Measurement of jet charge in dijet events from $\sqrt{s} = 8$ TeV $p p$ collisions with the ATLAS detector

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1103/PhysRevD.93.052003">http://dx.doi.org/10.1103/PhysRevD.93.052003</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Physical Society</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://hdl.handle.net/1721.1/113821">http://hdl.handle.net/1721.1/113821</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
</tbody>
</table>
Measurement of jet charge in dijet events from $\sqrt{s} = 8$ TeV $pp$ collisions with the ATLAS detector

G. Aad et al.*
(ATLAS Collaboration)
(Received 18 September 2015; published 2 March 2016)

The momentum-weighted sum of the charges of tracks associated to a jet is sensitive to the charge of the initiating quark or gluon. This paper presents a measurement of the distribution of momentum-weighted sums, called jet charge, in dijet events using 20.3 fb$^{-1}$ of data recorded with the ATLAS detector at $\sqrt{s} = 8$ TeV in $pp$ collisions at the LHC. The jet charge distribution is unfolded to remove distortions from detector effects and the resulting particle-level distribution is compared with several models. The $p_T$ dependence of the jet charge distribution average and standard deviation are compared to predictions obtained with several leading-order and next-to-leading-order parton distribution functions. The data are also compared to different Monte Carlo simulations of QCD dijet production using various settings of the free parameters within these models. The chosen value of the strong coupling constant used to calculate gluon radiation is found to have a significant impact on the predicted jet charge. There is evidence for a $p_T$ dependence of the jet charge distribution for a given jet flavor. In agreement with perturbative QCD predictions, the data show that the average jet charge of quark-initiated jets decreases in magnitude as the energy of the jet increases.

DOI: 10.1103/PhysRevD.93.052003

1. INTRODUCTION

Quarks and gluons produced in high-energy particle collisions hadronize before their electric charge can be directly measured. However, information about the electric charge is embedded in the resulting collimated sprays of hadrons known as jets. One jet observable sensitive to the electric charge of quarks and gluons is the momentum-weighted charge sum constructed from charged-particle tracks in a jet [1]. Called the jet charge, this observable was first used experimentally in deep inelastic scattering studies [2–8] to establish a relationship between the quark model and hadrons. Since then, jet charge observables have been used in a variety of applications, including tagging the charge of $b$-quark jets [9–19] and hadronically decaying $W$ bosons [20–25] as well as distinguishing hadronically decaying $W$ bosons from jets produced in generic quantum chromodynamic (QCD) processes [25] and quark jets from gluon jets [24,26–28].

The study presented in this paper is a measurement of the jet charge distribution in inclusive dijet events from $pp$ collisions at the LHC. Inclusive dijet events provide a useful environment for measuring the jet charge as they are an abundant source of gluon-initiated and quark-initiated jets. There are fewer theoretical ambiguities from close-by jets and large-angle radiation associated with assigning the jet flavor in events with two jets than in events with higher jet multiplicities. Furthermore, the transverse momentum ($p_T$) range accessible in dijet events is broad, $O(10)$ GeV up to $O(1000)$ GeV. Since the initial state at the LHC has a net positive charge, the probability for positively charged quarks to be produced in $pp$ collisions is higher than that for negatively charged quarks. The probability for collisions to involve a positively charged valence up quark in the proton increases with the parton center-of-mass energy $\sqrt{s}$. Thus the average jet charge in inclusive dijet events is expected to increase with $\sqrt{s}$ if it is correlated with the quark charge. The parton distribution functions (PDFs) encode the probabilities to find gluons and certain flavors of quarks at given momentum fractions $x$ of the proton. The momentum fractions of the two initial partons $x_1$, $x_2$, and the proton and parton center-of-mass energies $\sqrt{s}$ and $\sqrt{\hat{s}}$ are related by $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$. The PDFs are fairly well constrained [29–33] in the $x$ range relevant for this study, 0.005–0.5. However, if the jet charge is directly sensitive to the parton flavor, its $p_T$ dependence can provide a consistency check using new information beyond the jet $p_T$, which is currently used in PDF fits. The PDFs are not the only nonperturbative input needed to model the jet charge distribution and its evolution with $\sqrt{s}$. As a momentum-weighted sum over jet constituents, the jet charge is sensitive to the modeling of fragmentation. Previous studies have shown that there are qualitative differences between the charged-particle track multiplicities of jets in data and as predicted by the leading models of hadron production [26]. Thus, a measurement of the jet charge distribution...
with a range of quark/gluon compositions can provide a
constraint on models of jet formation.

While the change in jet parton flavor due to the PDF $x$
dependence predicts most of the variation in the jet charge
distribution as a function of $\sqrt{s}$, there is a second
contribution due to the energy dependence of the fragmenta-
tion functions. Ratios of the charge distribution moments
at different values of $\sqrt{s}$ can be calculated perturbatively.
Recent calculations [34,35] in the context of soft collinear
effective theory [36–39] show a significant reduction in the
magnitude of the average jet charge for a given jet flavor as
a function of jet energy. Information from PDFs can be
used to extract the energy dependence of the average jet
charge in the data for direct comparisons to the predictions.
This paper presents a measurement of the $p_T$ dependence
of the jet charge distribution’s mean and standard deviation
in dijet events in $pp$ collisions at $\sqrt{s} = 8$ TeV with the
ATLAS detector. The jet $p_T$ is a measurable quantity that is
strongly related to $\sqrt{s}$. The average jet charge is extracted
for both the leading and subleading jet and they are
distinguished based on their relative orientation in rapidity.

After a description of the ATLAS detector (Sec. II), the data
and simulated samples (Sec. III) and the detector- and
particle-level objects and selections used in the analysis
(Sec. IV), Sec. V details the construction of the jet charge
distribution as a function of $\theta$ and the
at the nominal interaction point (IP) in the center of the detector

\[ \eta = \ln \tan(\theta/2). \]
\[ \Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2} \]
\[ \Delta \eta = \frac{\Delta E}{E} \]

\[ \Delta \phi = \frac{\Delta p_T}{p_T} \]

The pseudorapidity is defined in terms of the polar angle $\theta$
as $\eta = -\ln \tan(\theta/2)$. The variable $\Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2}$ is a
measure of how close two objects are in the $(\eta, \phi)$ plane. The
rapidity of a four-vector is defined as $y = 0.5 \ln \left( \frac{E + p_T}{E - p_T} \right)$, where $E$
is the energy and $p_T$ is the component of the momentum parallel to
the beam axis.

II. THE ATLAS DETECTOR

ATLAS [40] is a general-purpose detector designed to
measure the properties of particles produced in high-energy
$pp$ collisions with nearly a full 4$\pi$ coverage in solid angle.
The innermost subsystem of the detector is a series of
tracking devices used to measure charged-particle trajec-
tories bent in a 2 T axial field provided by a solenoid whose
axis is parallel with the beam direction. This inner detector
(ID) consists of a silicon pixel detector surrounded by a
semiconductor microstrip detector (SCT) and a straw-tube
trackers. It has full coverage in $\phi$ and can detect particles
with $|\eta| < 2.5$. Charged-particle tracks are reconstructed
from all three ID components, providing measurements of the
transverse momentum of tracks with a resolution
$\sigma_{p_T}/p_T = 0.5% \times p_T/GeV \oplus 1%$. The track reconstruction
algorithm fits five track parameters: $d_0$, $z_0$, $\phi$, $\theta$, and $q/p$,
where $d_0$ and $z_0$ are the transverse and longitudinal impact
parameters, respectively, $q$ is the track charge and $p$ is the
track momentum. Excellent spatial precision is required to
maintain a well-performing track reconstruction out to and
exceeding charged-particle $p_T$ of 1 TeV, where track
sagittas are $\lesssim 0.2$ mm.

Surrounding the ID and solenoid are electromagnetic and
hadronic calorimeters to measure showers from charged
and neutral particles. The high-granularity liquid-argon
(LAr) sampling electromagnetic calorimeter is located just
beyond the solenoid and spans the range $|\eta| < 3.2$. Beyond
the electromagnetic calorimeter is the two-component
hadronic calorimeter that uses scintillator-tile sampling
technology in the range $|\eta| < 1.7$ and LAr sampling
technology for $1.5 < |\eta| < 3.2$. Additional calorimeters
are located in the forward region. Surrounding the calo-
rimeters is a muon spectrometer, with trigger and precision
chambers, and incorporating three large toroid magnets
composed of eight coils each.

Due to the large event rate, not every collision can be
recorded for processing offline. Events are selected using a
three-level trigger system [41] that is hardware based at the
first level and software based for the two following levels.
An event must satisfy all three trigger levels to be recorded
for further processing. At each stage of the trigger, energy
thresholds are placed on jet-like objects, with the similarity
between online and offline jets increasing with each level.
The first level makes decisions based on low-granularity
calorimeter towers with thresholds that are typically less
than half of the energy required by jets at the second level.
A simple jet reconstruction algorithm is used at the second
level in regions around the jets identified by the first level.
The third level (known as the Event Filter) clusters jets with
the same algorithm as used offline over the entire detector
with thresholds that are typically 20–30 GeV higher than at
level two. The single-jet trigger thresholds increase at each
level due not only to differences in the jet reconstruction
and calibrations, but also to meet the different bandwidth
requirements at each trigger level. For low-$p_T$ dijets, the
event rate is far too large to save every event that passes the
trigger selection and so most of the jet triggers are
prescaled to artificially lower their recording rate.

III. DATA AND SIMULATED SAMPLES

This measurement uses the full data set of $pp$ collisions recorded by the ATLAS detector in 2012, corresponding to
an integrated luminosity of 20.3 fb$^{-1}$ at a center-of-mass
energy of $\sqrt{s} = 8$ TeV. Events are only considered if they
are collected during stable beam conditions and satisfy all
data-quality requirements [42]. To reject noncollision events, there must be a primary vertex reconstructed from at least two tracks each with $p_T > 400$ MeV [43]. Due to the high instantaneous luminosity and the large total inelastic proton-proton cross section, on average there are about 21 simultaneous (pileup) collisions in each bunch crossing.

A set of single-jet triggers is used to collect dijet events with high efficiency. Table I shows the collected luminosity for each trigger as well as the offline jet $p_T$ ranges used, chosen such that the trigger is fully efficient. The highest-$p_T$ trigger is not prescaled.

Monte Carlo (MC) simulated events are generated in $p_T$ slices in order to ensure a large number of events over a broad range of reconstructed jet $p_T$, given constraints on the available computing resources. The $p_T$ slices span the interval 0 to 5 TeV in ranges that approximately double with each increasing slice, starting with a range of size 8 GeV and ending with a range of size 2240 GeV. The baseline sample used for the measurement is generated with PYTHIA 8.175 [44] with the AU2 [45] set of tuned parameters (tune) and the next-to-leading-order (NLO) PDF set\(^2\) CT10 [48,49]. Another large sample of events is generated with HERWIG++ 2.63 [50,51] with tune EE3 [52] and leading-order (LO) PDF set CTEQ6L1 [53] (particle-level samples with CT10 and EE4 are also used for comparisons). Both PYTHIA and HERWIG++ are LO in perturbative QCD for the $(2 \rightarrow 2)$ matrix element and resum the leading logarithms in the parton shower. However, the ordering of emissions in the MC resummation in the shower differs between these two generators: PYTHIA implements $p_T$-ordered showers [54] whereas HERWIG++ uses angular ordering [55]. The phenomenological modeling of the nonperturbative physics also differs between PYTHIA and HERWIG++. In addition to different underlying-event models (Ref. [56] for PYTHIA and an eikonal model [57] for HERWIG++ the hadronization models differ between PYTHIA (Lund string model [58]) and HERWIG++ (cluster model [59]). These two schemes are known [26] to predict different numbers of charged particles within jets and different distributions of the charged-particle energies within jets, both of which are important for the jet charge. All tunes of the underlying event that are used with PYTHIA and HERWIG++ in this analysis use LHC data as input. As discussed in Sec. I, the corrected data are compared to models with various PDF sets; for consistency, each set has a dedicated underlying-event tune constructed in the same way from a fixed set of data inputs (AU2) described in detail in Ref. [45]. The PDF sets include LO sets CTEQ6L1 [53] and MSTW2008LO [29] as well as NLO sets CT10 [48,49], NNPDF21 NLO [60], and MSTW2008NLO [29]. A sample generated with a NLO matrix element from POWHEG-BOX R2262 [61–64] (henceforth referred to as POWHEG) with PDF set CT10 interfaced with PYTHIA 8.175 and the AU2 tune is also used for comparisons.

All MC samples are processed using the full ATLAS detector simulation [65] based on GEANT4 [66].

### IV. OBJECT RECONSTRUCTION AND EVENT SELECTION

The reconstructed objects used for the jet charge as well as for the event selection are described in Sec. IVA. The fiducial definition of the measurement, unfolded to particle level, is given in Sec. IV B.

#### A. Object reconstruction at detector level

Jets are clustered using the anti-$k_T$ jet algorithm [67] with radius parameter $R = 0.4$ implemented in fastjet [68] from topological calorimeter-cell clusters [69], calibrated using the local cluster weighting algorithm [70,71]. An overall jet energy calibration accounts for residual detector effects as well as contributions from pileup [72] in order to make the reconstructed jet energy an unbiased measurement of the particle-level jet energy. Jets are required to be central ($|\eta| < 2.1$) so that their charged particles are within the $|\eta| < 2.5$ coverage of the ID.

When more than one primary vertex is reconstructed, the one with the highest $\sum p_T^2$ of tracks is selected as the hard-scatter vertex. Events are further required to have at least two jets with $p_T > 50$ GeV and only the leading two jets are considered for the jet charge measurement. To select dijet topologies, the two leading jets must have $p_T^{\text{lead}}/p_T^{\text{sublead}} < 1.5$, where $p_T^{\text{lead}}$ and $p_T^{\text{sublead}}$ are the transverse momenta of the jets with the highest and second-highest $p_T$, respectively. The jet with the smaller (larger)
The measurement is carried out within a fiducial volume matching the experimental selection to avoid extrapolation into unmeasured kinematic regions that have additional model dependence and related uncertainties. Particle-level definitions of the reconstructed objects are chosen to be as close as possible to those described in Sec. IV A. Particle-level jets are clustered from generated stable particles with a mean lifetime $\tau > 30$ ps, excluding muons and neutrinos. As with the detector-level jets, particle-level jets are clustered with the anti-$k_t$, $R = 0.4$ algorithm. In analogy to the ghost association of tracks to jets performed at detector level, any charged particle clustered in a particle-level jet is considered for the jet charge calculation.\footnote{There is no $p_T > 500$ MeV threshold applied to charged particles. The impact of applying such a threshold is negligible for all $p_T$ bins except the first two where effects of up to 1% are observed in the mean and standard deviation of the jet charge.} There must be at least two jets with $|\eta| < 2.1$ and $p_T > 50$ GeV. The two highest-$p_T$ jets must satisfy the same $p_T$-balance requirement between the leading and subleading jet as at detector level ($p_T^{lead}/p_T^{sublead} < 1.5$). Due to the high-energy and well-separated nature of the selected jets, the hard-scatter quarks and gluons can be cleanly matched to the outgoing jets. While it is possible to classify jets as quark- or gluon-initiated beyond leading order in $m_{jet}/E_{jet}$\footnote{G. Aad et al. and more central jet charge distributions can exploit the freedom (resulting from the track fit) less than 3.0. Additional quality criteria are applied to select tracks originating from the collision vertex and reject fake tracks reconstructed from random hits in the detector. In particular, tracks must be well matched to the hard-scatter vertex with $|z_0 \sin(\theta)| < 1.5$ mm and $|d_0| < 1$ mm, where $z_0$ and $d_0$ are calculated with respect to the primary vertex. Tracks must furthermore have at least one hit in the pixel detector and at least six hits in the SCT. The matching of tracks with the calorimeter-based jets is performed via the ghost-association technique\footnote{There is no unique way to define the jet charge. The most naive construction is to add up the charge of all tracks associated with a jet. However, this scheme is very sensitive to lost or extraneous soft radiation. Therefore, a weighting scheme is introduced to suppress fluctuations. Using the tracks assigned to a jet by ghost association, the jet charge $Q_J$ of a jet $J$ is calculated using a transverse-momentum-weighting scheme\footnote{There is no $p_T > 500$ MeV threshold applied to charged particles. The impact of applying such a threshold is negligible for all $p_T$ bins except the first two where effects of up to 1% are observed in the mean and standard deviation of the jet charge.}}: the jet clustering process is repeated with the addition of ghost versions of measured tracks that have the same direction but infinitesimally small $p_T$, so that they do not change the properties of the calorimeter jets. A track is associated with a jet if its ghost version is contained in the jet after reclustering. The distribution of the number of tracks in two representative jet $p_T$ ranges is shown in Fig. 1. The number of tracks increases with jet $p_T$ and the data fall between the predicted distributions of PYTHIA and HERWIG++.}

V. CONSTRUCTING THE JET CHARGE

There is no unique way to define the jet charge. The most naive construction is to add up the charge of all tracks associated with a jet. However, this scheme is very sensitive to lost or extraneous soft radiation. Therefore, a weighting scheme is introduced to suppress fluctuations. Using the tracks assigned to a jet by ghost association, the jet charge $Q_J$ of a jet $J$ is calculated using a transverse-momentum-weighting scheme\footnote{There is no $p_T > 500$ MeV threshold applied to charged particles. The impact of applying such a threshold is negligible for all $p_T$ bins except the first two where effects of up to 1% are observed in the mean and standard deviation of the jet charge.}:

$$Q_J = \frac{1}{(p_{TJ})^\kappa} \sum_{i \in \text{Tracks}} q_i \times (p_{TJ})^\kappa,$$  

where Tracks is the set of tracks associated with jet $J$, $q_i$ is the charge (in units of the positron charge) of track $i$ with associated transverse momentum $p_{TJ}$, $\kappa$ is a free regularization parameter, and $p_{TJ}$ is the transverse momentum of the calorimeter jet. The distributions of $Q_J$ for various jet flavors are shown in Fig. 3 for $\kappa = 0.3$.

B. Object definitions at particle level

The measurement is carried out within a fiducial volume matching the experimental selection to avoid extrapolation into unmeasured kinematic regions that have additional model dependence and related uncertainties. Particle-level absolute pseudorapidity $|\eta|$ is classified as the more central (more forward) jet. A measurement of the more forward and more central jet charge distributions can exploit the rapidity dependence of the jet flavor to extract information about the jet charge for a particular flavor. This is discussed in more detail in Sec. IV B.
there is a clear relationship between the jet charge and the initiating parton’s charge, as up-quark jets tend to have a larger jet charge than gluon jets. Furthermore, gluon jets tend to have a larger jet charge than down-quark jets. However, the jet charge distribution is already broad at particle level and the jet charge response \((Q_{\text{particle-level}} - Q_{\text{detector-level}})\) resolution is comparable to the differences in the means of the distributions for different flavors, so one can expect only small changes in the inclusive jet charge distribution for changes in the jet flavor composition. The three narrow distributions on top of the bulk response distribution in Fig. 3(b) are due to cases in which only one or two charged particles dominate the jet charge calculation at particle level. The two off-center peaks are due to cases in which one of the two high-\(p_T\)-fraction tracks is not reconstructed and the widths of the two off-center and central peaks are due to the (single) track and jet \(p_T\) resolutions. The bulk response is fit to a Gaussian function with standard deviation \(\sigma \approx 0.5\ e\), where \(e\) is the positron charge.

The parameter \(\kappa\) in Eq. (1) regulates the sensitivity of the jet charge to soft radiation. For \(\kappa > 0\), the jet charge is infrared safe.\(^4\) Low values of \(\kappa\) enhance the contribution to the jet charge from low-\(p_T\) particles while in the \(\kappa \to \infty\) limit the jet charge is never collinear safe for \(\kappa > 0\) and not even the sign of the jet charge is Lorentz invariant, although it is clearly invariant under longitudinal boosts.

\(^4\)The jet charge is never collinear safe for \(\kappa > 0\) and not even the sign of the jet charge is Lorentz invariant, although it is clearly invariant under longitudinal boosts.
limit, only the highest-$p_T$ track contributes to the sum in Eq. (1). The dependence on the highest-$p_T$ tracks is demonstrated with the plots in Fig. 4 with the variable $Q_{J,n}$, which is the jet charge in Eq. (1), but built from the leading $n$ tracks. The variable $Q_{J,1}$ is simply the weighted fragmentation function of the leading-track $p_T$ to the jet $p_T$ with weight $\kappa$. The usual $Q_j$ is recovered in the limit $n \to \infty$. Figure 4 shows the sequence $Q_{J,n}$ for $\kappa = 0.3$ and $\kappa = 0.7$. For lower values of $\kappa$, many tracks are required for the sequence of distributions to converge to the full jet charge. However, for $\kappa = 0.7$, the distribution converges quickly, indicating that only the highest-$p_T$ tracks are contributing. All reconstructed tracks are henceforth used when computing the jet charge, but the plots in Fig. 4 give an indication of the contribution of (relatively) high- and low-$p_T$ tracks. Dedicated studies [24] agree with theoretical predictions [34] that suggest that $\kappa \sim 0.5$ is the most sensitive to the charge of the parton initiating a jet. Therefore, the measurement presented in this paper uses $\kappa = 0.5$ in addition to $\kappa = 0.3$ and $\kappa = 0.7$ in order to maintain a broad sensitivity to both hard and soft radiation inside jets. The reconstructed jet charge distributions for $\kappa = 0.3$ and $\kappa = 0.7$ are shown in Fig. 5 for events passing the selection described in Sec. IV. Section VI describes how the jet charge moments are corrected for detector resolution and acceptance effects through an unfolding procedure.

VI. UNFOLDING

The particle-level jet charge distribution’s mean and standard deviation are measured as a function of jet $p_T$. This is accomplished by unfolding a discretized two-dimensional distribution of the jet charge and the jet $p_T$ and then computing the first two moments of the jet charge distribution in each bin of $p_T$. The jet charge distribution is discretized into 15 bins in each of ten bins of the jet $p_T$. The jet charge mean is robust against the bias introduced from the discretization procedure, where fewer than five charge bins per $p_T$ bin is required to ensure negligible bias after recovering the mean from the discretized distribution. However, the standard deviation of the jet charge distribution is sensitive to the discretization and requires about 15 charge bins$^5$ in order for the inherent bias due to discretization to be negligible. The jet charge spans the range $|Q_j| < 1.8$ for $\kappa = 0.3$, $|Q_j| < 1.2$ for $\kappa = 0.5$ and $|Q_j| < 0.9$ for $\kappa = 0.7$. Events in the overflow of the jet charge distribution are placed in the first or last bins. The $p_T$ binning is given by $[50,100)$, $[100,200)$, $[200,300)$, $[300,400)$, $[400,500)$, $[500,600)$, $[600,800)$, $[800,1000)$, $[1000,1200)$, and $[1200,1500)$ GeV. Figure 6 displays the $p_T$ dependence of the jet charge distribution’s mean and standard deviation for detector-level data and simulation and for particle-level simulation. The differences between the simulated detector- and particle-level distributions give an indication of the corrections required to account for detector acceptance and resolution effects in the unfolding procedure. The growing difference between the particle- and detector-level average jet charge is due to the loss of charged-particle momentum inside jets as a result of track merging. At particle level, the standard deviation of the jet charge distribution decreases with increasing $p_T$, but at detector level it increases with $p_T$ due to resolution effects.

There is no unique way to extract the particle-level distribution of the jet charge and jet $p_T$ from the reconstructed distributions. An iterative Bayesian (IB) technique [75], implemented in the RooUnfold framework [76], is used

$^5$Fifteen is the number used in the unfolding and differs from the number shown for illustration in e.g. Fig. 5.
to unfold the two-dimensional jet charge and jet $p_T$ distribution. In the IB unfolding technique, the number of iterations and the prior distribution are the input parameters. In the first step, the raw data are corrected using the simulation to account for events that pass the fiducial selection at detector level, but not the corresponding selection at particle level; this correction is the fake factor. Then, the IB method iteratively applies Bayes’ theorem using the response matrix to connect the prior to posterior at each step, with the nominal PYTHIA sample used for the initializing prior. The response matrix describes the bin migrations between the particle-level and detector-level two-dimensional jet charge and jet $p_T$ distributions. While the response matrix is nearly diagonal, the resolution degrades at high $p_T$ where more bin-to-bin migrations from particle level to detector level occur. The number of iterations in the IB method trades off unfolding bias with statistical fluctuations. An optimal value of four iterations is obtained by minimizing the bias when unfolding pseudodata HERWIG++ with PYTHIA as a test of the methodology. The last step of the unfolding applies another correction from simulation to the unfolded data to account for the differential rate of events passing the particle-level selection but not the detector-level selection; this correction is the inefficiency factor.

VII. SYSTEMATIC UNCERTAINITIES

All stages of the jet charge measurement are sensitive to sources of potential bias. The three stages of the measurement are listed below, with an overview of the systematic uncertainties that impact the results at each stage:

Correction factors: Fake and inefficiency factors are derived from simulation to account for the fraction of events that pass either the detector-level or particle-level fiducial selection, but not both. These factors are generally...
between 0.9 and 1.0 except in the first \( p_T \) bin, where threshold effects introduce corrections that can be as large as 20%. Experimental uncertainties correlated with the detector-level selection acceptance, such as the jet energy scale uncertainty, result in uncertainties in these correction factors. An additional source of uncertainty on the correction factors is due to the explicit dependence on the particle-level jet charge and jet \( p_T \) spectra. A comparison of particle-level models (PYTHIA and HERWIG++) is used to estimate the impact on the correction factors.

Response matrix: For events that pass both the detector-level and particle-level fiducial selections, the response matrix describes migrations between bins when moving between the detector level and the particle level. The response matrix is taken from simulation and various experimental uncertainties on the jet charge and jet \( p_T \) spectra result in uncertainties in the matrix. Uncertainties can be divided into two classes: those impacting the calorimeter jet \( p_T \) and those impacting track reconstruction inside jets.

Unfolding procedure: A data-driven technique is used to estimate the potential bias from a given choice of prior and number of iterations in the IB method [77]. The particle-level spectrum is reweighted using the response matrix so that the simulated detector-level spectrum has significantly improved agreement with 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).

The following two subsections describe the impact of the detector-related sources of systematic uncertainty in more detail. Uncertainties on the calorimeter jet \( p_T \) are described in Sec. VII A and the uncertainties related to tracking are described in Sec. VII B. Summaries of the systematic uncertainties for the more forward jet and \( \kappa = 0.5 \) are found in Tables II and III for the average jet charge and the jet charge distribution’s standard deviation, respectively. The uncertainties for the more central jet are similar.

### A. Calorimeter jet uncertainties

Jets are calibrated so that the detector-level \( p_T \) is an unbiased measurement of the particle-level jet \( p_T \) and various data-driven techniques are used to derive in situ estimates of the difference in this calibration between the data and the simulation. Uncertainties in the energy scale and resolution of calibrated jets impact the jet charge in the normalization of Eq. (1) (but preserve the jet charge sign) as well as the binning for the two-dimensional distribution. Complete details of this source of uncertainty can be found in Ref. [78]. There are many components of the jet energy scale uncertainty. The in situ correction is derived from data using the momentum balance in events with Z bosons (low \( p_T \)) or photons (moderate \( p_T \)) produced in association with jets as well as the balance of multijet (high \( p_T \)) and dijet (high |\( \eta | \)) systems. Uncertainties on this method stem from

---

\( \kappa = 0.5 \).  

-----

6The uncertainties on the first \( p_T \) bin of the average jet charge are much larger than on the other bins because the mean is small compared to the resolution.
TABLE II. A summary of all the systematic uncertainties and their impact on the mean jet charge for $\kappa = 0.5$ and the more forward jet. The correction factors are the fake and inefficiency corrections applied before/after the response matrix. The Other Tracking category includes uncertainty on the track reconstruction efficiency, track momentum resolution, charge misidentification, and fake track rate. All numbers are given in percent.

<table>
<thead>
<tr>
<th>Systematic Uncertainty [%]</th>
<th>[0.5,1]</th>
<th>[1.2]</th>
<th>[2.3]</th>
<th>[3.4]</th>
<th>[4.5]</th>
<th>[5.6]</th>
<th>[6.8]</th>
<th>[8.10]</th>
<th>[10.12]</th>
<th>[12.15]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Jet Energy Scale</td>
<td>+8.4</td>
<td>+3.8</td>
<td>+0.9</td>
<td>+0.8</td>
<td>+1.1</td>
<td>+1.1</td>
<td>+1.1</td>
<td>+0.7</td>
<td>+0.7</td>
<td>+0.4</td>
</tr>
<tr>
<td>Jet Energy Resolution</td>
<td>-13.6</td>
<td>-3.5</td>
<td>-5.0</td>
<td>-0.3</td>
<td>-1.6</td>
<td>-1.1</td>
<td>-1.0</td>
<td>-0.9</td>
<td>-0.7</td>
<td>-0.3</td>
</tr>
<tr>
<td>Charged Energy Loss</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+1.7</td>
<td>+1.5</td>
<td>+1.5</td>
<td>+1.5</td>
<td>+1.6</td>
<td>+1.6</td>
</tr>
<tr>
<td>Other Tracking</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.0</td>
</tr>
<tr>
<td>Track Multiplicity</td>
<td>+3.3</td>
<td>+0.0</td>
<td>+0.9</td>
<td>+0.7</td>
<td>+0.7</td>
<td>+0.5</td>
<td>+1.4</td>
<td>+0.7</td>
<td>+1.2</td>
<td>+1.1</td>
</tr>
<tr>
<td>Correction Factors</td>
<td>-1.6</td>
<td>-0.4</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.4</td>
<td>-0.6</td>
<td>-0.6</td>
<td>-1.2</td>
<td>-1.3</td>
<td>-1.7</td>
</tr>
<tr>
<td>Unfolding Procedure</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+0.0</td>
<td>+0.0</td>
</tr>
<tr>
<td>Total Systematic</td>
<td>+39.0</td>
<td>-38.4</td>
<td>+5.1</td>
<td>+1.7</td>
<td>+1.6</td>
<td>+2.1</td>
<td>+2.4</td>
<td>+2.1</td>
<td>+2.3</td>
<td>+2.6</td>
</tr>
<tr>
<td>Data Statistics</td>
<td>28.0</td>
<td>7.4</td>
<td>1.4</td>
<td>0.7</td>
<td>0.3</td>
<td>0.6</td>
<td>0.9</td>
<td>2.0</td>
<td>4.2</td>
<td>7.0</td>
</tr>
<tr>
<td>Total Uncertainty</td>
<td>-47.0</td>
<td>-8.9</td>
<td>-5.4</td>
<td>-1.9</td>
<td>-1.9</td>
<td>-1.7</td>
<td>-2.3</td>
<td>-2.3</td>
<td>-3.0</td>
<td>-7.8</td>
</tr>
<tr>
<td>Measured Value [e]</td>
<td>0.014</td>
<td>0.024</td>
<td>0.049</td>
<td>0.065</td>
<td>0.076</td>
<td>0.082</td>
<td>0.092</td>
<td>0.100</td>
<td>0.108</td>
<td>0.115</td>
</tr>
</tbody>
</table>

TABLE III. A summary of all the systematic uncertainties and their impact on the jet charge distribution’s standard deviation for $\kappa = 0.5$ and the more forward jet. The correction factors are the fake and inefficiency corrections applied before/after the response matrix. The Other Tracking category includes uncertainty on the track reconstruction efficiency, track momentum resolution, charge misidentification, and fake track rate. All numbers are given in percent.

<table>
<thead>
<tr>
<th>Standard Deviation</th>
<th>Jet $p_T$ Range [100 GeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic Uncertainty [%]</td>
<td>[0.5,1]</td>
</tr>
<tr>
<td>Total Jet Energy Scale</td>
<td>+1.9</td>
</tr>
<tr>
<td>Jet Energy Resolution</td>
<td>-1.3</td>
</tr>
<tr>
<td>Charged Energy Loss</td>
<td>-0.0</td>
</tr>
<tr>
<td>Other Tracking</td>
<td>+0.0</td>
</tr>
<tr>
<td>Track Multiplicity</td>
<td>+0.0</td>
</tr>
<tr>
<td>Correction Factors</td>
<td>-0.2</td>
</tr>
<tr>
<td>Unfolding Procedure</td>
<td>+0.0</td>
</tr>
<tr>
<td>Total Systematic</td>
<td>+3.1</td>
</tr>
<tr>
<td>Data Statistics</td>
<td>-3.0</td>
</tr>
<tr>
<td>Total Uncertainty</td>
<td>-3.2</td>
</tr>
<tr>
<td>Measured Value [e]</td>
<td>0.410</td>
</tr>
</tbody>
</table>

the modeling of these processes in simulation. There is also a contribution from the response to single hadrons [79]. Additional sources of uncertainty are due to the modeling of the in-time and out-of-time pileup corrections to the jet energy scale as well as differences in the response due to the flavor of the jet. To assess the impact of each component of the jet energy scale uncertainty, the jet energies in simulation are shifted according to the $p_T$- and $\eta$-dependent ±1σ variations. For a fixed variation, the response matrix, and fake and inefficiency factors are recomputed and the unfolding procedure is repeated. The resulting uncertainty on the jet charge distribution’s mean and standard deviation is about 1% or less for jet $p_T$ above 200 GeV. The jet energy resolution uncertainty is derived using data-driven techniques in dijet events [80]. To assess the impact of a slightly larger jet energy resolution, jet energies are smeared according to $p_T$- and $\eta$-dependent factors and propagated through the entire unfolding procedure, as for the jet energy scale uncertainty. The jet energy resolution uncertainty is dominant to the jet energy scale uncertainty.

B. Tracking uncertainties

Uncertainties on tracking are broken down into contributions related to (i) the efficiency of reconstructing tracks and (ii) measurements of those tracks that are successfully reconstructed. The uncertainty on the inclusive track
reconstruction efficiency is dominated by the uncertainty in the material in the ID. The amount of material is known to within $\sim 5\%$ [81]. Simulated detector geometries with various levels of material in the ID within the measured uncertainties are used to estimate the track reconstruction efficiency uncertainty. These uncertainties are $\eta$- and $p_T$- dependent, ranging from $\lesssim 1\%$ for $|\eta| < 2.1$ to $\lesssim 4\%$ for $2.1 \leq |\eta| < 2.3$ and $\lesssim 7\%$ for $2.3 \leq |\eta| < 2.5$. The impact of the uncertainty is estimated by randomly removing tracks within the $p_T$- and $\eta$-dependent probabilities, leading to a $\lesssim 0.5\%$ uncertainty on the jet charge distribution’s mean and standard deviation. An additional uncertainty accounts for the difference in efficiency between data and simulation due to the modeling of the track $\chi^2$ per number of degrees of freedom (NDF) requirement. A requirement of $\chi^2$/NDF $< 3$ is more than 99% efficient across jet and track $p_T$, but the efficiency is generally higher in simulation than in data. The difference in the efficiency between data and simulation is $\lesssim 10\%$ of the inefficiency. The impact of this mismodeling is evaluated by independently removing tracks with a probability that is 10% of the $\chi^2$/NDF $< 3$ requirement inefficiency. As a result of this procedure, the jet charge distribution’s mean and standard deviation change by $\lesssim 0.1\%$ in most $p_T$ bins.

In addition to the loss of tracks due to the material in the ID, tracks can be lost due to the busy environment inside the cores of jets. This loss can be studied in simulation by comparing the reconstructed charged-particle momentum with the charged-particle momentum inside the corresponding particle-level jet. In order to remove the impact of the tracking resolution and the contribution from fake tracks already accounted for separately, reconstructed tracks in the simulation are matched with charged particles. The matching is performed by considering the energy deposited in the various layers of the ID by charged particles due to material interactions modeled with GEANT4. Weights are assigned to charged particles based on the energy deposited in detector elements that were used to reconstruct a given track. A match is declared if the weight for one charged particle is sufficiently high. Figure 7(a) shows the ratio in simulation of the sum of the $p_T$ of charged particles that were matched to reconstructed tracks to the sum of the $p_T$ of all the charged particles as a function of the jet $p_T$. At low $p_T$, the ratio increases with $p_T$ due to losses as a result of hadronic interactions with the material in the ID. Beyond about 200 GeV, the fraction monotonically decreases due to the loss of tracks in the core of the jet. A related quantity is $\langle \Sigma\text{track} p_T / \text{Jet} p_T \rangle$, where the denominator is the reconstructed calorimeter jet $p_T$ and the numerator is a sum over tracks associated with the jet. Since the particle-level charged-to-neutral fraction of the energy is independent of $p_T$, a degradation in this ratio can provide information about the loss of tracks inside the core of a jet in data. Figure 7(b) shows the distribution of $\langle \Sigma\text{track} p_T / \text{Jet} p_T \rangle$ as a function of jet $p_T$. It exhibits trends very similar to those in Fig. 7(a), and in fact the relative loss (fraction with respect to the peak) is similar. The MC underestimates the loss by $\lesssim 1\%$. The impact of the charged-particle momentum loss inside the cores of jets is estimated by randomly removing tracks with a $p_T$-dependent probability such that the relative
loss in the simulation matches that in the data. This uncertainty is negligible for jets with $p_T < 400 \text{ GeV}$, but is non-negligible for higher-$p_T$ jets, resulting in a $\lesssim 4\%$ uncertainty on the average jet charge in the highest $p_T$ bin.

The momentum resolution of isolated tracks has been well measured in $J/\psi \to \mu \mu$ and $Z \to \mu \mu$ events [82]. The scale and resolution of reconstructed muon candidates are shifted and smeared in the MC simulation to account for differences between the data and the simulation for $m_{\mu \mu}$. Generic tracks are not corrected in the same way as muon candidates reconstructed from the ID (and ignoring the muon spectrometer). Thus, the correction factors are taken here as the systematic uncertainty on the momentum resolution. The momentum resolution is parametrized as a sum in quadrature of a $p_T^{-1}$-dependent term, a constant term, and a term linear in the track $p_T$, with coefficients $r_0$, $r_1$, and $r_2$, respectively. The first term accounts for fluctuations in the energy loss in the detector material, the second term captures effects due to multiple scattering, and the third term accounts for the intrinsic resolution caused by misalignment and the finite spatial resolution of ID hits. Unlike muon spectrometer tracks, ID tracks do not traverse a significant amount of material and so the energy-loss coefficient $r_0$ and its uncertainty are neglected. The uncertainties on $r_1$, $r_2$ and the momentum scale $s$ are estimated by randomly smearing the $p_T$ of every track with $|\eta|$-dependent factors that are $\lesssim 2\%$ for track $p_T < 100 \text{ GeV}$ and increase to 10–20% at 1 TeV depending on $|\eta|$. Propagating these variations through the unfolding procedure results in uncertainties that are subdominant to other uncertainties, but non-negligible ($\sim 2\%$) in the highest $p_T$ bins for the average jet charge.

Aside from the track $p_T$, the other track parameter that is relevant for the jet charge is the track charge. Especially at high $p_T$ when the tracks are very straight, the probability for misidentifying the track charge can become non-negligible. Simulated particles matched to generator-level information, as described above, are used to study the charge-flipping probability (charge misidentification rate) for nonfake tracks originating from charged particles inside jets (mostly pions). The rate predicted from the simulation is $< 0.1\%$ for track $p_T < 100 \text{ GeV}$, $0.5\%$ for $100 \text{ GeV} \leq p_T < 200 \text{ GeV}$, $1\%$ for $200 \text{ GeV} \leq p_T < 300 \text{ GeV}$, $2\%$ for $300 \text{ GeV} \leq p_T < 400 \text{ GeV}$ and $4\%$ for $p_T \geq 400 \text{ GeV}$. Dedicated studies of charge flipping in searches for same-sign leptons [83] suggest that the mismodeling of the charge-flipping rate is (much) less than 50%. Therefore, the impact of charge flipping on the jet charge measurement is conservatively estimated by randomly flipping the charge of tracks at 50% of the charge misidentification rate. The impact on the measured jet charge mean and standard deviation is negligible.

Random combinations of hits in the detector can be combined to form a reconstructed track. Tracks resulting in particular from multiparticle trajectories that have kinks can result in very large reconstructed track $p_T$. The quality criteria are effective at mitigating the presence of fake tracks, which constitute less than $\lesssim 0.1\%$ of all reconstructed tracks. To determine the impact of fake tracks on the jet charge, fake tracks are randomly removed with a probability that is 50% of the rate in simulation. This results in a negligibly small uncertainty on the mean jet charge and a $\lesssim 0.5\%$ uncertainty on the standard deviation of the jet charge distribution.

The tracking uncertainties described so far take into account the resolution and efficiency of reconstruction of charged-particle momenta. One last source of systematic uncertainty is the number of charged particles. The unfolding procedure uncertainty takes into account the uncertainty on the prior due to the charged-particle multiplicity, but the jet charge resolution also changes with the charged-particle multiplicity. To assess the impact on the response matrix of the mismodeled charged-particle multiplicity, the distribution of $n_{\text{track}}$ is reweighted in the simulation per $p_T$ bin and the relative difference when unfolding the nominal PYTHIA distribution with the reweighted PYTHIA distribution is taken as a systematic uncertainty. This uncertainty is subdominant for the standard deviation across $p_T$ and for the mean at low to moderate jet $p_T$. For the mean jet charge, the largest uncertainty is with the smallest $\kappa$ and for large $p_T$, where it is 3–4% percent in the highest $p_T$ bin for $\kappa = 0.3$ and $\kappa = 0.5$.

VIII. RESULTS

The data satisfying the event selection criteria described in Sec. IV are unfolded according to the procedure in Sec. VI and the average and standard deviation of the jet charge distribution are computed as a function of the jet $p_T$. These results, along with the systematic uncertainties detailed in Sec. VII, are discussed in Sec. VIII A. The PDF uncertainty and jet formation uncertainties in the theory predictions are compared to the unfolded data in Secs. VIII B and VIII C, respectively. Using PDF information as input, the average charge per jet flavor is extracted in Sec. VIII E and its $p_T$ dependence is studied in Sec. VIII F.

A. Unfolded jet charge spectrum

The unfolded jet charge mean is shown as a function of the jet $p_T$ in the top plots of Fig. 8 for $\kappa = 0.3$, 0.5 and 0.7. The average charge increases with jet $p_T$ due to the increase in up-flavor jets from PDF effects. The average charge increases from 0.01e at $p_T \sim 100 \text{ GeV}$ to 0.15e at $p_T \sim 1.5 \text{ TeV}$. Systematic uncertainties are generally a few percent, except at low jet $p_T$ where the fractional uncertainty is large because the average jet charge in the denominator is small, and at high $p_T$ where the tracking

\footnote{Since the prior is also changed, this uncertainty at least partially includes the unfolding procedure uncertainty.}
uncertainties are not negligible. The statistical uncertainty is estimated by repeating the measurement on an ensemble of bootstrapped data sets: each event is used in each pseudo-data set \( n \) times, where \( n \) is a random number distributed according to a Poisson distribution with mean one. The first bin suffers from large statistical uncertainties (up to 170%), but for the higher \( p_T \) bins the systematic uncertainty is dominant, except at the highest \( p_T \) bin where statistical and systematic uncertainties are of similar size (about 7%). The jet charge distributions of the more forward and more central jet differ in shape, in particular at low \( p_T \), due to the different shape of the up-/down-flavor fractions in those bins as shown in Fig. 2(b).

Analogous results for the standard deviation of the jet charge distribution are shown in the bottom plots of Fig. 8. Even though the standard deviation of the reconstructed jet charge distribution increases with jet \( p_T \) (Fig. 6), the particle-level value decreases and approaches an asymptote for \( p_T \gtrsim 300 \) GeV.

**B. Sensitivity of PDF modeling**

Variations in the PDF set impact the relative flavor fractions and thus in turn change the jet charge distribution. Such changes do not vary much with \( \kappa \), since the PDF impacts the jet charge distribution mostly through the flavor fractions. Figures 9 and 10 compare the unfolded distributions of the jet charge distribution’s average and standard deviation with several PDF sets, with tuned predictions for PYTHIA for each PDF, and with the same AU2 family of tunes. The sampling of PDF sets results in a significant spread for the average jet charge, but has almost no effect on the standard deviation. CTEQ6L1 describes the data best, although the data/MC ratio has a stronger \( p_T \) dependence. In particular, the data/MC differences with CTEQ6L1 are up to 10% (15%) at moderate \( p_T \) for the more forward (central) jet. For high \( p_T \), differences between data and simulation are less significant. NLO PDFs such as CT10 are consistently below the data by about 10–15%.
C. Sensitivity of QCD models and tunes

The measurements presented in Sec. VIII A show that there are qualitative differences between the data and the MC simulations, and comparisons in Sec. VIII B suggest that variations in the PDF set cannot fully explain the differences. Differences in Sec. VIII A between PYTHIA and HERWIG++ suggest that some aspect of the modeling of fragmentation could lead to the observed differences between the simulation and the data. One possible source is the hadronization modeling, which differs between PYTHIA (Lund-string fragmentation) and HERWIG++ (cluster fragmentation). The modeling of final-state radiation (FSR) is expected to have an impact on the jet charge distribution because variations in the radiation lead to different energy flow around the initial parton and hence different fragmentation of the jet. The plots in

![Diagram A](image1)

![Diagram B](image2)

**FIG. 9.** The average jet charge ($\kappa = 0.5$) in units of the positron charge for (a) the more forward jet and (b) the more central jet compared with theory predictions due to various PDF sets. The crossed lines in the bars on the data indicate the statistical uncertainty and the full extent of the bars is the sum in quadrature of the statistical and systematic uncertainties.

**FIG. 10.** The standard deviation of the jet charge ($\kappa = 0.5$) distribution in units of the positron charge for (a) the more forward jet and (b) the more central jet compared with theory predictions due to various PDF sets. The crossed lines in the bars on the data indicate the statistical uncertainty and the full extent of the bars is the sum in quadrature of the statistical and systematic uncertainties.
Figs. 11 and 12 show the measured average jet charge and the jet charge distribution’s standard deviation, respectively, for $\kappa = 0.3$, 0.5, and 0.7, compared to various QCD MC models and tunes for the more forward jet. The crossed lines in the bars on the data indicate the statistical uncertainty and the full extent of the bars is the sum in quadrature of the statistical and systematic uncertainties.

Figs. 11 and 12 contain the predictions from PYTHIA 6 using the Perugia 2012 tune [84] and the radHi and radLo Perugia 2012 tune variations. These Perugia tune variations test the sensitivity to higher/lower amounts of initial- and final-state radiation (via the scaling of $\alpha_s$), although only variations of the FSR are important for the jet charge distribution. For the mean jet charge, PYTHIA 6 with the P2012 radLo tune is very similar to PYTHIA 8 with the AU2 tune. The spread in the average jet charge due to the difference between the radHi and radLo tunes increases with $\kappa$, since suppression of soft radiation makes the jet charge distribution more sensitive to the modeling of the energy fraction of the leading emissions. For the jet charge distribution’s standard deviation, the sensitivity to the $\alpha_s$ scaling is large at both high and low $\kappa$. However, the sensitivity is inverted: radHi gives a larger standard deviation for $\kappa = 0.3$, but a lower standard deviation for $\kappa = 0.7$. Other Perugia 2012 tunes have been...
studied, testing the sensitivity to color reconnection and multiple parton interactions, but the differences in the jet charge distribution’s mean and standard deviation are small. The Perugia 2012 tunes may not fully capture the spread in nonperturbative effects, which is also suggested by the increasing difference between PYTHIA 8 and HERWIG++ for decreasing $\kappa$.

**D. Model comparison overview**

Figures 13 and 14 show comparisons of the unfolded jet charge distribution’s mean and standard deviation for different QCD simulations using LO and NLO PDF sets. The predictions using the CT10 NLO PDF set as shown in Fig. 13 are generally about 10% below the data. Consistent with the expectation that the PDF and (nearly collinear) fragmentation are responsible for the jet charge distribution’s mean and standard deviation, there does not seem to be an effect from the POWHEG NLO matrix element. For the jet charge distribution’s standard deviation and $\kappa = 0.3$, the data falls between PYTHIA (larger standard deviation) and HERWIG++ (smaller standard deviation), but this trend is less evident for larger $\kappa$ values, suggesting a difference due...
to soft tracks. As seen in Sec. VIII B, comparisons with CTEQ6L1 show it to be a better model for the $p_T$ dependence of the mean jet charge than CT10. The analogous plots to Fig. 8 but using CTEQ6L1 instead of CT10 are shown in Fig. 14. Generally, there is agreement between the simulation and the data with only a $\lesssim 5\%$ difference in the lower $p_T$ bins.

E. Extraction of the average up-quark and down-quark jet charges

In addition to understanding the trends in the jet charge distribution from PDFs, one can use PDFs to extract information about jets of a particular flavor. These exclusive interpretations rely on flavor-fraction information in PDFs and matrix element calculations to extract the jet charge distribution for particular jet (anti)flavors in each $p_T$ bin. The required nonperturbative information is summarized in Fig. 2(a). Jets with flavors other than up/down/anti-up/anti-down/gluon are not included in Fig. 2(a) and give a negligible contribution ($\lesssim 2\%$) in the highest $p_T$ bins.

One way of extracting the up- and down-flavor average jet charges is to exploit the difference in flavor fractions shown in Fig. 2(a) between the more forward and the more central jets. Due to the $p_T$-balance requirement between the leading and subleading jet in the event selection, to a good approximation, the $p_T$ spectrum is the same for the more forward and the more central jet. Assuming that the average...
jet charge of the sum of flavors that are not up/down/anti-up/anti-down is zero, in each bin \(i\) of \(p_T\):

\[
\langle Q_{J_{\text{forward}}} \rangle_{i} = \left( f_{\text{forward}, \text{up}, i} - f_{\text{forward}, \text{anti-up}, i} \right) Q_{i}^{\text{up}} \\
+ \left( f_{\text{forward}, \text{down}, i} - f_{\text{forward}, \text{anti-down}, i} \right) Q_{i}^{\text{down}},
\]

\[
\langle Q_{J_{\text{central}}} \rangle_{i} = \left( f_{\text{central}, \text{up}, i} - f_{\text{central}, \text{anti-up}, i} \right) Q_{i}^{\text{up}} \\
+ \left( f_{\text{central}, \text{down}, i} - f_{\text{central}, \text{anti-down}, i} \right) Q_{i}^{\text{down}},
\]  

(2)

where \(Q_J\) is the jet charge from Eq. (1), \(f_{y,i}^x\) is the fraction of flavor \(y\) in \(p_T\) bin \(i\) for the jet \(x \in \{\text{more forward, more central}\}\) and \(Q_i^x\) is the average jet charge for such jets. The values \(f_{y,i}^x\) are taken from simulation (PYTHIA with CT10 PDF and AU2 tune), which then allows an extraction of \(Q_i^x\) by solving the system of equations in Eq. (2). This extraction is performed separately in each \(p_T\) bin. Figure 15 shows the extracted up- and down-flavor jet charges in bins of jet \(p_T\). At very high jet \(p_T\), the absolute quark-flavor fractions are large (Fig. 2), but the difference between the more forward and more central jets is small and the statistical uncertainty is large. At low jet \(p_T\), the difference between the more forward and more central jets is large (Fig. 2), but the absolute quark-flavor fraction is small and the statistical uncertainty is once again large because the mean jet charge is close to zero. In the limit that
the flavor fractions are identical for the more forward and more central jet, the equations become degenerate and it is not possible to simultaneously extract the average up- and down-flavor jet charges. The uncertainties on the flavor fractions and on the measured average jet charges are propagated through the solutions of Eq. (2). Generally, the uncertainty is larger for the down-flavor jets because the fraction of these jets is smaller than the fraction of up-flavor jets.

F. Dependence of the up-quark and down-quark jet charge on $p_T$

Using the methods of Sec. VIII E, one can examine the residual $p_T$ dependence of the average jet charge after accounting for PDF effects. The inclusive jet charge has been shown to increase with $p_T$ due to a mixing of jet flavors and the following subsection investigates the $p_T$ dependence of a fixed jet flavor. Section VIII F 1 describes the theory predictions and the extracted $p_T$ dependence from the data is discussed in Sec. VIII F 2.

I. Theory prediction

Recent theoretical studies have shown that the energy dependence of jet charge moments is calculable perturbatively [34,35]. At leading power (treating $m_{\text{jet}}/E_{\text{jet}}$ as the expansion parameter), the charge defined with jet and track energies is equivalent to the one given in Eq. (1), since the jet opening angle is small. For jets defined by a radius $R$ at an energy $E$, the average jet charge $\langle Q_j \rangle$ is given by

$$\langle Q_j \rangle = \left[1 + \mathcal{O}(\alpha_s) \right] \sum_h Q_h \tilde{D}_h^q(\kappa, E \times R). \quad (3)$$

where $Q_h$ is the charge of hadron $h$ and the functions $\tilde{D}_h^q(\kappa, \mu)$ are the Mellin moments of the fragmentation functions

$$\tilde{D}_h^q(\kappa, \mu) = \int_0^1 dx x^\kappa D_h^q(x, \mu). \quad (4)$$

The fragmentation functions $D_h^q(x, \mu)$ describe the probability for a hadron $h$ to carry a momentum fraction $x$ of a quark $q$ at the energy scale $\mu$ [85,86]. The $\mathcal{O}(\alpha_s)$ correction is small and is dominated by the uncertainty on the fragmentation functions [34]. Ratios of $\langle Q_j \rangle$ at different energies result in a cancellation of the leading corrections and associated uncertainties. Soft corrections to the average jet charge are small not only because of the additional suppression at $\kappa > 0$ compared to collinear radiation, but also because the leading soft emissions are made of gluons, which carry no electric charge. Although gluons can split into quark and antiquark pairs, the same number of quarks and antiquarks go into the jet so that on average the jet charge is unchanged.\textsuperscript{8}

The leading energy dependence of the average jet charge is due to the derivative of the fragmentation functions with respect to $E$, which is determined by the renormalization group equations for $\tilde{D}_h^q(\kappa, \mu)$:

$$\frac{d}{d\ln \mu} \tilde{D}_h^q(\kappa, \mu) = \frac{\alpha_s}{\pi} \tilde{P}_{qq}(\kappa) \tilde{D}_h^q(\kappa, \mu) + \mathcal{O}(\alpha_s^2), \quad (5)$$

where $\tilde{P}_{qq}(\kappa)$ is the moment of the leading-order splitting function

$$\tilde{P}_{qq}(\kappa) = C_F \int_0^1 dz (z^\kappa - 1) \frac{1 + z^2}{1 - z}. \quad (6)$$

Recalling that $E/p_T$ is constant at leading power for all particles in the jet, the prediction for the scale violation of a quark jet is given by

$$\left.\frac{p_T}{\langle Q_j \rangle} \frac{d}{dp_T} \langle Q_j \rangle \right|_{p_T} = \frac{\alpha_s}{\pi} \tilde{P}_{qq}(\kappa) \equiv c_\kappa \approx \begin{cases} -0.024 \pm 0.004 & \kappa = 0.3, \\ -0.038 \pm 0.006 & \kappa = 0.5, \\ -0.049 \pm 0.008 & \kappa = 0.7 \end{cases} \quad (7)$$

\textsuperscript{8}This is not strictly true, as the soft radiation pattern depends on the partonic colors, which are correlated with electric charge, but these correlations are negligible compared to the other uncertainties.
MEASUREMENT OF JET CHARGE IN DIJET EVENTS ...

where the last form consists of numerical approximations setting \( \mu = E \times R \) equal to 50 and 500 GeV with their average giving the central value, and using \( \alpha_s(50 \text{ GeV}) = 0.130 \) and \( \alpha_s(500 \text{ GeV}) = 0.094 \). Experimentally, one measures combinations of quark and gluon jets with the fractions of different partonic flavors varying with energy. The next subsection discusses how the scaling violation parameter, defined in Eq. (7), can be extracted from the data.

2. Extraction from the data

Since \( c_x \ll 1 \) from Eq. (7), one can approximate a linear dependence on \( c_x \):

\[
\langle Q_f \rangle(p_T) = \bar{Q}(1 + c_x \ln(p_T/\bar{p}_T)) + \mathcal{O}(c_x^2),
\]

where \( \bar{Q} = \langle Q_f \rangle(\bar{p}_T) \) for some fixed (but arbitrary) transverse momentum, \( \bar{p}_T \). Therefore, for a fixed \( p_T \) bin \( i \), the measured charge is given as a superposition of the average jet charge for various jet flavors:

\[
\langle Q_i \rangle \approx \sum_f \beta_{f,i} \bar{Q}_f (1 + c_x \ln(p_{T,i}/\bar{p}_T)),
\]

where \( \beta_{f,i} \) is the fraction of flavor \( f \) in bin \( i \), \( \bar{Q}_f \) is the average jet charge of flavor \( f \) and \( \bar{p}_T \) is a fixed transverse momentum. Fitting the model in Eq. (9) directly to the data to extract \( \bar{Q}_f \) is not practical because there are three parameters and only ten \( p_T \) bins, some of which have very little sensitivity due to low fractions \( \beta \) or large uncertainties on \( \langle Q_f \rangle \). One way around this is to extract \( \bar{Q}_f \) in one fixed bin of transverse momentum (denoted \( p_T \)) as described in Sec. VIII E. Then Eq. (9) is highly constrained, with only one parameter for which each other bin of \( p_T \) gives an estimate. The systematic uncertainties are propagated through the fit treated as fully correlated between bins and the statistical uncertainty is treated coherently by bootstrapping.\(^9\) A weighted average is performed across all \( p_T \) bins and for both the more forward and the more central jet. The procedure is summarized below:

1. In the bin 600 GeV < \( p_T < 800 \) GeV, extract the values \( \bar{Q}_{\bar{u}}^\text{up} \) and \( \bar{Q}_{\bar{d}}^\text{down} \). These values can be seen in the fifth \( p_T \) bin of Fig. 15.
2. With \( \bar{Q}_{\bar{u}}^\text{up} \) and \( \bar{Q}_{\bar{d}}^\text{down} \) fixed, extract the scale violation parameter estimate \( c_{x,i} \) in each \( p_T \) bin \( i \) by solving

\[
\langle Q_i \rangle_{\text{measured}} = \sum_f \beta_{f,i} \bar{Q}_f (1 + c_{x,i} \ln(p_{T,i}/\bar{p}_T))
\]

where \( \bar{p}_T = 700 \text{ GeV} \) is the bin center from the previous step.
3. Repeat the above procedure for all systematic variations and for all bootstrap pseudo-data sets to arrive at estimates of the uncertainty \( \sigma(c_{x,i}) \) for each \( p_T \) bin \( i \).
4. The central value for the extracted scale violation parameter is \( c_x = (\sum c_{x,i}/\sigma(c_{x,i}))/\sum(1/\sigma(c_{x,i})) \).
5. The uncertainty \( \sigma(c_x) \) is determined by repeating step 3 with the nominal values \( c_x \) replaced by their systematic varied versions or the bootstrap pseudo-data values for the statistical uncertainty estimate.

The results are presented in Fig. 16. The data support the prediction that \( c_x < 0 \) and \( \partial c_x/\partial \kappa < 0 \). Linear correlations between \( \kappa \) values can be determined using the bootstrapped data sets: about 0.9 between \( c_{0.3} \) and \( c_{0.5} \) as well as between \( c_{0.5} \) and \( c_{0.7} \), while the correlation is about 0.7 between \( c_{0.3} \) and \( c_{0.7} \). Thus, the three points are quite correlated, but there is additional information from considering more than one \( \kappa \) value.

IX. SUMMARY

This paper presents a measurement of the particle-level \( p_T \) dependence of the jet charge distribution’s mean and standard deviation in dijet events from 20.3 fb\(^{-1} \) of \( \sqrt{s} = 8 \text{ TeV} \).
8 TeV pp collision data recorded by the ATLAS detector at the LHC. The measured jet charge distribution was unfolded to correct for the detector acceptance and resolution for direct comparison to particle-level models. Comparisons were made at particle level between the measured jet charge distribution and various PDF sets and models of jet formation. Simulations with PYTHIA 8 using the CTEQ6L1 PDF set describe the average jet charge of the more forward jet within about 5% and the more central jet within about 10%. The jet charge distribution’s standard deviation is described within 2%. HERWIG++ shows a similarly good agreement for $p_T > 500$ GeV and $\kappa = 0.7$. However, the HERWIG++ predictions decrease systematically for both the average and the standard deviation for decreasing $\kappa$. Predictions with the CT10 NLO PDF are systematically below the data across jet $p_T$ for the average jet charge and systematically above for the jet charge distribution’s standard deviation. Taking the PDFs as inputs, the average up- and down-flavor jet charges were extracted as a function of $p_T$ and were compared with predictions for scale violation. The data show that the average up- and down-quark jet charges decrease slightly with $p_T$ and this decrease increases with $\kappa$, as predicted. The particle-level spectra are publicly available [87] for further interpretation and can serve as a benchmark for future measurements of the evolution of nonperturbative jet observables to validate QCD MC predictions and tune their free model parameters.

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; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF, DNSRC and Lundbeck Foundation, Denmark; IN2P3-CNRS, CEA-DSM/IRFU, France; GNSF, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; 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 BCKDF, the Canada Council, CANARIE, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, FP7, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Region Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BFS, GIF and Minerva, Israel; BRF, Norway; the Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA) and in the Tier-2 facilities worldwide.

measurement of $\sin^2\theta(W)$ from the charge asymmetry of hadronic events at the $Z^0$ peak, Phys. Lett. B 277, 371 (1992).


[16] F. Abe et al. (CDF Collaboration), Measurement of $B^0 - \bar{B}^0$ flavor oscillations using jet-charge and lepton flavor tagging in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV, Phys. Rev. D 60, 072003 (1999).


[42] ATLAS Collaboration, Data-quality requirements and event cleaning for jets and missing transverse energy reconstruction with the ATLAS detector in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 7$ TeV, https://cds.cern.ch/record/1277678.


MEASUREMENT OF JET CHARGE IN DIJET EVENTS ... PHYSICAL REVIEW D 93, 052003 (2016)
MEASUREMENT OF JET CHARGE IN DIJET EVENTS …

PHYSICAL REVIEW D 93, 052003 (2016)

†Deceased.

Also at School of Physics, Shandong University, Shandong, China.

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

Also at Section de Physique, Université de Genève, Geneva, Switzerland.

Also at International School for Advanced Studies (SISSA), Trieste, Italy.

Also at Department of Physics and Astronomy, University of South Carolina, Columbia, SC, USA.

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

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

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

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

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

Also at University of Malaya, Department of Physics, Kuala Lumpur, Malaysia.