Identification of charm jets at LHCb

To cite this article: The LHCb collaboration et al 2022 JINST 17 P02028

View the article online for updates and enhancements.
Identification of charm jets at LHCb

The LHCb collaboration

E-mail: dcraik@cern.ch

ABSTRACT: The identification of charm jets is achieved at LHCb for data collected in 2015–2018 using a method based on the properties of displaced vertices reconstructed and matched with jets. The performance of this method is determined using a dijet calibration dataset recorded by the LHCb detector and selected such that the jets are unbiased in quantities used in the tagging algorithm. The charm-tagging efficiency is reported as a function of the transverse momentum of the jet. The measured efficiencies are compared to those obtained from simulation and found to be in good agreement.

KEYWORDS: Analysis and statistical methods; Pattern recognition, cluster finding, calibration and fitting methods; Performance of High Energy Physics Detectors

ArXiv ePrint: 2112.08435
1 Introduction

Identification of charm jets, i.e. those originating from the hadronisation of a charm quark, is of interest in both the study of Standard Model (SM) processes and the search for new physics. For example, the production of events containing a Z boson and a c jet in the forward region provides a direct probe of the charm content of the proton at large parton momentum fractions \([1, 2]\). Such studies rely upon algorithms capable of distinguishing charm jets from beauty and light-parton jets.

An algorithm for identifying both charm and beauty jets reconstructed by the LHCb detector has been used for studies of the dataset recorded in 2011–2012 (Run 1) \([3]\). However, the higher particle multiplicity of 2015–2018 (Run 2) data has been found to degrade its performance \([4]\). Furthermore, for measurements that only involve charm jets, it is possible to achieve better performance using a dedicated charm-tagging algorithm.

Charm jets are defined as those that have a promptly produced and weakly decaying c hadron with transverse momentum \(p_T(c \text{ hadron}) > 5 \text{ GeV}\) within the jet cone.\(^1\) Therefore, the tagging of c jets is performed using displaced vertices (DVs) formed from the decays of such c hadrons. The choice of using DVs and not single-track or other non-DV-based jet properties, e.g. the number of particles in the jet, is driven by the need for a small misidentification probability of the copious light-parton jets in LHCb c-jet analyses. In addition, the properties of DVs from c-hadron decays

\(^1\)Natural units are used throughout this article.
are known to be well modeled by simulation, which means that only small corrections obtained from control samples are required. Since DVs can also be formed from the decays of beauty or strange hadrons, or due to artifacts of the reconstruction, the DV-tagged charm yields are obtained by fitting the distributions of DV features with good discrimination power between $c$, $b$, and light-parton jets.

This article presents a dedicated $c$-tagging procedure used to efficiently identify charm jets produced in proton-proton ($pp$) collisions at a centre-of-mass energy $\sqrt{s} = 13$ TeV and recorded by the LHCb detector. The procedure described produces a statistical separation of the different jet-flavour populations resulting in a measurement of the total yield of each category rather than event-by-event flavour tags as given by some other methods. Therefore, the ability to reject light-flavour jets is contained within the uncertainty on the charm-tagging efficiency rather than a mistag rate. The $c$-tagging efficiency is precisely determined using a sample of unbiased charm-enriched jets obtained from dijet events in a dataset corresponding to an integrated luminosity of $1.7 \text{fb}^{-1}$ collected in 2016. This efficiency is reported for jets with $20 < p_T(j) < 100$ GeV in the pseudorapidity range $2.2 < \eta(j) < 4.2$. The region below 20 GeV is not reported because the $c$-tagging efficiency varies rapidly there, whereas above 100 GeV the limited size of the calibration sample prohibits precisely determining the performance. The $\eta(j)$ range, which was first used in refs. [3, 5, 6], ensures a nearly uniform $c$-tagging efficiency of about 24%, with minimal $p_T(j)$ or $\eta(j)$ dependence.

2 Detector and simulation

The LHCb detector [7, 8] is a single-arm forward spectrometer covering $2 < \eta < 5$, designed for the study of particles containing $b$ or $c$ quarks. The detector includes a high-precision tracking system consisting of a silicon-strip vertex detector surrounding the $pp$ interaction region, a large-area silicon-strip detector located upstream of a dipole magnet with a bending power of about 4 Tm, and three stations of silicon-strip detectors and straw drift tubes placed downstream of the magnet. The tracking system provides a measurement of the momentum of charged particles with a relative uncertainty that varies from 0.5% at low momentum to 1.0% at 200 GeV. The minimum distance of a track to a primary $pp$ collision vertex (PV), the impact parameter (IP), is measured with a resolution of $(15 + 29/p_T) \mu$m, where $p_T$ is in GeV. Different types of charged hadrons are distinguished using information from two ring-imaging Cherenkov detectors. Photons, electrons, and hadrons are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers.

The online event selection is performed by a trigger, which consists of a hardware stage, based on information from the calorimeter and muon systems, followed by a software stage, which applies a full event reconstruction. At the hardware trigger stage, events are required to have a muon with high $p_T$ or a hadron, photon, or electron with high transverse energy in the calorimeters. For hadrons, the transverse energy threshold is 3.5 GeV. The software trigger requires at least one charged particle to be reconstructed with $p_T > 1.6$ GeV that is inconsistent with originating from any PV, as well as the presence of two jets. Both jets are reconstructed as described in section 3, and required to have $p_T > 17$ GeV. At least one jet is required to have a DV in the jet cone.

Simulation is required to model the effects of the detector acceptance and the imposed selection requirements. In the simulation, $pp$ collisions are generated using PYTHIA [9, 10] with a specific
LHCb configuration [11]. Decays of unstable particles are described by EvtGen [12], in which final-state radiation is generated using Photos [13]. The interaction of the generated particles with the detector, and its response, are implemented using the Geant4 toolkit [14, 15] as described in ref. [16].

3 Charm-jet identification

Jets are reconstructed from particle flow objects [17] using the FastJet [18] implementation of the anti-\(k_T\) algorithm [19] with a jet radius parameter of \(R = 0.5\). The same jet reconstruction algorithm is used online and offline; however, differences in the reconstruction routines for tracks and calorimeter clusters lead to minor differences between the online and offline jets. In addition to jets in the fiducial region, offline jets with \(15 < p_T(j) < 20\) GeV are retained for use when unfolding the detector response.

Charm jets are identified based on properties of DVs associated with the jets, reconstructed in a manner similar to that used in ref. [3]. DV candidates are reconstructed using good-quality tracks both within and outside of the jet, with \(p_T > 0.5\) GeV and \(\chi^2_{IP} > 9\), where \(\chi^2_{IP}\) is defined as the difference in the vertex-fit \(\chi^2\) of the PV reconstructed with and without the track under consideration. Tracks are combined into two- and three-body DVs, which are required to form a good-quality vertex, be downstream of the PV, and have an invariant mass greater than 0.4 GeV and less than that of the \(B^0\) meson. The corrected mass is required to satisfy

\[
m_{\text{cor}}(\text{DV}) \equiv \sqrt{m(\text{DV})^2 + (p(\text{DV}) \sin \theta)^2 + p(\text{DV}) \sin \theta > 0.6\text{ GeV}},
\]

where \(\theta\) is the angle between the DV momentum and its direction of flight, defined by the vector from the \(pp\) interaction point to the DV position. In addition, the uncertainty on the corrected mass, as computed from the covariances of the primary and displaced vertices and the DV momentum, is required to be less than 0.5 GeV. Two- and three-body DV candidates that pass these requirements and share one or more tracks are linked together to form \(n\)-body DVs. All DV candidates are subsequently required to have \(p_T > 2\) GeV and a significant separation from all PVs. To reduce backgrounds due to strange-hadron decays and material interactions, DVs are required to have a decay time consistent with a heavy-flavour hadron, and have a significant separation from all material within and around the vertex detector [20].

Given that the method presented here is only concerned with tagging charm jets and that DVs with more than four tracks originate predominantly from beauty decays, only DV candidates with two, three, or four tracks are retained. A DV is associated to a jet when \(\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2} < 0.5\) between the jet axis and the DV direction of flight, where \(\phi\) denotes the azimuthal angle. If more than one DV candidate is assigned to a given jet, which occurs for \(O(1\%)\) of charm jets, the candidate with the largest \(p_T\) is retained. The key differences between the DV candidates used in this study and those in ref. [3] are a less stringent \(\chi^2_{IP}\) requirement, the addition of the corrected mass uncertainty requirement, and the requirement for no more than four tracks.

To determine the number of DVs that originate from charm jets, a two-dimensional maximum-likelihood fit is performed to the \(m_{\text{cor}}(\text{DV})\) and \(N_{\text{trk}}(\text{DV})\) distributions, where the latter is the track multiplicity of the DV candidate. The fit procedure and the probability density functions, referred to as templates, used to describe the charm, beauty, and light-parton components are described
in section 4.3. The requirement of the presence of a reconstructed DV in the jet, along with the application of this fit, constitutes the charm-jet tagging, or $c$-tagging, algorithm.

In simulated data, it is possible to unambiguously determine the fraction of charm jets that contain a DV candidate without the need to perform a fit. While small discrepancies are expected between data and simulation, simulated DVs should reliably reproduce the efficiency. In the simulation, the $c$-tagging efficiency is about 24% and nearly uniform in the $20 < p_T(j) < 100$ GeV and $2.2 < \eta(j) < 4.2$ region.

4 Calibration in data

The $c$-tagging efficiency is measured using a sample of charm-enriched jets obtained from dijet events as described in section 4.1, which have been selected such that the jets remain unbiased with respect to the charm-tagging algorithm. Exclusive charm decays are used to determine the total number of $c$ jets in the sample. This is inefficient as the majority of charm hadrons do not decay to any given final state; however, the efficiency can be reliably modelled in simulation. In addition, charm-quark hadronisation into a $c$ hadron, followed by an exclusive decay, can be calculated in many cases with well known fragmentation and branching fractions. The $c$-tagging efficiency is determined as the ratio of the number of charm jets tagged to the total charm-jet yield in the sample:

$$\epsilon_{c\text{-tag}} = \frac{N_{c\text{-tag}}}{N_c}$$

(4.1)

where $N_{c\text{-tag}}$ is the number of $c$-tagged jets, i.e. the charm yield obtained by fitting the $[m_{c\text{cor}}(DV), N_{\text{trk}}(DV)]$ distribution for jets with an associated DV, and $N_c$ is the total charm-jet yield. The total $c$-jet yield is calculated separately using $D^0 \rightarrow K^-\pi^+$ and $D^+ \rightarrow K^-\pi^+\pi^+$ decays,\(^2\) collectively or generically referred to as $D$ decays hereafter, and a weighted average of the two results is used for the default $c$-tagging efficiency. The total charm-jet yield is determined from each decay channel as

$$N_c(D) = \frac{N_{\text{prompt}}(D)}{\epsilon_D f_{c\rightarrow D} B(D)}$$

(4.2)

where $N_{\text{prompt}}(D)$ is the observed number of promptly produced $D$ mesons obtained by fitting the $D$-meson candidate mass and $X_{\text{IP}}^2$ distributions, $\epsilon_D$ is the efficiency with which $D$ candidates are reconstructed and selected as determined from simulation, $f_{c\rightarrow D}$ is the fragmentation fraction for a charm quark to hadronise as the required $c$ hadron, and $B(D)$ is the corresponding branching fraction for the $D$-meson decay. The branching and fragmentation fractions, which are derived from refs. [21] and [22], respectively, are listed in table 1. The fragmentation fractions in ref. [22] are corrected to account for updated branching-fraction measurements [21], and include a small correction derived from simulation for the case where ground-state $c$ hadrons produced in the decays of excited charm states, e.g. $D^* \rightarrow D\pi$, have $p_T(D) < 5$ GeV despite the parent state being above the 5 GeV fiducial threshold. The total correction is $O(1\%)$ for both decay channels.

\(^2\)Note that the inclusion of charge-conjugate decay modes is implied.
Table 1. Branching and fragmentation fractions used to obtain the total charm yields from $D^0 \rightarrow K^- \pi^+$ and $D^+ \rightarrow K^- \pi^+ \pi^+$ decays. The PDG [21] averages are used for both branching fractions. Charm fragmentation fractions are based on the global averages reported in ref. [22], but have been updated as detailed in the text. The fragmentation fractions are inclusive of feed down from excited charm states.

<table>
<thead>
<tr>
<th>Decay</th>
<th>Branching fraction (%)</th>
<th>Fragmentation fraction (%)</th>
<th>$B(D) \times f_c \rightarrow D$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D^0 \rightarrow K^- \pi^+$</td>
<td>3.950 ± 0.031</td>
<td>60.86 ± 0.76</td>
<td>60.12 ± 0.77</td>
</tr>
<tr>
<td>$D^+ \rightarrow K^- \pi^+ \pi^+$</td>
<td>9.38 ± 0.16</td>
<td>24.04 ± 0.67</td>
<td>23.90 ± 0.68</td>
</tr>
</tbody>
</table>

4.1 Calibration datasets

The $c$-tagging efficiency is measured on a dijet control sample using a tag-and-probe method. Events are retained for further analysis if one of the two jets, henceforth called the tag jet and denoted by the symbol $j_{\text{tag}}$, is associated with a DV. The jets are required to be well separated in azimuthal angle, $\Delta \phi > 2$, and to have well balanced transverse momenta: $A_{\text{pt}}^{jj} \equiv (p_T(j) - p_T(j_{\text{tag}})) / (p_T(j) + p_T(j_{\text{tag}})) < 0.25$. Additionally, the trigger requirements introduced in section 2 are required to be fulfilled by the tag jet. These requirements enhance the fraction of heavy flavour, i.e. $b \bar{b}$ and $c \bar{c}$, events within the sample relative to those containing light-parton jets without biasing the properties of the probe jet. Further enriched sub-samples are also obtained by placing additional requirements on the tag-jet DV candidates. Specifically, the enriched charm-jet sub-sample requires a DV candidate with $m_{\text{cor}}(\text{DV}) < 2$ GeV and only two tracks, while the beauty-jet sub-sample requires a DV candidate with $m_{\text{cor}}(\text{DV}) > 2$ GeV and three or four tracks. The obtained charm-jet and beauty-jet sub-samples are approximately 30–40% and 60–70% pure, respectively, based on fits to the samples. These enriched sub-samples are used to perform data-driven corrections to the fit templates as described in section 4.3. An additional sample, with no DV requirements placed on the tag jet, is used to study light-parton jets. This sample is enriched in fake DVs, and hence mis-tagged light-parton jets, by retaining only DV candidates reconstructed upstream of the associated PV. The tagging requirements applied in the four samples derived from the dijet dataset are illustrated in figure 1.

4.2 $D$-meson decay selection and fits

The selection and fit procedures used for $D$-meson decays closely match those used in previous studies of prompt charm production at LHCb [23–25]. The $D$-meson decays are reconstructed by combining good-quality charged tracks with requirements placed on their momentum, $p_T$, and $\chi^2_{IP}$. Kaon candidates are also required to either pass a kaon particle-identification requirement or have high momentum. Requirements are also placed on the invariant mass and $p_T$ of the combination as well as the vertex quality, the significance of separation from the PV, and the angle between the flight direction from the PV and the momentum vector. For $D^0$ candidates, a requirement is also placed on the distance of closest approach of the two charged particle tracks. For $D^+$ candidates, an additional requirement is placed on the minimum decay time. All candidates are required to have momentum vectors that fall within $\Delta R < 0.5$ of the jet axis.

To distinguish promptly produced charm hadrons from those produced in $b$-hadron decays and from combinatorial background, a two-dimensional unbinned maximum-likelihood fit is performed

---

---
to the invariant mass and log $\chi^2_{IP}$ distributions of the $D$-meson candidates. The mass distributions of the prompt and from-$b$ components are each described by the sum of a Gaussian with a Crystal Ball function [26], while the background is described by a linear function. The log $\chi^2_{IP}$ distributions of the prompt and from-$b$ components are described by asymmetric Gaussian functions with exponential tails, while the background is described by a kernel density estimation derived from data in the mass-sideband regions. Various shape parameters of the fit components are fixed to values determined from simulation. To better describe the data, fits are performed simultaneously to five intervals of $p_T$, with some shape parameters allowed to vary independently in each interval.

Figure 2 shows the combinatorial-background-subtracted invariant mass and log $\chi^2_{IP}$ distributions for all $D^0$ and $D^+$ candidates associated with jets reconstructed in the efficiency-reporting region, along with projections of fits performed on these samples. Such fits are performed in each interval of jet $p_T$. The prompt signal yields extracted from these fits are scaled by an efficiency-correction factor, which is determined from simulated events as a function of the charm-hadron kinematics, and weighted according to the kinematic distribution of candidates in the signal region of invariant mass and log $\chi^2_{IP}$.

4.3 Displaced vertex fits

Candidate DVs are selected as described in section 3. A two-dimensional fit is performed to the corrected mass and track multiplicity distributions to extract the $c$-jet component. The template describing the light-parton-jet background is taken from jets with the DV displacement requirement reversed, such that reconstructed DV candidates are displaced backwards with respect to the PV. Templates describing the distributions of the $c$ and $b$ components are taken from simulation but corrected to match data using fits performed to subsets of the data that have been further enriched in charm and beauty as described in section 4.1. After each fit, the template of the enriched component
is modified to minimise the residuals. Fits are first performed to the enriched beauty sample and then to the enriched charm sample, and this process is repeated iteratively until the fit results change by less than 1%. In practice, a single iteration is found to be sufficient. Templates for the $m_{\text{cor}}(\text{DV})$ and $N_{\text{trk}}(\text{DV})$ distributions are shown in figure 3. Examples of $m_{\text{cor}}(\text{DV})$ and $N_{\text{trk}}(\text{DV})$ distributions together with fit projections to the charm- and beauty-enriched sub-samples and to the full heavy-flavour-enriched sample with corrections applied are shown in figure 4. As this study defines charm jets as those containing a $c$ hadron with $p_T > 5$ GeV, a correction must be applied to account for cases where a $p_T < 5$ GeV $c$ hadron produces a DV candidate. However, this correction is found to be $O(1\%)$ in all of the $p_T(j)$ ranges considered. Furthermore, this correction largely cancels when using the $c$-tagging efficiencies measured here to correct $c$-jet yields in analyses that employ the same charm-jet definition.

4.4 Unfolding

As $p_T(j)$ resolution effects may differ between jets containing a reconstructed $D$-meson decay or a DV candidate, $p_T(j)$ interval migration must be considered separately for the numerator and denominator in the $c$-tagging efficiency measurement. In both cases, unfolding is performed using an iterative Bayesian procedure [27] as implemented in RooUNFOLD[28] with two iterations. The unfolding matrices for DV-tagged charm jets as well as charm jets containing reconstructed $D^0$ and $D^*$ decays are shown in figure 5. These are determined from simulated data that have been
Figure 3. Probability density functions for (left) $m_{\text{cor}}(\text{DV})$ and (right) $N_{\text{trk}}(\text{DV})$ used in the fits for (solid green) charm, (dashed red) beauty, and (dotted blue) light-parton jets.

Figure 4. DV (left) corrected mass and (right) track multiplicity projections of fits to the flavour-enriched jet samples: (top-to-bottom) beauty-enriched sub-sample, charm-enriched sub-sample, and heavy-flavour-enriched sample fit with data-driven corrections.
Figure 5. Detector-response matrices for (left) $D^0$, (right) $D^+$ and (bottom) DV-tagged charm jets. The shading represents the interval-to-interval migration probabilities ranging from (white) 0 to (black) 1 such that each row sums to unity when the underflow and overflow bins are included. Jets with true (reconstructed) $p_T(j)$ in the 20–100 GeV region but whose reconstructed (true) $p_T(j)$ is either below 15 GeV or above 100 GeV are included in the unfolding but not shown graphically.

weighted to better describe the $p_T(j)$, $p_T(DV)$, and $p_T(D)$ distributions observed in data. In addition, the detector response is studied in data using the $p_T$-balance distribution $p_T(j)/p_T(Z)$ for $Z$+jet candidates that are nearly back-to-back in the transverse plane, using the same technique as in refs. [17, 29]. Small adjustments are applied to the $p_T(j)$ scale and resolution in simulation to obtain the best agreement with data.

4.5 Systematic uncertainties

Two categories of systematic uncertainty affect the $c$-tagging efficiency: those that affect both the efficiency measurement performed here and the $c$-tagged data samples used in subsequent measurements, e.g. uncertainties in the DV fitting procedure; and those that affect only the $c$-tagging efficiency, e.g. the $D$-meson fitting procedure. The former category of uncertainties partially cancel in any efficiency-corrected results, e.g. the measurement of $\sigma(Zc)/\sigma(Zj)$ [2], and therefore must
Table 2. Relative systematic uncertainties (%) on the tagging efficiencies determined using the $D^0$ and $D^+$ decays as well as their weighted combination. Ranges of uncertainties are given when the value depends on the $p_T(j)$ interval. The total systematic uncertainty is evaluated as the sum in quadrature of the uncertainties from all sources.

<table>
<thead>
<tr>
<th>Source</th>
<th>$D^0$</th>
<th>$D^+$</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$ fit models</td>
<td>4</td>
<td>5–18</td>
<td>3–6</td>
</tr>
<tr>
<td>$D$ efficiency method</td>
<td>1–2</td>
<td>3–8</td>
<td>1–2</td>
</tr>
<tr>
<td>Simulation sample size</td>
<td>1</td>
<td>2–4</td>
<td>1</td>
</tr>
<tr>
<td>Particle identification</td>
<td>1–2</td>
<td>4–7</td>
<td>1–2</td>
</tr>
<tr>
<td>Modeling detector response</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Fragmentation &amp; branching fractions</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2015–16 vs. 2017–18</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>5–6</td>
<td>9–21</td>
<td>5–7</td>
</tr>
</tbody>
</table>

be calculated separately for each study that uses this tagging method. Sources of uncertainty that fall into the latter category, which do not cancel in analyses, are considered below.

The systematic uncertainty due to the $D$-decay $[m(K^−π^+), \log \chi^2_IP]$ and $[m(K^−π^+π^−), \log \chi^2_IP]$ fits accounts for imperfect knowledge of the probability density functions used to model the three fit components. The uncertainty is assigned as the largest deviation from the default results based on several variations to the fit model. An uncertainty is assigned due to the efficiency-weighting procedure applied to the $D$-decay yields. Specifically, the integrated efficiency factor is replaced by a per-event efficiency correction and the differences from the default results are assigned as systematic uncertainties. Additional systematic uncertainties are assigned due to the limited size of simulation samples used to determine $D$-decay efficiencies, the procedure used to determine the efficiency of particle-identification requirements, and potential data–simulation discrepancies in modeling the detector response. Uncertainties on the fragmentation and branching fractions are propagated through to the $c$-tagging efficiency results accounting for correlations. The uncertainties for each source are listed in table 2.

As the error parameterisation used during track reconstruction was changed between data collected in 2015–16 and 2017–18, the tagging efficiency differs between these datasets. This leads to a need for a correction to the tagging efficiencies calculated from this study, which uses only data recorded in 2016. The impact of this change on the $c$-tagging efficiency is obtained from simulation. After accounting for the relative proportions of 2015–16 and 2017–18 data in the full LHCb Run 2 dataset, a 2% correction to the $c$-tagging efficiency is obtained, which is also assigned as a systematic uncertainty.

As discussed above, systematic uncertainties that affect both the $c$-tagging efficiency measurement and the $c$-tagged data samples used in subsequent measurements partially cancel in any efficiency-corrected results. Therefore, these uncertainties must be estimated separately for each measurement and are not reported here. Additional measurement-dependent sources of uncertainty include the unfolding procedure, jet reconstruction efficiency, jet energy scale and resolution, and the DV-fit templates. In ref. [2], these sources contribute an additional 4–5% relative uncertainty to the results.
Table 3. Charm-tagging efficiencies (%) determined in intervals of $p_T(j)$. First and second uncertainties are statistical and systematic, respectively. Note that systematic uncertainties that affect both the efficiency measurement and the $c$-tagged data samples used in subsequent measurements are not included because these will largely cancel.

<table>
<thead>
<tr>
<th>$p_T(j)$ interval [GeV]</th>
<th>(20, 30)</th>
<th>(30, 50)</th>
<th>(50, 100)</th>
<th>(20, 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>23.9 ± 0.7 ± 1.2</td>
<td>24.4 ± 1.4 ± 1.3</td>
<td>23.6 ± 3.7 ± 1.7</td>
<td>24.0 ± 0.6 ± 1.4</td>
</tr>
</tbody>
</table>

5 Results

The $c$-tagging efficiency is measured in intervals of $p_T(j)$ using eq. (4.1), i.e. as the ratio of the number of DV-tagged charm jets to the total charm-jet yield in the control sample. The DV-tagged charm yields in intervals of reconstructed $p_T(j)$ are obtained from the fits described in section 4.3. The total charm-jet yields, also in intervals of reconstructed $p_T(j)$, are obtained using eq. (4.2), which takes as input the $D$-meson fit results and efficiency corrections of section 4.2, and the fragmentation and branching fractions from table 1. Interval migration due to $p_T(j)$ resolution is accounted for separately for the DV-tagged and total charm-jet yields using the unfolding approach of section 4.4.

The measured Run 2 tagging efficiency is given in table 3 in intervals of $p_T(j)$ as well as integrated in $p_T(j)$. Comparing to simulation, the scale factors required to correct the $c$-tagging efficiency are determined to be $1.03 ± 0.06$, $1.01 ± 0.08$, and $1.09 ± 0.17$ in the 20–30, 30–50, and 50–100 GeV $p_T(j)$ intervals, respectively, which include both the statistical and systematic uncertainties. As described in section 4.5, systematic uncertainties that affect both the efficiency measurement and the $c$-tagged data samples used in subsequent measurements are not included because these will largely cancel. Figure 6 displays the $c$-tagging efficiencies and compares the results obtained from the two $D$-meson decays separately. For the measurement presented in ref. [2], which involved integrating over $p_T(j)$, the relative $c$-tagging efficiency uncertainty is 6% including both the statistical and systematic contributions.

6 Summary

In summary, the identification of charm jets is achieved at LHCb in Run 2 using a method based on the properties of displaced vertices reconstructed within the jets. The performance of this method is determined using an unbiased dijet calibration dataset recorded by the LHCb detector during the same data-taking period. The charm-tagging efficiency in data, which is found to be consistent with simulation, is reported as a function of the transverse momentum of the jet, and found to be about 24% for $20 < p_T(j) < 100$ GeV and $2.2 < \eta(j) < 4.2$.

Acknowledgments

We express our gratitude to our colleagues in the CERN accelerator departments for the excellent performance of the LHC. We thank the technical and administrative staff at the LHCb institutes. We acknowledge support from CERN and from the national agencies: CAPES, CNPq, FAPERJ and FINEP.
Figure 6. Charm-tagging efficiency in intervals of $p_T$ determined from (blue triangles) $D^0 \rightarrow K^- \pi^+$ and (red squares) $D^+ \rightarrow K^- \pi^+ \pi^+$ decays, as well as (black circles) the weighted average. The points are offset in each $p_T$ interval to aid visibility.

(Brazil); MOST and NSFC (China); CNRS/IN2P3 (France); BMBF, DFG and MPG (Germany); INFN (Italy); NWO (Netherlands); MNiSW and NCN (Poland); MEN/IFA (Romania); MSHE (Russia); MICINN (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United Kingdom); DOE NP and NSF (U.S.A.). We acknowledge the computing resources that are provided by CERN, IN2P3 (France), KIT and DESY (Germany), INFN (Italy), SURF (Netherlands), PIC (Spain), GridPP (United Kingdom), RRCKI and Yandex LLC (Russia), CSCS (Switzerland), IFIN-HH (Romania), CBPF (Brazil), PL-GRID (Poland) and NERSC (U.S.A.). We are indebted to the communities behind the multiple open-source software packages on which we depend. Individual groups or members have received support from ARC and ARDC (Australia); AvH Foundation (Germany); EPLANET, Marie Skłodowska-Curie Actions and ERC (European Union); A*MIDEX, ANR, IPhU and Labex P2IO, and Région Auvergne-Rhône-Alpes (France); Key Research Program of Frontier Sciences of CAS, CAS PIFI, CAS CCEPP, Fundamental Research Funds for the Central Universities, and Sci. & Tech. Program of Guangzhou (China); RFBR, RSF and Yandex LLC (Russia); GVA, XuntaGal and GENCAT (Spain); the Leverhulme Trust, the Royal Society and UKRI (United Kingdom).

References


[7] LHCb collaboration, Measurement of differential $b\bar{b}$- and $c\bar{c}$-dijet cross-sections in the forward region of $pp$ collisions at $\sqrt{s} = 13$ TeV, JHEP 02 (2021) 023 [arXiv:2/zero.alt31/zero.alt3./zero.alt39437].


Università di Urbino, Urbino, Italy

MSU - Iligan Institute of Technology (MSU-IIT), Iligan, Philippines

P.N. Lebedev Physical Institute, Russian Academy of Science (LPI RAS), Moscow, Russia

Novosibirsk State University, Novosibirsk, Russia

Deceased