Jet mass and substructure of inclusive jets in $s = 7$ TeV pp collisions with the ATLAS experiment


Published in:
The Journal of High Energy Physics

DOI: 10.1007/JHEP05(2012)128

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)

Download date: 06 Dec 2018
Jet mass and substructure of inclusive jets in \( \sqrt{s} = 7 \text{ TeV} \) \( pp \) collisions with the ATLAS experiment

The ATLAS collaboration

ABSTRACT: Recent studies have highlighted the potential of jet substructure techniques to identify the hadronic decays of boosted heavy particles. These studies all rely upon the assumption that the internal substructure of jets generated by QCD radiation is well understood. In this article, this assumption is tested on an inclusive sample of jets recorded with the ATLAS detector in 2010, which corresponds to 35 pb\(^{-1}\) of \( pp \) collisions delivered by the LHC at \( \sqrt{s} = 7 \text{ TeV} \). In a subsample of events with single \( pp \) collisions, measurements corrected for detector efficiency and resolution are presented with full systematic uncertainties. Jet invariant mass, \( k_t \) splitting scales and \( N \)-subjettiness variables are presented for anti-\( k_t \) \( R = 1.0 \) jets and Cambridge-Aachen \( R = 1.2 \) jets. Jet invariant-mass spectra for Cambridge-Aachen \( R = 1.2 \) jets after a splitting and filtering procedure are also presented. Leading-order parton-shower Monte Carlo predictions for these variables are found to be broadly in agreement with data. The dependence of mean jet mass on additional \( pp \) interactions is also explored.

KEYWORDS: Hadron-Hadron Scattering
1 Introduction

The ATLAS experiment observes proton-proton ($pp$) collisions provided by the Large Hadron Collider (LHC). The outcome of these collisions is frequently the production of large numbers of hadrons. In order to understand these collisions, studies usually group hadrons into jets defined by one of a number of standard algorithms [1–7]. The variables most often used in analyses are the jet direction and momentum transverse to the beam ($p_T$). However the jets remain composite objects and their masses and internal substructure contain additional information.

One strong motivation for studies of the internal substructure of jets is that at the LHC particles such as $W$ and $Z$ bosons and top quarks are produced abundantly with significant Lorentz boosts. The same may also be true for new particles produced at the LHC. When
such particles decay hadronically, the products tend to be collimated in a small area of the
detector. For sufficiently large boosts, the resulting hadrons can be clustered into a single
jet. Substructure studies offer a technique to extract these single jets of interest from the
overall jet background. Such techniques have been found promising for boosted $W$ decay
identification, Higgs searches and boosted top identification amongst others [8]. However,
many of these promising approaches have never been tested with collision data and rely on
the assumption that the internal structure of jets is well modelled by parton-shower Monte
Carlo approaches. It is therefore important to measure some of the relevant variables in a
sample of jets to verify the expected features.

In this paper, measurements are made with an inclusive sample of high-transverse
momentum jets produced in proton-proton collisions with a centre-of-mass energy ($\sqrt{s}$) of
7 TeV. This is a natural continuation of the studies in previous experiments [9–13]. It also
complements previous ATLAS studies [14] probing the shape of jets reconstructed with the
anti-$k_t$ algorithm [5] with smaller radius parameters $R = 0.4$ and 0.6.

This study focuses on two specific jet algorithms that are likely to be of interest for
future searches: anti-$k_t$ jets with an $R$-parameter of 1.0 and Cambridge-Aachen [3, 4] jets
with $R = 1.2$. Jets are required to be at high-transverse momentum ($p_T > 200$ GeV)
and central in rapidity\(^1\) ($|y| < 2$). The normalised cross-section as a function of jet mass,
taken from the jet four-momentum, is measured for both these algorithms. In addition
to the mass, two sets of substructure variables, $k_t$ splitting scales [15] and \(N\)-subjettiness
ratios [16], are measured. For the Cambridge-Aachen jets, the mass distribution after a
substructure splitting and filtering procedure [17] is also presented.

## 2 Definitions

### 2.1 Jet algorithms

Jets are constructed using two infrared and collinear-safe recombination jet algorithms,
anti-$k_t$ and Cambridge-Aachen, as implemented in the FastJet package [18]. Both act
by iteratively merging the nearest objects in the event but with different definitions of
“distance”. The anti-$k_t$ jet algorithm builds jets with a very regular shape, while the
Cambridge-Aachen algorithm builds less regular jets. This analysis uses $R$-parameters of
1.0 and 1.2, for anti-$k_t$ and Cambridge-Aachen jets respectively, in line with recent studies
of heavy boosted objects. All discussion of jet algorithms in this paper assumes that the
constituents from which the jet is composed are recombined using four-vector addition.

Jets in the real or simulated detector are constructed from clusters of energy deposited in
the calorimeter (see section 4). Jets in the “true” final state, or “hadron level”, are defined
as jets made from all particles with a proper lifetime longer than 10 ps, including neutrinos.

\(^1\) ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in
the centre of the detector and the $z$-axis along the beam pipe. The $x$-axis points from the IP to the centre
of the LHC ring, and the $y$-axis points upward. Cylindrical coordinates ($r,\phi$) are used in the transverse
plane, $\phi$ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the
polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Rapidity is $y = 1/2 \times \ln((E + p_z)/(E - p_z))$. 

- 2 -
2.2 The splitting and filtering procedure

The “splitting and filtering” procedure aims to identify relatively hard, symmetric splittings in a jet that contribute significantly to the jet’s invariant mass. This procedure is taken from recent Higgs search studies [17, 19]. The parameters are tuned to maximise sensitivity to a Standard Model Higgs boson decaying to $b\bar{b}$, but this procedure is suitable generally for identifying two-body decay processes. The effect of the procedure is to search for jets where clustering the constituents with Cambridge-Aachen combines two relatively low mass objects to make a much more massive object. This indicates the presence of a heavy particle decay. The procedure then attempts to retain only the constituents believed to be related to the decay of this particle. Because the procedure itself uses the Cambridge-Aachen algorithm, it is most natural to apply it to jets originally found with this algorithm.

Each stage in the clustering combines two objects $j_1$ and $j_2$ to make another object $j$. Use definitions $v = \frac{\min(p_{T,j_1}^2, p_{T,j_2}^2)}{m_j^2} \delta R_{j_1,j_2}^2$ and $\delta R_{j_1,j_2} = \sqrt{\delta y_{j_1,j_2}^2 + \delta \phi_{j_1,j_2}^2}$, where $\delta y$ and $\delta \phi$ are the differences in rapidities and azimuthal angles respectively. The procedure takes a jet to be the object $j$ and applies the following:

1. Undo the last clustering step of $j$ to get $j_1$ and $j_2$. These are ordered such that their mass has the property $m_{j_1} > m_{j_2}$. If $j$ cannot be unclustered (i.e. it is a single particle) or $\delta R_{j_1,j_2} < 0.3$ then it is not a suitable candidate, so discard this jet.

2. If the splitting has $m_{j_1}/m_j < \mu$ (large change in jet mass) and $v > v_{\text{cut}}$ (fairly symmetric) then continue, otherwise redefine $j$ as $j_1$ and go back to step 1. Both $\mu$ and $v_{\text{cut}}$ are parameters of the algorithm.

3. Recluster the constituents of the jet with the Cambridge-Aachen algorithm with an $R$-parameter of $R_{\text{filt}} = \min(0.3, \delta R_{j_1,j_2}/2)$ finding $n$ new subjets $s_1, s_2 \ldots s_n$ ordered in descending $p_T$.

4. Redefine the jet as the sum of subjet four-momenta $\sum_{i=1}^{\min(n,3)} s_i$.

The algorithm parameters $\mu$ and $v_{\text{cut}}$ are taken as 0.67 and 0.09 respectively [19].

The $\mu$ cut attempts to identify a hard structure in the distribution of energy in the jet, which would imply the decay of a heavy particle. The cut on $v$ further helps by suppressing very asymmetric decays of the type favoured by splittings of quarks and gluons. A notable modification of the original procedure [17] in this paper is the addition of the $\delta R_{j_1,j_2}$ cut in step 1. This cut is applied because with current techniques the correction for detector resolution at angular scales below 0.3 is not well controlled. Steps 3 and 4 filter out some of the particles in the candidate jet, the aim being to retain particles relevant to the hard process while reducing the contribution from effects like underlying event and pile-up. The 4-vector after step 4 can be treated like a new jet. This new jet has a $p_T$ and mass less than or equal to those of the original jet.
2.3 $k_t$ splitting scales, $\sqrt{d_{ij}}$

The $k_t$ splitting scales are defined by reclustering the constituents of the jet with the $k_t$ recombination algorithm \cite{1,2}. The $k_t$-distance of the final clustering step can be used to define a splitting scale variable $\sqrt{d_{ij}}$:

\[
\sqrt{d_{ij}} = \min(p_T_{j1}, p_T_{j2}) \times \delta R_{j1,j2},
\]

where 1 and 2 are the two jets before the final clustering step \cite{15}. The ordering of clustering in the $k_t$ algorithm means that in the presence of a two-body heavy particle decay the final clustering step will usually be to combine the two decay products. The parameter $\sqrt{d_{ij}}$ can therefore be used to distinguish heavy particle decays, which tend to be more symmetric, from the largely asymmetric splittings of quarks and gluons. The expected value for a heavy particle decay is approximately $m/2$, whereas inclusive jets will tend to have values $\sim p_T/10$, although with a tail extending to high values. The variable $\sqrt{d_{23}}$ is defined analogously but for the two objects combined in the penultimate clustering step.

2.4 $N$-subjettiness

The $N$-subjettiness variables $\tau_N$ \cite{16} are designed to be smooth, continuous observables related to the subjet multiplicity. Intuitively, the variables can be thought of as answering the question: “How much does this jet look like $N$ different subjets?” The variable $\tau_N$ is calculated by clustering the constituents of the jet with the $k_t$ algorithm and requiring $N$ subjets to be found. These $N$ subjets define axes within the jet around which the jet constituents may be concentrated. The variables $\tau_N$ are then defined as the following sum over all constituents $k$ of the jet:

\[
\tau_N = \frac{1}{d_0} \sum_k p_T_{k} \times \min(\delta R_{1,k}, \delta R_{2,k}, \ldots, \delta R_{N,k}) \quad (2.1)
\]

\[
d_0 = \sum_k p_T_{k} R, \quad (2.2)
\]

where $\delta R_{i,k}$ is the distance from the subjet $i$ to the constituent $k$ and $R$ is the $R$-parameter of the original jet algorithm.

Using this definition, $\tau_N$ describes how well the substructure of the jet is described by $N$ subjets by assessing the degree to which constituents are localized near the axes defined by the $k_t$ subjets. For two- and three-body decays, respectively, the ratios $\tau_2/\tau_1$ and $\tau_3/\tau_2$ have been shown to provide excellent discrimination for hadronic decays of $W$-bosons and boosted top quarks \cite{20}. These ratios will be referred to as $\tau_{21}$ and $\tau_{32}$ respectively. These variables mostly fall within the range 0 to 1. As an example, $\tau_{21} \approx 1$ corresponds to a jet which is narrow and without substructure; $\tau_{21} \approx 0$ implies a jet which is much better described by two subjets than one. Similarly low values of $\tau_{32}$ imply a jet which is much better described by three subjets than two. However, as can be seen from the definition, adding an additional subjet axis will tend to reduce the value of $\tau_N$ and therefore even narrow jets tend to have values of $\tau_{21}$ and $\tau_{32}$ slightly less than 1.
3 The ATLAS detector

The ATLAS detector [21] provides nearly full solid angle coverage around the collision point with tracking detectors, calorimeters and muon chambers. Of these subsystems the most relevant to this study are the inner detector, the barrel and endcap calorimeters, and the trigger system.

The inner detector is a tracking detector covering the range $|\eta| < 2.5$ and with full coverage in $\phi$. It is composed of a silicon pixel detector, a silicon microstrip detector and a transition radiation tracker. The whole system is immersed in a 2 T magnetic field. The information from the inner detector is used to reconstruct tracks and vertices.

The barrel and endcap calorimeters cover the regions $|\eta| \lesssim 1.5$ and $1.5 \lesssim |\eta| < 3.2$, respectively. Electromagnetic measurements are provided by a liquid-argon (LAr) sampling calorimeter. The granularity of this detector ranges from $\delta \eta \times \delta \phi = 0.025 \times 0.025$ to $0.1 \times 0.1$. Hadronic calorimetry in $|\eta| < 1.7$ is provided by a scintillating-tile detector, while in the endcaps, coverage is provided by a second LAr system. The granularity of the hadronic calorimetry ranges from $0.1 \times 0.1$ to $0.2 \times 0.2$.

The trigger system [22] is composed of three consecutive levels. Only the Level-1 (L1) trigger is used in this study, with higher levels not rejecting any events. The L1 trigger is based on custom-built hardware that processes events with a fixed latency of 2.5 $\mu$s. Events in this analysis are selected based on their L1 calorimeter signature. The L1 calorimeter trigger uses coarse detector information to identify interesting physics objects above a given transverse energy ($E_T$) threshold. The jet triggers use a sliding window algorithm taking square $\delta \eta \times \delta \phi = 0.2 \times 0.2$ jet elements as input. The window size is $0.8 \times 0.8$.

4 Dataset and reconstruction

The data analysed here come from the 2010 $\sqrt{s} = 7$ TeV $pp$ dataset. Data are used in this study only if the detector conditions were stable, there was a stable beam present in the LHC, the luminosity was reliably monitored and the trigger was operational. The selected data set corresponds to an integrated luminosity of $35.0 \pm 1.1$ pb$^{-1}$ [23, 24].

Events in this analysis are first selected by the L1 calorimeter trigger system. The efficiency of this trigger was evaluated in data and found to contain no significant biases for the selection used here. For the lowest $p_T$ bin (200–300 GeV) a trigger is used which was only available for part of the dataset. As a result some plots are presented with the lower integrated luminosity of $2.0 \pm 0.1$ pb$^{-1}$.

To reject events that are dominated by detector noise or non-collision backgrounds, events are required to contain a primary vertex consistent with the LHC beampot, reconstructed from at least five tracks with $p_T > 150$ MeV. Additionally, jets are reconstructed with the anti-$k_t$ algorithm using an $R$-parameter of 0.6. Events are discarded if any such jet with transverse momentum greater than 30 GeV fails to satisfy a number of quality criteria, including requirements on timing and calorimeter noise [25]. This selection removes approximately 3% of events in this dataset.
Additional proton-proton collisions (pile-up) can have a significant impact on quantities like jet mass and substructure [8]. The primary results in this paper are therefore presented only in events where the number of reconstructed primary vertices ($N_{\text{PV}}$) composed of at least five tracks is exactly one. This requirement selects approximately 22% of events in the 2010 dataset. As vertex finding is highly efficient, this approach is expected to be very good at rejecting pile-up, and no additional systematic uncertainties as a result of this requirement are considered. The effects of pile-up are discussed in more detail in section 10.

Calorimeter cells are clustered using a three-dimensional topological algorithm. These clusters provide a three-dimensional representation of energy depositions in the calorimeter with a nearest neighbour noise suppression algorithm [26]. The resulting clusters are made massless and then classified as either electromagnetic or hadronic in origin based on their shape, depth and energy density. Cluster energies are corrected with calibration constants, which depend on the cluster classification to account for calorimeter non-compensation [25]. The clusters are then used as input to a jet algorithm.

As part of this study, specific calibrations for these jet algorithms have been devised. Calibrations for the mass, energy and $\eta$ of jets are derived from Monte Carlo (specifically PYTHIA [27]). Hadron-level jets (excluding muons and neutrinos) are matched to jets reconstructed in the simulated calorimeter. The matched pairs are used to define functions for these three variables, dependent on energy and $\eta$, which on average correct the reconstructed quantities back to the true scale. This correction is of the order 10-20% for mass and energy and 0.01 for $\eta$.

Jets constructed from tracks are used for systematic studies in this paper. These track-jets are constructed using the same algorithms as calorimeter jets. The input constituents are inner-detector tracks originating only from the selected $pp$ collision of interest as selected by the criteria $p_T > 500$ MeV, $|\eta| < 2.5$, $|z_0| < 5$ mm and $|d_0| < 1.5$ mm [28]. Here $z_0$ and $d_0$ are the longitudinal and transverse impact parameter of the track at closest approach to the $z$-axis, relative to the primary vertex.

The measurements presented in this paper are for jets that have $|y| < 2$ in four 100 GeV $p_T$ bins spanning 200 to 600 GeV. This selection is not biased by trigger effects and the jets it selects are contained entirely within the barrel and end-cap subdetectors.

5 Monte Carlo samples

Samples of inclusive jet events were produced using several Monte Carlo (MC) generators including PYTHIA 6.423 [27] and HERWIG++ 2.4 [29]. These programs implement leading-order (LO) perturbative QCD (pQCD) matrix elements for $2 \to 2$ processes. Additionally, ALPGEN 2.13 [30] and SHERPA 1.2.3 [31] are used for some cross-checks. SHERPA and ALPGEN implement $2 \to n$ processes such as explicit QCD multijet production. Parton-shower are calculated in a leading-logarithm approximation. Showers are $p_T$ ordered in PYTHIA and angular ordered in HERWIG++. Fragmentation into particles is implemented in PYTHIA following the string model [32] and in HERWIG++ the cluster [33] model. ALPGEN is interfaced with HERWIG [34, 35] for parton-shower and fragmentation and JIMMY [36] for an underlying event model. Unless otherwise specified, PYTHIA samples
use the AMBT1 tune [28]. In some figures the Perugia2010 PYTHIA tune is used [37], which has been found to describe jet shapes more accurately at ATLAS [14]. Leading-order parton density functions are taken from the MRST2007 LO* set [38, 39], unless stated otherwise. No pile-up was included in any of these samples.

The MC generated samples are passed through a full simulation [40] of the ATLAS detector and trigger, based on GEANT4 [41]. The Quark Gluon String Precompound (QGSP) model is used for the fragmentation of nuclei, and the Bertini cascade (BERT) model for the description of the interactions of the hadrons in the medium of the nucleus [42].

6 Detector-level distributions

Detector-level distributions for jet $p_T$, $\eta$, mass, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, $\tau_{21}$ and $\tau_{32}$ are shown in figures 1–6. The statistical uncertainty represented in ratios is that from Monte Carlo and data added in quadrature. Representative distributions of the substructure variables are shown for the 300–400 GeV bin only. The Monte Carlo is normalised to the data separately in each plot. The properties of these jets are observed to be reasonably well modelled by leading-order parton-shower Monte Carlo. There are approximately four times fewer split and filtered jets (e.g. figure 3) because many jets fail the splitting criteria described above.

7 Systematic uncertainties

The modelling of the calorimeter response is the biggest systematic uncertainty for this analysis. The key issue therefore is to validate the Monte Carlo-based jet calibration described in section 4. As the results here use jet algorithms with larger $R$-parameters, the ATLAS jet energy scale uncertainty [25] for anti-$k_t$ $R = 0.4$ and 0.6 jets cannot be applied.
The primary systematic uncertainties considered in the present study are those relating to scales and resolutions, such as jet $p_T$ scale (JES) and jet $p_T$ resolution (JER). For each substructure variable, the scale and resolution of the variable itself are also considered, for example the jet mass scale (JMS) and jet mass resolution (JMR). The scale uncertainties are primarily constrained by in-situ validation using track-jets. The inner detector and calorimeter have largely uncorrelated systematic effects, therefore comparison of variables such as jet mass and energy between the two sub-detectors allows for some separation of physics and detector effects. This technique is limited to a precision of around 3-5% by systematic uncertainties arising from the inner-detector tracking efficiency and confidence in Monte Carlo modelling of the relative behaviour of the charged and neutral components of jets.
Jets composed from tracks are matched to calorimeter-jets if they are within \( \delta R < 0.3 \) of each other. The split and filtered calorimeter-jets are matched to Cambridge-Aachen \( R = 1.2 \) track-jets. Ratios are defined between track- and calorimeter-jets for each variable \( X (p_T, \text{mass}, \sqrt{d_{12}}, \sqrt{d_{23}}, \tau_{21}, \tau_{32}) \):

\[
r^X = \frac{X_{\text{calorimeter-jet}}}{X_{\text{track-jet}}}.
\]

Example distributions of some of the ratio variables are shown in figure 7. It can be seen that the ratios are in broad agreement between data and Monte Carlo. To quantify the level of agreement, double ratios are defined:

\[
\rho^X = \frac{r^X_{\text{data}}}{r^X_{\text{MC}}}.
\]

Figure 4. Mass distributions for jets with \(|y| < 2.0\) in the 300–400 GeV \( p_T \) bin. Jets shown are Cambridge-Aachen (top left), Cambridge-Aachen after splitting and filtering (top right) and anti-\(k_t\) (bottom).
Figure 5. Distributions for $\sqrt{d_{12}}$ (left) and $\sqrt{d_{23}}$ (right) of anti-$k_t$ $R = 1.0$ jets with $|y| < 2.0$ in the 300–400 GeV $p_T$ bin.

Figure 6. Distributions for $\tau_{21}$ (left) and $\tau_{32}$ (right) of jets with $|y| < 2.0$ in the 300–400 GeV $p_T$ bin for anti-$k_t$ (top) and Cambridge-Aachen jets (bottom).
Figure 7. The ratio of a jet property determined by the calorimeter to that determined by tracks versus the calorimeter jet mass for jets with 300–400 GeV in $p_T$. Shown are the data and a variety of Monte Carlo models. The bottom frame shows the ratio of the Monte Carlo models to data. The top left, top right and bottom left figures show the ratio for jet mass for three different jet algorithms. The bottom-right figure shows the ratios for $\sqrt{d_{12}}$ in anti-$k_T$ jets.

where again, X can be $p_T$, mass or any of the substructure variables. The distributions of the variables $X_{\text{calorimeter-jet}}$ themselves are not necessarily expected to be correctly modelled by Monte Carlo. However, if the simulation correctly models the effect of the detector on these variables, the double ratios $\rho^X$, are expected to be consistent with unity. Figure 7 also shows below each plot the corresponding double ratio. In order to account for possible uncertainties due to different fragmentation and hadronisation models, these double ratios are also calculated with a variety of Monte Carlo programs.

Final scale uncertainties are determined by adding in quadrature the estimated uncertainty on the inner-detector measurement with the deviation from unity observed in the double ratios. The resulting scale uncertainties on $p_T$, mass and substructure variables are in the range 3-6%. The highest $p_T$ bins contain fewer events and therefore suffer from sta-
tical fluctuations when calculating the double ratio deviation. These scale uncertainties tend to dominate the systematic uncertainties on the final measurements.

As an additional cross-check, Monte Carlo-based tests are used to determine the dependence of the detector response on a number of different variables. These include samples produced with modified detector geometry, different GEANT hadronic physics models and different Monte Carlo generators. These tests indicate variations of a similar order of magnitude to those observed in the in-situ studies. The in-situ track-jet study is limited by inner-detector acceptance and only extends as far as $|\eta| < 1.0$, which corresponds to $\approx 75\%$ of the jets in the measured distributions. However, the Monte Carlo-based tests also indicate no strong $\eta$-dependence from any of the different possible types of mismodelling examined. Based on this, the systematic uncertainty is applied to the entire sample.

In-situ tests of the JER \[43\] for anti-$k_t$ jets with $R = 0.4$ and 0.6 indicate that the jet $p_T$ resolution predicted by simulation is in good agreement with that observed in the data. Here, the resolution uncertainties are taken from the Monte Carlo tests described above only, primarily because the mass and substructure variable resolutions are difficult to validate in-situ with this dataset. From studying the variations in resolution created by varying the detector geometry, GEANT hadronic physics model and Monte Carlo generator, resolution uncertainties of around 20% are conservatively estimated, except for $\tau_{21}$ and $\tau_{32}$ where they are around 10%.

8 Data correction

To compare the measurements directly to theoretical predictions the final distributions in this study are corrected for detector resolution and acceptance effects. The procedure here is a matrix-based unfolding technique called Iterative Dynamically Stabilised (IDS) unfolding \[44, 45\].

In this procedure truth jets and reconstructed jets in Monte Carlo simulated events are matched using the criterion $\delta R < 0.2$, which leads to a match for $> 99\%$ of jets. Matched pairs of jets are used to construct a transfer matrix corresponding to the effect of the detector. A true jet can be matched with a reconstructed jet that fails the $p_T$ cut and vice-versa. As such, the efficiency for matching a true jet to a reconstructed jet in the same $p_T$ bin is recorded as a function of the variable of interest. The reverse quantity is also defined for reconstructed jets. The data are then scaled by the reconstructed matching efficiency, multiplied by the transfer matrix and finally divided by the truth matching efficiency. There is also an iterative optimisation step, where the rows of the matrix are scaled to match the corrected result. PYTHIA is used to provide the central value. Each $p_T$ bin is unfolded independently. The systematic uncertainty is assessed by repeating the procedure using SHERPA samples.

9 Results

Using the analysis techniques outlined above, measured normalised cross-sections are shown in figures 8–16. In ratio plots, the statistical uncertainty on Monte Carlo predictions does
Figure 8. Normalised cross-sections as functions of mass of Cambridge-Aachen jets with $R = 1.2$ in four different $p_T$ bins.

not include the data statistical uncertainty. Although in some cases the Monte Carlo predictions are not in agreement with the data, the shapes of the distributions are correctly reproduced. For jet mass the distributions produced by PYTHIA tend to be too soft, while those from HERWIG++ are too hard. Notably, the Cambridge-Aachen jet mass after splitting and filtering, as shown in figure 9, is the only variable for which the Monte Carlo predictions are in agreement to within statistical uncertainties, both with each other and the data. The substructure variables exhibit generally better agreement with Monte Carlo predictions than mass, with all but a few bins correctly described by both PYTHIA and HERWIG++. In the higher $p_T$ bins statistical fluctuations begin to limit the precision of the measurements, but the level of agreement in all variables appears to remain approximately constant between $p_T$ bins.
Figure 9. Normalised cross-sections as functions of mass of Cambridge-Aachen jets with $R = 1.2$ after splitting and filtering in four different $p_T$ bins.

The unfolding technique used introduces correlations between the bins. The statistical uncertainty in these results represents the diagonal element of the covariance matrix only; therefore, comparison to alternative predictions requires use of the full covariance matrices. These matrices are available, along with the full results presented here, in HepData [46]. In particular the structure at a jet mass of 150–180 GeV in the 500–600 GeV bin of figures 8 and 10 is consistent with statistical fluctuations.
Figure 10. Normalised cross-sections as functions of mass of anti-$k_t$ jets with $R = 1.0$ in four different $p_T$ bins.

10 Mean mass with multiple proton-proton interactions

The results presented so far have been for events containing only one $pp$ interaction; however even in this early period of running, the data contain events with multiple simultaneous $pp$ interactions (pile-up) [47]. These additional collisions are uncorrelated with the hard-scattering process that typically triggers the event. They therefore present a background of soft, diffuse radiation that offsets the energy measurement of jets and will impact jet-shape and substructure measurements. It is essential that future studies involving jet-substructure variables, such as those investigated here, be able to understand and correct for the effects of pile-up. Methods to mitigate these effects will be essential for jet multiplicity and energy scale measurements.
Substructure observables are expected to be especially sensitive to pile-up [8]. This is true in particular for the invariant mass of large-size jets. Techniques such as the splitting and filtering procedure used in this study reduce the effective area of large jets and are therefore expected to reduce sensitivity to pile-up.

The sensitivity of mean jet mass to pile-up is tested in this dataset. The correlation of the mean jet mass of anti-$k_t$ jets with the number of reconstructed primary vertices is presented in figure 17 (left). All jets with a $p_T$ of at least 300 GeV in the rapidity range $|y| < 2$ are considered. The mean mass of jets in the absence of pile-up and the variation with pile-up activity show the expected dependence on the jet size. The mean mass in the
Figure 12. Normalised cross-sections as functions of $\sqrt{d^2_{23}}$ of anti-$k_t$ jets with $R = 1.0$ in four different $p_T$ bins.

$N_{PV} = 1$ bin increases linearly with $R$. The ratios of the fitted slopes $s_R$ are found to be:

\[
s_{1.0}/s_{0.6} = 4.3 \pm 0.5
d_{1.0}/d_{0.4} = 13 \pm 3
d_{0.6}/d_{0.4} = 3.0 \pm 0.8
\]

in good agreement with the ratio of the third power of the jet $R$-parameter. This is in agreement with predictions of scaling of the mean mass [48, 49]. This behaviour can also be qualitatively explained by two factors. Firstly the jet area in the $y-\phi$ plane grows roughly as $R^2$. Moreover, the contribution of these particles to the jet mass scales with the distance between them approximately as $R/2$, giving another power of $R$. 

\[
(1.0/0.6)^3 = 4.6, \\
(1.0/0.4)^3 = 15.6, \\
(0.6/0.4)^3 = 3.4
\]
Figure 13. Normalised cross-sections as functions of $\tau_{21}$ of Cambridge-Aachen jets with $R = 1.2$ in four different $p_T$ bins.

Figure 17 (right) shows the dependence on $N_{PV}$ of the mean jet mass before and after the splitting and filtering procedure for Cambridge-Aachen jets. Since the angular requirement $R_{jj} > 0.3$ is imposed, the splitting steps of this procedure naturally select more massive jets. Since the splitting procedure selects a kinematically biased subset of jets, a third line shows the mean mass prior to filtering of jets that pass the splitting. The filtering step significantly reduces the impact of pile-up on mean jet mass. In fact, the slope of the straight line fitted to the filtered jet data points is statistically consistent with zero.

Altogether, this demonstrates that the pile-up dependence of mean jet mass in real LHC conditions matches expectations. Additionally, jet substructure techniques that reduce the area of jets are promising for suppressing the effects of pile-up.
11 Conclusions

Jet mass and several jet substructure variables have been measured. This is the first particle-level measurement of these variables at the LHC and in many cases the first at any experiment. There is broad agreement between data and leading-order parton-shower Monte Carlo predictions from Pythia and Herwig++, although there is some scope to improve this. Jet mass has generally been found to exhibit the largest disagreements with Monte Carlo simulations. However, in contrast to this, the masses of jets after the Cambridge-Aachen splitting and filtering procedure display good agreement both with and
between Monte Carlo simulations. The substructure variables $\sqrt{d_{12}}, \sqrt{d_{23}}, \tau_{21}$ and $\tau_{32}$ are all reasonably well reproduced by Monte Carlo predictions. Additionally, the effects of pile-up on mean jet mass have been found to match phenomenological expectations for $R$-parameter dependence. Splitting and filtering has also been found to reduce the impact of pile-up significantly.

Generally these results show that jet mass and substructure quantities can be successfully reproduced by leading-order parton-shower Monte Carlo. This result bodes well for future analyses aiming to make use of jet substructure techniques.

Figure 15. Normalised cross-sections as functions of $\tau_{21}$ of anti-$k_t$ jets with $R = 1.0$ in four different $p_T$ bins.
Figure 16. Normalised cross-sections as functions of $\tau_{32}$ of anti-$k_t$ jets with $R = 1.0$ in four different $p_T$ bins.

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, 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; ARTEMIS and ERC, European Union; IN2P3-CNRS, CEA-DSM/IRFU, France; GNAS, Georgia; BMBF, DFG, HGF,
**Figure 17.** The mean mass for jets with $p_T > 300$ GeV as a function of the number of primary vertices identified in the event. Comparisons show the effect for anti-$k_t$ jets with different $R$-parameters (left) and Cambridge-Aachen $R = 1.2$ jets with and without splitting and filtering procedure (right). Each set of points is fitted with a straight line.

MPG and AvH Foundation, Germany; GSRT, Greece; ISF, MINERVA, GIF, DIP and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; 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 MVZT, 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 (U.K.) and BNL (U.S.A.) 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**


[34] G. Corella et al., HERWIG 6.5 release note, hep-ph/0210213 [SPIRE].


The ATLAS collaboration

1 University at Albany, Albany NY, U.S.A.
2 Department of Physics, University of Alberta, Edmonton AB, Canada
3 (a) Department of Physics, Ankara University, Ankara, Turkey
   (b) Department of Physics, Dumlupinar University, Kutahya, Turkey
   (c) Department of Physics, Gazi University, Ankara, Turkey
   (d) Division of Physics, TOBB University of Economics and Technology, Ankara, Turkey
   (e) Turkish Atomic Energy Authority, Ankara, Turkey
4 LAPP, CNRS/IN2P3 and Université de Savoie, Annecy-le-Vieux, France
5 High Energy Physics Division, Argonne National Laboratory, Argonne IL, U.S.A.
6 Department of Physics, University of Arizona, Tucson AZ, U.S.A.
7 Department of Physics, The University of Texas at Arlington, Arlington TX, U.S.A.
8 Physics Department, University of Athens, Athens, Greece
9 Physics Department, National Technical University of Athens, Zografou, Greece
10 Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
11 Institut de Física d’Altes Energies and Departament de Física de la Universitat Autònoma de Barcelona and ICREA, Barcelona, Spain
12 (a) Institute of Physics, University of Belgrade, Belgrade, Serbia
    (b) Vinca Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia
13 Department for Physics and Technology, University of Bergen, Bergen, Norway
14 Physics Division, Lawrence Berkeley National Laboratory and University of California, Berkeley CA, U.S.A.
15 Department of Physics, Humboldt University, Berlin, Germany
16 Albert Einstein Center for Fundamental Physics and Laboratory for High Energy Physics, University of Bern, Bern, Switzerland
17 School of Physics and Astronomy, University of Birmingham, Birmingham, U.K.
18 (a) Department of Physics, Bogazici University, Istanbul, Turkey
    (b) Division of Physics, Dogus University, Istanbul, Turkey
    (c) Department of Physics Engineering, Gaziantep University, Gaziantep, Turkey
    (d) Department of Physics, Istanbul Technical University, Istanbul, Turkey
19 (a) INFN Sezione di Bologna, Italy
    (b) Dipartimento di Fisica, Università di Bologna, Bologna, Italy
20 Physikalisches Institut, University of Bonn, Bonn, Germany
21 Department of Physics, Boston University, Boston MA, U.S.A.
22 Department of Physics, Brandeis University, Waltham MA, U.S.A.
23 (a) Universidade Federal do Rio De Janeiro COPPE/EE/IF, Rio de Janeiro, Brazil
    (b) Federal University of Juiz de Fora (UFJF), Juiz de Fora, Brazil
    (c) Federal University of Sao Joao del Rei (UFSJ), Sao Joao del Rei, Brazil
    (d) Instituto de Fisica, Universidade de Sao Paulo, Sao Paulo, Brazil
24 Physics Department, Brookhaven National Laboratory, Upton NY, U.S.A.
25 (a) National Institute of Physics and Nuclear Engineering, Bucharest, Romania
    (b) University Politehnica Bucharest, Bucharest, Romania
    (c) West University in Timisoara, Timisoara, Romania
26 Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina
<table>
<thead>
<tr>
<th>Number</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>Cavendish Laboratory, University of Cambridge, Cambridge, U.K.</td>
</tr>
<tr>
<td>28</td>
<td>Department of Physics, Carleton University, Ottawa ON, Canada</td>
</tr>
<tr>
<td>29</td>
<td>CERN, Geneva, Switzerland</td>
</tr>
<tr>
<td>30</td>
<td>Enrico Fermi Institute, University of Chicago, Chicago IL, U.S.A.</td>
</tr>
<tr>
<td>31</td>
<td>(a) Departamento de Física, Pontificia Universidad Católica de Chile, Santiago, Chile</td>
</tr>
<tr>
<td></td>
<td>(b) Departamento de Física, Universidad Técnica Federico Santa María, Valparaíso, Chile</td>
</tr>
<tr>
<td>32</td>
<td>(a) Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, China</td>
</tr>
<tr>
<td></td>
<td>(b) Department of Modern Physics, University of Science and Technology of China, Anhui, China</td>
</tr>
<tr>
<td></td>
<td>(c) Department of Physics, Nanjing University, Jiangsu, China</td>
</tr>
<tr>
<td></td>
<td>(d) School of Physics, Shandong University, Shandong, China</td>
</tr>
<tr>
<td>33</td>
<td>Laboratoire de Physique Corpusculaire, Clermont Université and</td>
</tr>
<tr>
<td></td>
<td>Université Blaise Pascal and CNRS/IN2P3, Aubiere Cedex, France</td>
</tr>
<tr>
<td>34</td>
<td>Nevis Laboratory, Columbia University, Irvington NY, U.S.A.</td>
</tr>
<tr>
<td>35</td>
<td>Niels Bohr Institute, University of Copenhagen, Kobenhavn, Denmark</td>
</tr>
<tr>
<td>36</td>
<td>(a) INFN Gruppo Collegato di Cosenza, Italy</td>
</tr>
<tr>
<td></td>
<td>(b) Dipartimento di Fisica, Università della Calabria, Arcavata di Rende, Italy</td>
</tr>
<tr>
<td>37</td>
<td>AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Krakow, Poland</td>
</tr>
<tr>
<td>38</td>
<td>The Henryk Niewodniczanski Institute of Nuclear Physics, Polish Academy of Sciences, Krakow, Poland</td>
</tr>
<tr>
<td>39</td>
<td>Physics Department, Southern Methodist University, Dallas TX, U.S.A.</td>
</tr>
<tr>
<td>40</td>
<td>Physics Department, University of Texas at Dallas, Richardson TX, U.S.A.</td>
</tr>
<tr>
<td>41</td>
<td>DESY, Hamburg and Zeuthen, Germany</td>
</tr>
<tr>
<td>42</td>
<td>Institut für Experimentelle Physik IV, Technische Universität Dortmund, Dortmund, Germany</td>
</tr>
<tr>
<td>43</td>
<td>Institut für Kern- und Teilchenphysik, Technical University Dresden, Dresden, Germany</td>
</tr>
<tr>
<td>44</td>
<td>Department of Physics, Duke University, Durham NC, U.S.A.</td>
</tr>
<tr>
<td>45</td>
<td>SUPA - School of Physics and Astronomy, University of Edinburgh, Edinburgh, U.K.</td>
</tr>
<tr>
<td>46</td>
<td>Fachhochschule Wiener Neustadt, Johannes Gutenbergstrasse 3 2700 Wiener Neustadt, Austria</td>
</tr>
<tr>
<td>47</td>
<td>INFN Laboratori Nazionali di Frascati, Frascati, Italy</td>
</tr>
<tr>
<td>48</td>
<td>Fakultät für Mathematik und Physik, Albert-Ludwigs-Universität, Freiburg i.Br., Germany</td>
</tr>
<tr>
<td>49</td>
<td>Section de Physique, Université de Genève, Geneva, Switzerland</td>
</tr>
<tr>
<td>50</td>
<td>(a) INFN Sezione di Genova; (b) Dipartimento di Fisica, Università di Genova, Genova, Italy</td>
</tr>
<tr>
<td>51</td>
<td>(a) E. Andronikashvili Institute of Physics, Tbilisi State University, Tbilisi, Georgia</td>
</tr>
<tr>
<td></td>
<td>(b) High Energy Physics Institute, Tbilisi State University, Tbilisi, Georgia</td>
</tr>
<tr>
<td>52</td>
<td>II Physikalisches Institut, Justus-Liebig-Universität Giessen, Giessen, Germany</td>
</tr>
<tr>
<td>53</td>
<td>SUPA - School of Physics and Astronomy, University of Glasgow, Glasgow, U.K.</td>
</tr>
</tbody>
</table>
Department of Physics, University of Massachusetts, Amherst MA, U.S.A.
Department of Physics, McGill University, Montreal QC, Canada
School of Physics, University of Melbourne, Victoria, Australia
Department of Physics, The University of Michigan, Ann Arbor MI, U.S.A.
Department of Physics and Astronomy, Michigan State University, East Lansing MI, U.S.A.

(a) INFN Sezione di Milano, Italy
(b) Dipartimento di Fisica, Università di Milano, Milano, Italy

B.I. Stepanov Institute of Physics, National Academy of Sciences of Belarus, Minsk, Republic of Belarus
National Scientific and Educational Centre for Particle and High Energy Physics, Minsk, Republic of Belarus
Department of Physics, Massachusetts Institute of Technology, Cambridge MA, U.S.A.
Group of Particle Physics, University of Montreal, Montreal QC, Canada
P.N. Lebedev Institute of Physics, Academy of Sciences, Moscow, Russia
Institute for Theoretical and Experimental Physics (ITEP), Moscow, Russia
Moscow Engineering and Physics Institute (MEPhI), Moscow, Russia
Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
Fakultät für Physik, Ludwig-Maximilians-Universität München, München, Germany
Max-Planck-Institut für Physik (Werner-Heisenberg-Institut), München, Germany
Nagasaki Institute of Applied Science, Nagasaki, Japan
Graduate School of Science, Nagoya University, Nagoya, Japan
(a) INFN Sezione di Napoli, Italy
(b) Dipartimento di Scienze Fisiche, Università di Napoli, Napoli, Italy
Department of Physics and Astronomy, University of New Mexico, Albuquerque NM, U.S.A.
Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen/Nikhef, Nijmegen, Netherlands
Nikhef National Institute for Subatomic Physics and University of Amsterdam, Amsterdam, Netherlands
Department of Physics, Northern Illinois University, DeKalb IL, U.S.A.
Budker Institute of Nuclear Physics, SB RAS, Novosibirsk, Russia
Department of Physics, New York University, New York NY, U.S.A.
Ohio State University, Columbus OH, U.S.A.
Faculty of Science, Okayama University, Okayama, Japan
Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman OK, U.S.A.
Department of Physics, Oklahoma State University, Stillwater OK, U.S.A.
Palacký University, RCPTM, Olomouc, Czech Republic
Center for High Energy Physics, University of Oregon, Eugene OR, U.S.A.
LAL, Univ. Paris-Sud and CNRS/IN2P3, Orsay, France
Graduate School of Science, Osaka University, Osaka, Japan
SLAC National Accelerator Laboratory, Stanford CA, U.S.A.

(a) Faculty of Mathematics, Physics & Informatics, Comenius University, Bratislava, Slovak Republic
(b) Department of Subnuclear Physics, Institute of Experimental Physics of the Slovak Academy of Sciences, Kosice, Slovak Republic

(a) Department of Physics, University of Johannesburg, Johannesburg, South Africa
(b) School of Physics, University of the Witwatersrand, Johannesburg, South Africa

(a) Department of Physics, Stockholm University, Stockholm, Sweden
(b) The Oskar Klein Centre, Stockholm, Sweden

Physics Department, Royal Institute of Technology, Stockholm, Sweden

Departments of Physics & Astronomy and Chemistry, Stony Brook University, Stony Brook NY, U.S.A.

Department of Physics and Astronomy, University of Sussex, Brighton, U.K.

School of Physics, University of Sydney, Sydney, Australia

Institute of Physics, Academia Sinica, Taipei, Taiwan

Department of Physics, Technion: Israel Inst. of Technology, Haifa, Israel

Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel

Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece

International Center for Elementary Particle Physics and Department of Physics, The University of Tokyo, Tokyo, Japan

Graduate School of Science and Technology, Tokyo Metropolitan University, Tokyo, Japan

Department of Physics, Tokyo Institute of Technology, Tokyo, Japan

Department of Physics, University of Toronto, Toronto ON, Canada

(a) TRIUMF, Vancouver BC, Vancouver, Canada
(b) Department of Physics and Astronomy, York University, Toronto ON, Canada

Institute of Pure and Applied Sciences, University of Tsukuba, Ibaraki 305-8571, Japan

Science and Technology Center, Tufts University, Medford MA, U.S.A.

Centro de Investigaciones, Universidad Antonio Narino, Bogota, Colombia

Department of Physics and Astronomy, University of California Irvine, Irvine CA, U.S.A.

(a) INFN Gruppo Collegato di Udine, Udine, Italy
(b) ICTP, Trieste, Italy
(c) Dipartimento di Chimica, Fisica e Ambiente, Università di Udine, Udine, Italy

Department of Physics, University of Illinois, Urbana IL, U.S.A.

Department of Physics and Astronomy, University of Uppsala, Uppsala, Sweden

Instituto de Física Corpuscular (IFIC) and Departamento de Física Atómica, Molecular y Nuclear and Departamento de Ingeniería Electrónica and Instituto de Microelectrónica de Barcelona (IMB-CNM), University of Valencia and CSIC, Valencia, Spain

Department of Physics, University of British Columbia, Vancouver BC, Canada
Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary

Also at California Institute of Technology, Pasadena CA, U.S.A.

Also at Institute of Physics, Jagiellonian University, Krakow, Poland

Also at LAL, Univ. Paris-Sud and CNRS/IN2P3, Orsay, France

Also at Department of Physics and Astronomy, University of Sheffield, Sheffield, U.K.

Also at Department of Physics, Oxford University, Oxford, U.K.

Also at Institute of Physics, Academia Sinica, Taipei, Taiwan

Also at Department of Physics, The University of Michigan, Ann Arbor MI, U.S.A.

Also at Laboratoire de Physique Nucléaire et de Hautes Energies, UPMC and Université Paris-Diderot and CNRS/IN2P3, Paris, France

* Deceased