Properties of jet fragmentation using charged particles measured with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV

Aad, G.; ATLAS Collaboration

DOI
10.1103/PhysRevD.100.052011

Publication date
2019

Document Version
Final published version

Published in
Physical Review D. Particles and Fields

License
CC BY

Citation for published version (APA):
Properties of jet fragmentation using charged particles measured with the ATLAS detector in $pp$ collisions at $\sqrt{s} = 13$ TeV

G. Aad et al.$^*$

(ATLAS Collaboration)

(Received 24 June 2019; published 19 September 2019)

This paper presents a measurement of quantities related to the formation of jets from high-energy quarks and gluons (fragmentation). Jets with transverse momentum $100 \text{ GeV} < p_T < 2.5 \text{ TeV}$ and pseudorapidity $|\eta| < 2.1$ from an integrated luminosity of $33 \text{ fb}^{-1}$ of $\sqrt{s} = 13 \text{ TeV}$ proton-proton collisions are reconstructed with the ATLAS detector at the Large Hadron Collider. Charged-particle tracks with $p_T > 500 \text{ MeV}$ and $|\eta| < 2.5$ are used to probe the detailed structure of the jet. The fragmentation properties of the more forward and the more central of the two leading jets from each event are studied. The data are unfolded to correct for detector resolution and acceptance effects. Comparisons with parton shower Monte Carlo generators indicate that existing models provide a reasonable description of the data across a wide range of phase space, but there are also significant differences. Furthermore, the data are interpreted in the context of quark- and gluon-initiated jets by exploiting the rapidity dependence of the jet flavor fraction. A first measurement of the charged-particle multiplicity using model-independent jet labels (topic modeling) provides a promising alternative to traditional quark and gluon extractions using input from simulation. The simulations provide a reasonable description of the quark-like data across the jet $p_T$ range presented in this measurement, but the gluon-like data have systematically fewer charged particles than the simulation.

DOI: 10.1103/PhysRevD.100.052011

I. INTRODUCTION

Jets are collimated sprays of particles resulting from high-energy quark and gluon production. The details of the process that underlies the fragmentation of quarks and gluons with net quantum chromodynamic (QCD) charge into net neutral hadrons are not fully understood. Jet formation is a complex multiscale problem, including important contributions from QCD effects that cannot be described by perturbation theory. Measuring basic quantities related to fragmentation is therefore essential to furthering our understanding of the emergent properties of QCD.

Perturbative and nonperturbative physically inspired models have free parameters that are tuned to data in order to best describe the radiation pattern inside jets [1]. This is in turn an important input to all analyses at the Large Hadron Collider (LHC) due to the ubiquity of jets. Measurements of jet substructure in proton-proton ($pp$) collisions at a center-of-mass energy of $\sqrt{s} = 7 \text{ TeV}$ [2-5] have already been used by the ATLAS collaboration for parameter optimizations (tunes) of the *PYTHIA* 8 Monte Carlo (MC) generator [6]. A measurement of the average number of charged particles inside jets at $\sqrt{s} = 8 \text{ TeV}$ [7] was also used as input to recent developments in the *HERWIG* 7 MC program [8]. Further measurements of jet constituent multiplicity and energy sharing will provide powerful constraints for future generator optimizations.

Quark- and gluon-initiated jets (henceforth quark and gluon jets) have different radiation patterns (see e.g., Ref. [9]). As many analyses at the LHC target either quark-enriched or gluon-enriched processes, these radiation-pattern differences can be useful for jet tagging [10,11]. Measurements of jet structure can be used to calibrate quark-versus-gluon jet taggers. By exploiting the rapidity dependence of the relative quark and gluon jet rates, ATLAS [7] extracted the average charged-particle multiplicity for quark and gluon jets separately. This was then combined with detector-level systematic uncertainties to provide quark/gluon tagger uncertainties at $\sqrt{s} = 13 \text{ TeV}$ [12]. A more complex tagger based on several jet shapes could be calibrated in a similar manner using extended results. The benefit of a particle-level measurement is that a portion of the calibration can be independent of ATLAS and LHC operating conditions. Uncertainties in detector effects can be updated with the changing detector environment. Adding more observables and measuring their differential distributions will improve this calibration.
Although the full radiation pattern inside jets is not calculable from first principles, the energy dependence of many observables can be calculated in perturbation theory. There have been significant theoretical advances in soft-collinear effective theory (SCET) [13–16] to derive factorization theorems that describe the evolution of universal nonperturbative functions [17–20]. This was applied to the measurement of jet charge at $\sqrt{s} = 8$ TeV [21]. There have also been predictions and comparisons with the jet transverse momentum ($p_T$) dependence of the average number of charged particles inside jets (see Ref. [7] and references therein). This quantity does not have a perturbative expansion in the usual sense (as a series in $\alpha_S$); instead there is a series expansion in $\sqrt{\alpha_S}$ [22,23]. This behavior is predicted for a wide class of 
Sudakov safe observables [24]. At least for the case of charged-particle multiplicity, this nonstandard expansion seems to be an excellent model of the data [7].

The goal of this paper is to measure properties of jet fragmentation using charged-particle tracks. Such properties have been measured at many colliders at various center-of-mass energies, including the SPS [25–27], PETRA [28,29], PEP [30–33], TRISTAN [34], CESR [35], LEP [36–47], HERA [48,49], and the Tevatron [50–53]. Previous measurements by the ATLAS and CMS collaborations were performed at $\sqrt{s} = 2.76$ TeV [54,55], $\sqrt{s} = 5.02$ TeV [56–58], $\sqrt{s} = 7$ TeV [25,59,60] and $\sqrt{s} = 8$ TeV [7,21,61] in $pp$ collisions and are also compared with jet fragmentation measured in Pb + Pb collisions [54,56–58, 62,63] and $p + $ Pb collisions [56]. The measurement presented here represents a significant extension of previous work. In particular, the accessible jet energy range is increased due to the larger $\sqrt{s} = 13$ TeV. There are enough events in the 2016 dataset to probe the substructure of jets with $p_T$ up to 2.5 TeV. Next, the precision of the measurement has improved due to advances in track reconstruction inside jets during the long shutdown between LHC runs 1 and 2, including the additional insertable B-layer (IBL) detector [64,65] and new algorithms for tracking inside dense environments [66–68]. Furthermore, detailed experimental studies to derive uncertainties in all aspects of tracking inside jets extend the capabilities of previous measurements to a wider region of phase space and also allow differential analyses [67,69]. These new data therefore probe broader and deeper aspects of the radiation pattern inside jets across an extended phase space.

The paper is organized as follows. Section II introduces the observables to be measured. Then, following a brief description of the ATLAS detector in Sec. III, the data and simulation samples are documented in Sec. IV. Charged-particle track, jet, and event reconstruction are detailed in Sec. V. Corrections for detector effects (unfolding) are documented in Sec. VI. A description of the corresponding systematic uncertainties can be found in Sec. VII and the results are presented in Sec. VIII. Section IX provides conclusions and future outlook.

II. OBSERVABLES

This analysis builds upon the previous ATLAS jet structure measurements presented in Refs. [7,21,59]. The fundamental quantity is the fragmentation function $D_{h^+}(z,E_T)$, which describes the probability of finding a hadron $h$ with energy fraction $z$ of the parton $p$ that has energy $E$. At a hadron collider, the jet transverse momentum, $p_T$, is a better proxy for the starting scale ($\mu$) of jet evolution. To avoid confusion with previous measurements of similar observables, the transverse momentum fraction is denoted in this paper by the symbol $\zeta = p_T^\text{particle} / p_T^\text{jet}$. The fragmentation function itself, like parton distribution functions (PDFs), cannot be calculated from first principles in perturbation theory. However, it has a Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution [70–72] and so the $p_T$ dependence of many observables can be calculated. In particular,

$$
\mu \frac{\partial}{\partial \mu} D_{h^+}(\zeta,\mu) = \sum_{p} \int_{\zeta}^{1} \frac{d\zeta'}{\pi} p_{T^*} \frac{\alpha_S(\mu)}{\zeta} D_{h^+}(\zeta',\mu)
$$

(1)

where $p_{T^*} = p(\zeta,\mu)$ are the Altarelli-Parisi splitting functions [70] and depend on the scale $\mu$ through $\alpha_S$. Charged particles are studied because they provide a way to measure single hadrons inside the jet (as opposed to calorimeter energy deposits, which can result from multiple particles) which gives access to $\sum_{h} D_{h^+}(\zeta)$. A basic quantity related to the fragmentation function is the charged-particle multiplicity. The average charged-particle multiplicity is an integral over $\zeta$ and a sum over $h$ and $p$ of $D_{h^+}(\zeta)$. An extension of the multiplicity is the set of $\zeta$ moments of $D$. The zeroth moment is the average multiplicity. The full distribution of multiplicity depends on (multihadron) fragmentation functions in a complicated way; a more direct probe of $D$ is to measure hadron production as a function of $\zeta$, which is a sum of $D$ over $p$ and $h$ (but no integral over $\zeta$). Additional observables are also studied in order to probe the angular spread of jet fragmentation beyond the collinear limit. All of the observables are described below.

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

G. AAD et al. 
PHYS. REV. D 100, 052011 (2019)
A. Charged-particle multiplicity \( \langle n_{\text{ch}} \rangle \)

The number of charged particles inside a jet with \( p_T \) above some threshold. In terms of the fragmentation function,

\[
\langle n_{\text{ch}} \rangle (p_T^{\text{jet}}) = \sum_p f_p (p_T^{\text{jet}}) \sum_h \int_{\text{threshold}/p_T^{\text{jet}}}^{1} d\zeta D_p^h (\zeta, p_T^{\text{jet}}),
\]

where \( f_p \) is the fraction of parton type \( p \) at a given jet \( p_T \). The multiplicity is not calculable in perturbation theory, but to lowest order in \( \sqrt{a_S} \), the ratio of the multiplicity for quark-initiated jets to that for gluon-initiated jets is the ratio of color factors \( C_A/C_F = 9/4 \). The fraction of quark jets increases with \( p_T \), which decreases the inclusive multiplicity. However, this is compensated by an inherent increase in the multiplicity with \( p_T \), which is carried by particles of a given momentum fraction \( \zeta \). Above some threshold, in log space, the inelastic charged-particle multiplicity \( \langle n_{\text{ch}} \rangle \) is considered.

B. Summed fragmentation function

The distribution of the momentum fraction \( \zeta \) is studied inside jets summed over charged-hadron types. The quantity that is measured is \( F(\zeta, p_T^{\text{jet}}) = \langle 1/N_{\text{jet}} \rangle dn_{\text{ch}}/d\zeta \). In terms of the fragmentation function,

\[
F(\zeta, p_T^{\text{jet}}) = \sum_p f_p (p_T^{\text{jet}}) \sum_h D_p^h (\zeta, p_T^{\text{jet}}).
\]

By definition, \( \int d\zeta F(\zeta) = \langle n_{\text{ch}} \rangle \). In addition to measuring the probability of \( F(\zeta) \) in bins of jet \( p_T \), summary statistics of the \( F(\zeta) \) distribution are extracted to show how the distribution evolves with \( p_T \). The following properties of the \( \zeta \) distribution are extracted:

- **Summation**
  \[
  \sum_{x \text{ charged}} \langle \zeta \rangle = \langle 1/N_{\text{jet}} \rangle dn_{\text{ch}}/d\zeta = n_{\text{ch}} (\zeta < X) / n_{\text{ch}} \text{ to show how much of the jet energy is carried by particles of a given } p_T \text{ fraction. For illustration, the values considered are } X \in \{0.1, 0.01, 0.001\}. \text{ As } X \rightarrow 1, \text{ these partial fractions become a constant value of 1.0, independent of the jet } p_T. \]

- **Moments of \( F(\zeta) \)**
  \[
  \zeta^\kappa = \int \zeta^\kappa F(\zeta) d\zeta / \int F(\zeta) d\zeta. \]
  The distribution of \( F(\zeta) \) is nearly normally distributed in log space, which means that it is defined by its first two moments [4]. For this reason, \( \kappa = 2 \) is measured as a function of the jet \( p_T \). For illustration, the case \( \kappa = 1/2 \) is also considered.

- **Weighted sums over the jet**
  \[
  \sum_{\text{jet}} \zeta^\kappa = \int \zeta^\kappa F(\zeta) d\zeta. \]
  The values considered are \( \kappa \in \{1/2, 2\} \). The observable \( \sum_{\text{jet}} \zeta^\kappa \) is often called \( p_T^\kappa \) and can be used for quark/gluon jet tagging [4]. For a given jet type, these observables increase monotonically with increasing jet \( p_T \) for \( \kappa \lesssim 1 \) and decrease monotonically for \( \kappa \gtrsim 1 \) (see Sec. VIII B); the \( \kappa \) values chosen are representative of these trends.

Each of these derived quantities is extracted from the measured \( F(\zeta) \) distribution. More details about the procedure for unfolding these derived quantities are presented in Sec. VI.

C. Transverse momentum

\[
p_T^{\text{rel}} \equiv p_T^{\text{charged particle}} \sin \Delta \phi, \text{ where } \Delta \phi \text{ is the angle between the momentum of the constituent charged particle and the jet in the transverse plane. The quantity that is measured is } f(p_T^{\text{rel}}, p_T^{\text{jet}}) = (1/N_{\text{jet}})dn_{\text{ch}}/dp_T^{\text{rel}}. \text{ The average value is defined by } \langle p_T^{\text{rel}} \rangle = \int p_T^{\text{rel}} f(p_T^{\text{rel}})/\int f(p_T^{\text{rel}}).
\]

D. Radial profile

The number of charged particles in various annuli around the jet axis. The quantity that is measured is \( \rho_{\text{ch}}(r, p_T^{\text{jet}}) = (1/N_{\text{jet}})dn_{\text{ch}}/2\pi r dr, \text{ where } r = \Delta R(\text{charged particle, jet}). \text{ The average value is defined by } \langle r \rangle = \int \rho_{\text{ch}}(r)/\int \rho_{\text{ch}}(r). \text{ The last two quantities are not simple derivatives of the fragmentation function as they additionally depend on finite opening angles encoded in the } d\theta/d\phi \text{ emission phase space. Since quantities are measured as a function of jet } p_T \text{ which is defined using charged and neutral particles, the observables are sensitive to the charged-to-neutral fraction inside jets. However, this fraction is robust to mismodeling as isospin is an approximate symmetry of the strong force.}

III. ATLAS DETECTOR

The ATLAS detector [75] at the LHC covers nearly the entire solid angle around the collision point. It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadronic calorimeters, and a muon spectrometer incorporating three large superconducting toroidal magnets. The inner-detector system (ID) is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range \(|\eta| < 2.5\). The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track, the first hit being normally in the IBL. It is followed by the silicon microstrip tracker which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker, which enables radially extended track reconstruction up to \(|\eta| \approx 2.0\).

The calorimeter system covers the pseudorapidity range \(|\eta| < 4.9\). Within the region \(|\eta| < 3.2\), electromagnetic calorimetry is provided by barrel and end cap high-granularity lead/liquid-argon (LAr) calorimeters, with an additional thin LAr presampler covering \(|\eta| < 1.8\), to correct for energy loss in material upstream of the calorimeters. Hadronic calorimetry is provided by the steel/scintillating-tile calorimeter, segmented into three barrel structures within \(|\eta| < 1.7\), and two copper/LAr hadronic end cap calorimeters. The solid angle coverage is...
completed with forward copper/LAr and tungsten/LAr calorimeter modules optimized for electromagnetic and hadronic measurements respectively.

Interesting events are selected to be recorded by the first-level trigger system implemented in custom hardware, followed by selections made by algorithms implemented in software in the high-level trigger [76]. The first-level trigger reduces the 40 MHz bunch crossing rate to below 100 kHz, which the high-level trigger further reduces in order to record events to disk at about 1 kHz.

IV. DATASETS AND SIMULATED SAMPLES

These measurements use the dataset of \( pp \) collisions recorded by the ATLAS detector in 2016, corresponding to an integrated luminosity of 33 fb\(^{-1} \) at a center-of-mass energy of \( \sqrt{s} = 13 \) TeV. Events are only considered if they are collected during stable beam conditions and satisfy all data quality requirements. Due to the high instantaneous luminosity and the large total inelastic proton-proton cross section, on average there are about 25 simultaneous (pileup) collisions in each bunch crossing.

The measurements presented in this paper use a variety of MC samples for estimating correction factors as well as for comparison with the corrected data. Dijet events were generated at leading order (LO) with PYTHIA 8.186 [77], with the \( 2 \rightarrow 2 \) matrix element convolved with the NNPDF2.3LO PDF set [78] and using the A14 tune of multiple-parton-interaction and shower parameters [6]. SHERPA uses a \( p_T \)-ordered parton shower model. Additional dijet events were simulated using different generators, in order to study the impact of modeling uncertainties. SHERPA 2.1 [79] events were generated using multileg \( 2 \rightarrow 2 \) and \( 2 \rightarrow 3 \) matrix elements, which were matched to parton showers following the Catani-Krauss-Kuhn-Webber prescription [80]. These SHERPA events were simulated using the CT10 PDF set [81] and the default SHERPA event tune. HERWIG++ 2.7 [82,83] was used to provide a sample of events with an angle-ordered parton shower model. These events were generated with the \( 2 \rightarrow 2 \) matrix element, convolved with the CTEQ6L1 PDF set [84] and configured with the UE-EE-5 tune [85].

All simulated events were passed through a full simulation of the ATLAS detector [86] implemented in GEANT 4 [87], which describes the interactions of particles with the detector and the subsequent digitization of analog signals. The effects of multiple simultaneous \( pp \) collisions were simulated with inelastic \( pp \) collisions using the PYTHIA 8.186 generator with the A2 [88] set of tuned parameters and the MSTW2008LO [89] PDF set; these events were overlaid on the nominal dijet events.

V. OBJECT AND EVENT SELECTION

Since the data are unfolded to particle level, it is necessary to define both the particle-level and detector-level objects used in the measurement. The former are chosen to be as close as possible to the latter in order to minimize the model dependence caused by an extrapolation from the measured phase space at detector level to the phase space at particle level. Section VA describes the definition of charged-particle tracks and jets. Following the discussion of objects, Sec. VB describes the particle-level and detector-level event selection criteria.

A. Object reconstruction

While it is not possible to separate the underlying event from the hard scatter at particle level, it is possible to remove the contribution from pileup. Therefore, the unfolding target is particle-level distributions produced in single proton-proton interactions. However, at detector level, there is ambiguity about which \( pp \) collision vertex corresponds to the hard-scatter event. Collision vertices are reconstructed from tracks in the inner detector. Each vertex is required to be associated with at least two tracks with \( p_T > 0.4 \) GeV. The primary hard-scattering vertex of the event is chosen to be the one with the highest \( \sum p_T^2 \) calculated using all tracks associated with the vertex.

Particle-level jets are built from MC-simulated stable particles \( (\epsilon_T > 10 \) mm\) excluding muons and neutrinos. By definition, particles from pileup and from interactions with the detectors are not included. These jets are clustered using the anti-\( k_T \) [90] algorithm with radius parameter \( R = 0.4 \) as implemented in FASTJET [91]. Detector-level jets are built from topological calorimeter-cell energy clusters [92] using the same algorithm as is used at particle level. A series of simulation- and data-based correction and calibration factors are applied to ensure that the resulting jet \( p_T \) is the same as the particle-level value on average [93]. Jets are required to have \( p_T > 60 \) GeV so that the rate of jets originating from pileup is negligible. The detector-level phase space includes one bin at low jet \( p_T \) (60–100 GeV) which is not in the fiducial phase space of the measurement due to the large impact of migrations into and out of the acceptance.

Charged particles are used to compute the particle-level definitions of all observables if they are clustered within a particle-level jet and have \( p_T > 500 \) MeV and \( |\eta| < 2.5 \). The detector-level analog to charged particles is tracks. Tracks are reconstructed from hits in the inner detector (see e.g., Ref. [67]) and a series of quality criteria are applied to the selected tracks to reject those originating from hits due to multiple charged particles (fake tracks) and from pileup. The transverse momentum resolution is approximately \( \sigma(p_T)/p_T \approx 0.05\% \times p_T/\text{GeV} \oplus 1\% \), with a significant degradation in the core of high-\( p_T \) jets due to challenges associated with pattern recognition. Tracks are required to pass the tight primary selection as well as the

\[^2\] For example, the \( p_T \) resolution is approximately 30\% at 100 GeV when five or more particles are within \( \Delta R < 0.015 \).
loose track-to-vertex association [94]. In particular, they must have $p_T > 500$ MeV and $|\eta| < 2.5$, and the number of pixel and strip clusters associated with the track is required to be at least 9 (11) for $|\eta| < 1.65$ ($\geq 1.65$). In addition, the transverse impact parameter $d_0$ relative to the beam line must be less than 2 mm and the longitudinal impact parameter, $z_0$, is required to satisfy $|z_0 \sin \theta| < 3$ mm. Tracks are matched to jets via ghost association [95]. This matching procedure creates ghost versions of the tracks with the same direction but infinitesimal $p_T$. Jet clustering is repeated and tracks are assigned to the jet that contains their ghosted version. For the isolated, high $p_T$ jets used in this measurement, ghost association is nearly identical to a geometric matching based on $\Delta R < 0.4$.

### B. Event selection

Particle-level events are required to have at least two jets with $|\eta| < 2.1$ (within the tracking detector acceptance) and the leading two such jets must satisfy $p_T^{\text{lead}} / p_T^{\text{sublead}} < 1.5$. This jet-$p_T$ balance requirement simplifies the interpretation of the final state in terms of a $2 \rightarrow 2$ scattering process.

Detector-level events are selected using single-jet triggers. Due to the large cross section for jet production, most of the jet triggers are prescaled: events that pass the trigger are randomly discarded with a fixed probability. The trigger used for a particular jet $p_T$ is chosen to ensure that the trigger is 100% efficient (for the measurement phase space and prior to prescaling) and has the lowest prescale factor. Events in data are weighted by the prescale. The lowest-threshold unprescaled jet trigger is used for jets with $p_T > 600$ GeV. Detector-level events are required to pass the same selection requirements as particle-level events: there must be at least two off-line calibrated jets with $|\eta| < 2.1$ and the leading two of these jets must satisfy $p_T^{\text{lead}} / p_T^{\text{sublead}} < 1.5$. Figure 1 shows the basic kinematic properties of the two leading jets passing this event selection compared with various MC predictions at detector level.

The substructure of the two leading jets is used in the analysis. Figure 2 shows detector-level distributions for a selection of the observables that were introduced in Sec. II. For the jets with $p_T \sim 1$ TeV shown in Fig. 2, the most probable number of tracks is about 15 and the most probable momentum fraction is about 1%. The radiation pattern is peaked at the center of the jet, so both the $p_T^{\text{lead}}$ and $r$ distributions are peaked at zero. The PYTHIA, HERWIG++, and SHERPA distributions generally bracket the data and are accurate to within about 20%.

In order to expose differences between quark and gluon jets, the more forward and more central of the two jets are distinguished and measured separately. Figure 3 shows the gluon-jet fraction as a function of jet $p_T$ and jet $\eta$ (more details about quark/gluon definitions are given in Sec. VIII B). For a fixed jet $p_T$, higher-$|\eta|$ jets are more often quark initiated due to valence quarks scattering off gluons. For a fixed $\eta$, the quark fraction increases with jet $p_T$ due to the relative increase in valence-quark scattering off a quark or gluon compared with gluon-gluon scattering.

Table I summarizes the object and event selections from Secs. VA and VB.

### VI. UNFOLDING

The data are corrected for resolution and acceptance effects, and the fiducial phase space of the measurement is described by the particle-level object and event selection in Sec. V. Equation (2) symbolically summarizes the unfolding procedure for a binned distribution $x$:
FIG. 2. The detector-level distributions of (top left) the number of charged-particle tracks $n_{\text{track}}$, (top right) the transverse momentum fraction $\zeta$, (bottom left) the transverse momentum $p_{T}^{\text{rel}}$, and (bottom right) the radial profile in bins of the distance $r$ from the jet axis for jets with transverse momentum $1 \text{ TeV} < p_{T}^{\text{jet}} < 1.2 \text{ TeV}$. Error bars include the statistical uncertainty only.

FIG. 3. Left: The gluon-jet fraction as a function of jet transverse momentum $p_{T}$ and jet pseudorapidity $\eta$. Right: the fraction of the more forward and the more central jets that are gluon initiated. The error bars in the right plot represent the uncertainty computed from the 100 NNPDF2.3LO replicas [78]. See Sec. VIII B for more details about quark/gluon definitions and uncertainties.
TABLE I. A summary of the object and event selection criteria at particle level and detector level.

<table>
<thead>
<tr>
<th>Particle level</th>
<th>Detector level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pileup</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Jet algorithm</td>
<td>Anti-kt, R = 0.4</td>
</tr>
<tr>
<td>Jet requirements</td>
<td></td>
</tr>
<tr>
<td>Jet constituents</td>
<td>Particles with</td>
</tr>
<tr>
<td>Measurement inputs</td>
<td>ct &gt; 10 mm prior to detector</td>
</tr>
<tr>
<td></td>
<td>interactions excluding μ and ν</td>
</tr>
<tr>
<td>Event selection</td>
<td>At least two jets, with the leading two satisfying</td>
</tr>
<tr>
<td>Jet selection</td>
<td></td>
</tr>
</tbody>
</table>

\[
x_{\text{unfolded},i} = \frac{1}{n_{\text{jets,unfolded}}} \sum_{j=1}^{N_{\text{bins}}} \theta_{ij} x_{\text{detected},j} \left( \frac{1 - \epsilon_{\text{reco not true},j}}{1 - \epsilon_{\text{true not reco},j}} \right).
\]

(2)

where \( n_{\text{jets,unfolded}} \) is the unfolded number of forward or central jets (depending on the bin), determined by the number of entries in the \( n_{\text{ch}} \) unfolding (as there is one entry per jet). The symbols \( \theta \) and \( \epsilon \) represent the unfolding matrix and correction factors, described in more detail below.

The jet substructure observables are simultaneously unfolded with the jet \( p_T \) and for the more forward and the more central jets at the same time. For an observable with \( n_{\text{bins}} \) bins in a given \( p_T \) bin, this results in a total of \( N_{\text{total}} = 2 \times (n_{\text{bins}}) \times (p_{\text{T}} \text{ bins}) \) bins. All of these bins are concatenated to form a one-dimensional input. To begin the unfolding, the data are corrected for the fraction of events that pass the detector-level selection but not the particle-level selection, \( \epsilon_{\text{true not reco}} \). This also corrects for non-dijet events, but their rate is negligible. Then, an iterative Bayesian (IB) unfolding technique [96] is used as a regularized matrix inversion to correct for the detector resolution in events that pass both the detector-level and particle-level selections. The IB method is implemented in the ROOUnfold framework [97] with the unfolding matrix \( \theta \) and one iteration, is chosen to minimize the total uncertainty. After the application of the response matrix, a final correction is applied to account for the fraction of events that pass the particle-level but not detector-level selection, \( \epsilon_{\text{true not reco}} \). The resulting unfolded measurement is reorganized into individual distributions with \( n_{\text{bins}} \) per \( p_{\text{T}} \) bin for each of the more forward and more central jets. The jet \( p_{\text{T}} \) is also unfolded in parallel and each \( p_{\text{T}} \) bin of the jet substructure observable is normalized by the number of measured jets in that bin. For \( n_{\text{ch}} \), this renders the distributions normalized to unity per jet \( p_{\text{T}} \) bin; for the other observables, the normalization in each \( p_{\text{T}} \) bin is (up to acceptance effects) \( \langle n_{\text{ch}} \rangle \), as discussed in Sec. II.

To illustrate the jet-\( p_{\text{T}} \) dependence of the measured observables, the evolution with the jet \( p_{\text{T}} \) of various moments (\( \kappa \)) is computed using Eq. (3):

\[
\langle x^\kappa \rangle_{\text{unfolded}} = \frac{1}{K} \sum_{i=1}^{n_{\text{bins}}} x_{\text{unfolded},i} \times (\text{bin center } i)^{\kappa},
\]

(3)

where the sum is over all \( i \) that correspond to \( p_{\text{T}} \) bin \( j \). Since the bin center is used to calculate the average, a correction \( c_{\text{binning}} \) is applied to account for the difference between the bin center and the mean of the distribution within the bin. This correction is calculated using PYTHIA, and is computed by reweighting PYTHIA so that it agrees with the unfolded distribution. For \( \zeta, p_{\text{T}}^\text{lead}, \) and \( r \), Eq. (3) represents the \( \kappa \) moment for individual particles. For \( \zeta \), the jet-based moments are also computed: \( \langle \zeta^\kappa \rangle_{\text{bin}} \). For these jet-based moments, Eq. (3) is modified by removing the denominator \( \sum_{i=1}^{n_{\text{bins}}} x_{\text{unfolded},i} \). By construction, the \( \kappa = 0 \) jet-based moment of \( \zeta \) is the \( \kappa = 1 \) moment of \( n_{\text{ch}} \). The binning correction factor is mostly near unity, deviating by less than 1% for \( n_{\text{ch}} \) and up to about 10% for the other observables.

Figure 4 shows the response matrix normalized per particle-level bin. As stated above, the observable bins are concatenated with the jet \( p_{\text{T}} \) and for both the more forward and more central jets to form a one-dimensional distribution that is unfolded. A diagonal stripe represents events where the detector-level jet \( p_{\text{T}} \) is the same as the particle-level value; off-diagonal components represent jet \( p_{\text{T}} \) migrations. Within a jet \( p_{\text{T}} \) bin, there is a small dependence on \( \zeta \), with a worse resolution at high \( \zeta \) due to the deteriorating momentum resolution at high track \( p_{\text{T}} \). The diagonal strips in the upper left and lower right quadrants correspond to events where the more forward particle-level jet is the more central detector-level jet and vice versa. This migration happens in about 1% of events. Within a given jet \( p_{\text{T}} \) bin, the migrations to neighboring \( \zeta \) bins are small. Except at high \( \zeta \) and high jet \( p_{\text{T}} \) where the migrations can reach 50%, the off-diagonal components of the response matrix are about 10%.

VII. UNCERTAINTIES

Systematic and statistical uncertainties are assessed for each step of the analysis, including the acceptance
with the tracking uncertainties matching in size in the highest jet-$p_T$ bins. The tracking and jet energy uncertainties are about the same size for $\langle p_T^{\text{rel}} \rangle$. Fragmentation modeling uncertainties are large for $\langle n_{\text{ch}} \rangle$ at low jet $p_T$ and for $\langle \zeta \rangle$ at high jet $p_T$. While the size of the binning correction uncertainty is less than 2% for $\langle p_T^{\text{rel}} \rangle$ and $\langle r \rangle$, it is still the dominant uncertainty for these observables. Further details about each uncertainty source are given below and in the full covariance matrices, including all correlation information, are made available in Ref. [99].

A. Track and jet reconstruction

Except for $\zeta$, the jet energy is only used to determine the $p_T$ bin. Since the fragmentation properties vary slowly with jet $p_T$, the resulting impact of jet energy scale and resolution uncertainties on the analysis is often less important than other sources of uncertainty. Nonetheless, the impact of a 19-parameter decomposition of the jet energy scale uncertainty was evaluated [93]. Six of these 19 components are due to in situ constraints on the jet energy scale from various multiobject balance studies, such as $Z \rightarrow j$ jets. Additional sources of uncertainty are related to pileup, jet flavor, and extrapolations to high $p_T$. The total uncertainty in the jet energy scale is about 1% for jets with $p_T$ between 100 and 1000 GeV and the impact on this measurement is much less than 1% except at high $\zeta$, where it can reach as high as 2%. The impact of the jet energy resolution is determined from an ensemble of event samples with jet energies smeared within the uncertainty.

The most important experimental uncertainties are related to track reconstruction and cover the track reconstruction efficiency, the rate of fake and secondary tracks, the momentum scale, and density effects from pixel and strip cluster merging. In the PYTHIA simulation, approximately 60% of the charged particles/tracks inside jets are charged pions that are well matched, $3 10\%$ are well-matched kaons, $5\%$ are well-matched protons, $15\%$ are charged particles that are not matched to reconstructed tracks (inefficiency), $5\%$ are secondaries (split equally between photon conversions and nuclear interactions), $1\%$ are not well-matched tracks (fake tracks), and about $O(0.1\%)$ are pileup tracks wrongly matched to the primary hard-scattering vertex. The pileup contribution decreases with jet $p_T$ and momentum fraction, but increases with jet cone size (reaching $1\%$ at $\Delta = 0.4$). In contrast, the fake-track rate increases slightly with jet $p_T$ and has a contribution at high-momentum fraction of a few percent from

---

3Reconstructed tracks are matched to charged particles by examining the pattern of sensors where energy was deposited. If over 50% of the weighted number of measurements on a track are due to one charged particle, it is declared matched to the track. The weights are chosen to reflect the amount of information present in each detector and are ten for the pixel detector, five for the strip detector, and one for the straw tube tracker.
kinked tracks reconstructed with a very high $p_T$. The reconstruction inefficiency grows with jet $p_T$, and is peaked at both low and high radial distance from the center of the jet and is reduced at high momentum fraction. This is because tracks with a larger radial distance from the jet axis tend to have lower $p_T$ (larger material effects and thus lower efficiency), while tracks in the core of the jet suffer from an inefficiency in the pattern recognition in the dense environment.

The uncertainty in the inclusive track reconstruction efficiency is dominated by the uncertainty in the amount of material in the inner detector. Variations in the amount of material that are consistent with detector construction knowledge and measurements from secondary vertices [100] result in an uncertainty of 0.5% for $|\eta| < 0.1$, which grows to 2.7% for $2.3 < |\eta| < 2.5$. This uncertainty is applied in the simulation by randomly removing tracks with a $p_T$- and $|\eta|$-dependent probability. This uncertainty dominates the $n_{ch}$ measurement for jet $p_T \lesssim 1$ TeV.

Since the ATLAS pixel detector measures the charge collected from ionization, it is possible to constrain the inefficiency from density effects by looking for single tracks with pixel charge consistent with two minimum-ionizing particles [67]. The resulting uncertainty is about 0.4% for tracks with $\Delta R < 0.1$ and is validated with additional studies related to the charged-to-neutral ratio in the jet as well as the geometric orientation of pixel clusters [68]. This uncertainty is most important for the radial energy measurement at small radii from the jet axis and for the $n_{ch}$ measurement in the highest jet-$p_T$ bins.

The rate of fake tracks is studied inside jets by inverting some of the track quality criteria such as the fit $\chi^2$/NDF and

---

**FIG. 5.** An overview of the statistical and systematic uncertainties for the average value of (top left) charged-particle multiplicity $n_{ch}$, (top right) transverse momentum fraction $\zeta$, (bottom left) transverse momentum $p_T$, and (bottom right) radial profile as a function of jet transverse momentum $p_T$. The MC statistical uncertainties are negligible and are not shown. When the uncertainties go through zero (as for the unfolding uncertainty in the top left), the signed uncertainty has changed from positive to negative or vice versa. Most of the lines are the sum in quadrature of individual sources of uncertainty in each category, such as the various sources of tracking uncertainties as described in Sec. VII A.
is found to agree between data and simulation at the 30% level [68]. A related source of uncertainty is due to the rate of secondary tracks. These tracks originate from real charged particles, but are the result of interactions in detector material and not direct fragmentation processes. The rate of secondaries is estimated by fitting the track $d_0$ distribution and is found to agree with simulation within about 30%. These rates are then varied to determine an uncertainty in the measurement. The fake-track rate is the leading source of uncertainty for $n_{ch}$ when $p_T \gtrsim 1$ TeV and when $\zeta \sim 1$ or $r \lesssim 0.05$ for all jet-$p_T$ bins. Uncertainties related to the modeling of pileup have a negligible impact.

The leading source of uncertainty in the track parameters is in the $q/p_T$ ($q$ is the electric charge) from a potential sagitta distortion due to detector misalignment weak modes [94]. This bias is corrected and the uncertainty in the correction is about 0.1/TeV except at $\phi \approx 0$ and $|\eta| \sim 2.5$ where the correction can reach 1/TeV.

The impact on the measurement is smaller than the other tracking uncertainties.

**B. Unfolding method and fragmentation modeling**

An uncertainty resulting from the unfolding method described in Sec. VI is determined by unfolding the prediction from a reweighted simulation with the nominal procedure. The reweighted simulation is constructed by modifying the nominal PYTHIA 8 particle-level spectrum so that the simulated detector-level spectrum, from propagating the reweighted particle-level spectrum through the response matrix, has significantly improved agreement with the data. The modified detector-level distribution is unfolded with the nominal response matrix and the difference between this and the reweighted particle-level spectrum is an indication of the bias due to the unfolding method (in particular, the choice of prior) [101].
The weights are chosen by comparing the PYTHIA 8 particle-level spectrum with the unfolded data. After applying the reweighting, the $\chi^2/\text{NDF}$ calculated using only the statistical uncertainties improves significantly in each jet $p_T$ bin. The resulting systematic uncertainties are generally much smaller than the detector-level differences between the data and simulation, as desired.

The unfolded result depends on the modeling of jet fragmentation through the prior, the response matrix, and the correction factors. Variations in the prior are already accounted for in the data-driven nonclosure uncertainty described above. The other contributions are evaluated by comparing the result using PYTHIA 8 with the result using the alternative HERWIG++ sample described in Sec. IV. A similar uncertainty is obtained when using HERWIG++ or SHERPA as the alternative model. This comparison is decomposed into components corresponding to varying only the response matrix or only the initial/final correction factors, $\epsilon_{\text{reco not true}}$ and $c_{\text{true not reco}}$ in Eq. (2). All three components are added in quadrature to determine the total uncertainty due to fragmentation modeling. Even though these sources of uncertainty are correlated, they were treated as independent because the level of correlation is unknown given that there are only two alternative models. The resulting uncertainty is much smaller than the difference between PYTHIA 8 and HERWIG++ at particle level. For $n_{ch}$, the response matrix is the dominant contribution to this uncertainty, except in the first jet-$p_T$ bin where the correction factors and their uncertainty are also important. For the per-particle observables ($\zeta$, $r$, $p_T^{\text{rel}}$), the correction factors dominate the uncertainty because acceptance effects are much more important.

VIII. RESULTS

The unfolded data are presented in two ways. Section VIII A focuses on the inclusive spectra for both
jets together, while Sec. VIII B uses the differences between forward and central jets to determine the unique features of quark-initiated and gluon-initiated jets, some of which can be compared with perturbative QCD calculations. These sections show a selection of jet \( p_T \) bins; a complete set of results can be found in Ref. [99].

**A. Inclusive distributions**

The unfolded averages of the measured observables are presented as a function of the jet \( p_T \) in Fig. 6 for the more forward and more central jets separately and then combined in Fig. 7. All other figures in this section combine measurements of both jets. The more central jets show properties that are more gluon-like than the more forward jets: they have a larger charged-particle multiplicity and a softer momentum-fraction spectrum. The modeling of the all-jet spectra is very similar to that of the more forward/backward jets and is described in detail for the all-jet spectra only.

As the jet \( p_T \) increases, the average charged-particle multiplicity increases, the average momentum fraction decreases, the average \( p_T^{\text{rel}} \) increases, and the average multiplicity-weighted radius decreases. Charged-particle multiplicity increases from about 10 at jet \( p_T \) of 100 GeV to just over 20 at 2.5 TeV. In most cases, **PYTHIA 8** and **SHERPA** bracket the data, and are accurate to better than 10%; **HERWIG++** is often between these two and closer to the data. As the distribution of \( n_{\text{ch}} \) is almost Poissonian, nearly all of the information about the distribution is encoded in the mean value. In contrast, the distribution of \( \zeta \) is more complicated.\(^4\) The average

\[ n_{\text{ch}}(\zeta < 1.5^S = 0.1) / n_0, \]

\[ (\chi^2 / n_0) < 1.5 \]

\[ \text{Data (stat. uncert.) syst. uncert. Stat. } \]

\[ \text{Pythia 8.186 A14 Herwig++ Sherpa 2.1 } \]

\[ \text{All selected jets } \]

\[ \text{ATLAS } \]

\[ (S = 13 \text{ TeV, 33 fb}^{-1}) \]

\[ \text{Jet } p_T [\text{GeV}] \]

\[ 500 1000 1500 2000 2500 \]

\[ \text{MC / Data} \]

\[ \text{FIG. 8. The unfolded fraction of charged particles carrying a fraction } \zeta < 10\% \text{ (top left), } 1\% \text{ (top right), and } 0.1\% \text{ (bottom) of the jet transverse momentum } p_T. \text{ The exact fractions are given in the figures and correspond to powers of 1.5 so that the values align precisely with bin edges and thus no binning correction is required. The lower panels show the ratio of various MC predictions to the data, with the total uncertainty band centered on the data at unity.} \]

\(^4\)The distribution is nearly Gaussian in \( \log \zeta \), so it is well specified by two parameters instead of one [73].
momentum fraction is about 5% at jet $p_T$ of 100 GeV and decreases to about 2.5% at 2.5 TeV (the most probable value, shown below, is lower). The distributions of $p_{rel}^T$ and the radial profiles fall steeply (nearly exponentially) away from zero and the average values in Fig. 7 give a sense of how fast they fall (exponential distributions are uniquely specified by their mean). The average $p_{rel}^T$ at $p_T^{jet} = 100$ GeV is about 0.35 GeV and increases to about 0.55 GeV at $p_T^{jet} = 2.5$ TeV. If the angular distribution about the jet axis is independent of $p_T$, the average value of $p_{rel}^T$ should be proportional to $⟨ζ⟩(p_T^{jet}) × p_T^{jet}$. This would suggest an increase by a factor of $(2.5%/5%)/(2500/100) ∼ 12.5$ across the measured range; instead it only increases by a factor of about 1.5. This means that the angular distribution is not independent of $p_T$ and in particular, the jets become more collimated. This is also consistent with direct measurement of the radial profile, where the average value drops from about 0.06 at $p_T^{jet} = 100$ GeV to about 0.03 at $p_T^{jet} = 2.5$ TeV. While PYTHIA 8, SHERPA, and HERWIG++ agree well with the data for $p_{rel}^T$, SHERPA provides a poorer model of the average radial profile as a function of the jet $p_T$.

As noted above, the distribution of $ζ$ cannot be described simply by its average value, in contrast to $n_{ch}$, $p_{rel}^T$ and $r$, which are nearly Poisson or exponentially distributed. Therefore, it is useful to summarize the $p_T^{jet}$ dependence of other aspects of the $ζ$ distribution. Figure 8 shows partial integrals of the $ζ$ distribution and Fig. 9 shows the average values of $ζ^{1/2}$, $ζ$, $∑_{i ℕ} ζ_i^{1/2}$ and $∑_{i ℕ} ζ_i^2$. Figure 8 illustrates how the average fraction of charged particles...
with a given momentum fraction evolves with $p_T^{\text{jet}}$. There is no correction for binning effects, as the measured $\zeta$ distribution has bin edges which nearly align with 0.1%, 1%, and 10%. In particular, the $\zeta$ bins are $1/1.5^n$, for $n = 0, \ldots, 21$, and the fractions in Fig. 8 are estimated as $0.1 \% \approx 1/1.5^{17}, 1 \% \approx 1/1.5^{11}$, and $10 \% \approx 1/1.5^5$. The fraction of particles carrying 10% or less of the momentum changes very little across the entire $p_T^{\text{jet}}$ range and is also near unity ($>90\%$ for all $p_T^{\text{jet}}$). A strong $p_T^{\text{jet}}$ dependence is introduced when the $\zeta$ threshold is lowered to 1% and to 0.1%. Since charged particles are required to have $p_T > 500$ MeV, only jets with $p_T^{\text{jet}} > 500$ GeV can have particles with $\zeta < 0.1\%$. The fraction of particles with $\zeta < 1\%$ has a logarithmic increase while the fraction of particles with $\zeta < 0.1\%$ appears to increase faster than linearly with $p_T^{\text{jet}}$. Both of these general trends are reproduced by PYTHIA 8, SHERPA, and HERWIG++, although for example, PYTHIA 8 disagrees with the exact value at low $p_T^{\text{jet}}$ for $\zeta < 1\%$ and all $p_T^{\text{jet}}$ for $\zeta < 0.1\%$. For the $\zeta < 0.1\%$ case, SHERPA and PYTHIA 8 bracket the data, with SHERPA predicting more particles with a lower $\zeta$ fraction, while HERWIG++ is much closer to the data. The average values of $\sqrt{\zeta}$ and $\zeta^2$ for individual particles as a function of jet $p_T^{\text{jet}}$ in the top panel of Fig. 9 show a decreasing trend that is qualitatively similar to the trend for the average $\zeta$ in Fig. 7. For $\sqrt{\zeta}$, PYTHIA 8/HERWIG++ and SHERPA bracket the data, although PYTHIA 8 agrees with the data within the uncertainty. SHERPA predicts a significantly higher average $\zeta^2$ than is present in the data. As with $\langle n_{ch}\rangle$, the average value of $\langle \sum_{i \in \text{jet}} \zeta_i^{1/2}\rangle$ increases with the jet $p_T$, while the $p_T$ dependence of $\langle \sum_{i \in \text{jet}} \zeta_i^2\rangle$ is more complicated as it first decreases and then slowly increases with jet $p_T$. This trend is well reproduced by PYTHIA and HERWIG++, but not by SHERPA.
To present more differential information, the full unfolded distributions for $n_{	ext{ch}}$, $\zeta$, $p_{T}^\text{rel}$, and $r$ are shown in Figs. 10–13, respectively, for representative $p_T^\text{jett}$ bins. Many of the relevant trends are captured in the above discussion about the $p_T^\text{jett}$ dependence of the moments. However, finer information that may be useful for generator tuning is provided by the differential distributions.

### B. Quark and gluon distributions

As discussed in Sec. V B, the more forward and the more central of the two selected jets can be separated to study differences between the radiation patterns within quark and gluon jets. Using the fraction of quark jets $f_q$ in the two jet samples (forward $f$ and central $c$), one can extract the quark ($h_q^i$) and gluon ($h_g^i$) jet fragmentation properties separately by solving a system of equations per bin $i$ of an observable:

\begin{align}
    h_q^i &= f_q h_q^0 + (1 - f_q) h_g^0, \\
    h_g^i &= f_q h_q^0 + (1 - f_q) h_g^0,
\end{align}

where $f_q$ is the fraction of quark jets in sample $x$ (see Fig. 3 for the gluon fraction) and the nominal fractions are taken from the default PYTHIA simulation described in Sec. IV. The flavor of a jet is defined as the type of the highest-energy parton from the event record (all partons prior to hadronization) matched to the jet via ghost association. This definition is not unique because quark and gluon labels are not universal due to color connections with other partons in the event.\(^5\) In addition to the uncertainty in $h_q^i$.

\(^5\)However, for isolated jets, the topology dependence is predicted to be much smaller than the difference between quark and gluon jets [102].
and $h^f$ from the unfolding method, uncertainties in the extracted $h^q$ and $h^g$ distributions arise from the PDF choice, from the matrix elements, from the fragmentation model (due to flavor changing), and from the method nonclosure. The determination of the uncertainty from the choice of PDF uses the NNPDF uncertainty set [NNPDF 2.3 at LO in QCD and QED with $\alpha_s(m_Z) = 0.119$] and the matrix-element uncertainty is estimated by comparing the nominal fractions from PYTHIA with those from HERWIG.\textsuperscript{6} The nonclosure uncertainty is due to the small (subpercent level) differences between forward and central quark jets, as well as forward and central gluon jets, resulting from an $\eta$ dependence in the jet fragmentation at a fixed $p_T$ \cite{Cacciari:2011ma}. When presenting the average properties in bins of jet $p_T$, the binning correction described in Sec. VI is also applied and the corresponding uncertainty contributes to the total uncertainty (though it is smaller than other sources of uncertainty).

The matrix-element uncertainty dominates the total uncertainty in the extraction procedure, resulting in an uncertainty that is about 1\% at high jet $p_T$ and about 5\% at low to moderate jet $p_T$ for quark jets, with the inverse trend for gluon jets (low uncertainty at low jet $p_T$ and large uncertainty at high jet $p_T$). The extractions presented here use leading-order matrix elements and leading-logarithm parton shower programs; higher-order effects that modify the fractions $f$ are not included in this leading-order extraction. Figure 14 shows the extracted quark and gluon distributions for jets with $1000 \text{ GeV} < p_T^\text{jet} < 1200 \text{ GeV}$. 

---

\textsuperscript{6}These two generators also use different PDF sets, so this uncertainty is double counted in the overall uncertainty.
To reinforce the simulation dependence of these extractions, the data distributions are referred to as “extracted quark-like” and “extracted gluon-like.”

A key challenge with the extraction based on Eqs. (4) and (5) is that it strongly depends on simulation for the fractions \( f_g \) and \( f_q \). A new approach that does not require the input of any fractions is topic modeling [103,104], which holds great promise for learning about quark- and gluon-like jets with less input from theory. In this approach, one can extract distributions of “topics” \( T_1 \) and \( T_2 \):

\[
h^T_1 = \frac{h^f_i - (\min_j \{h^f_j/h^c_j\}) \times h^c_i}{1 - \min_j h^f_j/h^c_j},
\]

\[
h^T_2 = \frac{h^c_i - (\min_j \{h^f_j/h^c_j\}) \times h^f_i}{1 - \min_j h^f_j/h^c_j},
\]

in the limit that \( \min_j \{h^f_j/h^c_j\} = \min_j \{h^f_j/h^g_j\} = 0 \), \( h^T_1 = h^c \) and \( h^T_2 = h^f \). When this is not exactly the case, the topics are universal but not pure combinations of quarks and gluons. The extracted topics using \( n_{\text{ch}} \) in two jet \( p_T \) bins are shown in Fig. 15. The very low \( n_{\text{ch}} \) region is dominated by quarks and the very high \( n_{\text{ch}} \) region is dominated by gluons and therefore \( n_{\text{ch}} \) nearly has the property that \( \min_j \{h^f_j/h^c_j\} \approx \min_j \{h^f_j/h^g_j\} \approx 0 \). Therefore, the first topic is well aligned with quarks and the second topic is more gluon-like. This alignment is better for quarks than for gluons, but the second topic does converge to the gluon distribution at high jet \( p_T \). Other observables aside from \( n_{\text{ch}} \) are not considered for topic modeling because there are no bins where \( h^f_j/h^c_j = 0 \) or \( h^f_j/h^g_j = 0 \) is approximately true and therefore the topics do not align with quark- and gluon-like quantities.
While the full quark and gluon distributions presented in Fig. 14 cannot be predicted from perturbative QCD, it is possible to model the jet dependence of the moments of the $\zeta$ distribution. Positive moments of the fragmentation function have a perturbative evolution with a proper $\alpha_S$ power series via DGLAP-like equations. In general, there are two terms that contribute to the right-hand side of Eq. (1) that prevent an analytic solution: one term proportional to $D_{gh}$ and one term proportional to $D_{hq}$, where the coefficients for the $\kappa$ sums are the Mellin transforms $\tilde{P}_{p'\rightarrow p}(\kappa) = \int_0^\infty d\zeta \zeta^\kappa P_{p'\rightarrow p}(\zeta)$ for $p' = g$ and $p' = q$, respectively. For gluon jets ($p = g$), the $g \rightarrow qq'$ splitting function is finite,\footnote{The splitting function $\tilde{P}_{q\rightarrow g}$ is also finite, but is not numerically small compared with $\tilde{P}_{g\rightarrow g}$ except when $\kappa$ is very small so this case is not considered further.} so $|\tilde{P}_{q\rightarrow g}| \gg |\tilde{P}_{g\rightarrow q}|$ for $\kappa \neq 0$ where $\tilde{P}_{g\rightarrow q}$ switches sign. Therefore, away from $\kappa \approx 0.8$ and in the modified leading-logarithm approximation\footnote{This means resummation that includes the leading-order splitting functions and the first-order running of the strong coupling. A more refined calculation [18,19] using SCET [13–16] and fragmenting jet functions [105–107] is possible. However, the deviations from this simple approach are higher-order corrections and do not qualitatively change the comparisons in this section.} [73,108,109],

$$\left\langle \sum_{i \in \text{jet}} \delta_{\text{gluons}} \right\rangle \sim \log(p_T^2/\Lambda^2)^{2P_{g\rightarrow q}(\kappa)/\beta_0},$$

where $\beta_0$ is the first term in the QCD $\beta$-function and $\Lambda$ is a nonperturbative parameter (of order $\Lambda_{QCD}$). The predictions...
FIG. 15. The jet topics extracted using charged-particle multiplicity $n_{ch}$ for jets in two transverse momentum, $p_T$, bins together with the topics and quark and gluon distributions from PYTHIA. The uncertainty bands on the data include both statistical and systematic uncertainties while the error bars are due to statistical uncertainties only.

FIG. 16. The dependence on transverse momentum $p_T$ of the average extracted gluon-like transverse-momentum-fraction weighted sum $\sum_{i\in \text{jet}} \zeta_{\kappa}^{i}$ for (top left) $\kappa = 0.5$, (top right) $\kappa = 1.0$, and $\kappa = 2.0$ (bottom). For comparison, the results from PYTHIA and a simple leading-logarithm (LL) calculation are also presented. The prediction is normalized to the data in the sixth jet $p_T$ bin, called the anchor bin and indicated by an arrow. The uncertainty band on the calculation is from varying $\Lambda$ in Eq. (6) up and down by a factor of 2 from its nominal value of 400 MeV (in most regions, this band is not much wider than the linewidth and thus not visible). The uncertainty bands on the data include both statistical and systematic uncertainties while the error bars are due to statistical uncertainties only.
are scaled to match the data in the sixth jet $p_T$ bin (referred to as the “anchor bin”). There is no a priori reason to select any particular bin as the anchor bin so one of the first bins after the lowest-threshold unscaled jet trigger is selected. Figure 16 shows the distributions of the average $\sum_{i \in \text{jet}} S_i^\kappa$ for $\kappa = 0.5, 1.0,$ and $2.0$ for gluon jets.

As mentioned above, $\bar{P}_{g\rightarrow g}(1)$ is predicted to change sign at $\kappa = 0.8$, a trend which is supported by the data: for low $\kappa$, the average value increases with $p_T$ and when $\kappa$ is large, the average decreases with $p_T$. For $\kappa = 1$, momentum conservation and isospin symmetry predict that the average value of $\sum_{i \in \text{jet}} S_i^\kappa$ should be constant and approximately $2/3$, the ratio of charged pions to all pions. The leading-logarithm (LL) calculation predicts $\bar{P}_{g\rightarrow g}(1) \approx 0$ so the $p_T$ dependence is already negligible compared with the $\kappa = 0.5$ and $\kappa = 2$ cases.

When $\kappa \to 0$, both the quark and gluon fragmentation-function Mellin transforms diverge and so the analysis with Eq. (6) is not accurate. The $\kappa \to 0$ limit is $\langle n_{ch} \rangle$ and there is no known series in $\alpha_s$ to describe its $p_T$ dependence. Despite this, the anomalous dimension for the $p_T^{\text{jet}}$ dependence of $\langle n_{ch} \rangle$ has been calculated to “$N^3\text{LO}$” where the series is in $\sqrt{\alpha_S}$ instead of $\alpha_S$ [22,23]. Figure 17 shows $\langle n_{ch} \rangle$ as a function of $p_T^{\text{jet}}$ for both extracted quark-like and gluon-like jets as well as the topic extraction along with the prediction for the pure quark/gluon case. Gluon jets from data deviate significantly from simulation and from the calculation at high jet $p_T$; this is also true to a lesser extent for quark jets, which seem to have a different slope that is most prominent at low jet $p_T$. A similar trend was first observed in Ref. [7], albeit with lower precision in the highest $p_T$ bins. There are several possibilities for this discrepancy, such as an unaccounted for potential source of bias in the quark/gluon jet fractions. The data in the right panel of Fig. 17 do not yet conclusively support or reject this hypothesis; with more data, it may be possible to determine if the data match topic 2 in PYTHIA or deviate as is the case for gluons in the left panel of Fig. 17.

The $p_T$ dependence of the average $\zeta$, $p_T^{\text{rel}}$, and $r$ are shown in Fig. 18. Gluon jets have more constituents than quark jets on average so their average $\zeta$ is lower. For both quark and gluon jets, $\langle \zeta \rangle$ decreases with the jet $p_T$ in part because constituent multiplicity increases with $p_T$. Gluon jets are wider than quark jets on average, but both quark and gluon jets become denser with increasing jet $p_T$. The data show nearly the same trends as PYTHIA in all cases.

\footnote{The measured value is not exactly 2/3 because a jet’s energy is only about 60% due to pions.}
This paper documents a measurement of track-based jet fragmentation functions in $pp$ collisions at $\sqrt{s} = 13$ TeV. The analysis uses a dataset corresponding to an integrated luminosity of 33 fb$^{-1}$ recorded by the ATLAS detector at the LHC. Multiple jet properties, including the charged-particle multiplicity, the momentum fraction carried by charged particles, and angular properties of the radiation pattern inside jets are studied. There are key areas where there are significant disagreements between the ATLAS default MC simulation (PYTHIA 8.2 with the A14 tune, HERWIG++, and SHERPA) and the data, especially for the radial profiles and momentum distributions in SHERPA. The radial profile is systematically broader in data than in simulation, but the momentum transverse to the jet axis and the momentum fraction are well modeled within the precision of this measurement. Near 1 TeV in jet $p_T$, these measurements have achieved percent-level uncertainties for a variety of observables. In addition to measuring the forward, central, and combined jet distributions, the forward and central jet spectra are considered separately to study quark- and gluon-like distributions. A first measurement of topic modeling for the charged-particle multiplicity provides a promising alternative to traditional methods of extracting quark- and gluon-jet distributions that use input from simulation. The simulations provide a reasonable description of the quark-like data across the jet $p_T$ range presented in this measurement, but the gluon-like data have systematically fewer charged particles than the simulations by about 10%.

The unfolded data are made public through HEPDATA to provide input to help improve both perturbative and non-perturbative aspects of fragmentation modeling in the future.
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; BMWFFW 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 and DNSRC, Denmark; IN2P3-CNRS, CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MINERVA and MIZ, Korea; LNC09, Lithuania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC Ki, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BKDF, CANARIE, CRC and Compute Canada, Canada; COST, ERC, ERDF, Horizon 2020, and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programs cofinanced by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; CERCA Programme Generalitat de Catalunya, Spain; The Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, 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), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [110].


[35] CLEO Collaboration, Study of gluon versus quark fragmentation in $Y \rightarrow gg$ and $e^+e^- \rightarrow q\bar{q}T$ events at $\sqrt{s}=10$ GeV, Phys. Rev. D 56, 17 (1997).


[53] CDF Collaboration, Measurement of the $k_T$ Distribution of Particles in Jets Produced in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. Lett. 102, 232002 (2009).


[74] CMS Collaboration, Search for a Higgs boson in the decay channel $H \rightarrow ZZ^{(*)} \rightarrow q\bar{q}\ell^+\ell^-$ in pp collisions at $\sqrt{s} = 7$ TeV, J. High Energy Phys. 04 (2012) 036.


G. Aad,101 B. Abbott,128 D. C. Abbott,102 O. Abdi...,122 M. Abdo...,128 J. Ab...
PROPERTIES OF JET FRAGMENTATION USING CHARGED... Phys. Rev. D 100, 052011 (2019)
PROPERTIES OF JET FRAGMENTATION USING CHARGED … PHYS. REV. D 100, 052011 (2019)

140 Dep Física and CEFITEC of Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal
141 Institute of Physics of the Czech Academy of Sciences, Prague, Czech Republic
142 Czech Technical University in Prague, Prague, Czech Republic
143 Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic
144 Particle Physics Department, Rutherford Appleton Laboratory, Didcot, United Kingdom
145 IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
146 Santa Cruz Institute for Particle Physics, University of California Santa Cruz, Santa Cruz, California, USA

147 Departamento de Física, Pontificia Universidad Católica de Chile, Santiago, Chile
148 Department of Physics, Universidad Técnica Federico Santa María, Valparaíso, Chile
149 Department of Physics, University of Washington, Seattle, Washington, USA
150 Department of Physics and Astronomy, University of Sheffield, Sheffield, United Kingdom
151 Department of Physics, Shinshu University, Nagano, Japan
152 Department Physik, Universität Siegen, Siegen, Germany
153 SLAC National Accelerator Laboratory, Stanford, California, USA
154 Physics Department, Royal Institute of Technology, Stockholm, Sweden
155 Departments of Physics and Astronomy, Stony Brook University, Stony Brook, New York, USA
156 Department of Physics and Astronomy, University of Sussex, Brighton, United Kingdom
157 School of Physics, University of Sydney, Sydney, Australia
158 Institute of Physics, Academia Sinica, Taipei, Taiwan
159a E. Andronikashvili Institute of Physics, Iv. Javakhishvili Tbilisi State University, Tbilisi, Georgia
159b High Energy Physics Institute, Tbilisi State University, Tbilisi, Georgia
160 Department of Physics, Technion, Israel Institute of Technology, Haifa, Israel
161 Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel
162 Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece
163 International Center for Elementary Particle Physics and Department of Physics, University of Tokyo, Tokyo, Japan
164 Graduate School of Science and Technology, Tokyo Metropolitan University, Tokyo, Japan
165 Department of Physics, Tokyo Institute of Technology, Tokyo, Japan
166 Tomsk State University, Tomsk, Russia
167 Department of Physics, University of Toronto, Toronto, Ontario, Canada
168a TRIUMF, Vancouver, British Columbia, Canada
168b Department of Physics and Astronomy, York University, Toronto, Canada, Canada
169 Division of Physics and Tomonaga Center for the History of the Universe, Faculty of Pure and Applied Sciences, University of Tsukuba, Tsukuba, Japan
170 Department of Physics and Astronomy, Tufts University, Medford, Massachusetts, USA
171 Department of Physics and Astronomy, University of California Irvine, Irvine, California, USA
172 Department of Physics and Astronomy, University of Uppsala, Uppsala, Sweden
173 Department of Physics, University of Illinois, Urbana, Illinois, USA
174 Instituto de Física Corpuscular (IFIC), Centro Mixto Universidad de Valencia—CSIC, Valencia, Spain
175 Department of Physics, University of British Columbia, Vancouver, British Columbia, Canada
176 Department of Physics and Astronomy, University of Victoria, Victoria, British Columbia, Canada
177 Fakultät für Physik und Astronomie, Julius-Maximilians-Universität Würzburg, Würzburg, Germany
178 Department of Physics, University of Warwick, Coventry, United Kingdom
179 Waseda University, Tokyo, Japan
180 Department of Particle Physics, Weizmann Institute of Science, Rehovot, Israel
181 Department of Physics, University of Wisconsin, Madison, Wisconsin, USA
182 Fakultät für Mathematik und Naturwissenschaften, Fachgruppe Physik, Bergische Universität Wuppertal, Wuppertal, Germany
183 Department of Physics, Yale University, New Haven, Connecticut, USA
184 Yerevan Physics Institute, Yerevan, Armenia

aDeceased.
bAlso at Department of Physics, King’s College London, London, United Kingdom.
cAlso at Istanbul University, Dept. of Physics, Istanbul, Turkey.
dAlso at Instituto de Física Teorica, IFT-UAM/CSIC, Madrid, Spain.
eAlso at TRIUMF, Vancouver, British Columbia, Canada.
fAlso at Department of Physics and Astronomy, University of Louisville, Louisville, Kentucky, USA.
Also at Physics Department, An-Najah National University, Nablus, Palestine.

Also at Department of Physics, California State University, Fresno, California, USA.

Also at Department of Physics, University of Fribourg, Fribourg, Switzerland.

Also at Physics Dept, University of South Africa, Pretoria, South Africa.

Also at Departament de Fisica de la Universitat Autonoma de Barcelona, Barcelona, Spain.

Also at Tomsk State University, Tomsk, and Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia.

Also at The Collaborative Innovation Center of Quantum Matter (CICQM), Beijing, China.

Also at Departamento de Fisica, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal.

Also at Universita di Napoli Parthenope, Napoli, Italy.

Also at Institute of Particle Physics (IPP), Canada.

Also at Department of Physics, University of Adelaide, Adelaide, Australia.

Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg, Russia.

Also at Department of Financial and Management Engineering, University of the Aegean, Chios, Greece.

Also at Department of Physics, California State University, East Bay, California, USA.

Also at Institucio Catalana de Recerca i Estudis Avancats, ICREA, Barcelona, Spain.

Also at Department of Physics, University of Michigan, Ann Arbor, Michigan, USA.

Also at LAL, Université Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France.

Also at Graduate School of Science, Osaka University, Osaka, Japan.

Also at Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany.

Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan.

Also at Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen/Nikhef, Nijmegen, Netherlands.

Also at Institute of Theoretical Physics, Ilia State University, Tbilisi, Georgia.

Also at CERN, Geneva, Switzerland.

Also at Department of Physics, Stanford University, Stanford, California, USA.

Also at Manhattan College, New York, New York, USA.

Also at Joint Institute for Nuclear Research, Dubna, Russia.

Also at Hellenic Open University, Patras, Greece.

Also at The City College of New York, New York, New York, USA.

Also at Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, China.

Also at Department of Physics, California State University, Sacramento, USA.

Also at Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia.

Also at Département de Physique Nucléaire et Corpusculaire, Université de Genève, Genève, Switzerland.

Also at Department of Physics and Astronomy, University of Sheffield, Sheffield, United Kingdom.

Also at Louisiana Tech University, Ruston, Louisiana, USA.

Also at School of Physics, Sun Yat-sen University, Guangzhou, China.

Also at Institute for Nuclear Research and Nuclear Energy (INRNE) of the Bulgarian Academy of Sciences, Sofia, Bulgaria.

Also at Faculty of Physics, M.V. Lomonosov Moscow State University, Moscow, Russia.

Also at Department of Applied Physics and Astronomy, University of Sharjah, Sharjah, United Arab Emirates.

Also at Institut für Experimentalphysik, Universität Hamburg, Hamburg, Germany.

Also at CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France.

Also at National Research Nuclear University MEPhI, Moscow, Russia.

Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary.

Also at Giresun University, Faculty of Engineering, Giresun, Turkey.

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

Also at LPNHE, Sorbonne Université, Paris Diderot Sorbonne Paris Cité, CNRS/IN2P3, Paris, France.

Also at Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan, USA.