop criterion and subsequent iterative declustering is not applied to \( j_2 \). The values of \( m_{\text{cut}} \) and \( R \) studied are summarized in table 2. \( R \) values of 1.5 and 1.8, somewhat larger than used generally in mass-drop filtering, are chosen based on previous studies [32, 55]. When the iterative process of declustering the jet is complete, there must exist at least three subjets (\( N_i \geq 3 \)), otherwise the jet is discarded.

**Filtering:** all possible combinations of three subjets are formed, and each triplet is filtered one at a time. The constituents of the subjets in a given triplet are reclustered into \( N_j \) new subjets using the C/A algorithm with a size parameter \( R_{\text{filt}} = \min[0.3, \Delta R_{j_1,j_2}/2] \), where \( \Delta R_{j_1,j_2} \) is the minimum separation between all possible pairs in the current triplet. It is therefore possible that \( N_j > 3 \) after the reclustering step. All energy deposits not in the \( N_j \) subjets are discarded.

**Top mass window requirement:** if the invariant mass of the four-momentum determined by summing the constituents of the \( N_j \) subjets is not in the range \( 140 \text{ GeV} \leq m^{\text{jet}} < 200 \text{ GeV} \), the triplet combination is ignored. If more than one triplet satisfies the criterion, only the one with mass closest to the top-quark mass, \( m_t \), is used. This triplet (which consists of \( N_j \geq 3 \) subjets) is thus identified as the top-candidate triplet.

**Reclustering of subjets:** from the \( N_j \) subjets of the top-candidate triplet, \( N_{\text{subjet}} \) leading-\( p_T \) subjets are chosen, where \( N_{\text{subjet}} \) is a parameter satisfying \( 3 \leq N_{\text{subjet}} \leq N_j \). From this set of subjets, exactly three jets are built by re-applying the C/A algorithm to the constituents of the \( N_{\text{subjet}} \) subjets, which are exclusively clustered using a distance parameter \( R_{\text{jet}} \) listed in table 2. This latter step reflects the hypothesis that this is likely to be a top-jet candidate. These subjets are calibrated as described in section 3.5.

**W boson mass requirements:** relations listed in eqs. (A.1) of ref. [32] are defined using the total invariant mass of the three subjets \( (m_{123}) \) and the invariant mass \( m_{ij} \) formed from combinations of two of the three subjets ordered in \( p_T \). These conditions include:

\[
R_- < \frac{m_{23}}{m_{123}} < R_+
\]

\[
0.2 < \arctan \frac{m_{13}}{m_{12}} < 1.3.
\]

Here, \( R_{\pm} = (1 \pm f_W)(m_W/m_t) \), \( f_W \) is a resolution variable (given in table 2), and the quantities \( m_W \) and \( m_t \) denote the W boson and top-quark masses, respectively. If at least one of the criteria in eqs. (A.1) of [32] is met, the four-momentum addition of the three subjets is considered a candidate top quark.
(a) Every object encountered in the declustering process is considered a ‘substructure object’ if it is of sufficiently low mass or has no clustering history.

(b) The mass-drop criterion is applied iteratively, following the highest subjet-mass line through the clustering history, resulting in $N_i$ substructure objects.

(c) For every triplet-wise combination of the substructure objects found in (b), recluster the constituents into subjets and select the $N_{\text{subjet}}$ leading-$p_T$ subjets, with $3 \leq N_{\text{subjet}} \leq N_i$ (here, $N_{\text{subjet}} = 5$).

(d) Recluster the constituents of the $N_{\text{subjet}}$ subjets into exactly three subjets to make the top candidate for this triplet-wise combination of substructure objects.

Figure 6. The HEPTopTagger procedure.

2 The ATLAS detector and data samples

2.1 The ATLAS detector

The ATLAS detector [56, 57] provides nearly full solid angle coverage around the collision point with an inner tracking system covering $|\eta| < 2.5$, electromagnetic and hadronic calorimeters covering $|\eta| < 4.9$, and a muon spectrometer covering $|\eta| < 2.7$. Of the multiple ATLAS subsystems, the most relevant to this analysis are the barrel and endcap calorimeters [58, 59] and the trigger [60].

The calorimeter comprises multiple sub-detectors of various designs, spanning the pseudorapidity range up to $|\eta| = 4.9$. The measurements presented here are performed using data predominantly from the central calorimeters, comprising the liquid argon (LAr) barrel electromagnetic calorimeter ($|\eta| < 1.475$) and the tile hadronic calorimeter ($|\eta| < 1.7$). Three additional calorimeter subsystems are located in the higher-$\eta$ regions of the detector: the LAr electromagnetic endcap calorimeter, the LAr hadronic endcap calorimeter, and the forward calorimeter with separate components for electromagnetic and hadronic showers.

Dedicated trigger and data acquisition systems are responsible for the online event selection, which is performed in three stages: Level 1, Level 2, and the Event Filter. The measurements presented in this paper rely primarily on the single-jet and multi-jet triggers implemented at the Event Filter level, which has access to the full detector granularity, and finds multi-jet events with high efficiency. The intermediate trigger levels provide coarser jet finding and sufficient rate reduction to satisfy the trigger and offline selection requirements.
2.2 Data and Monte Carlo samples

Data from the entire 2011 ATLAS data-taking period are used, corresponding to $4.7 \pm 0.1 \text{ fb}^{-1}$ of integrated luminosity [61]. All data are required to have met baseline quality criteria and were taken during periods in which almost all of the detector was fully functional. Data quality criteria reject events with significant contamination from detector noise or with issues in the read-out, and are based on assessments for each subdetector individually. Multiple proton-proton collisions, or pile-up, result in several reconstructed primary vertices per event. The inclusive jet sample that is used for many studies in this paper is selected from the data using a single high-$p_T$ jet trigger that requires the leading jet in the event to have $p_T^{\text{jet}} > 350 \text{ GeV}$. This trigger threshold was used for the entire 2011 data-taking period and thus represents the full integrated luminosity with negligible inefficiency.

These data are compared to inclusive jet events, which are dominated by light-quark or gluon jets whose properties are defined primarily by soft gluon emission, that are generated by three Monte Carlo (MC) simulation programs: PYTHIA 6.425 [10], HERWIG++ [62], and POWHEG-BOX 1.0 [63–65] (patch 4) interfaced to PYTHIA 6.425 for the parton shower, hadronization, and UE models. Both PYTHIA and HERWIG++ use the modified-LO parton distribution function (PDF) set MRST LO* [66]. POWHEG+PYTHIA uses the CT10 NLO PDF [67] in the matrix element and CTEQ6L1 PDF set [68] for the PYTHIA parton shower. For both cases, PYTHIA is used with the corresponding ATLAS AUET2B tune [69, 70] and HERWIG++ uses the so-called UE7-2 tune [71], which is tuned to UE data from the LHC experiments. PYTHIA or HERWIG++ with POWHEG+PYTHIA provide an important comparison, at least at the matrix-element level, between leading-order (LO) (PYTHIA and HERWIG++) and next-to-leading-order (NLO) (POWHEG) calculations. Furthermore, PYTHIA and HERWIG++ offer distinct approaches to the modelling of the parton shower, hadronization, and the UE.

Samples of $t\bar{t}$ events are generated with MC@NLO v4.01 [72] using the CT10 NLO PDF, interfaced to HERWIG v6.520 [73] and JIMMY v4.31 [74]. Alternative samples for the study of systematic uncertainties are generated with POWHEG, with showering provided by either HERWIG or PYTHIA. Samples generated with the AcerMC v3.8 [75] package, using CTEQ6L1 PDFs, with showering provided by PYTHIA are also used. In these samples PYTHIA parameters have been tuned to increase or decrease the amount of initial- and final-state radiation. Single-top-quark events in the s-channel and $Wt$ processes are also generated with MC@NLO using the CT10 NLO PDF set, with only leptonically decaying $W$ bosons allowed in the final state. Single-top-quark events in the t-channel, where all $W$ boson decay channels are produced, are generated with AcerMC using the CTEQ6L1 PDF set, and are showered using PYTHIA with the AUET2B tune.

Samples of $W$+jets and $Z$+jets events are produced with the ALPGEN v2.13 [76] generator, using CTEQ6L1 PDFs, interfaced to HERWIG for parton showering and hadronization. Samples of diboson production processes ($WW$, $WZ$ and $ZZ$) are produced with the HERWIG generator.
After simulation of the parton shower and hadronization, as well as of the UE, events are passed through the full Geant4 [77] detector simulation [78]. Following this, the same trigger, event, data quality, jet, and track selection criteria are applied to the Monte Carlo simulation events as are applied to the data.

Boosted particles decaying to hadrons are used for direct comparisons of the performance of the various reconstruction and jet substructure techniques. For two-pronged decays, a sample of hadronically decaying Z bosons is generated using the HERWIG v6.510 [73] event generator interfaced with JIMMY v4.31 [74] for the UE. A sample of hadronically decaying W bosons produced using the same configuration as for the Z boson sample is also used for comparisons of the HEPTopTagger performance in section 5.3.3. In order to test the performance of techniques designed for three-pronged decays, t¯t events from a non-Standard-Model heavy gauge boson (Z′ with m_{Z′} = 1.6 TeV) are generated using the same PYTHIA 6.425 tune as above. This model provides a relatively narrow t¯t resonance and top quarks with p_T ≃ 800 GeV.

Pile-up is simulated by overlaying additional soft pp collisions, or minimum bias events, which are generated with PYTHIA 6.425 using the ATLAS AUET2B tune [70] and the CTEQ6L1 PDF set. The minimum bias events are overlaid onto the hard scattering events according to the measured distribution of the average number ⟨μ⟩ of pp interactions. The proton bunches were organized in trains of 36 bunches with a 50 ns spacing between the bunches. Therefore, the simulation also contains effects from out-of-time pile-up, i.e. contributions from the collision of bunches neighbouring those where the events of interest occurred. Simulated events are reweighted such that the MC distribution of ⟨μ⟩ agrees with the data, as measured by the luminosity detectors in ATLAS [61].

3 Jet reconstruction and calibration

3.1 Inputs to jet reconstruction

The inputs to jet reconstruction are either stable particles with a lifetime of at least 10 ps (excluding muons and neutrinos) in the case of MC generator-level jets (also referred to as particle jets), charged particle tracks in the case of so-called track-jets [53], or three-dimensional topological clusters (topo-clusters) in the case of fully reconstructed calorimeter-jets. Stable particles, such as pions or protons in the simulation, retain their respective masses when input to the jet reconstruction algorithm. Tracks are assigned the pion mass when used as input to the jet algorithm. Quality selections are applied in order to ensure that good quality tracks that originate from the reconstructed hard scattering vertex are used to build track-jets. The hard scattering vertex is selected as the primary vertex that has the largest \( \sum (p_T^{\text{track}})^2 \) in the event and that contains at least two tracks. The selection criteria are:

- transverse momentum: \( p_T^{\text{track}} > 0.5 \) GeV;
- transverse impact parameter: \( |d_0| < 1.0 \) mm;
- longitudinal impact parameter: \( |z_0| \times \sin(\theta) < 1.0 \) mm;
silicon detector hits on tracks: hits in pixel detector ≥ 1 and in the silicon strip detector ≥ 6;

where the impact parameters are computed with respect to the hard scattering vertex, and θ is the angle between the track and the beam. In the reconstruction of calorimeter jets, calorimeter cells are clustered together using a three-dimensional topological clustering algorithm that includes noise suppression [79]. The resulting topo-clusters are considered as massless four-momenta, such that \( E = |\vec{p}| \). They are classified as either electromagnetic or hadronic based on their shape, depth and energy density. In the calibration procedure, corrections are applied to the energy in order to calibrate the clusters to the hadronic scale.

### 3.2 Jet quality criteria and selection

All jets in the event reconstructed with the anti-\( k_t \) algorithm with \( R = 0.4 \) and a measured \( p_T^{\text{jet}} > 20 \text{ GeV} \) are required to satisfy the looser requirements discussed in detail in ref. [80]. These selections are designed to retain good quality jets while rejecting as large a fraction as possible of those from non-collision beam background and calorimeter noise. Jets are required to deposit at least 5\% of their measured total energy in the electromagnetic (EM) calorimeter as well as not more than 99\% of their energy in a single calorimeter layer.

To prevent contamination from detector noise, these jet quality criteria are extended by several requirements applied in a specific detector region. Any event with an anti-\( k_t \) \( R = 0.4 \) jet with \( p_T^{\text{jet}} > 20 \text{ GeV} \) that fails the above non-collision beam background or noise rejection requirements is removed from the analysis.

### 3.3 Jet calibration and systematic uncertainties

The precision and accuracy of energy measurements made by the calorimeter system are integral to every physics analysis, and the procedures to calibrate jets are described in ref. [53]. The baseline energy scale of the calorimeters derives from the calibration of the electronic signal arising from the energy deposited by electromagnetic showers measured in beam tests, known as the electromagnetic scale. The hadronic calorimeter has been calibrated with electrons, pions, and muons in beam tests and the energy scale has been validated using muons produced by cosmic rays with the detector in situ in the experimental hall [59]. The invariant mass of the Z boson in \( Z \to ee \) events measured in situ in the same data sample studied here is used to adjust the calibration for the EM calorimeters.

#### 3.3.1 Monte Carlo based calibration

The MC hadronic calibration scheme starts from the measured calorimeter energy at the electromagnetic (EM) energy scale [81–89], which correctly measures the energy deposited by electromagnetic showers. A local cluster weighting (LCW) calibration method classifies topological clusters along a continuous scale as being electromagnetic or hadronic, using shower shapes and energy densities. Energy corrections are applied to hadronic clusters based on this classification scheme, which is derived from single pion MC simulations and tested in situ using beam tests. These corrections account for the effects of non-compensation, signal losses due to noise suppression and out-of-cluster effects in building
topo-clusters, and energy lost in non-instrumented regions of the calorimeters. The results shown here use LCW clusters as input to the jet algorithm.

The final jet energy scale (JES) calibration is derived as a correction relating the calorimeter’s response to the true jet energy. The jet energy response is defined as the ratio of the reconstructed calorimeter jet energy (or transverse energy) to that of its matched truth particle jet energy. The inverse of this response ratio defines the multiplicative jet energy scale factor that is dependent on the jet energy and pseudorapidity. This JES factor is then applied to the full jet four-momentum. It can be applied to EM scale jets, with the resulting calibrated jets referred to as EM+JES, or to LCW calibrated jets, with the resulting jets referred to as LCW+JES jets. More details regarding the evaluation and validation of this approach for standard anti-$k_t \ R = 0.4, 0.6$ jets can be found in ref. [53].

The JES correction used here for large-$R$ jets is derived from a PYTHIA MC sample including pile-up events. There is no explicit offset correction for pile-up contributions, as in the standard JES procedure [53]. For standard jet algorithms, the dependence of the jet response on the number of primary vertices ($N_{PV}$) and the average number of interactions ($\langle \mu \rangle$) is removed by applying a pile-up offset correction to the EM or LCW scale before applying the JES correction. However, no explicit pile-up correction is applied to large-$R$ jets or to jets with the various grooming algorithms applied.

Since one of the primary goals of the use of large-$R$ and groomed jet algorithms is to reconstruct the masses of jets accurately and precisely, a last step is added to the calibration procedure of large-$R$ jets wherein the mass of the jet is calibrated based on the MC simulation of dijet events. An explicit jet mass calibration is important when using the individual invariant jet mass in physics analyses since it is particularly susceptible to soft, wide-angle contributions that do not otherwise significantly impact the jet energy scale. The procedure measures the jet mass response for jets built from LCW clusters after the standard JES calibration. The mass response is determined from the mean of a Gaussian fit to the core of the distribution of the reconstructed jet mass divided by the corresponding generator-level jet mass.

Figure 7 shows the jet mass response ($m_{\text{reco}}/m_{\text{true}}$) for several values of jet energy as a function of $\eta$ for anti-$k_t \ R = 1.0$ jets, before and after calibration to the true jet mass and without jet grooming. In each case, the jet energy itself has been calibrated by applying the JES correction. One can see from this figure that even very high-energy jets near the central part of the detector can have a mean mass scale (or JMS) differing by up to 20% from the particle level true jet mass. In particular, the reconstructed mass is, on average, greater than that of the particle-level jet due in part to noise and pile-up in the detectors. Furthermore, the finite resolution of the detector has a differential impact on the mass response as a function of $\eta$. Following the jet mass calibration, performed also as a function of $\eta$, a uniform mass response can be restored within 3\% across the full energy and $\eta$ range. Results for C/A jets with $R = 1.2$ are similar, but with an additional $\sim 1\%$ non-closure due to the increased contamination from pile-up, as these jets have a larger distance parameter compared to the anti-$k_t \ R = 1.0$ jets. After applying grooming to the C/A jets, such as the mass-drop filtering procedure, the pile-up contamination is reduced and the jet mass response variation is reduced to less than 3\%.
Figure 7. Mass response ($m_{\text{reco}}/m_{\text{true}}$) (a) before and (b) after mass calibration for ungroomed anti-$k_t$, $R=1.0$ jets. The dotted lines shown in (b) represent a ±3% envelope on the precision of the final jet mass scale calibration. In each case, the jet energy itself has been calibrated by applying the JES correction.

3.3.2 Jet mass scale validation in inclusive jet events using track-jets

In order to validate the jet mass measurement made by the calorimeter, calorimeter-jets are compared to track-jets. Track-jets have a different set of systematic uncertainties and allow a reliable determination of the relative systematic uncertainties associated with the calorimeter-based measurement. Performance studies [90] have shown that there is excellent agreement between the measured positions of clusters and tracks in data, indicating no systematic misalignment between the calorimeter and the inner detector.

The use of track-jets reduces or eliminates the impact of additional $pp$ collisions by requiring the jet inputs (tracks) to come from the hard-scattering vertex. The inner detector and the calorimeter have largely uncorrelated instrumental systematic effects, and so a comparison of variables such as jet mass and energy between the two systems allows a separation of physics (correlated) and detector (uncorrelated) effects. It is therefore possible to validate the JES and JMS, and also to estimate directly the pile-up energy contribution to jets. This approach was used extensively in the measurement of the jet mass and substructure properties of jets in the 2010 data [25] where pile-up was significantly less important and the statistical reach of the measurement was smaller than with the full integrated luminosity of 4.7 fb$^{-1}$ for the 2011 dataset.

The relative uncertainty is determined using the ratio of the transverse momentum of the calorimeter jet, $p_T^{\text{jet}}$, to that of the track-jet, $p_T^{\text{track jet}}$. The same procedure is repeated for the jet mass, $m^{\text{jet}}$, by using the track-jet mass, $m^{\text{track jet}}$. The ratios are defined as

\[
\frac{r^{p_T}}{\text{track jet}} = \frac{p_T^{\text{jet}}}{p_T^{\text{track jet}}}, \quad \frac{r^{m}}{\text{track jet}} = \frac{m^{\text{jet}}}{m^{\text{track jet}}},
\]

where the matching between calorimeter and track-jets is performed using a matching criterion of $\Delta R < 0.3$. The mean values of these ratios are expected to be well described
by the detector simulation if detector effects are well modelled. That is to say, even if some underlying physics process is unaccounted for in the simulation, as long as this process affects both the track-jet and calorimeter-jet $p_T$ or masses in a similar way, then the ratio of data to simulation should be relatively unaffected when averaged over many events.

Double ratios of $r_{\text{track jet}}^m$ and $r_{\text{track jet}}^{P_T}$ are constructed in order to evaluate this agreement. These double ratios, $R_{P_T}^r_{\text{track jet}}$ and $R_m^r_{\text{track jet}}$, are defined as:

$$R_{P_T}^r_{\text{track jet}} = \frac{r_{P_T,\text{data}}^m_{\text{track jet}}}{r_{P_T,\text{MC}}^m_{\text{track jet}}}$$

$$R_m^r_{\text{track jet}} = \frac{r_{m,\text{data}}^m_{\text{track jet}}}{r_{m,\text{MC}}^m_{\text{track jet}}}$$

The dependence of $R_{P_T}^r_{\text{track jet}}$ and $R_m^r_{\text{track jet}}$ on $p_T$ and $m_{\text{jet}}$ provides a test of the deviation of simulation from data, thus allowing an estimate of the uncertainty associated with the Monte Carlo derived calibration.

Figure 8 shows the distribution of $r_{\text{track jet}}^m$ for two jet algorithms and for jets in the range $500 \text{ GeV} \leq p_T^{\text{jet}} < 600 \text{ GeV}$ in the central calorimeter region, $|\eta| < 0.8$. Comparisons between MC simulation and the data are made using PYTHIA, HERWIG++, and POWHEG+PYTHIA, where the distributions are normalized to the number of events observed in the data. This $p_T$ range is chosen for illustrative purposes and because of its relevance to searches for boosted vector bosons and top quarks, as the decay products of both are expected to be fully merged into a large-$R$ jet in this transverse momentum range. The peak near $r_{\text{track jet}}^m \approx 2$ and the shape of the distribution are both generally well described by the Monte Carlo simulations. Both the ungroomed and the trimmed anti-$k_t$, $R = 1.0$ distributions show some discrepancies at very low $r_{\text{track jet}}^m$, where the description of very soft radiation and hadronization is important, and at high values of $r_{\text{track jet}}^m$ above $r_{\text{track jet}}^m \gtrsim 4$. The differences are approximately 20%. However, these spectra are used
primarily to test the overall scale, so that the important comparison is of the mean values of the distributions, which are quite well described.

The relative systematic uncertainty is first estimated for each MC generator sample as the weighted average absolute deviation of the double ratio, $R_{m_{\text{track jet}}}$, from unity. Measurements of $R_{m_{\text{track jet}}}$ are performed in exclusive $p_T$ and $\eta$ ranges. The statistical uncertainty is used as the weight in this case. The final relative uncertainty is then determined by the maximum of the weighted average deviation among the MC samples considered. Comparisons are made using PYTHIA, HERWIG++, and POWHEG+PYTHIA.

Figure 9 presents the distributions of both $\langle r_{m_{\text{track jet}}} \rangle$ and the double ratio with respect to MC simulation, $\langle R_{m_{\text{track jet}}} \rangle$, for the same algorithms and grooming configurations as shown in figure 8. In the peak of the jet mass distribution, logarithmic soft terms dominate [8] and lower-$p_T$ particles constitute a large fraction of the calorimeter-jet mass. These particles are bent more by the magnetic field than higher-$p_T$ particles, or are not reconstructed as charged-particle tracks, and thus contribute more to the calorimeter-jet mass than the track-jet mass. At much lower calorimeter-jet masses, charged particles can be completely bent out of the jet acceptance, thus reducing the calorimeter-jet mass for a fixed track-jet mass. These effects result in the shape observed in the $r_{m_{\text{track jet}}}$ distribution in this region. Higher-mass jets tend to be composed of multiple higher-$p_T$ particles that are less affected by the magnetic field and therefore contribute more similarly to the calorimeter-based and track-based mass reconstruction. This results in a flatter and fairly stable $r_{m_{\text{track jet}}}$ ratio. This flat $r_{m_{\text{track jet}}}$ distribution is present across the mass range for trimmed and mass-drop filtered (not shown) jet masses, as both these algorithms are designed to remove softer particles. Although there is a difference in the phase space of
emissions probed at low and high mass, the mean calorimeter response relative to the tracker response is well modelled by each of the three MC simulations.

The weighted average deviation of $R_{\text{track}}^{p_{T}}$ from unity ranges from approximately 2% to 4% for the set of jet algorithms and grooming configurations tested for jets in the range $500 \text{ GeV} \leq p_{T}^\text{jet} < 600 \text{ GeV}$ and in the central calorimeter, $|\eta| < 0.8$. The results are fairly stable for the slightly less central $\eta$ range $0.8 \leq |\eta| < 1.2$.

Figure 10 presents the full set of jet mass scale systematic uncertainties for various jet algorithms estimated using the calorimeter-to-track-jet double ratios. The total relative

---

**Figure 10.** Summary of the jet mass scale (JMS) relative systematic uncertainties as a function of $p_{T}^\text{jet}$. These uncertainties are determined from track-jet double ratios. For each jet algorithm (a) anti-$k_{t}$, $R = 1.0$ without trimming, (b) anti-$k_{t}$, $R = 1.0$ with trimming ($f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.3$), (c) C/A, $R = 1.2$ without filtering, and (d) C/A, $R = 1.2$ with filtering, the two lines shown represent the uncertainty evaluated at $|\eta| = 0.0$ (solid) and $|\eta| = 1.0$ (dashed). These estimates include a 3% relative non-closure uncertainty on the MC-based mass scale calibration factors, as well as systematic uncertainties on the impact of the tracking efficiency on the track measurements.
uncertainty includes the 3% uncertainty on the precision of the jet mass scale calibration (see figure 7(b)) as well as the uncertainty on the track measurements themselves. The latter uncertainty takes into account the knowledge of tracking inefficiencies and their impact on the $p_{T}^{\text{jet}}$ and $m_{\text{jet}}$ measurements using track-jets. Each of these two additional components is assumed to be uncorrelated and added in quadrature with the uncertainty determined solely from the calorimeter-to-track-jet double ratios. The uncertainties are smoothly interpolated between the multiple discrete $\eta$ ranges in which they are estimated. Figure 10 shows the uncertainty evaluated at two points $|\eta| = 0.0$ (solid) and $|\eta| = 1.0$ (dashed) as a function of $p_{T}^{\text{jet}}$.

The impact of the tracking efficiency systematic uncertainty on $r_{\text{track jet}}^{m}$ is evaluated by randomly rejecting tracks used to construct track-jets according to the efficiency uncertainty. This is evaluated as a function of $\eta$ and $m_{\text{jet}}$ for various $p_{T}^{\text{jet}}$ ranges. Typically, this results in a 2–3% shift in the measured track-jet kinematics (both $p_{T}$ and mass) and thus a roughly 1% contribution to the resulting total uncertainty, since the tracking uncertainty is taken to be uncorrelated to that determined from the double ratios directly.

The total systematic uncertainty on the jet mass scale is fairly stable near 4–5% for all jet algorithms up to $p_{T}^{\text{jet}} \approx 800$ GeV. At low $p_{T}^{\text{jet}}$, in the range $200$ GeV $\leq p_{T}^{\text{jet}} < 300$ GeV, the average uncertainty for some jet algorithms rises to approximately 5–7%. The estimated uncertainty is similar for both the ungroomed and the trimmed or mass-drop filtered jets, except for trimmed anti-$k_{t}$ jets (see figure 10(b)) for which the uncertainty in the range $900$ GeV $\leq p_{T}^{\text{jet}} < 1000$ GeV and $|\eta| = 1.0$ is approximately 8%.

### 3.3.3 Jet mass scale validation using hadronic W decays in $t\bar{t}$ events

An alternative approach to validating the jet mass scale is to study a hadronically decaying particle with a known mass. The most accessible source of hadronically decaying massive particles is events containing top-quark pairs ($t\bar{t}$). The $t\bar{t}$ process at the LHC has a relatively large cross-section and the final state contains two $W$ bosons. About 15% of the time, one of the $W$ bosons decays to a muon and neutrino, while the other decays to hadrons. This leads to an abundant source of events with a distinctive leptonic signature and a hadronically decaying known heavy particle.

Signal muons are defined as having $p_{T} > 25$ GeV and $|\eta| < 2.5$, as well as passing a number of quality criteria. Events are required to contain a signal muon and no additional muons or electrons. The missing transverse momentum ($E_{T}^{\text{miss}}$) is calculated as the negative of the vector sum of the transverse momenta of all physics objects, at the appropriate energy scale, and the transverse momentum of any remaining topo-clusters not associated with physics objects in the event. Events are required to have $E_{T}^{\text{miss}} > 25$ GeV.

In events passing this leptonic selection, a $W$ candidate is constructed from the signal muon and $E_{T}^{\text{miss}}$. In order to reject multi-jet background, this candidate is required to have transverse mass ($m_{T}$) greater than 40 GeV.3

Boostered hadronically decaying $W$ boson candidates are defined as single large-$R$ jets with $p_{T} > 200$ GeV. The hadronic and leptonic $W$ candidates are required to be separated

---

3Transverse mass $m_{T}$ is defined as $\sqrt{E_{T}^{2} - p_{T}^{2}}$ of the vector sum of the four-momentum of the signal muon and $E_{T}^{\text{miss}}$, assumed to be due to the neutrino.
Figure 11. Jet mass distributions for jets with $p_T > 200$ GeV in events containing a $W \rightarrow \mu\nu$ candidate and a $b$-tagged anti-$k_t$ jet. The jets shown are (a) Cambridge–Aachen $R = 1.2$ after mass-drop filtering, and (b) anti-$k_t$ $R = 1.0$ after trimming.

by at least 1.2 radians in $\phi$ to minimize potential overlap between the decay products of the two $W$ bosons. In order to further enhance the fraction of top-quark production processes, an anti-$k_t$ jet with $R = 0.4$, $p_T > 20$ GeV and with $\Delta R > 1.0$ to the hadronic $W$ candidate is required. This additional jet is also required to be tagged as a $b$-jet by the MV1 algorithm at the 70% efficient working point [91].

The resulting mass distributions of hadronic $W$ candidates for two jet algorithms can be seen in figure 11. A peak near the $W$ mass is clearly observed for both the mass-drop filtering and trimming algorithms in data and simulated Monte Carlo events.

A fitting procedure is used to extract the features of the jet mass distribution. The peak produced by the hadronic decays of $W$ bosons is modelled by a Voigtian function, which is the convolution of Gaussian and Lorentzian functions. In this function the width of the Lorentzian component is fixed to the world-average $W$ boson width. The width of the Gaussian function is a measure of the mass resolution although this is not explored in this study. The background shape has no simple analytic form and is assumed to be modelled by a quadratic polynomial. In order to simplify the background modelling, the $W$+jets MC prediction and multi-jet prediction are both subtracted from the data and the resulting distributions are compared to the sum of the $t\bar{t}$, single-top quark, and $WW$ MC samples. The results of this fit for Cambridge–Aachen jets can be seen in figure 12.

After fitting, the parameter controlling the mean of the Voigtian function ($\mu$) is a measure of the reconstructed mass of hadronically decaying $W$ bosons. As shown in figure 12, this scale is observed to be $\mu_{\text{data}} = 86.9 \pm 0.8$ GeV and $\mu_{\text{MC}} = 87.4 \pm 0.2$ GeV. The departure from the world-average value of $m_W = 80.385 \pm 0.015$ GeV [92] is due to clustering and detector effects, as well as physics effects such as contributions from the UE and hadroniza-

---

$4$The Lorentzian function is defined as $L(x) = \frac{1}{\pi \sigma} \frac{e^{-x-x_0}}{\left(1+(x-x_0)^2/\sigma^2\right)}$, where $x_0$ specifies the mean and $\sigma$ specifies the width.
Figure 12. Results of fitting the mass distributions of Cambridge–Aachen $R = 1.2$ jets after mass-drop filtering. The $W$+jets and multi-jet predictions have been subtracted from the distributions. The two figures show (a) the Monte Carlo prediction and (b) data. The dashed line represents the polynomial fit to the background shape, while the solid line represents the combined signal and background fit.

Figure 13. The ratio $\mu_{\text{data}}/\mu_{\text{MC}}$ as a function of jet $|\eta|$ for (a) Cambridge–Aachen $R = 1.2$ after mass-drop filtering and (b) anti-$k_t$, $R = 1.0$ after trimming.
It is also possible to repeat the fitting procedure after dividing the data and Monte Carlo simulation samples into bins of $|\eta|$. This allows a test of any potential $|\eta|$ dependence in a range outside that covered by track-jet techniques. The result of these fits can be seen in figure 13. The results extracted in individual bins show no statistically significant deviation from the average.

Furthermore, systematic uncertainties are evaluated for this technique. The Monte Carlo modelling is assessed by replacing the $t\bar{t}$ and single-top samples with alternative samples as described in section 2.2. The impact of mismodelling of the jet resolution is assessed by applying additional smearing. The effect of mismodelling of the $p_T$ scale is studied by shifting the scale in the Monte Carlo predictions. Other systematic uncertainties considered are: a possible bias in the fitting procedure; cross-section uncertainties; muon reconstruction uncertainties; energy, resolution and $b$-tagging uncertainties associated with the additional jet; $E_T^{\text{miss}}$ reconstruction uncertainties. The most significant of these are the uncertainties associated with Monte Carlo modelling and jet resolution.

The final relative scales including all uncertainties are, for Cambridge–Aachen jets ($-0.6 \pm 1.0(\text{stat})^{+1.9}_{-1.7}(\text{syst})\%$) and for anti-$k_t$ jets ($+0.5 \pm 1.2(\text{stat})^{+2.7}_{-2.7}(\text{syst})\%$). There is therefore no observed discrepancy between data and Monte Carlo simulation for jets originating from $W$ bosons with $p_T \gtrsim 200$ GeV.

3.4 In situ validation of the subjet energy scale

The trimming and mass-drop filtering procedures both rely heavily on the energy scale of subjets in order to evaluate $f_{\text{cut}}$ or $\mu_{\text{frac}}$, respectively. A similar approach based on double ratios as described in section 3.3.2 is used in order to determine the uncertainty on the energy scale of these subjets. Tracks measured by the inner detector are utilized as an independent reference with which to compare calorimeter measurements. Each subjet measured by the calorimeter has a set of tracks associated with it. The following momentum ratio is defined using the calorimeter $p_T (p_T^{\text{subjet}})$ and the track $p_T (p_T^{\text{track}})$ for each subjet:

$$r^{\text{subjet}}_{\text{trk}} = \frac{\sum p_T^{\text{track}}}{p_T^{\text{subjet}}}$$

It is useful to identify the general features of the subjet structure of large-$R$ jets in dijet events in order to guide the study of the kinematic properties of subjets. These jets are typically characterized by a highly energetic leading-$p_T^{\text{subjet}}$ subjet located close to the parent jet axis as shown in figure 14. These leading subjets (with $R_{\text{sub}} = 0.3$) carry a large fraction of the parent jet energy, and this fraction increases with the parent jet $p_T$: $p_T^{\text{subjet,lead}}/p_T^{\text{jet}} \approx 0.71$ for $100 \text{ GeV} \leq p_T^{\text{jet}} < 150$ GeV and $p_T^{\text{subjet,lead}}/p_T^{\text{jet}} \approx 0.86$ for $400 \text{ GeV} \leq p_T^{\text{jet}} < 500$ GeV. The second leading-$p_T^{\text{subjet}}$ subjet carries approximately 10% of the energy of the parent jet, for jets in the range $100 \text{ GeV} \leq p_T^{\text{jet}} < 500$ GeV. The leading-$p_T^{\text{subjet}}$ subjet is located on average at $\Delta R \leq 0.07$ from the axis of the parent jet, while less energetic subjets are more distant from the axis of the parent jet: $\Delta R \geq 0.5$. These observations are consistent with the increased radial collimation expected in jets produced from gluons and light quarks as the jet $p_T$ rises [93]. Therefore, the subjet structure of jets from dijet events can be characterized by looking only at the leading and sub-leading subjets.
Given the crowded nature of the subjet topology within a jet, many standard approaches to studying the subjet energy scale begin to fail. Associating individual tracks with calorimeter subjets may begin to suffer from the proximity of multiple such subjets to a given track, which can in turn impact the measurement of $r_{\text{subj}}^{\text{trk}}$. Geometrical matching relies on assumptions that are not fulfilled in high-density environments. This association assumes a cone-like shape and an area of $\pi R_{\text{sub}}^2$ for all subjets, and that all tracks within this area belong to the subjet. This assumption generally works well in the case of an isolated anti-$k_t$ jet. Conversely, $k_t$, $C/A$, and even anti-$k_t$ subjets in high-multiplicity environments often have very irregular boundaries and the question of which tracks to associate becomes more difficult to answer. Ghost-association [94, 95] provides a much more appropriate matching of the tracks to the calorimeter subjets for this scenario. In this technique, tracks are treated as infinitesimally soft, low-$p_T$ particles by setting their $p_T$ to 1 eV. These tracks are then added to the list of inputs for jet finding. The low scale means the tracks do not affect the reconstruction of calorimeter-jets. However, after jet finding, it is possible to identify which tracks are clustered into which subjets. This technique shows a more stable dependence of the ratio $r_{\text{subj}}^{\text{trk}}$ on the angular separation between subjets. Generally, this approach facilitates the measurement of the effective area of a jet, or the so-called ghost area. Instead of tracks, a uniform, fixed density (one per $\Delta y \times \Delta \phi = 0.01 \times 0.01$) of infinitesimally soft particles is distributed within the event and are allowed to participate in the jet clustering algorithm. Instead of identifying tracks associated with the resulting jets, the number of such ghost particles present in the jet after reconstruction defines the effective area of that jet.

Figure 15 shows the mean ratio $r_{\text{subj}}^{\text{trk}}$ as a function of the distance between the subjet and its closest neighbour subjet in the $\eta - \phi$ plane ($\Delta R_{\text{min}}$). The numerator of $r_{\text{subj}}^{\text{trk}}$ corresponds to the total transverse momentum of tracks associated with the subjet using either a standard geometric association (figure 15(a)) or the novel ghost-association tech-
Figure 15. Mean ratio of track $p_T$ to calorimeter $p_T$ ($r_{\text{subj}}$) for sub-leading subjets as a function of the distance of the subjet to the nearest subjet within the jet for (a) geometrical association and (b) ghost association techniques.

For standard geometric track association, a rise in $r_{\text{subj}}$ is observed for close-by subjets with $\Delta R_{\text{min}} \leq 2 \times R_{\text{sub}}$. The impact of the track association scheme is more significant for the second leading subjet, which often has a very energetic subjet nearby (i.e. the leading-$p_T$ subjet). For these cases geometrical association of tracks fails dramatically, and the ratio $r_{\text{subj}}$ is a factor of two smaller at $\Delta R_{\text{min}} = 0.6$ than at $\Delta R_{\text{min}} = 0.3$, which is approximately the smallest separation between jets. The double ratio $R_{\text{subj}} = \langle r_{\text{subj}} \rangle_{\text{data}} / \langle r_{\text{subj}} \rangle_{\text{MC}}$ provides an estimate of the calibration uncertainty. Any difference is well within 5% for the leading-$p_T$ subjet and 20% for the second leading subjet, independent of the track matching scheme.

3.5 Calibration of subjets

The HEPTopTagger [32] also relies on the energy scale of subjets to analyse the structure of $R=1.5$ jets and to reconstruct a top-quark candidate four-momentum. This section describes a dedicated calibration procedure for C/A subjets with a radius parameter $R = 0.2 - 0.5$, which are used in the study of HEPTopTagger performance. Both the compatibility of the structure with hadronic top-quark decay and the dependence of the reconstructed four-momentum on the calibration are discussed here.

C/A jets — reconstructed as independent jets and not as constituents of parent jets, as above — are first calibrated using a simulation of the calorimeter response to jets by comparing the energy and pseudorapidity of a generator-level jet to that of a matched calorimeter jet. Calibration constants obtained from this procedure are then applied to the actual subjets reconstructed by the HEPTopTagger. For this reason, these C/A jets are referred to using the same subjet notation as in the previous section. Prior to calibration, the reconstructed jet energy is lower than the particle jet’s energy and is corrected as a
Figure 16. Comparison of the calibrated Cambridge–Aachen $R = 0.4$ calorimeter jet $p_T$ with the $p_T$ of tracks matched to the jet for ATLAS data recorded in 2011 and for dijet simulations with PYTHIA. The average number $\mu$ of interactions per bunch crossing ranges from 4 to 7. (a) The average ratio $\langle r_{\text{subjet}}^{\text{trk}} \rangle$ as a function of the calibrated jet $p_T$. (b) The double ratio $R_{\text{trk}} = \langle r_{\text{subjet}}^{\text{trk}} \rangle_{\text{data}} / \langle r_{\text{subjet}}^{\text{trk}} \rangle_{\text{MC}}$. The horizontal line indicates the uncertainty-weighted average.

function of $p_T$ and in bins of $\eta$. For example, the correction for C/A jets with $R = 0.4$ and $|\eta| < 0.1$ is +9% at low $p_T$ and up to +2.5% for $p_T > 500$ GeV. The corrected $p_T$ matches that of the particle jet to within 2% for all energies and pseudorapidities. For C/A $R = 0.2$ jets this closure test is 4% at $p_T = 20$ GeV and better for higher $p_T$.

Uncertainties on the jet calibration are determined from the quality of the modelling of the calorimeter-jet $p_T$. The direct ratio $p_T^{\text{jet}}(\text{MC}) / p_T^{\text{jet}}(\text{data})$ is sensitive to mismodelling of jets at the hadron level. To reduce this effect, the calorimeter-jet $p_T$ is normalized to the $p_T$ of the tracks within the jet. This is done because the uncertainty of the track-jet $p_T$ tends to be small compared to the calorimeter-jet $p_T$ in the kinematic regime considered here. Tracks are matched to calorimeter-jets using ghost-track association. The jets are required to be within $|\eta| < 2.1$ to ensure coverage of the associated tracks by the tracking detector.

The average $r_{\text{subjet}}^{\text{trk}}$ for a subset of the data characterized by an average number of interactions per bunch crossing in the range $4 < \mu < 7$ is shown in figure 16(a) as a function of $p_T^{\text{jet}}$ for both data and simulation for C/A $R = 0.4$ jets with $|\eta| < 0.8$. The double ratio $R_{\text{trk}} = \langle r_{\text{trk}}^{\text{subjet}} \rangle_{\text{data}} / \langle r_{\text{trk}}^{\text{subjet}} \rangle_{\text{MC}}$ is shown in figure 16(b). Deviations of $R_{\text{trk}}$ from unity serve as an estimate of the uncertainty of the MC calibrated calorimeter-jet $p_T$. The largest deviation from unity is seen at low $p_T$ and is 4% with a statistical uncertainty of 1%. The statistical uncertainty-weighted average double ratio is indicated by the horizontal line.

Similar results are obtained when varying the jet radius parameter between 0.2 and 0.5 and for higher pile-up conditions (evaluated using a subset of the data characterized by $13 < \mu < 15$). A jet energy uncertainty of 3.5% is assigned.

The imperfect knowledge of the material distribution in the tracking detector constitutes the dominating systematic uncertainty. It results in an additional uncertainty in $R_{\text{trk}}$ of $\approx 2\%$ for $|\eta| < 1.4$ and $\approx 3\%$ for $1.4 < |\eta| < 2.1$, although it does not introduce a measurable shift.
The jet $p_T$ systematic uncertainty is taken to be the absolute deviation of the central weighted-average $R_{\text{trk}}$ from unity, with the shifts introduced by the systematic variations added in quadrature. The uncertainty varies between 2.3% and 6.8%, depending on the jet $p_T$, $\eta$, and the jet radius parameter. The uncertainty has been determined independently in samples with low and high pile-up $\mu$-values and no significant difference has been found.

A sample of boosted top quarks is used to study the impact of heavy flavour and close-by jet topologies on the systematic uncertainties estimated above. The track-based validation is applied to a sample of events that contains $\simeq 50\%$ semileptonically decaying $t\bar{t}$ pairs, in which the top quark with the hadronically decaying $W$ boson has $p_T > 200$ GeV. The remaining events are dominated by $W$+jets production. Figure 17 shows $r_{\text{subj}}$ and $R_{\text{trk}}$ for C/A $R = 0.4$ jets in these events. The jet $p_T$ uncertainty in this sample varies between 2.4% and 5.7%.

4 Jet substructure and grooming in the presence of pile-up

4.1 Impact of pile-up on the jet energy scale and the jet mass scale

This section elaborates on the impact of pile-up on the jet mass and other observables, and the extent to which trimming, mass-drop filtering, and pruning are able to minimize these effects. In particular, these measures of performance are used as some of the primary figures of merit in determining a subset of groomed jet algorithms on which to focus for physics analysis in ATLAS.

Figure 18 shows the dependence of the mean uncalibrated jet mass, $\langle m_{\text{jet}} \rangle$, on the number of reconstructed primary vertices, $N_{\text{PV}}$, for a variety of jet algorithms in the central region $|\eta| < 0.8$. The events used for these comparisons are obtained with the
Figure 18. Evolution of the mean uncalibrated jet mass, $\langle m^{\text{jet}} \rangle$, for jets in the central region $|\eta| < 0.8$ as a function of the reconstructed vertex multiplicity, $N_{\text{PV}}$, for jets in the range $200 \text{ GeV} \leq p_T^{\text{jet}} < 300 \text{ GeV}$ (left) and for leading-$p_T^{\text{jet}}$ jets ($\langle m^{\text{jet}} \rangle$) in the range $600 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{ GeV}$ (right). (a)-(b) show trimmed anti-$k_t$ jets with $R = 1.0$. (c)-(d) show pruned anti-$k_t$ jets with $R = 1.0$, and (e)-(f) show mass-drop filtered C/A jets with $R = 1.2$. The error bars indicate the statistical uncertainty on the mean value in each bin.
inclusive selection described in section 2.2 and contain an admixture of light quark and gluon jets. This dependence is shown in two $p_T^{\text{jet}}$ ranges of interest for jets after trimming (figures 18(a)–18(b)), pruning (figures 18(c)–18(d)), and mass-drop filtering (figures 18(e)–18(f)). For these comparisons, only the final period of data collection from 2011 is used, which corresponds to approximately 1 fb$^{-1}$ of integrated luminosity but is the period with the highest instantaneous luminosity recorded at $\sqrt{s} = 7$ TeV, where $\langle \mu \rangle = 12$, higher than the average over the whole 2011 data-taking period. The lower range, 200 GeV $\leq p_T^{\text{jet}} < 300$ GeV, represents the threshold for most hadronic boosted-object measurements and searches, whereas the range 600 GeV $\leq p_T^{\text{jet}} < 800$ GeV is expected to contain top quarks for which the decay products are fully merged within an $R = 1.0$ jet nearly 100% of the time. In each figure, the full set of grooming algorithm parameter settings is included for comparison. As noted in table 1, two values of the subjet radius, $R_{\text{sub}}$, are used for trimming, three $R_{\text{cut}}$ factors for pruning are tested (using the $k_t$ algorithm with the procedure in all cases), and three $\mu_{\text{frac}}$ settings are evaluated using the filtering algorithm.

Several observations can be made from figure 18. Of the grooming configurations tested, trimming and filtering both significantly reduce the rise with pile-up of $\langle m^{\text{jet}} \rangle$ seen for ungroomed jets, whereas pruning does not. For at least one of the configurations tested, trimming and filtering are both able to essentially eliminate this dependence. Furthermore, the trimming configurations tested provide a highly tunable set of parameters that allow a relatively continuous adjustment from small to large reduction of the pile-up dependence of the jet mass. The trimming configurations with $R_{\text{sub}} = 0.2, f_{\text{cut}} = 0.03$ and $R_{\text{sub}} = 0.3, f_{\text{cut}} = 0.05$ exhibit good stability for both small and large $p_T^{\text{jet}}$, with the $f_{\text{cut}} = 0.05$ configuration exhibiting a slightly smaller impact from pile-up at high $N_{\text{PV}}$ for low $p_T^{\text{jet}}$ (not shown). The other parameter settings either do not reduce the pile-up dependence at low $p_T^{\text{jet}}$ (e.g. $f_{\text{cut}} = 0.01$) or result in a downward slope of $\langle m^{\text{jet}} \rangle$ as a function of pile-up at high $p_T^{\text{jet}}$ (e.g. $f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.2$).

Pruning, on the other hand, exhibits the smallest impact on the pile-up dependence of the jet mass for these large-$R$ jets. Only by increasing the $z_{\text{cut}}$ parameter from $z_{\text{cut}} = 0.05$ to $z_{\text{cut}} = 0.10$ can any reduction on the dependence of $\langle m^{\text{jet}} \rangle$ on pile-up be observed. This is equivalent to reducing the low-$p_T$ contributions during the jet recombination; in the language of trimming, this is analogous to raising $f_{\text{cut}}$. This change slightly reduces the magnitude of the variation of the mean jet mass as a function $N_{\text{PV}}$ for low $p_T^{\text{jet}}$. The $R_{\text{cut}}$ parameter has very little impact on the performance, with nearly all of the differences observed being due to the change in $z_{\text{cut}}$. This observation holds for both small and large $p_T^{\text{jet}}$.

The mass-drop filtering algorithm can be made to affect $\langle m^{\text{jet}} \rangle$ significantly solely via the mass-drop criterion, $\mu_{\text{frac}}$. A drastic change in $\langle m^{\text{jet}} \rangle$ is observed for all configurations of the jet filtering, with the strictest $\mu_{\text{frac}} = 0.20$ setting rejecting nearly 90% of the jets considered and resulting in a slightly negative slope in the mean jet mass versus $N_{\text{PV}}$. Nevertheless, the other two settings of $\mu_{\text{frac}}$ tested exhibit no significant variation as a function of the number of reconstructed vertices, and the optimum value of $\mu_{\text{frac}} = 0.67$ found previously seems to have the best stability. Studies from 2010 [25] demonstrate
that this reduction in the sensitivity to pile-up is due primarily to the filtering step in the algorithm as opposed to the jet selection itself.\footnote{The performance of grooming also depends on the radius parameter, $R$, which was not varied in these studies for a single jet algorithm. In particular, the relative efficacy of the various grooming algorithms and configurations in mitigating the effects of pile-up can change with $R$ and will be studied in a future analysis.}

Figure 19 presents the pile-up dependence of the mean leading-$p_T$ jet mass, $\langle m_{\text{jet}} \rangle$, in data compared to the three simulations. Here, only the range $600 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{ GeV}$ for ungroomed and trimmed anti-$k_t$ jets is shown for brevity, but similar conclusions apply in all $p_T^{\text{jet}}$ ranges. The comparison is made using the full 2011 dataset. PYTHIA, HERWIG++, and POWHEG+PYTHIA all model the data fairly accurately, with a slight 5\%–10\% discrepancy appearing in the predictions from PYTHIA and HERWIG++ for the trimmed jets. Most importantly, the impact of pile-up is very well modelled, with the slope of the dependence of $\langle m_{\text{jet}} \rangle$ on $N_{PV}$ in data agreeing within 3\% with the POWHEG+PYTHIA prediction for both the ungroomed and trimmed jets.

Beyond simply providing a pile-up-independent average jet mass, the optimal grooming configurations render the full jet mass spectrum insensitive to high instantaneous luminosity. Figure 20 demonstrates this by comparing the jet mass spectrum for leading-$p_T^{\text{jet}}$ ungroomed and trimmed anti-$k_t$ jets for various values of $N_{PV}$. The comparison is performed both in an inclusive data sample and using the $Z' \rightarrow t\bar{t}$ MC sample, which produces a characteristic peak at the top-quark mass. The inclusive jet sample obtained from data shows that a nearly identical trimmed $m_{\text{jet}}$ spectrum is obtained regardless of the level of pile-up. The peak of the leading-$p_T^{\text{jet}}$ jet mass distribution for events with $N_{PV} \geq 12$ is shifted comparatively more due to trimming: from $m_{\text{jet}} \approx 125 \text{ GeV}$ to $m_{\text{jet}} \approx 45 \text{ GeV}$.
Jet mass spectra for four primary vertex multiplicity ranges for anti-\(k_t\) jets with \(R = 1.0\) in the range 600 \(\text{GeV} \leq p_T^\text{jet} < 800\) \(\text{GeV}\). Both untrimmed (left) and trimmed (right) anti-\(k_t\) jets are compared for the various \(N_{PV}\) ranges in data (top) and for a \(Z' \rightarrow t\bar{t}\) Monte Carlo sample (bottom).

Figure 20. Jet mass spectra for four primary vertex multiplicity ranges for anti-\(k_t\) jets with \(R = 1.0\) compared to an initial peak position of \(m^\text{jet} \approx 90\) \(\text{GeV}\) for events with \(1 \leq N_{PV} \leq 4\). Nonetheless, the resulting trimmed jet mass spectra exhibit no dependence on \(N_{PV}\).

Comparisons performed using the simulated \(Z' \rightarrow t\bar{t}\) sample demonstrate the same performance of the trimming algorithm, but in the context of the reconstruction of highly boosted top quarks. Figures 20(c)-(d) indicate that the ability to render the full jet mass distribution independent of pile-up does not come at the cost of the mass resolution or scale. Prior to jet trimming, a variation in the peak position of the jet mass of nearly 15 \(\text{GeV}\) is observed between the lowest and the highest ranges of \(N_{PV}\) studied. After jet trimming, the resulting mass spectra for the various \(N_{PV}\) ranges are narrower and
lie directly on top of one another, even in the case of a jet containing a highly boosted, and thus very collimated, top-quark decay. This observation, combined with that above, demonstrates that the trimming algorithm is working as expected by removing soft, wide-angle contributions to the calculation of the jet mass while retaining the relevant hard substructure of the jet.

Finally, although not shown explicitly, the mass-drop filtered jet mass is also stable with respect to pile-up.

The track-jet approach used to evaluate the jet mass uncertainty (see section 3.3.2) is also used to understand the effects of pile-up. It is observed that $r_{\text{track jet}}^m$ is nearly equal for the various trimming configurations in the case of little or no in-time pile-up (i.e. $N_{\text{PV}} \approx 1$) whereas filtering shows a significant, although small, difference between the configurations using $\mu_{\text{frac}} = 0.67, 0.33, 0.20$. This shows that the filtering method does affect the magnitude of $m_{\text{jet}}$ and $m_{\text{track jet}}$ slightly differently, resulting in an approximate 12% relative drop in $\langle r_{\text{track jet}}^m \rangle$ after filtering. These distributions are nonetheless well modelled by the simulation, resulting in double ratios of data to simulation very close to one. The trimming configuration with $R_{\text{sub}} = 0.3, f_{\text{cut}} = 0.01$ has almost no impact on the dependence of $r_{\text{track jet}}^m$ with pile-up.

4.2 Impact of pile-up on jet substructure properties

Figures 21 and 22 show the variation with pile-up observed in data for the splitting scales and $N$-subjettiness observables for jets in the range $600 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{ GeV}$. In this case, the focus is on jet trimming since the mass-drop filtering algorithm makes a pre-defined choice to search for properties of a jet characteristic of a two-body decay. The constraints placed on subjet multiplicity by the filtering procedure are not appropriate for calculating generic jet shapes given the strict substructure requirements they place on a jet. Furthermore, pruning of jets with $R = 1.0$ does not seem to mitigate the effects of pile-up. The trimming configurations with $R_{\text{sub}} = 0.3$ and $f_{\text{cut}} = 0.03, 0.05$ yield the most stable jet substructure properties with the smallest deviation in their observed mean values at low $N_{\text{PV}}$. This conclusion holds for all other jet $p_T^{\text{jet}}$ ranges as well, with larger differences between $f_{\text{cut}} = 0.03, 0.05$ appearing at low $p_T^{\text{jet}}$.

Figure 23 presents a comparison of data and Monte Carlo simulation for $\langle \sqrt{d_{12}} \rangle$ and $\langle \tau_{32} \rangle$ for ungroomed and trimmed ($R_{\text{sub}} = 0.3, f_{\text{cut}} = 0.05$) anti-$k_t$ jets with $R = 1.0$. The slope of these observables as a function of $N_{\text{PV}}$ is well modelled by the simulation.

4.3 Impact of pile-up on signal and background in simulation

In addition to the comparisons between data and simulation, and between the various grooming configurations, a comparison of how grooming impacts signal-like events versus background-like events in searches for resonances decaying to boosted jets is crucial.

Figure 24 shows the variation of the average leading-$p_T^{\text{jet}}$ jet mass, $\langle m_1^{\text{jet}} \rangle$, with $N_{\text{PV}}$ for events with $600 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{ GeV}$ for ungroomed and trimmed anti-$k_t$, $R = 1.0$ jets, for both the $Z' \rightarrow tt$ sample and the POWHEG+PYTHIA dijet sample. The average ungroomed leading-$p_T^{\text{jet}}$ jet mass in the sample of gluons and light quarks in the inclusive POWHEG+PYTHIA dijet events exhibits a slope of approximately $d\langle m_1^{\text{jet}} \rangle/dN_{\text{PV}} \approx 3 \text{ GeV}/N_{\text{PV}}$. The leading-$p_T^{\text{jet}}$ jets in the $Z'$ sample are typically entirely composed of fully
jets with $R < 800$ GeV before and after trimming. The error bars indicate the statistical uncertainty on the mean value in each bin.

![Figure 21](image1.png)

**Figure 21.** Variation with the number of reconstructed primary vertices, $N_{\text{PV}}$, of the mean splitting scales (a) $\langle \sqrt{d_{12}} \rangle$ and (b) $\langle \sqrt{d_{23}} \rangle$ measured in data for anti-$k_t$ jets with $R = 1.0$ in the range $600 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{ GeV}$ before and after trimming. The error bars indicate the statistical uncertainty on the mean value in each bin.

![Figure 22](image2.png)

**Figure 22.** Variation with the number of reconstructed primary vertices, $N_{\text{PV}}$, of the mean $N$-subjettiness ratios (a) $\langle \tau_{21} \rangle$ and (b) $\langle \tau_{32} \rangle$ measured in data for anti-$k_t$ jets with $R = 1.0$ in the range $600 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{ GeV}$ before and after trimming. The error bars indicate the statistical uncertainty on the mean value in each bin.

hadronic boosted top-quark decays contained in a single jet. The mass reconstruction in this case proceeds as usual (four-momentum recombination) and the mass distribution is highly peaked near the top-quark mass of approximately 175 GeV. Jets in this peak but without grooming exhibit a slope of roughly $d(m_{\text{jet}}^2)/dN_{\text{PV}} \approx 2.15 \text{ GeV}/N_{\text{PV}}$, or about 30% smaller than in the inclusive jet sample. In the case of trimmed jets, the slopes as a function of $N_{\text{PV}}$ for both signal-like jets and jets in dijet events are consistent with zero.
Figure 23. Variation with the number of reconstructed primary vertices, $N_{PV}$, of (a) the mean $N$-subjettiness ratio ($\langle \tau_{32} \rangle$) for ungroomed anti-$k_t$ jets with $R=1.0$ and (b) the mean splitting scales ($\langle \sqrt{d_{12}} \rangle$) for trimmed jets ($R_{\text{sub}}=0.3, f_{\text{cut}}=0.05$), for data and simulation. The error bars indicate the statistical uncertainty on the mean value in each bin. The lower panels show the ratio of the mean values measured in data to simulation.

Figure 24. Variation with the number of reconstructed primary vertices, $N_{PV}$, of the average leading-$p_T$ jet mass, $\langle m_{\text{jet}} \rangle$, in the range $600 \text{ GeV} \leq p_{\text{jet}}^T < 800 \text{ GeV}$ for (a) ungroomed and (b) trimmed jets. The error bars indicate the statistical uncertainty on the mean value in each bin.

Most importantly, the average separation between the mean jet mass for signal-like jets in the $Z'$ sample and those in the POWHEG+PYTHIA dijet sample increases by nearly 50% after trimming and remains stable across the full range of $N_{PV}$. This allows for much better discrimination between the two processes. The separation shown here is significant since the widths of the peaks of each of the distributions are also simultaneously

\[ \text{JHEP09(2013)076} \]
narrowed by the grooming algorithm, as shown in figure 20. This differential impact of trimming is again due to the design of the algorithm: soft, wide-angle contributions to the jet mass that are ubiquitous in jets produced from light quarks and gluons are suppressed whereas the hard components present in a jet with true substructure — as in the case of the top-quark jets here — are preserved.

5 Jet substructure and grooming with boosted objects in data and simulation

Comparisons between jets containing signal-like boosted objects and a light-quark or gluon jet background are presented here. Boosted objects are divided into two categories depending on the event topology: two-pronged, such as hadronically decaying W or Z bosons, and three-pronged, such as the top quark decaying into a b-jet and a hadronically decaying W boson. Performance measures are shown for both simulated samples of $Z \rightarrow q\bar{q}$ and top quarks (from $Z' \rightarrow t\bar{t}$), as well as for inclusive jet data and events enriched in boosted top quark pairs. In addition to the event and object selection listed in section 2.2, the large-\(R\) ungroomed leading-p$_T$ jet axis is required to be within $|\eta| < 2.0$. For the signal distributions, a $\Delta R < 1.5$ match between the four-momentum of the hadronically decaying boosted object in the truth record and the reconstructed ungroomed leading-p$_T$ jet is made to minimize the contamination from light-quark or gluon jets (or top quarks with a leptonically decaying W boson in the $Z'$ sample).

5.1 Expected performance of jet substructure and grooming in simulation

5.1.1 Jet mass resolution for background

The fractional jet mass resolution is defined as the width of a Gaussian fit to the central part of the distribution that is generated by taking the difference between the generator-level jet mass and the reconstructed jet mass, divided by the same generator-level jet mass. Here, the generator-level jet is the simulated particle shower that has been groomed according to the same grooming algorithms used after jet reconstruction. Large-\(R\) generator-level and reconstructed jets before grooming are matched if they are within $\Delta R < 0.7$. The matching is performed only once and comparisons between generator-level and reconstructed jets after grooming are made using the groomed versions of the matched ungroomed jets. The mass-drop filtering method is not applied to anti-\(k_t\) jets, as discussed in section 1.2.

Figure 25 shows the fractional mass resolution for leading-p$_T^{\text{jet}}$ jets in the POWHEG+PYTHIA dijet sample with the same pile-up conditions as were observed in the data. Abbreviated versions of the groomed algorithm names used to label the figures are listed in table 3. In general, the groomed jets have better resolution than the ungroomed large-\(R\) jets, with improvements of up to $\sim 10\%$ (absolute) in some cases. The trimmed and pruned jet resolution improves with increasing $p_T$, where the calibrated jets gain $\sim 3-5\%$ over the range $300 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{ GeV}$. The pruning algorithm, especially with C/A jets, produces larger tails in the resolution distribution compared to the trimmed algorithm, worsening the overall fractional resolution in comparison. The resolution is fairly stable for the mass-drop filtering algorithm over a large range of $p_T$. It is important to
Algorithm and Parameters

<table>
<thead>
<tr>
<th>Label</th>
<th>Algorithm and Parameters</th>
<th>Label</th>
<th>Algorithm and Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trm PtF1 R30</td>
<td>Trimmed, $f_{cut} = 1%$, $R_{sub} = 0.3$</td>
<td>Prn Rc10 Zc5</td>
<td>Pruned, $R_{cut} = 0.1$, $z_{cut} = 0.3$</td>
</tr>
<tr>
<td>Trm PtF3 R30</td>
<td>Trimmed, $f_{cut} = 3%$, $R_{sub} = 0.3$</td>
<td>Prn Rc10 Zc10</td>
<td>Pruned, $R_{cut} = 0.1$, $z_{cut} = 10%$</td>
</tr>
<tr>
<td>Trm PtF5 R30</td>
<td>Trimmed, $f_{cut} = 5%$, $R_{sub} = 0.3$</td>
<td>Prn Rc20 Zc5</td>
<td>Pruned, $R_{cut} = 0.2$, $z_{cut} = 5%$</td>
</tr>
<tr>
<td>Trm PtF1 R20</td>
<td>Trimmed, $f_{cut} = 1%$, $R_{sub} = 0.2$</td>
<td>Prn Rc20 Zc10</td>
<td>Pruned, $R_{cut} = 0.2$, $z_{cut} = 10%$</td>
</tr>
<tr>
<td>Trm PtF3 R20</td>
<td>Trimmed, $f_{cut} = 3%$, $R_{sub} = 0.2$</td>
<td>Prn Rc30 Zc5</td>
<td>Pruned, $R_{cut} = 0.3$, $z_{cut} = 5%$</td>
</tr>
<tr>
<td>Trm PtF5 R20</td>
<td>Trimmed, $f_{cut} = 5%$, $R_{sub} = 0.2$</td>
<td>Prn Rc30 Zc10</td>
<td>Pruned, $R_{cut} = 0.3$, $z_{cut} = 10%$</td>
</tr>
<tr>
<td>MD Filt mf20</td>
<td>Mass-drop Filtered, $\mu_{frac} = 0.2$</td>
<td>MD Filt mf33</td>
<td>Mass-drop Filtered, $\mu_{frac} = 0.67$</td>
</tr>
<tr>
<td>MD Filt mf33</td>
<td>Mass-drop Filtered, $\mu_{frac} = 0.33$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Labels used in figures to represent the various configurations of the grooming algorithms.

†Groomed jets have been calibrated for both anti-$k_t$ and C/A jets. ‡Groomed jets have been calibrated for anti-$k_t$ only. §Groomed jets have been calibrated for C/A only.

![Fractional mass resolution comparing various grooming algorithms](image)

**Figure 25.** Fractional mass resolution comparing the various grooming algorithms (with labels defined in Table 3) for the leading-$p_T^{jet}$ jet in POWHEG+PYTHIA dijet simulated events. Here, nominal refers to jets before grooming is applied. Three ranges of the nominal jet $p_T^{jet}$ are shown. The uncertainty on the width of the Gaussian fit is indicated by the error bars.

It is noted that the efficiency is considerably lower for jets resulting from this algorithm compared with jets produced in other grooming procedures (~30%) due to the strict mass-drop requirement, which is often not met for jets without boosted object substructure.

A summary of the fractional mass resolution for jets before and after grooming in the presence of various pile-up conditions is shown in Figure 26. Trimming in both anti-$k_t$ and C/A jets reduces the dependence of the jet mass on pile-up (spread in the points) compared to the ungroomed jet, as does the mass-drop filtering procedure in the case of C/A jets, while pruning has little impact. In all cases, no pile-up subtraction is applied to the ungroomed jet kinematics. In particular, the trimming parameters $f_{cut} = 0.03$ and 0.05 slightly outperform the looser $f_{cut} = 0.01$ setting in events with a mean number of
Figure 26. Fractional mass resolution comparing the various grooming algorithms (with labels defined in table 3) for the leading-$p_T$ jet in the range 500 GeV $\leq p_T^{\text{jet}} < 600$ GeV in dijet events, simulated with POWHEG+PYTHIA. Nominal refers to jets before grooming is applied. Various ranges of the average number of interactions ($\langle \mu \rangle$) in the events are shown. The uncertainty on the width of the Gaussian fit is indicated by the error bars.

interactions greater than 12. They also exhibit a significantly reduced overall variation between various instantaneous luminosities.

Based on the above comparisons of mass resolution in different $p_T^{\text{jet}}$ ranges and under various pile-up conditions, two configurations, trimmed anti-$k_t$ jets ($f_{\text{cut}} = 0.05$, $R_{\text{sub}} = 0.3$) with $R = 1.0$ and filtered C/A jets ($\mu_{\text{frac}} = 0.67$) with $R = 1.2$, are chosen for detailed comparisons between data and simulation and are presented in section 5.2.

5.1.2 Jet mass resolution for simulated signal events

Figures 27 and 28 show the fractional mass resolution for the two-pronged and three-pronged cases, respectively. The mass-drop filtering algorithm is shown only for the simulated two-pronged signal events with C/A jets. In the two-pronged case, as for the case of jets in the inclusive jet events shown in figure 25, the C/A mass-drop filtering algorithm performs the best, but with a signal reconstruction efficiency of $\sim 45\%$ in $Z \rightarrow q\bar{q}$ events (for $\mu_{\text{frac}} = 0.67$). In both the two-pronged and three-pronged configurations, the trimmed jets have better fractional mass resolution ($\sim 5 - 10\%$) than the pruned jets, especially for those jets with grooming applied after the C/A algorithm. The trimmed jet mass resolution also remains fairly stable across a large $p_T^{\text{jet}}$ range, with equivalent performance for anti-$k_t$ and C/A jets.

5.1.3 Signal and background comparisons with and without grooming

Leading-$p_T^{\text{jet}}$ jet distributions of mass, splitting scales and $N$-subjettiness are compared for jets in simulated signal and background events in the range $600$ GeV $\leq p_T^{\text{jet}} < 800$ GeV. As seen in figures 29–31, showing distributions for the two-pronged decay case, and in
 Fractional mass resolution of the leading-$p_T^\text{jet}$ jet in $Z \to q\bar{q}$ simulated events comparing the various grooming algorithms. Here, nominal refers to jets before grooming is applied. Three ranges of the nominal jet $p_T^\text{jet}$ are shown. The uncertainty on the width of the Gaussian fit is indicated by the error bars.

(a) $\text{anti-}k_t$, $R = 1.0$

(b) C/A, $R = 1.2$

**Figure 27.**

Fractional mass resolution of the leading-$p_T^\text{jet}$ jet in $Z \to q\bar{q}$ simulated events comparing the various grooming algorithms. Here, nominal refers to jets before grooming is applied. Three ranges of the ungroomed jet $p_T^\text{jet}$ are shown. The uncertainty on the width of the Gaussian fit is indicated by the error bars.

(a) $\text{anti-}k_t$, $R = 1.0$

(b) C/A, $R = 1.2$

**Figure 28.**

figures 32–35 showing comparisons for the three-pronged decay case, better discrimination between signal and background is obtained after grooming. In these figures, the ungroomed distributions are normalized to unit area, while the groomed distributions have the efficiency with respect to the ungroomed large-$R$ jets folded in for comparison. This is especially conspicuous in the C/A jets with mass-drop filtering applied as mentioned previously.

The mass resolution of the simulated $Z \to q\bar{q}$ signal events shown in figure 29 dramatically improves after trimming or mass-drop filtering for anti-$k_t$ jets with $R = 1.0$ and
Figure 29. Leading-$p_T$ jet mass for simulated HERWIG+JIMMY $Z \rightarrow q\bar{q}$ signal events (red) compared to POWHEG+PYTHIA dijet background events (black) for jets in the range $600 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{GeV}$. The dotted lines show the ungroomed jet distributions, whereas the solid lines show the (a) trimmed and (b) mass-drop filtered jet distributions. The trimming parameters are $f_{\text{cut}} = 0.05$ and $R_{\text{sub}} = 0.3$ and the mass-drop filtering parameter is $\mu_{\text{frac}} = 0.67$. The groomed distributions are normalized with respect to the ungroomed distributions, which are themselves normalized to unity.

Figure 30. Leading-$p_T$ jet splitting scale $\sqrt{d_{12}}$ for simulated HERWIG+JIMMY $Z \rightarrow q\bar{q}$ signal events (red) compared to POWHEG+PYTHIA dijet background events (black) for jets in the range $600 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{GeV}$. The dotted lines show the ungroomed jet distributions, while the solid lines show the (a) trimmed and (b) mass-drop filtered jet distributions. The trimmed parameters are $f_{\text{cut}} = 0.05$ and $R_{\text{sub}} = 0.3$ and the mass-drop filtering parameter is $\mu_{\text{frac}} = 0.67$. The groomed distributions are normalized with respect to the ungroomed distributions, which are themselves normalized to unity.
Figure 31. Leading-$p_T$ jet $N$-subjettiness $\tau_{21}$ for simulated HERWIG+JIMMY $Z \rightarrow q\bar{q}$ signal events (red) compared to POWHEG+PYTHIA dijet background events (black) for jets in the range $600 \text{ GeV} \leq p_T \leq 800 \text{ GeV}$. The dotted lines show the ungroomed jet distributions, while the solid lines show the (a) trimmed and (b) mass-drop filtered jet distributions. The trimmed parameters are $f_{\text{cut}} = 0.05$ and $R_{\text{sub}} = 0.3$ and the mass-drop filtering parameter is $\mu_{\text{fac}} = 0.67$. The groomed distributions are normalized with respect to the ungroomed distributions, which are normalized to unity.

C/A jets with $R = 1.2$, respectively. Mass-drop filtering has an efficiency of approximately 55% and therefore fewer jets remain in this figure. After trimming or mass-drop filtering, the mass peak corresponding to the $Z$ boson is clearly seen at the correct mass. Note that the dijet background is pushed much lower in mass after grooming as was demonstrated in figure 20, while the constituents of signal jets have higher $p_T$ and survive the grooming procedure, thus improving discrimination between signal and background. The small excess of signal events below 50 GeV is the result of one of the two quarks from the decay of the $Z$ boson being removed by the jet grooming, thus leaving only one quark reconstructed as the jet and making it indistinguishable from the background. Figure 30 shows the splitting scale $\sqrt{d_{12}}$ in $Z \rightarrow q\bar{q}$ events. The signal exhibits a splitting scale roughly equal to half the mass of the jet, whereas the splitting scale distribution for jets produced in dijet events peaks at smaller values of $\sqrt{d_{12}}$ and falls more steeply. This effect is enhanced after grooming, especially in the case of C/A jets after mass-drop filtering. In figure 31, the $N$-subjettiness variable $\tau_{21}$ is observed to have improved discrimination between signal and background with anti-$k_t$ trimmed jets compared to C/A mass-drop filtered jets, where the discrimination is worsened after applying the mass-drop filtering criteria. The filtering step is explicitly reconstructing a fixed number of final subjets (three, in this case), thereby shaping the background and worsening the resulting separation.

The three-pronged hadronic top-jet mass distributions from $Z' \rightarrow t\bar{t}$ events are shown in figure 32, where the signal peak is relatively unshifted between groomed and ungroomed jets, especially with anti-$k_t$ jets. Again the mass resolution for the signal improves after grooming, where the $W$-mass peak can also be seen after trimming is applied. The en-
Figure 32. Leading-\(p_T\) jet mass for simulated PYTHIA \(Z' \rightarrow t \bar{t} (m_{Z'} = 1.6 \text{ TeV})\) signal events (red) compared to POWHEG+PYTHIA dijet background events (black) for jets in the range \(600 \text{ GeV} < p_T < 800 \text{ GeV}\). The dotted lines show the ungroomed leading-\(p_T\) jet distribution, while the solid lines show the corresponding trimmed (\(f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.3\)) jets. The groomed distributions are normalized with respect to the ungroomed distributions, which are themselves normalized to unity.

Enhancement of the \(W\)-mass peak is seen especially in jets with lower \(p_T\), as the jet from the \(b\)-quark decay falls outside the radius of the large-\(R\) jet. Figures 33 and 34 show the variables \(\sqrt{d_{12}}\) and \(\sqrt{d_{23}}\), respectively, for \(Z' \rightarrow t \bar{t}\) events compared to jets produced in dijet events. As in the two-pronged case, signal discrimination with the splitting scales is enhanced after jet trimming.

One of the primary applications of \(N\)-subjettiness is as a discriminating variable in searches for highly boosted top quarks [18]. A common method of comparing the performance of such discriminating variables or tagging algorithms is to compare the rate at which light-quark or gluon jets are selected (the mis-tag rate) to the efficiency for retaining jets containing the hadronic particle decay of interest [13, 14]. This comparison is performed for both ungroomed and trimmed jets in order to assess the impact of grooming on the discrimination power of this observable. Figure 35 shows the \(\tau_{32}\) distribution before and after trimming. Here, trimming of anti-\(k_t\) and C/A jets results in similar discrimination between signal and background. In order to understand the utility of the \(\tau_{32}\) selection criterion and the potential impact of jet grooming, trimmed anti-\(k_t\), \(R = 1.0\) jets are compared to their ungroomed counterparts in a boosted top sample for two jet momentum ranges. The signal mass range is defined as that which contains a large fraction of the boosted top signal. For ungroomed jets this fraction is set to 90%, and the mass range that satisfies this requirement is 100 GeV \(< m < 250\) GeV. A slightly lower signal fraction of 80% for the same mass range is required for groomed jets; this is motivated by the tendency for trimmed jets to populate an additional small peak around the \(W\) mass, as shown in figure 32.
Figure 33. Leading-p_T jet splitting scale \( \sqrt{d_{12}} \) for simulated PYTHIA \( Z' \rightarrow t\bar{t} \) \((m_{Z'} = 1.6\,\text{TeV})\) signal events (red) compared to POWHEG+PYTHIA dijet background events (black) for jets in the range 600 GeV \( \leq p_T^{\text{jet}} \) < 800 GeV. The dotted lines show the ungroomed leading-\( p_T^{\text{jet}} \) jet distribution, while the solid lines show the corresponding trimmed \((f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.3)\) jets. The groomed distributions are normalized with respect to the ungroomed distributions, which are themselves normalized to unity.

The mis-tag rate is defined as the fraction of the POWHEG dijet sample that remains in the mass window after a simple selection based on \( \tau_{32} \). The signal top jet efficiency is defined as the fraction of top jets selected in the \( Z' \) sample with the same \( \tau_{32} \) selection.
Figure 35. Leading-$p_T$ jet $N$-subjettiness $\tau_{32}$ for simulated PYTHIA $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1.6$ TeV) signal events (red) compared to POWHEG+PYTHIA dijet background events (black) for jets in the range $600$ GeV $\leq p_T^\text{jet} < 800$ GeV. The dotted lines show the ungroomed leading-$p_T^\text{jet}$ jet distribution, while the solid lines show the corresponding trimmed ($f_{\text{cut}} = 0.05$, $R_{\text{sub}} = 0.3$) jets. The groomed distributions are normalized with respect to the ungroomed distributions, which are themselves normalized to unity.

Figure 36. Jet mis-tag rate for dijet events vs. top-jet efficiency curves using $\tau_{32}$ as a top tagger for (a) $600$ GeV $\leq p_T^\text{jet} < 800$ GeV and (b) $800$ GeV $\leq p_T^\text{jet} < 1000$ GeV with masses in the range $100$ GeV $\leq m^\text{jet} < 250$ GeV.

Figure 36 shows the performance of the $N$-subjettiness tagger for jets with $600$ GeV $\leq p_T^\text{jet} < 800$ GeV and $800$ GeV $\leq p_T^\text{jet} < 1000$ GeV. In both cases, for a fixed top-jet efficiency, the reduction in high-invariant-mass jets due to trimming results in a relative reduction of several percent in the mis-tag rate. Moreover, in the case of very high $p_T^\text{jet}$, as in figure 36(b), the slightly more aggressive trimming configuration results in a slight performance gain as well.
5.2 Inclusive jet data compared to simulation with and without grooming

Previous studies conducted by ATLAS [25] and CMS [26] suggest that even complex jet substructure observables are fairly well modelled by the MC simulations used by the LHC experiments. This section reviews the description provided by PYTHIA, HERWIG++, and POWHEG+PYTHIA of the jet grooming techniques introduced above, and of the substructure of the ungroomed and groomed jets themselves.

Figure 37 presents a comparison of the jet invariant mass for ungroomed, trimmed, and filtered jets in the range $600 \text{ GeV} \leq m_{\text{jet}} \leq 800 \text{ GeV}$ and in the central calorimeter, $|\eta| < 0.8$. Similar performance is observed in all other $p_T$ regions in the range $p_T > 300 \text{ GeV}$, and
Figure 38. Mass of jets in the range 600 GeV ≤ p_T^jet < 800 GeV and in the central calorimeter (|η| < 0.8). In (a) anti-kt jets with R = 1.0 are compared before (ungroomed) and after trimming with several configurations of p_T fraction (f_{cut}) and subjet size (R_{sub}). In (b) C/A jets with R = 1.2 are compared before and after filtering with three different values of mass-drop fraction (µ_{frac}).
Figure 39. Average area of jets in the range $500 \text{ GeV} \leq p_{\text{jet}}^{T} < 600 \text{ GeV}$ as a function of jet mass. Shown are (a) ungroomed and (b) trimmed ($f_{\text{cut}} = 0.05$, $R_{\text{sub}} = 0.3$) anti-$k_t$ jets with $R = 1.0$. The ratios between data and MC distributions are shown in the lower section of each figure.

tested, with the large $f_{\text{cut}}$, small $R_{\text{sub}}$ setting for trimming and the small $\mu_{\text{frac}}$ setting for mass-drop filtering exhibiting the most dramatic changes. For jet masses in the range of $50 \text{ GeV} \leq m_{\text{jet}} < 300 \text{ GeV}$, which is expected to be the most relevant in searches for new physics, the trimming configurations exhibit efficiencies in the range of 30%–70%, defined as the ratio of the yield after grooming to that prior to grooming. In particular, the trimming configuration with $f_{\text{cut}} = 0.05$ and $R_{\text{sub}} = 0.3$ yields an approximate 47% efficiency in this mass range. Mass-drop filtering provides a more stringent selection, yielding efficiencies in the same $50 \text{ GeV} \leq m_{\text{jet}} < 300 \text{ GeV}$ mass range of 20%, 12%, and 3% for $\mu_{\text{frac}} = 0.67$, 0.33, 0.30, respectively.

The significant change observed in the jet mass distribution is due primarily to a reduction in the effective area of each jet (see section 3.4 for a detailed description of the jet area). Soft and wide-angle jet constituents are removed from the jet, thereby reducing the overall catchment area. This has the desirable effect of also reducing the impact of pile-up on the jet properties.

Figure 39 shows the effect of grooming on the average jet area as a function of the jet mass for both the ungroomed and trimmed anti-$k_t$, $R = 1.0$ jets. Prior to jet trimming, the anti-$k_t$, $R = 1.0$ area is very close to $\pi$ (i.e. $\pi R^2$, with $R = 1.0$). A small rise in the ungroomed jet area is observed for jets with very large mass, characterized by small additional clusters near the edge of the jet. For trimmed jets at low mass, the average jet area is reduced by a factor of 3–5 and continuously rises as a function of the jet mass to a maximum of approximately half of the original jet area for high-mass ungroomed jets. These features are very well described by the MC simulations across the entire spectrum of jet mass, both before and after trimming. Similar observations are made with respect to lower and higher $p_{\text{T}}^{\text{jet}}$ bins, as well as for pruning and filtering.

As discussed in section 1.2, the splitting scales $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ are designed to give approximate measures of the relative mass of the leading, sub-leading, and third leading
subjets, ordered by $p_T^{\text{subj}}$. These observables are therefore dominated by contributions originating from energetic partons, either from the parton shower or from massive particle decay. It is therefore not surprising that the jet grooming, which removes low-$p_T$ components of the jet, does not significantly affect $\sqrt{d_{12}}$, shown in figure 40. It is also not surprising that POWHEG+PYTHIA describes this variable better than PYTHIA, especially at large values. Interestingly, HERWIG++ also models this substructure characteristic well, despite providing only a LO description of the hard process. In the case of events selected only for the presence of a single high-$p_T$ jet, a second hard splitting is not highly probable, as also evidenced by the spectrum of $\sqrt{d_{23}}$, shown in figure 41, falling more steeply than $\sqrt{d_{12}}$. This demonstrates that the trimming tends to affect this splitting scale slightly more when a third $-p_T^{\text{subj}}$ subjet within the parent jet is modified during the trimming procedure.

$N$-subjettiness helps to discriminate between jets that have well-formed substructure and those that do not. Figures 42 and 43 demonstrate that the MC simulations model the distributions of $\tau_{21}$ and $\tau_{32}$ observed in the data within about 20%. The jets in this case are selected to have $600 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{ GeV}$. Both PYTHIA and POWHEG+PYTHIA exhibit a shift in the $\tau_{21}$ distribution towards larger values of $\tau_{21}$ with respect to the data for both the ungroomed and trimmed jets. HERWIG++, however, provides a much better description of $\tau_{21}$ for ungroomed jets, and shows a slight shift towards smaller values of $\tau_{21}$ compared to the data. The fraction of events with a small value of $\tau_{ij}$ is predicted to be slightly smaller than in data, while it is the opposite for high values of $\tau_{ij}$. In addition, the distributions of both $\tau_{21}$ and $\tau_{32}$ are broadened for trimmed jets (figures 42(b) and 43(b)) and shifted slightly towards smaller values of $\tau_{ij}$ and into the region expected to be populated by boosted hadronic particle decays. These observations suggest that the use of
shape observables for ungroomed jets, and of jet mass and substructure observables (like $\sqrt{\Delta z}$) may lead to better discrimination between signal and background.

For $\tau_{32}$, the MC distributions slightly underestimate the fraction of jets with $0.3 < \tau_{32} < 0.7$, which is the signal range for boosted top-quark candidates. Since these observables are intended to be used as discriminants between boosted object signal events and the inclusive jet background, such differences are important for the resulting estimation of signal efficiency compared to background rejection. However, the variations observed
in the distribution of each observable translate into much smaller differences in efficiency and rejection.

The modelling of the background with respect to massive boosted objects can be tested by evaluating, for example, the evolution of the mean of substructure variables as a function of jet mass. This is shown for $\langle \tau_{32} \rangle$ in figure 44 for the same $600 \text{ GeV} \leq p_T^{\text{jet}} < 800 \text{ GeV}$ range as used above. Three important observations can be made. Values of $\langle \tau_{32} \rangle$ are slightly lower in data than those predicted by the MC simulations, and the trimmed values are lower compared to ungroomed jets. Furthermore, $\langle \tau_{32} \rangle$ is a slowly varying function of the jet mass for both the ungroomed and trimmed jets. This variation is slightly reduced for trimmed jets.

5.3 Performance of jet grooming in boosted top-quark events

5.3.1 Semi-leptonic $t\bar{t}$ selection

A selection of $t\bar{t} \rightarrow (Wb)(W\bar{b}) \rightarrow (\mu\nu b)(q\bar{q}\bar{b})$ events is used to demonstrate in data the effect of grooming on large-$R$ jets with substructure. The semileptonic $t\bar{t}$ decay mode in which one $W$ boson decays into a neutrino and a muon is chosen in order to tag the $t\bar{t}$ event and reduce the overwhelming multi-jet background so that the top-quark signal is visible. This provides a relatively pure sample of top quarks and is also very close to the selection used in searches for resonances that decay to pairs of boosted top quarks [96, 97]. The following event-level and physics object selection criteria are applied to data and simulation:

- *Event-level trigger and data quality selection:* the standard data quality and vertex requirements described in section 2.2 are applied. Events are selected if they satisfy the single-muon Event Filter trigger with muon $p_T > 18 \text{ GeV}$.
Figure 44. Mean \( \tau_{32} \) as a function of jet mass for (a) ungroomed and (b) trimmed \((f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.3)\) anti-\( k_t \) jets with \( R = 1.0 \) in the range \( 600 \text{ GeV} < p_T^{\text{jet}} < 800 \text{ GeV} \).

- **Event-level jet selection:** events are required to have at least four anti-\( k_t \) jets with \( R = 0.4 \) having \( p_T^{\text{jet}} > 25 \text{ GeV} \) and jet vertex fraction \( |JVF| > 0.75 \). The jet vertex fraction is a discriminant that contains information regarding the probability that a jet originated from the selected primary vertex in an event [98].

- **Lepton selection:** muons must be reconstructed in both the inner detector and the muon spectrometer and have \( p_T > 20 \text{ GeV} \) and \(|\eta| < 2.5\). The opening angle between the muon and any \( R = 0.4 \) jet with \( p_T > 25 \text{ GeV} \) and jet vertex fraction \( |JVF| > 0.75 \) must be greater than \( \Delta R = 0.4 \) to be well isolated. Events with one or more electrons passing standard criteria as described in ref. [97] are rejected.

- **Event-level neutrino and leptonic W-decay requirement:** to tag events with a leptonically decaying W boson from a top-quark decay, events are required to have missing transverse momentum \( E_T^{\text{miss}} > 20 \text{ GeV} \). Additionally, the scalar sum of \( E_T^{\text{miss}} \) and the transverse mass of the leptonic W boson candidate must satisfy \( E_T^{\text{miss}} + m_W^T > 60 \text{ GeV} \), where \( m_W^{T} = \sqrt{2p_T E_T^{\text{miss}}(1 - \cos \Delta \phi)} \) is calculated from the muon \( p_T \) and \( E_T^{\text{miss}} \) in the event, and \( \Delta \phi \) is the azimuthal angle between the charged lepton and the \( E_T^{\text{miss}} \), which is assumed to be due to the neutrino.

After these selection requirements, the W+jets process constitutes the largest background, with smaller contributions from Z+jets and single-top-quark processes.

### 5.3.2 Performance of trimming in t\( \bar{t} \) events

Figure 45 shows the leading-\( p_T^{\text{jet}} \) jet mass of anti-\( k_t \) jets with \( R = 1.0 \) having \( p_T > 350 \text{ GeV} \) before and after trimming \((f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.3)\) after the above selection criteria are applied. The data and simulation agree within statistical uncertainty. The \( W \rightarrow \mu \nu \) events produced in association with jets form the largest background. Since large-\( R \) jets in W
events are formed from one or more random light-quark or gluon jets, trimming causes the mass spectrum to fall more steeply and the peak of the distribution to lie at smaller masses, similar to the multi-jet background in figure 38(a). However, trimming does not alter the signal mass spectrum drastically, and any signal loss near the top-mass peak is due to events in which the top quark is not boosted enough to have all three hadronic decay products fall within \( R = 1.0 \).

In order to look at events with a reduced \( W + \) jets background, a \( b \)-tagging requirement on at least one \( \text{anti-} k_t \) jet in the event with \( R = 0.4 \) is applied in addition to the selection criteria described in section 5.3.1. Figure 46 shows the effect of trimming on the mass of the leading-\( p_T^{\text{jet}} \) \( \text{anti-} k_t \) jet with \( R = 1.0 \) in a sample of nearly pure \( \bar{t}t \) events. Trimming clearly enhances the mass discrimination compared to the ungroomed case, with a peak at low mass corresponding to large-\( R \) jets containing one quark or gluon (probably from a fully leptonic \( \bar{t}t \) event) and a peak around the top mass, where all three top decay products, the \( b \)-jet and hadronic \( W \)-decay daughters, fall inside the large-\( R \) jet radius.
Figures 47 and 48 show the distributions of $\sqrt{d_{12}}$ and $\sqrt{d_{32}}$, respectively, before and after trimming. Again, the top-quark distribution remains relatively unaffected by trimming, while the $W+$ jets background is pushed to lower values; however, the effect is smaller for these variables than for the mass. Also shown are the distributions with the simulated background subtracted. Comparing the shape before and after grooming with the signal distributions shown in figures 33 and 34 for very high-$p_T$ top quarks, it is apparent that not all top quarks in this data sample were sufficiently boosted to have their decay products fall within a single $R = 1.0$ jet.

Figure 49 shows the signal-enriched distributions of $\tau_{32}$ before and after trimming. Here, there is less discrimination between the $W+$ jets background and the top-quark signal. This is due to the fact that multiple quark and gluon jets in $W$ events can be reconstructed as one $R = 1.0$ anti-$k_t$ jet and mimic the subjettiness signature of the large-$R$ jet containing the hadronic decay products of the top quark. For shape comparisons with figure 35, the background-subtracted plots are also shown.

5.3.3 Application of the HEPTopTagger in $t\bar{t}$ events

Basic kinematic distributions for large-$R$ jets before and after applying the HEPTopTagger algorithm, but without any $b$-tagging requirement, are shown in this section. Top-quark
shows the variables used in the requirements imposed by the HEPTopTagger, respectively.

50(b) boson t (c) 2f threshold. The large-W (d) anti-

subjets are the decay products of a anti-

jet configurations (see table 5.3.1). The large-

W. (d) anti-

and shows the top-candidate mass distribution after applying the HEPTopTagger. Splitting scale (b) anti-

Z is expected, due to the larger jet radius and

selection is obtained for top-quark candidate masses above 120 GeV.

For all shows the jet mass distributions of C/A jets show the jet mass distributions of C/A jets

with four different filtering and large-R jet configurations (see table 2 for details). For all

settings, the top-mass peak shape is generally well described by MC simulation and a

relatively pure t\bar{t} selection is obtained for top-quark candidate masses above 120 GeV.

The good agreement between data and simulation both before and after applying the

HEPTopTagger shows that the exploited substructure is modelled well, even for jets with a very large radius.

Figure 52 shows the variables used in the requirements imposed by the HEPTopTagger

on the subjet mass ratios, defined in eq. (1.6) and eq. (1.7). For example in figure 52(b), if

the sub-leading p_T and the sub-sub-leading p_T subjets are the decay products of a W boson
then \( m_{23}/m_{123} \) peaks at \( m_W/\langle p_t \rangle \sim 0.46 \). The \( m_W \) distribution in figure 52(c) is obtained by taking the sub jet pair with the invariant mass closest to the true \( m_W \), which also influences the shapes of the background distributions. All distributions are well modelled by the simulation.
Figure 51. Top candidate mass distribution after the HEPTopTagger procedure (before applying a \( m_t \) window). The distance parameter for large-\( R \) jet finding is \( R = 1.5 \) in (a), (c), and (d), and \( R = 1.8 \) in (b). Distributions (a) and (b) use the default filtering settings, whereas (c) and (d) have been optimized for high purity and high signal efficiency, respectively.

The efficiency of the HEPTopTagger is measured as a function of the transverse momentum of the generated top quark, and is the product of the large-\( R \) jet finding efficiency and the efficiency to tag the jet correctly: \( \varepsilon(\text{total}) = \varepsilon(\text{large-} R \text{ jet}) \cdot \varepsilon(\text{tag}) \). Figure 53 shows \( \varepsilon(\text{total}) \) for four different filtering configurations of the HEPTopTagger as a function of the generator-level true top-quark \( p_T \) for the \( t\bar{t} \) MC sample. The efficiency for the default settings is 20\% at 250 GeV and reaches a plateau of 40\% at 500 GeV. Below 400 GeV the efficiency can be improved by 5\% by using a larger radius parameter of \( R = 1.8 \). The maximum efficiency for the tight filtering settings is 30\%.

The fake efficiency, shown in figure 54(a), is defined in exactly the same way but is evaluated using the PYTHIA inclusive jet sample. The \( p_T \) of the leading-\( p_T^{\text{jet}} \) anti-\( k_T \) jet with \( R = 0.4 \) has been chosen to compute the efficiency as it provides a measure for the energy available in the event and is easily comparable between different tagging approaches. The fake efficiency shows a sharp turn-on around 200 GeV with efficiencies below 0.5\% below and a plateau of 4\% (2.5\%) for the default and loose (tight) filtering settings.

Figure 54(b) shows the fake tagging efficiency as a function of the top-jet \( p_T \) in a multi-jet background sample and for events with a hadronically decaying \( W \) boson. The
fake efficiency rises sharply at 300 GeV and reaches a plateau of 2.5% at a large-\(R\) jet \(p_T\) of 400 GeV.

The efficiency of the HEPTopTagger to select jets from boosted top quarks can be increased by varying the filtering parameters. Since this also increases the fake efficiency, the optimal working-point depends on the analysis in question.

6 Conclusions

It has been demonstrated experimentally in this paper that jet grooming algorithms can improve the identification of Lorentz-boosted physics objects that decay to jets, as well as increase sensitivity to several new physics processes. The performance of large-\(R\) jets is improved overall, and the dependence on pile-up and the underlying event is reduced.

Jet mass calibrations have been derived in simulation for various large-\(R\) jet algorithms, subjets, as well as for jets with grooming applied. These have been validated for boosted \(W\) bosons in \(t\bar{t}\) events and using calorimeter-jet versus track-jet double ratios. Uncertainties
Figure 53. Tagging efficiency per top quark as a function of the generator-level top-quark $p_T$ for various filtering settings of the HEPTopTagger, evaluated using the semileptonic $t\bar{t}$ MC sample.

Figure 54. (a) Per-event fake efficiency as a function of the leading-$p_T$ anti-$k_t$ $R=0.4$ jet $p_T$ in the event for different filtering settings, measured using the dijet MC sample and (b) a comparison of the per-jet fake efficiency between the dijet sample (PYTHIA) and the $W \rightarrow q\bar{q}$ sample (HERWIG+JIMMY) taken from $t\bar{t}$ events, both using the default filtering.

on the jet energy scale and the jet mass scale have been provided over a wide range of large-$R$ jet momentum.

The mass distributions observed in data for large-$R$ jets in the inclusive jet sample, before and after grooming, are well reproduced by the ATLAS simulation, especially using the POWHEG NLO generator. The substructure variables presented here also show good agreement between data and simulation, typically within 5% for key observables for modern
NLO plus parton shower Monte Carlo programs such as POWHEG+PYTHIA, as well as for the LO MC program HERWIG++. The parameters of trimmed, pruned and mass-drop filtered jet algorithms have been optimized for searches and precision Standard Model measurements using various performance measures. Among the configurations tested here, the trimming algorithm exhibits better performance than the pruning algorithm, with superior mass resolution and reduced dependence on pile-up. In particular, the anti-$k_t$ algorithm with $R = 1.0$ and trimming parameters $f_{cut} = 0.05$ and $R_{sub} = 0.3$ is recommended for boosted top physics analyses, where a minimum $p_T$ requirement of 200 GeV is typical. It is important to note, though, that only the $k_t$-pruning for large-$R$ ($R = 1.0$) jets has been tested in this work, and that future studies should expand the comparisons to include the C/A-pruning as well. Additionally, C/A jets with $R = 1.2$ using the mass-drop filtering parameter $\mu_{trac} = 0.67$ are recommended for boosted two-pronged analyses such as $H \rightarrow b\bar{b}$ or searches using $W \rightarrow q\bar{q}$.

The benefit of using these grooming algorithms along with substructure variables has been demonstrated in top-tagging studies, where the efficiency of finding a boosted top quark for a given background jet mis-tag rate is greatly increased after grooming is applied due to the improved mass resolution. Grooming has been shown to leave the boosted signal mass peak relatively unaffected while systematically shifting the light-quark and gluon jet background lower in mass, thus increasing the discrimination of signal from background.

The HEPTopTagger has been demonstrated to be a robust and versatile tool to reconstruct hadronically-decaying top quarks in the presence of the underlying event and pile-up, using jet grooming and substructure techniques. A comparison to data shows that the algorithm is well modelled by the simulations. The HEPTopTagger performance (efficiency, rejection, mass resolution) can be optimized for a given analysis by varying the algorithm parameters.

From the studies presented here, groomed jets and substructure variables are ready to be used in further ATLAS physics analyses. These techniques will become extremely beneficial tools in upcoming searches for boosted physics objects in supersymmetric and exotic models, measurements of boosted Higgs topologies, and in detailed precision Standard Model measurements of QCD and electroweak processes with jets and boosted hadronic objects.

Acknowledgments

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF, DNSRC and Lundbeck Foundation, Denmark; EPLANET, ERC and NSRF, European Union; IN2P3-CNRS, CEA-DSM/IRFU, France; GNSF, Georgia; BMBF, DFG, HGF, MPG and AvH Foundation, Germany; GSRT and NSRF, Greece; ISF, MINERVA,
GIF, DIP and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; BRF and RCN, Norway; MNiSW, Poland; GRICES and FCT, Portugal; MERSYS (MECTS), Romania; MES of Russia and ROSATOM, Russian Federation; JINR; MSTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SER, SNSF and Cantons of Bern and Geneva, Switzerland; NSC, Taiwan; TAEK, Turkey; STFC, the Royal Society and Leverhulme Trust, United Kingdom; DOE and NSF, United States of America.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA) and in the Tier-2 facilities worldwide.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References


[38] Y. Bai and J. Shelton, Composite octet searches with jet substructure, JHEP 07 (2012) 067 [arXiv:1107.3563] [INSPIRE].


ATLAS collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, 2008 *JINST* 3 S08003 [nSPIRE].


[96] ATLAS collaboration, A search for $t\bar{t}$ resonances in lepton+jets events with highly boosted top quarks collected in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, JHEP 09 (2012) 041 [arXiv:1207.2409] [inSPIRE].


Institut de Física d’Altes Energies and Departament de Física de la Universitat Autónoma de Barcelona, Barcelona, Spain

(a) Institute of Physics, University of Belgrade, Belgrade; (b) Vinca Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia

Department for Physics and Technology, University of Bergen, Bergen, Norway

Physics Division, Lawrence Berkeley National Laboratory and University of California, Berkeley CA, United States of America

Department of Physics, Humboldt University, Berlin, Germany

Albert Einstein Center for Fundamental Physics and Laboratory for High Energy Physics, University of Bern, Bern, Switzerland

School of Physics and Astronomy, University of Birmingham, Birmingham, United Kingdom

(a) Department of Physics, Bogazici University, Istanbul; (b) Department of Physics, Dogus University, Istanbul; (c) Department of Physics Engineering, Gaziantep University, Gaziantep, Turkey

(a) INFN Sezione di Bologna; (b) Dipartimento di Fisica e Astronomia, Università di Bologna, Bologna, Italy

Physikalisches Institut, University of Bonn, Bonn, Germany

Department of Physics, Boston University, Boston MA, United States of America

Department of Physics, Brandeis University, Waltham MA, United States of America

(a) Universidade Federal do Rio De Janeiro COPPE/EE/IF, Rio De Janeiro; (b) Federal University of Juiz de Fora (UFJF), Juiz de Fora; (c) Federal University of Sao Joao del Rei (UFJJ), Sao Joao del Rei; (d) Instituto de Fisica, Universidade de Sao Paulo, Sao Paulo, Brazil

Physics Department, Brookhaven National Laboratory, Upton NY, United States of America

(a) National Institute of Physics and Nuclear Engineering, Bucharest; (b) University Politehnica Bucharest, Bucharest; (c) West University in Timisoara, Timisoara, Romania

Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina

Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom

Department of Physics, Carleton University, Ottawa ON, Canada

CERN, Geneva, Switzerland

Enrico Fermi Institute, University of Chicago, Chicago IL, United States of America

(a) Departamento de Física, Pontificia Universidad Católica de Chile, Santiago; (b) Departamento de Física, Universidad Técnica Federico Santa María, Valparaíso, Chile

(a) Institute of High Energy Physics, Chinese Academy of Sciences, Beijing; (b) Department of Modern Physics, University of Science and Technology of China, Anhui; (c) Department of Physics, Nanjing University, Jiangsu; (d) School of Physics, Shandong University, Shandong; (e) Physics Department, Shanghai Jiao Tong University, Shanghai, China

Laboratoire de Physique Corpusculaire, Clermont Université and Université Blaise Pascal and CNRS/IN2P3, Clermont-Ferrand, France

Nevis Laboratory, Columbia University, Irvington NY, United States of America

Niels Bohr Institute, University of Copenhagen, Kobenhavn, Denmark

(a) INFN Gruppo Collegato di Cosenza; (b) Dipartimento di Fisica, Università della Calabria, Rende, Italy

(a) AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Krakow; (b) Marian Smoluchowski Institute of Physics, Jagiellonian University, Krakow, Poland

The Henryk Niewodniczanski Institute of Nuclear Physics, Polish Academy of Sciences, Krakow, Poland

Physics Department, Southern Methodist University, Dallas TX, United States of America

Physics Department, University of Texas at Dallas, Richardson TX, United States of America

DESY, Hamburg and Zeuthen, Germany

Institut für Experimentelle Physik IV, Technische Universität Dortmund, Dortmund, Germany

Institut für Kern- und Teilchenphysik, Technische Universität Dresden, Dresden, Germany

Department of Physics, Duke University, Durham NC, United States of America