Quark versus Gluon Jet Tagging Using Charged-Particle Constituent Multiplicity with the ATLAS Detector

The ATLAS Collaboration

Distinguishing quark-initiated from gluon-initiated jets is useful for many measurements and searches at the LHC. This note presents a quark-initiated versus gluon-initiated jet tagger using the number of reconstructed charged particles inside the jet. For an efficiency of 60% to select quark-initiated jets, the efficiency to select gluon-initiated jets is between 10 and 20% across a wide range in jet transverse momentum ($p_T$) up to $p_T = 1.5$ TeV, with an absolute systematic uncertainty of about 5%. 

© 2017 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.
1 Introduction

Since the discovery of gluons at PETRA [1–4], jet properties have been used to differentiate quark-initiated (quark) jets from gluon-initiated (gluon) jets. Gluons are in the adjoint representation of SU(3)C while quarks are in the fundamental representation. Therefore, gluons carry both a color and anti-color while quarks only carry a single color charge. More explicitly, the Altarelli-Parisi splitting functions [5] contain a factor of $C_A$ for gluon radiation off a gluon and a factor of $C_F$ for gluon radiation off a quark ($C_A/C_F = 9/4 \sim 2$). As a result, on average gluon jets have more constituents than quark jets and the radiation pattern within gluon jets tends to be broader than quark jets. Both ATLAS [6, 7] and CMS [8, 9] have developed taggers designed to distinguish quark jets from gluon jets based on these properties. This note presents the discrimination power and the derivation of the associated uncertainties for a simplified tagger based solely on the number of tracks ($n_{\text{track}}$), a proxy for the charged constituent multiplicity. The uncertainty derivation is based on a novel technique that exploits the rapidity dependence of the quark/gluon jet fraction in dijet events to calibrate the particle-level charged constituent multiplicity distribution. The tagger is simple and robust. The calibration technique can potentially be extended in the future to more sophisticated observables. Nonetheless, particle multiplicity is one of the most powerful single features for quark/gluon jet tagging, accounting for a significant fraction of the performance of a classifier trained on the full jet radiation pattern [11].

This note is organized as follows. Section 2 briefly introduces the reconstruction of the trajectory of charged particles (tracks) with the ATLAS detector and the performance of the track-based quark/gluon tagger. The procedure for deriving systematic uncertainties is described in Section 3 and the note concludes in Section 4.

2 Performance

The ATLAS experiment [12] features a multi-purpose particle detector with nearly $4\pi$ coverage in solid angle. Tracks are measured in a 2 T axial field generated by a solenoid magnet which surrounds the inner detector (ID) consisting of silicon pixels, silicon micro-strips, and a transition radiation tracking detector. Tracks are reconstructed from all three ID subsystems with a full coverage in azimuth ($\phi$), a pseudorapidity coverage of $|\eta| < 2.5$, and with transverse momentum $p_T > 400$ MeV. The charge $Q$ of a track is determined as part of the reconstruction procedure, which uses a fit with five parameters: the transverse and longitudinal impact parameters, $\phi$, the polar angle $\theta$, and $Q/p$, where $p$ is the track momentum. To suppress the impact of multiple overlapping proton–proton (pp) collisions (pile-up), tracks are required to originate from the hard-scatter collision vertex, which is defined as the vertex with the largest $\sum p_T^2$ computed from associated tracks. A series of additional criteria improves the quality of the selected tracks. These requirements differ between Runs 1 and 2 and will be described when relevant below.

---

1 Previous analyses have calibrated a multiplicity-based tagger for $W$ boson jets using the jet mass resonance peak [10]. Such a procedure does not extend to generic quark and gluon jets, necessitating the technique described in this note.

2 ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the $z$-axis along the beam pipe. The $x$-axis points from the IP to the centre of the LHC ring, and the $y$-axis points upward. Polar coordinates $(r, \phi)$ are used in the transverse plane, $\phi$ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Transverse momentum and energy are defined in the $x$–$y$ plane as $p_T = p \cdot \sin(\theta)$ and $E_T = E \cdot \sin(\theta)$. 

---
Surrounding the ID are electromagnetic and hadronic calorimeters, which are used to build jets. Jets are clustered using the anti-$k_t$ jet algorithm [13] with distance parameter $R = 0.4$, as provided by FastJet [14]. Topological calorimeter-cell clusters [15] are used as inputs. An overall jet calibration corrects the detector-level jet $p_T$ to the particle-level jet $p_T$ on average [16]. Tracks are matched to jets using the ghost-association technique [17]. Particle-level jets in Monte Carlo (MC) simulation are built 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. The charged particle multiplicity of a particle-level jet is simply the number of charged particles clustered within the particle-level jet. As quarks and gluons carry color charge and jets are color neutral, there is some ambiguity in the labeling of jets in simulation as quark or gluon. In this note, jets in simulation are labeled based on the highest energy parton emerging from the hard-scatter collision within the jet catchment area [17]. Uncertainties due to this choice are discussed in Section 3.1.

The distribution of $n_{\text{track}}$ for the highest $p_T$ jet in generic dijet events is shown in Figure 1. The average $n_{\text{track}}$ increases with $p_T$ and, as described in Section 1, gluon jets have more charged constituents than quark jets on average. It is hence possible to construct a tagger to discriminate between quark and gluon jets by labelling jets with $n_{\text{track}} < n_{\text{track}}^{\text{max}}$ as quark-initiated, and those above the threshold to be gluon-initiated. Since $n_{\text{track}}$ is an integer, the possible working points are quantized. The goal is to identify quark jets (signal) and reject gluon jets (background). For example, the requirement $n_{\text{track}} < 10$ is satisfied by 61% (21%) of quark (gluon) jets with $p_T$ between 100 GeV and 200 GeV. It is possible to repurpose the $n_{\text{track}}$-based tagger to identify gluon jets and reject quark jets by applying a minimum threshold $n_{\text{track}} > n_{\text{track}}^{\text{min}}$ instead of a maximum one. Figure 2 quantifies the performance of the $n_{\text{track}}$-based quark-jet tagger for three different $p_T$ bins. The horizontal axis is the quark jet efficiency (Pr(tagged|quark-jet)) and the vertical axis is the gluon jet mis-tag rate (Pr(tagged|gluon-jet)). The performance improves with increasing $p_T$, partly because the total number of tracks inside quark and gluon jets increases with $p_T$ while scaling as the Casimir ratio $C_A/C_F = 9/4$. For a 50% quark jet efficiency, the gluon jet mistag rate is about 20% for jets with $p_T \sim 50$ GeV and reaches about 10% for jets with $p_T \sim 500$ GeV.
Figure 1: Distribution of the jet reconstructed track multiplicity ($n_{\text{track}}$) in different $p_T$ ranges with the Pythia 8 generator using the A14 tune [18], the NNPDF2.3 [19] PDF set, and processes with a full simulation of the ATLAS detector [20, 21]. Jets must be fully within the tracking acceptance ($|\eta| < 2.1$) and tracks are required to have $p_T > 500$ MeV and pass additional quality criteria described in Section 3.2.

Figure 2: Gluon jet mistag rate as a function of the quark jet efficiency using the jet track multiplicity as discriminant, with the same simulation setup as in Figure 1. Each line corresponds to jets in different $p_T$ ranges. Jets must be fully within the tracking acceptance ($|\eta| < 2.1$) and tracks are required to have $p_T > 500$ MeV and pass additional quality criteria described in Section 3.2. Each curve was generated by placing an upper threshold on $n_{\text{track}}$. A threshold of 0 corresponds to zero efficiency for both quark and gluon jets while a threshold of $\infty$ corresponds to 100% for both types of jets.
3 Systematic Uncertainties

The core idea of this note is to use the measurement of the particle-level charged-particle multiplicity at $\sqrt{s} = 8$ TeV for quark and gluon jets [22] to derive uncertainties for our simulation at $\sqrt{s} = 13$ TeV. The uncertainties on the corresponding particle-level distribution are described in Section 3.1. Additional uncertainties related to the detector-level track reconstruction at $\sqrt{s} = 13$ TeV are highlighted in Section 3.2.

3.1 Modeling Uncertainties

Dijet events, $pp \rightarrow jj$ ($j =$ parton), are predominantly $pp \rightarrow gg$ at low jet $p_T$ and mostly $pp \rightarrow qq'$ at high $p_T$ ($q =$ quark, $g =$ gluon) as shown in the left plot of Figure 3. For intermediate jet transverse momenta, there is a mix of the $gg, qg, qq$ final states. When the outgoing jets are well-balanced in $p_T$, the one that is more forward (higher $|\eta|$) has a higher energy and is more likely to be a scattering involving a valence quark. For the two highest $p_T$ jets in a dijet event, define the fraction in MC simulation that are labeled as a quark or gluon jet by $f_{f,c}^{q,g}$, where $f$ ($c$) denotes the jet of the pair at the higher (lower) $\eta$.

The fractions $f_{f,c}^{q,g}$ are due to parton distribution functions (PDF) convolved with matrix element (ME) calculations. The difference $f_{g}^{c} - f_{g}^{f}$ approaches zero at high and low $p_T$ and peaks above zero at $p_T \approx 400$ GeV (Figure 3). One can describe $\langle n_{\text{charged}} \rangle$ separately for quarks and gluons by the following system of equations:

$$
\langle n_{\text{charged}}^{f} \rangle = f_{q}^{f} \langle n_{\text{charged}}^{q} \rangle + f_{g}^{f} \langle n_{\text{charged}}^{g} \rangle
$$

$$
\langle n_{\text{charged}}^{c} \rangle = f_{q}^{c} \langle n_{\text{charged}}^{q} \rangle + f_{g}^{c} \langle n_{\text{charged}}^{g} \rangle
$$

in each $p_T$ bin. If the distribution of the charged-particle multiplicity inside jets is independent of the rapidity, then Eq. (1) can be used to extract the average number of charged particles for quark- and gluon-initiated jets in a given $p_T$ range. Symbolically, for any $q/g$ discriminant $X$ ($X = n_{\text{track}}$ in this case), the method presented in this note will work given that the following is satisfied:

$$
\Pr(X | f) = \Pr(X | f, q) \Pr(q | f) + \Pr(X | f, g) \Pr(g | f)
= \Pr(X | q) \Pr(q | f) + \Pr(X | g) \Pr(g | f), \quad \text{(and the same for c)}
$$

where the first line is true by the law of total probability. Equation 1 is a special case of Eq. (2) when taking the mean. To demonstrate that Eq. (1) is valid, one has to show that extracting $\langle n_{\text{charged}}^{q,g} \rangle$ using Eq. (1) or directly from labeled jets results in the same answer. This is shown to be true to much better than 1% across nearly the entire $p_T$ range in Figure 4 - the open stars, circles, and up triangles are all on top of each other and separately the open crosses, squares, and down triangles are also on top of each other.

In other words, Figure 4 demonstrates that the charged particle multiplicity inside jets (to an excellent approximation) only depends on the $p_T$ and type (quark or gluon) of the initiating parton.

1 These definitions only involve the relative position of the two jets. No explicit requirement on the absolute $|\eta|$ is applied.
2 A similar method was used by UA1 to study quark and gluon jet properties [23]. The key difference, discussed below, is that this analysis only uses rapidity which is largely independent of jet structure (as opposed to e.g. the jet $p_T$).
3 This plot shows the closure for Pythia, but a similar level of closure is also observed for Herwig++.
Figure 3: (a) The simulated fraction of jets originating from gluons as a function of jet $p_T$ for the more forward jet (down triangle), the more central jet (up triangle), and the difference between these two fractions (circle). The fractions are derived from Pythia 8 with the CT10 PDF set [24] and the error bars represent the PDF and matrix element uncertainties, as discussed in the text. The uncertainties on the fraction difference are computed from propagating the uncertainties on the more forward and more central fractions, treating as fully correlated. (b) The measured values of $\langle n_{q,g}^{\text{charged}} \rangle$. The next-to-next-to-leading-order (N$^3$LO) pQCD calculation [25, 26] is normalized to the data in the second $p_T$ bin. Both plots are reproduced from Ref. [22].

In this note the approach taken in Ref. [22] is used to calibrate the systematic uncertainties on the tagger described above. In brief, Ref. [22] presents a measurement of the average charged-particle multiplicity inside jets as a function of jet $p_T$. Those data were further analyzed to extract the average quark and gluon jet charged-particle multiplicity separately, by exploiting the rapidity dependence of the quark/gluon jet fractions. The results of the extraction procedure applied to the unfolded data are shown in the right plot of Figure 3. The nominal fractions $f_{q,g}^{L,C}$ are taken from Pythia 8 [27, 28] with the CT10 [24] PDF set. In addition to the experimental and statistical uncertainties associated with the measurement, two additional uncertainties are included in the right plot of Figure 3. The CT10 eigenvector variations are used as an estimate of the PDF uncertainty and the comparison between Pythia 8 and Herwig++ [29] is used as an estimate of the uncertainty on the quark/gluon fraction due to the ME calculation. Both of these uncertainties are added in quadrature to the other uncertainties to form the uncertainty bands in the right plot of Figure 3. Since the PDF set in Herwig++ is CTEQ6L1 [30] and the default in Pythia 8 is CT10, the LHAPDF library [31] is used to re-weight Pythia 8 from CT10 to CTEQ6L1 in the estimation of this uncertainty. The uncertainties that go into the right plot of Figure 3 are summarized in Tables 1 and 2 for gluons and quarks, respectively.
Figure 4: An illustration of the closure test from the central-forward method for jets with $p_T > 50$ GeV. The filled squares and circles in the upper panel are value of $\langle n_{\text{charged}}^{c,f} \rangle$. In the same panel, the open blue points show $\langle n_{\text{charged}}^{q,g} \rangle$ extracted from Eq. (1) while the open red and black points show $\langle n_{\text{charged}}^{q,g} \rangle$ for the more forward and more central jets extracted from labels directly in simulation. The middle panel shows the ratio of $\langle n_{\text{charged}}^{q,g} \rangle$ for the forward versus central jets and the lower panel shows the ratio for the values extracted from Eq. (1) and the ones taken directly from simulation for the more forward jets. This extraction is possible because the filled red squares and circles are slightly displaced from each other; closure is given by the fact that the open stars, circles, and up triangles are on top of each other and separately the open crosses, squares, and down triangles are on top of each other.
Figure 5 illustrates the impact of these uncertainties on the 60% quark jet efficiency working point\(^6\) of the quark/gluon tagger. To translate a shift in the average charged-particle multiplicity to a change in the efficiency (via the full charged particle distributions), per-jet weights are assigned and the quark and gluon uncertainties are treated as fully correlated. The per-jet event weight for a jet with a given \(n_{\text{track}}\) are calculated from the ratio of Poisson probabilities taking \(n\) and \(n \pm \sigma\) as means. The per-jet weights are given by

\[
w_i(n_{\text{charged}}) = \frac{\Pr_i(n_{\text{charged}} | \langle n_{\text{charged}} \rangle \pm \sigma \langle n_{\text{charged}} \rangle)}{\Pr_i(n_{\text{charged}} | \langle n_{\text{charged}} \rangle)}.
\]

where \(i \in \{q, g\}\).

Several additional variations are studied to ensure that the size of the uncertainties are reasonable. First, the Pythia 8 sample is re-weighted to the central NNPDF 2.3 set [19] to see if the difference in these two unrelated PDF sets gives a similar size uncertainty to the CT10 variations. This seems to be true, with uncertainties for both ranging from 0.1 charged particles at low \(p_T\) to as much as \(\sim 1\) particles at high \(p_T\). Further tests are performed to assess the impact of the definition of quark and gluon jets on the results in part due to the ambiguity in identifying jets as quark or gluon. First, instead of using \(\Delta R \leq 0.4\) in the matching, a half cone of \(\Delta R \leq 0.2\) was used. The impact of this change is quite small, never exceeding 0.02 particles for gluons and 0.03 for quarks. A second check is performed by taking the outgoing partons from the ME and assigning them to the outgoing jets based on which ones are closer in \(\Delta R\). The impact has a similar (small) size as for the half-cone test. All three of these tests are also summarized in Tables 1 and 2 for gluons and quarks, respectively. Both these uncertainties are found to be negligible and are therefore ignored.

Note that even though the Pythia 8 sample with the A14 tune [18] is a good model of the charged particle multiplicity inside jets, this is not the case for all modern Pythia 8 sets of tuned parameters. For example, a tune widely used by Higgs analyses is the AZNLO tune [32]. As with the default Monash Pythia 8 tune [33], AZNLO over-estimated the multiplicity, as shown in Figure 6. As a result of the factorized approach presented in this note, it is possible to use the same procedure for any tune.

This section has presented modeling uncertainties for the dijet topology described in Ref. [22]. Quark versus gluon tagging is inherently topology dependent because quarks and gluons carry color charge while the hadrons in jets do not. However, due to the collinear factorization of QCD amplitudes, tagging well-isolated small-radius jets should be relatively universal. Previous studies in Ref. [6] suggest that for \(n_{\text{track}}\) the topology dependence is small, but this is an important consideration for any analysis using the tagger presented in this note.

---

\(^6\) For each \(p_T\) bin, the \(n_{\text{track}}\) requirement corresponding to the efficiency closest to 60% is used. Any efficiency is possible; 60% is merely chosen for illustration.
### Table 1: Upper table: a summary of the systematic uncertainties on the average charged multiplicity extraction for gluons. Lower table: additional checks on the size of the uncertainties (see text for details). The uncertainties are in units of $n_{\text{charged}}$.

<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 exp.</td>
<td>+0.44</td>
<td>+0.29</td>
<td>+0.15</td>
<td>+0.24</td>
<td>+0.21</td>
<td>+0.37</td>
<td>+0.48</td>
<td>+1.01</td>
<td>+2.20</td>
<td>+6.09</td>
</tr>
<tr>
<td></td>
<td>−0.34</td>
<td>−0.24</td>
<td>−0.24</td>
<td>−0.17</td>
<td>−0.33</td>
<td>−0.43</td>
<td>−0.58</td>
<td>−1.03</td>
<td>−2.39</td>
<td>−6.16</td>
</tr>
<tr>
<td>ME</td>
<td>0.04</td>
<td>0.06</td>
<td>0.05</td>
<td>0.12</td>
<td>0.14</td>
<td>0.16</td>
<td>0.06</td>
<td>0.01</td>
<td>0.05</td>
<td>0.22</td>
</tr>
<tr>
<td>PDF</td>
<td>−0.01</td>
<td>−0.05</td>
<td>−0.10</td>
<td>−0.19</td>
<td>−0.27</td>
<td>−0.34</td>
<td>−0.48</td>
<td>−0.60</td>
<td>−1.01</td>
<td>−0.81</td>
</tr>
<tr>
<td>PDF comparison</td>
<td>0.03</td>
<td>0.09</td>
<td>0.00</td>
<td>0.04</td>
<td>0.01</td>
<td>0.10</td>
<td>0.33</td>
<td>0.84</td>
<td>1.76</td>
<td>1.69</td>
</tr>
<tr>
<td>Half cone</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>HS Label</td>
<td>0.06</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.01</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
</tr>
</tbody>
</table>

### Table 2: Upper table: a summary of the systematic uncertainties on the average charged multiplicity extraction for quarks. Lower table: additional checks on the size of the uncertainties (see text for details). The uncertainties are in units of $n_{\text{charged}}$.

<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 exp.</td>
<td>+0.82</td>
<td>+0.36</td>
<td>+0.26</td>
<td>+0.22</td>
<td>+0.25</td>
<td>+0.30</td>
<td>+0.32</td>
<td>+0.41</td>
<td>+0.69</td>
<td>+1.42</td>
</tr>
<tr>
<td></td>
<td>−1.16</td>
<td>−0.41</td>
<td>−0.28</td>
<td>−0.30</td>
<td>−0.32</td>
<td>−0.35</td>
<td>−0.36</td>
<td>−0.47</td>
<td>−0.67</td>
<td>−1.70</td>
</tr>
<tr>
<td>ME</td>
<td>0.06</td>
<td>0.23</td>
<td>0.19</td>
<td>0.23</td>
<td>0.22</td>
<td>0.25</td>
<td>0.26</td>
<td>0.22</td>
<td>0.23</td>
<td>0.16</td>
</tr>
<tr>
<td>PDF</td>
<td>+0.02</td>
<td>+0.11</td>
<td>+0.17</td>
<td>+0.27</td>
<td>+0.33</td>
<td>+0.38</td>
<td>+0.44</td>
<td>+0.47</td>
<td>+0.62</td>
<td>+0.45</td>
</tr>
<tr>
<td></td>
<td>−0.02</td>
<td>−0.10</td>
<td>−0.16</td>
<td>−0.24</td>
<td>−0.27</td>
<td>−0.28</td>
<td>−0.30</td>
<td>−0.28</td>
<td>−0.33</td>
<td>−0.21</td>
</tr>
<tr>
<td>PDF comparison</td>
<td>0.04</td>
<td>0.01</td>
<td>0.17</td>
<td>0.23</td>
<td>0.17</td>
<td>0.10</td>
<td>0.01</td>
<td>0.21</td>
<td>0.44</td>
<td>0.39</td>
</tr>
<tr>
<td>Half cone</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>HS Label</td>
<td>0.07</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Figure 5: The systematic uncertainties on the particle-level charged particle distribution for the 60% quark jet efficiency working point of the quark/gluon tagger.

Figure 6: A comparison of the particle-level charged particle multiplicity for the more forward jet from data and Pythia 8 with the A14 and AZNLO tunes. The data with uncertainties are from Ref. [22] via HepData [34].
3.2 Experimental Uncertainties

The particle-level analysis described in Section 3.1 provides a detector-independent set of uncertainties on the (average) charged-particle multiplicity inside quark and gluon jets. In order to connect charged-particle multiplicity with track multiplicity, an additional set of detector-specific uncertainties are required. There are three categories of uncertainties associated with track reconstruction: reconstruction efficiency, fake rate, track fit parameter bias (scale) and resolution. One contribution to the reconstruction efficiency uncertainty is from the description of the ID material in simulation and the modeling of charged-particle interactions with this material. This uncertainty is estimated by varying the material within its measured uncertainty and by varying the Geant4 physics list [35]. A second contribution to the reconstruction efficiency is from the high particle density in the cores of jets where ID clusters in the same layer can merge. The fraction of lost tracks in the core of jets is measured by looking at one-track clusters in the core of jets with a deposited charge consistent with two minimum ionizing particles [36]. Fake tracks are reconstructed trajectories that cannot be (mostly) associated to a single charged particle. The rate of fake tracks increases with pile-up due to the higher hit rate in the ID. Quality criteria on the reconstructed track can significantly reduce the fake rate, which has been measured for various track selections [35].

Figure 7 summarizes the impact of the tracking uncertainties on the quark/gluon tagger for the 60% quark jet efficiency working point. Tracks are required to have $p_T > 0.5$ GeV and to satisfy quality criteria designed to reject poorly measured and fake tracks [35]. Tracks are assigned to primary vertices based on the track-to-vertex association resulting from the vertex reconstruction. Tracks not included in vertex reconstruction are assigned to the nearest vertex based on the distance $|\Delta z \times \sin \theta|$, up to a maximum distance of 3.0 mm. Tracks not matched to any vertex are not considered.

Figure 7: The experimental systematic uncertainties on the track reconstruction for the 60% quark jet efficiency working point of the quark/gluon tagger.
Figure 8 shows the total uncertainty for the 60% quark jet efficiency working point of the quark/gluon tagger. The modeling and experimental uncertainties are considered uncorrelated and summed in quadrature.
4 Conclusions

This note presents a novel approach to derive uncertainties for a $n_{\text{track}}$-based quark/gluon jet tagger. For a quark jet efficiency of about 60%, a gluon efficiency of less than 20% can be achieved with an uncertainty on particle- and detector-level modeling that is $\lesssim 5\%$ over a wide range in $p_T$. Particle-level modeling uncertainties are derived for a particular dijet topology. As quark versus gluon tagging is topology dependent, users must consider potential residual systematic effects when extending to other final states. Uncertainties are derived for jet $p_T$ well beyond 1 TeV, though the overflow bin begins at 1.2 TeV, at which point the uncertainties are extrapolated assuming a relatively constant uncertainty starting from the last bin. The methods presented here can be adapted for more complex taggers in the future, exploiting more features of the radiation pattern inside quark and gluon jets. Even though the method is currently used only to provide uncertainties that cover the data/MC differences, it can also be used in principle to calibrate in order to correct data/MC differences.

References


[29] M. Bähr, S. Gieseke and M. H. Seymour, 


[34] http://hepdata.cedar.ac.uk/.

[35] ATLAS Collaboration, 