First observation of the decay $D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-$ in the $\rho^0-\omega$ region of the dimuon mass spectrum

LHCb Collaboration

1. Introduction

Rare charm decays may proceed via the highly suppressed $c \rightarrow u \mu^+ \mu^-$ flavour changing neutral current process. In the Standard Model such processes can only occur through loop diagrams, where in charm decays the GIM cancellation [1] is almost complete. As a consequence, the short-distance contribution to the inclusive $D \rightarrow X \mu^+ \mu^-$ branching fraction is predicted to be as low as $O(10^{-9})$ [2], making these decays interesting for searches for new physics beyond the Standard Model. However, taking into account long-distance contributions through tree diagrams involving resonances such as $D \rightarrow XV (\rightarrow \mu^+ \mu^-)$, where $V$ represents a $\phi$, $\rho^0$ or $\omega$ vector meson, the total branching fraction of these rare charm decays can reach $O(10^{-6})$ [2–4]. Their sensitivity to new physics therefore is greatest in regions of the dimuon mass spectrum away from these resonances, where the main contributions to the branching fraction may come from short-distance amplitudes. Angular asymmetries are sensitive to new physics both in the vicinity of these resonances and away from them [4–8] and could be as large as $O(1%)$.

This Letter focuses on the measurement of the decay\(^1\) $D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-$. This will provide an important reference channel for measurements of the $c \rightarrow u \mu^+ \mu^-$ processes $D^0 \rightarrow \pi^+ \mu^- \mu^-$ and $D^0 \rightarrow K^- K^- \mu^+ \mu^-$. Precise branching fractions are easier to obtain if they are compared with a normalisation mode that has similar features. When restricted to the dimuon mass range $675 < m(\mu^+ \mu^-) < 875$ MeV/c\(^2\), where the $\rho^0$ and $\omega$ resonances are expected to dominate, it can also be used to normalise the decays $D^0 \rightarrow K^- \pi^+ \eta(\rightarrow \mu^+ \mu^-)$. Measuring their branching fractions allows the coupling $\eta(\rightarrow \mu^+ \mu^-)$ to be determined. This contains crucial information for various low energy phenomena, and is an input to the prediction of the anomalous magnetic moment of the muon [9–11]. Focusing on this dimuon mass range also simplifies the analysis, which does not have to account for the variation of the selection efficiency as a function of $m(\mu^+ \mu^-)$. From previous measurements the most stringent 90% confidence level upper limits on the decay $D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-$ are set by the E791 experiment [12]: $B(D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-) < 35.9 \times 10^{-5}$ in the full $K^- \pi^+$ mass region and $B(D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-) < 2.4 \times 10^{-5}$ in the region of the $K^0$ resonance.

The study presented here is based on data collected by the LHCb detector in proton–proton collisions at a centre-of-mass energy of 8 TeV, corresponding to an integrated luminosity of 2.0 fb\(^{-1}\). A subsample corresponding to an integrated luminosity of 1.6 fb\(^{-1}\) has been used to measure $B(D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-)$. The remainder of the data set was used to optimise the selection. The branching fraction $B(D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-)$ is measured relative to that of the normalisation decay $D^0 \rightarrow K^- \pi^- \pi^+ \pi^-$. The most accurate recent measurement of this branching fraction is used, $B(D^0 \rightarrow K^- \pi^+ \pi^- \pi^-) = (8.287 \pm 0.043 \pm 0.200) \times 10^{-2}$, obtained by the CLEO experiment [13].

---

\(^1\) The inclusion of charge conjugate decays is implied.
2. Detector and simulation

The LHCb detector [14,15] is a single-arm forward spectrometer covering the pseudorapidity range 2 < η < 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 momentum, p, of charged particles with a relative uncertainty that varies from 0.5% at low momentum to 1.0% at 200 GeV/c. The minimum distance of a track to a primary vertex, the impact parameter (IP), is measured with a resolution of (15 + 29/pt) μm, where pt is the component of the momentum transverse to the beam, in GeV/c.

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 calorimeter and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers [16].

The online event selection is performed by a trigger [17], 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. In the offline selection, requirements are made on whether the trigger decision was due to the signal candidate or to other particles produced in the pp collision. Throughout this Letter, these two non-exclusive categories of candidates are referred to as Trigger On Signal (TOS) and Trigger Independent of Signal (TIS) candidates.

Simulated samples of $D^0 \rightarrow K^{-}\pi^{+}\mu^{+}\mu^{-}$ and $D^0 \rightarrow K^{-}\pi^{+}\pi^{+}\pi^{-}$ decays have been produced. In the simulation, pp collisions are generated using Pythia [18] with a specific LHCb configuration [19]. Decays of hadronic particles are described byEvtGen [20], in which final-state radiation is generated using Phiroto [21]. The interaction of the generated particles with the detector, and its response, are implemented using the GEANT4 toolkit [22] as described in Ref. [23]. No theoretical model or experimental measurement provides a reliable decay model for $D^0 \rightarrow K^{-}\pi^{+}\mu^{+}\mu^{-}$. This decay mode is therefore modelled as an incoherent sum of resonant and non-resonant contributions, such as $K_{S0}^0 \rightarrow K^{-}\pi^{+}$ and $\rho^{0} \rightarrow \omega \rightarrow \mu^{+}\mu^{-}$, motivated by the resonant structure observed in $D^0 \rightarrow K^{-}\pi^{+}\pi^{-}$ and $D^0 \rightarrow K^{-}\pi^{+}\pi^{+}\pi^{-}$ decays [24], and by the theoretical predictions of Ref. [4]. In the case of $D^0 \rightarrow K^{-}\pi^{+}\pi^{-}$, a decay model reproducing the data was implemented using the MINT software package [25].

3. Event selection

The criteria used to select the $D^0 \rightarrow K^{-}\pi^{+}\mu^{+}\mu^{-}$ and $D^0 \rightarrow K^{-}\pi^{+}\pi^{+}\pi^{-}$ decays are as similar as possible to allow many systematic uncertainties to cancel in the efficiency ratio. At trigger level, only events that are TIS with respect to the hadron hardware trigger, which has a transverse energy threshold of 3.7 GeV, are kept. In the offline selection, the only differences between the signal and normalisation channels are the muon identification criteria.

The first-level software trigger selects events that contain at least one good quality track with high $p_T$ and $\chi^2_\text{IP}$, where the latter is defined as the difference in $\chi^2$ of the closest primary pp interaction vertex (PV) reconstructed with and without the particle under consideration. The offline selection requires that at least one of these tracks originates from either the $D^0 \rightarrow K^{-}\pi^{+}\mu^{+}\mu^{-}$ or the $D^0 \rightarrow K^{-}\pi^{+}\pi^{+}\pi^{-}$ decay candidates. The second-level software trigger uses two dedicated selections to reconstruct $D^0 \rightarrow K^{-}\pi^{+}\mu^{+}\mu^{-}$ or $D^0 \rightarrow K^{-}\pi^{+}\pi^{+}\pi^{-}$ candidates originating from the PV. These combine good quality tracks that satisfy $p_T > 350$ MeV/c and $p > 3000$ MeV/c. A muon ($D^0 \rightarrow K^{-}\pi^{+}\mu^{+}\mu^{-}$) or charged hadron ($D^0 \rightarrow K^{-}\pi^{+}\pi^{+}\pi^{-}$) pair is required to form a good quality secondary vertex that is significantly displaced from the PV. In events where such a pair is found, two charged hadrons are subsequently added. The resulting four-particle candidate must have a good quality vertex and its invariant mass must be consistent with the known $D^0$ mass [24]. The momentum vector of this $D^0$ candidate must be consistent with having originated from the PV.

A preselection follows the trigger selections. Four charged particles are combined to form $D^0$ candidates. Tracks that do not correspond to actual trajectories of charged particles are suppressed by using a neural network optimisation procedure. To reject the combinatorial background involving tracks from the PV, only high-p and high-$p_T$ tracks that are significantly displaced from any PV are used. This background is further reduced by requiring that the four decay products of the D meson form a good quality vertex that is significantly displaced from the PV and that $p_T(D^0) > 3000$ MeV/c. These three criteria also reject candidates formed from partially reconstructed charm hadron decays, combined with either random tracks from the PV or with tracks from the decay of another charged hadron in the same event. This type of background is further reduced by requiring the $D^0$ momentum vector is within 14 mrad of the vector that joins the PV with the $D^0$ decay vertex, ensuring that the $D^0$ candidate originates from the PV. Finally, the invariant mass of the $D^0$ candidate, which is reconstructed with a resolution of about 7 MeV/$c^2$, is required to lie within 65 MeV/$c^2$ of the known $D^0$ mass. In the case of $D^0 \rightarrow K^{-}\pi^{+}\mu^{+}\mu^{-}$, $m(\mu^{+}\mu^{-})$ is restricted to the range 675–875 MeV/$c^2$. The two backgrounds described above are referred to as the non-peaking background throughout this Letter.

After the preselection, a multivariate selection based on a boosted decision tree (BDT) [26,27] is used to further suppress the non-peaking background. The GradBoost algorithm is used [28]. The BDT uses the following variables: the $p_T$ and $\chi^2_\text{IP}$ of the final state particles; the $p_T$ and $\chi^2_\text{IP}$ of the $D^0$ candidate as well as the $\chi^2$ per degree of freedom of its vertex fit; the significance of the distance between this vertex and the PV; the largest distance of closest approach between the tracks that form the $D^0$ candidate; the angle between the $D^0$ momentum vector and the vector that joins the PV with its decay vertex. The cut on the BDT response used in the selection discards more than 80% of the non-peaking candidates and retains more than 80% of the signal candidates that have passed the preselection.

Finally, the information from the RICH, the calorimeters and the muon systems are combined to assign probabilities for each decay product to be a pion, a kaon or a muon, as described in Ref. [15]. A loose requirement on the kaon identification probability rejects about 90% of the backgrounds that consist of $\pi^{-}\pi^{+}\mu^{+}\mu^{-}$ or $\pi^{-}\pi^{+}\pi^{+}\pi^{-}$ combinations while preserving 98% of the signal candidates. In the case of $D^0 \rightarrow K^{-}\pi^{+}\mu^{+}\mu^{-}$ decays, the muon identification criteria have an efficiency of 90% per signal muon and reduce the rate of misidentified pions by a factor of about 150. In the absence of muon identification, $D^0 \rightarrow K^{-}\pi^{+}\pi^{+}\pi^{-}$ decays with two misidentified pions would outnumber signal decays by four orders of magnitude. After these particle identification requirements, this background is reduced to around 50% of the signal yield and is dominated by decays involving two pion decays in...
flight \((\pi^+ \rightarrow \mu^+\nu_\mu)\). It is referred to as the peaking background throughout this Letter.

In addition to \(D^0 \rightarrow K^-\pi^+\pi^+\pi^-\) decays with two misidentified pions, backgrounds due to the decays of \(D^+, D^+_s, D^{*+}, \tau, A^{\pm}\), and \(\Sigma^0\) are considered. These are studied using simulated events and found to be negligible.

The selection is optimised using data and simulated samples. The BDT is trained using simulated \(D^0 \rightarrow K^-\pi^+\mu^+\mu^-\) events to model the signal. The sample used to represent the background consists of candidates with \(m(K^+\pi^-\mu^+\mu^-) > 1890\text{ MeV}/c^2\), drawn from 2\% of the total data sample. Candidates on the low-mass side of the signal peak are not used due to the presence there of peaking background decays, whose features are very close to those of signal decays. Optimal selection criteria on the BDT response and muon identification are found using another independent data sample corresponding to 20\% of the total dataset. The fit described in Sect. 4 is used to estimate the yields of \(D^0 \rightarrow K^-\pi^+\mu^+\mu^-\) signal \((S)\), peaking background \((B_{\text{pk}})\), and non-peaking background \((B_{\text{npk}})\) present in this sample in the region of the signal peak, defined as \(1840 < m(K^+\pi^-\mu^+\mu^-) < 1890\text{ MeV}/c^2\). The requirements on the muon identification and BDT response are chosen to maximise \(S/\sqrt{S+B_{\text{pk}}+B_{\text{npk}}}\).

The two samples described above consist of events chosen randomly from the 2012 data and are not used for the subsequent analysis. The remainder of the dataset (78\%), which corresponds to an integrated luminosity of 1.6 fb\(^{-1}\), is used to measure \(B(D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\). The final \(D^0 \rightarrow K^-\pi^+\mu^+\mu^-\) sample obtained with this selection consists of 5411 candidates. In the case of \(D^0 \rightarrow K^-\pi^+\pi^-\pi^-\), the large value of \(B(D^0 \rightarrow K^-\pi^+\pi^-\pi^-)\) allows us to use a small sample (3 pb\(^{-1}\)), drawn randomly from the total dataset. The final \(D^0 \rightarrow K^-\pi^+\pi^-\pi^-\) sample consists of 121922 candidates.

4. Determination of the \(D^0 \rightarrow K^-\pi^+\mu^+\mu^-\) and \(D^0 \rightarrow K^-\pi^+\pi^-\pi^-\) yields

A simultaneous binned maximum likelihood fit to the \(m(K^-\pi^+\mu^+\mu^-)\) and \(m(K^-\pi^+\pi^-\pi^-)\) distributions is performed to measure \(B(D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\).

In each sample, the probability density function (PDF) fitted to the signal peak is a Gaussian function with power law tails. It is defined in the following way:

\[
\begin{align*}
    f(m; m_{\text{p0}}, \sigma, \alpha_L, n_L, \alpha_R, n_R) &= \frac{1}{\sqrt{2\pi}a}\left(\frac{1}{m-m_{\text{p0}}}\right)^{\frac{1}{2}} e^{-\frac{(m-m_{\text{p0}})^2}{2\sigma^2}} \quad \text{if } m-m_{\text{p0}} \leq -|\alpha_L|, \\
    &\quad \frac{1}{\sqrt{2\pi}a}\left(\frac{1}{m-m_{\text{p0}}}\right)^{\frac{1}{2}} e^{-\frac{(m-m_{\text{p0}})^2}{2\sigma^2}} \quad \text{if } m-m_{\text{p0}} \geq |\alpha_L|, \\
    &\quad \exp\left(-\frac{(m-m_{\text{p0}})^2}{2\sigma^2}\right) \quad \text{otherwise},
\end{align*}
\]

where \(m_{\text{p0}}\) and \(\sigma\) are the mean and width of the peak, and \(\alpha_L, n_L, \alpha_R, n_R\) parameterise the left and right tails. This function was found to describe accurately the \(m(K^-\pi^+\mu^+\mu^-)\) and \(m(K^-\pi^+\pi^-\pi^-)\) distributions obtained with the simulation, which exhibit non-Gaussian tails on both sides of the peaks. The tail on the left-hand side is dominated by final-state radiation and interactions with matter, while the right-hand side tail is due to non-Gaussian effects in the reconstruction.

The non-peaking background in the \(D^0 \rightarrow K^-\pi^+\pi^+\pi^-\) sample is described by a first-order polynomial. In the case of \(D^0 \rightarrow K^-\pi^+\mu^+\mu^-\), a second-order polynomial is used.

Three peaking backgrounds due to misidentified \(D^0 \rightarrow K^-\pi^+\pi^+\pi^-\) decays are categorised by the presence of candidates involving misidentified pions that did not decay in flight before reaching the most downstream tracking stations, or candidates where one or two pions decayed upstream of those tracking stations. Candidates from the first category are described by a one-dimensional kernel density estimate [29]. This PDF is derived from the \(m(K^-\pi^+\mu^+\mu^-)\) distribution obtained using simulated \(D^0 \rightarrow K^-\pi^+\pi^+\pi^-\) decays reconstructed under the \(D^0 \rightarrow K^-\pi^+\pi^+\pi^-\) hypothesis. Candidates from the remaining two categories appear as tails on the lower-mass side of the \(m(K^-\pi^+\mu^+\mu^-)\) distribution and must be accounted for to avoid biases in the non-peaking background and in the signal yield measured by the fit. Due to the small number of such candidates in the simulated sample, simulated \(D^0 \rightarrow K^-\pi^+\pi^+\pi^-\) candidates where no pion decays in flight are altered to reproduce the effect of such decays, and the corresponding \(m(K^-\pi^+\mu^+\mu^-)\) distribution is determined. This is achieved by modifying the momentum vectors of either one or two of the pions present in the \(D^0 \rightarrow K^-\pi^+\pi^+\pi^-\) final state according to the kinematics of \(\pi^+ \rightarrow \mu^+\nu_\mu\) decays. The \(m(K^-\pi^+\mu^+\mu^-)\) distributions obtained after this modification are converted into one-dimensional kernel density estimates.

The fit model involves 5 yields: the signal yield, \(N_{\text{sig}}\), the yield of normalisation decays, \(N_{D^0 \rightarrow K^-\pi^+\pi^-\pi^-}\), the peaking and non-peaking background yields, \(N_{\text{pk}}\) and \(N_{\text{npk}}\), and the yield of background candidates in the \(D^0 \rightarrow K^-\pi^+\pi^-\pi^-\) sample, \(N_{D^0 \rightarrow K^-\pi^+\pi^-\pi^-}\). They are all free parameters in the fit. It also involves 15 parameters to define the shapes of the PDFs. The parameters describing the widths and upper-mass tails are free parameters in the fit but are common between the PDFs for the \(D^0 \rightarrow K^-\pi^+\pi^-\pi^-\) and \(D^0 \rightarrow K^-\pi^+\pi^-\pi^-\) peaks. The lower-mass tail parameters are determined separately. Those used for \(D^0 \rightarrow K^-\pi^+\pi^-\pi^-\) candidates are allowed to vary in the fit. This is not possible for \(D^0 \rightarrow K^-\pi^+\pi^-\pi^-\) candidates because of the overlap between the signal and the \(D^0 \rightarrow K^+\pi^-\pi^-\pi^-\) peaking background and therefore the parameters are fixed to the values obtained from the simulated sample. In total, there are 15 free parameters in the fit.

The relative yields of the three peaking background categories described above are fixed to values obtained by a fit to a large control sample. It consists of \(D^0 \rightarrow K^-\pi^+\mu^+\mu^-\) candidates that are in the TOS category with respect to the muon hardware trigger, in contrast to the signal and normalisation samples that are in the TIS category with respect to the hadron trigger. All of the other selection requirements are the same as those described in Sect. 3. This TOS signal control sample consists of 28835 candidates and contains approximately six times more \(D^0 \rightarrow K^-\pi^+\pi^-\pi^-\) than the nominal TIS sample.

The fit results are summarised in Table 1 and the observed mass distributions are shown in Fig. 1, with fit projections overlaid. The main difficulties in this procedure are the similarities in the shape of the signal, peaking background and non-peaking background, and the overlap between their distributions in \(m(K^-\pi^+\mu^+\mu^-)\). However, their impact on the measurement presented in this Letter is limited, as can also be seen in Table 1.

5. Branching fraction measurement

The branching fraction of the decay \(D^0 \rightarrow K^-\pi^+\mu^+\mu^-\) is obtained by combining the quantities presented in Table 2 with the branching fraction of the \(D^0 \rightarrow K^-\pi^+\pi^-\pi^-\) decay according to
Table 1

Summary of the results of the fit described in Sect. 4. The yields measured in the \( D^0 \to K^-π^+μ^+γ^- \) sample and the correlations between them, the yields measured in the normalisation sample, the common width fitted to the \( D^0 \to K^-π^+μ^- \) and \( D^0 \to K^-π^+π^- \) yields, and the relative uncertainty on \( B(D^0 \to K^-π^+μ^-) \) are presented. Uncertainties on the fitted parameters are statistical. The variation of the uncertainty on \( B(D^0 \to K^-π^+μ^-) \) when the background yields are fixed indicates to what extent it is enhanced by the need to separate contributions in overlap and which shapes present some similarities.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_{ggg} )</td>
<td>2357 ± 67</td>
</tr>
<tr>
<td>( N_{gk} )</td>
<td>1047 ± 84</td>
</tr>
<tr>
<td>( N_{gkk} )</td>
<td>2007 ± 116</td>
</tr>
<tr>
<td>( \Delta N_{ggg}^{nππ} )</td>
<td>83575 ± 334</td>
</tr>
<tr>
<td>( \Delta N_{gk}^{nππ} )</td>
<td>38346 ± 257</td>
</tr>
<tr>
<td>( \sigma_{N_{ggg}} )</td>
<td>7.17 ± 0.03 MeV/c²</td>
</tr>
<tr>
<td>( \sigma_{N_{gk}} )</td>
<td>−78%</td>
</tr>
<tr>
<td>( C_{N_{ggg},N_{gk}} )</td>
<td>27%</td>
</tr>
<tr>
<td>( \sigma_{B(D^0 \to K^-π^+μ^-)} )</td>
<td>−48%</td>
</tr>
<tr>
<td>( \sigma_{B(D^0 \to K^-π^+μ^-)} ), if ( N_{ggg} ) fixed</td>
<td>2.9%</td>
</tr>
<tr>
<td>( \sigma_{B(D^0 \to K^-π^+μ^-)} ), if ( N_{ggg} ) and ( N_{gk} ) fixed</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

Fig. 1. Mass distributions of (a) \( D^0 \to K^-π^+μ^- \) and (b) \( D^0 \to K^-π^+π^-μ^- \) candidates. The data are shown as points (black) and the total PDF (blue solid line) is overlaid. In (a), the two corresponding components of the fit model are the \( D^0 \to K^-π^+μ^- \) decays (red solid line) and the non-peaking background (violet dashed line). In (b), the components are the \( D^0 \to K^-π^+μ^- \) (long-dashed green line), the peaking background due to misidentified \( D^0 \to K^-π^+π^-μ^- \) decays (red solid line), and the non-peaking background (violet dashed line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

\[
B(D^0 \to K^-π^+μ^-) = \frac{N_{D^0 \to K^-π^+μ^-}}{N_{D^0 \to K^-π^+π^-}} \times \frac{\varepsilon_{D^0 \to K^-π^+π^-}}{\varepsilon_{D^0 \to K^-π^+μ^-}} \times B(D^0 \to K^-π^+μ^-),
\]

where \( N_{D^0 \to K^-π^+π^-} \), \( N_{D^0 \to K^-π^+μ^-} \), \( \varepsilon_{D^0 \to K^-π^+π^-} \), and \( \varepsilon_{D^0 \to K^-π^+μ^-} \) are the yields and selection efficiencies for the signal and normalisation decays. The branching fraction of the signal decay for dimuon invariant masses in the range 675–875 MeV/c² is measured to be \( B(D^0 \to K^-π^+μ^-) = (4.17 ± 0.12) \times 10^{-6} \), where the uncertainty is statistical.

5.1. Systematic uncertainties

The systematic uncertainties on \( B(D^0 \to K^-π^+μ^-) \) are summarised in Table 3. Those related to reconstruction and selection efficiencies are minimised thanks to the efficiency ratio in Eq. (1) and to the similarities between \( D^0 \to K^-π^+μ^- \) and \( D^0 \to K^-π^+π^-μ^- \) decays. This is illustrated in Fig. 2, which shows the distributions of the BDT response for the \( D^0 \to K^-π^+μ^-μ^- \) and \( D^0 \to K^-π^+π^-μ^- \) decays, both in data and simulated samples. In data, the background contributions are removed using the sPlot technique [30]. Also shown in this figure are the ratios between the \( D^0 \to K^-π^+μ^-μ^- \) and \( D^0 \to K^-π^+π^-μ^- \) distributions. The BDT response, which combines all the offline selection variables (with the exception of muon identification criteria), is very similar for both kinds of decay and the differences are well described by the simulation. In cases where selection criteria depend on the nature of the decay products, data-driven methods are used, as described below.

The uncertainty on the charged hadron reconstruction efficiency is dominated by the uncertainty on the probability to undergo a nuclear interaction in the detector. This inefficiency is evaluated using simulated events. The corresponding uncertainty is derived from the 10% uncertainty on the modelling of the detector material [31].

The selection efficiencies based on the kinematical and geometrical requirements are derived from simulation. A systematic uncertainty to take into account imperfect track reconstruction...
modelling is estimated by smearing track properties to reproduce those observed in data. Similarly, a systematic uncertainty on the efficiency of the BDT selection is assigned as the difference between the efficiency obtained in data and simulation.

The uncertainties in the decay models are estimated separately for the signal and normalisation channels. For the signal, this is carried out by reweighting simulated $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ decays to reproduce the distributions of $m_{K^-\pi^+}$ and $m_{\mu^+\mu^-}$ observed in data, with the difference in efficiency relative to the default being assigned as the systematic uncertainty. For $D^0 \rightarrow K^-\pi^+\pi^-\pi^-$, the sensitivity to the decay model is studied by comparing the default efficiency with that obtained in an extreme case in which the decay model provided by the MINT package is replaced by an incoherent sum of the resonances involved in the decay, as given in Ref. [24].

To avoid dependence on the modelling of the hardware trigger in simulation, its efficiency is determined in data. The efficiency to be TIS with respect to hadron hardware trigger is determined as the fraction of $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ decays that fulfil this requirement among $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ candidates that are TOS with respect to the muon hardware trigger. It is measured in 12 different regions defined in the $(p_T(D^0), N_t)$ plane, where $N_t$ is the track multiplicity of the event. The overall hardware trigger efficiency for $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ decays is the average of these 12 efficiencies weighted according to the distributions of $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ candidates observed in data. The efficiency of the normalisation mode is obtained by weighting the same 12 efficiencies according to the distributions of $D^0 \rightarrow K^-\pi^+\pi^-\pi^-$ candidates. This procedure assumes that the probability for $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ decays to fulfil the TIS requirement is not enhanced by the requirement to also be in the TOS category and that this TIS efficiency is the same in every region for $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ and $D^0 \rightarrow K^-\pi^+\pi^-\pi^-$ decays. No difference is found in simulation between the $e_{\text{TIS}}$ and $e_{\text{TOS}}$ ratio obtained with this method and the ratio of true efficiencies, obtained by directly counting the number of simulated $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ and $D^0 \rightarrow K^-\pi^+\pi^-\pi^-$ decays that fulfil the hadron trigger TIS requirement. To determine the systematic uncertainty associated with the hardware trigger efficiency, the uncertainty on this comparison is combined with the statistical uncertainties on the 12 measurements performed in data in $(p_T(D^0), N_t)$ regions.

A similar approach is employed in the case of the first level of the software trigger. A sample of $D^0 \rightarrow K^-\pi^+\pi^-\pi^-$ candidates is selected from data that satisfied the trigger requirements independently of these candidates. The fraction of $D^0 \rightarrow K^-\pi^+\pi^-\pi^-$ decays where at least one of the decay products also satisfies the requirements of this trigger is measured using this sample. This efficiency is measured in regions of $p_T(D^0)$ and weighted according to distributions of this variable in simulated $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ and $D^0 \rightarrow K^-\pi^+\pi^-\pi^-$ events. The variation in the efficiency ratio when these distributions are corrected to match the data is used to evaluate the corresponding systematic uncertainty.

The efficiency of the second-level software trigger for the signal decay is calculated relative to that of the normalisation decay. This ratio is measured using $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ decays in data and simulation and consistent results are obtained. The uncertainty on this comparison is therefore assigned as the systematic uncertainty on this trigger efficiency.

The efficiency of the muon identification criteria is determined in data using a large and pure sample of $B \rightarrow J/\psi(\rightarrow \mu^+\mu^-)X$. Efficiencies measured in several regions of $p_T(\mu)$, $\eta(\mu)$ and $N_t$ are weighted according to the distribution observed for the muon candidates from $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ decays. Several definitions of these domains are considered, with varying binnings. The different efficiencies obtained this way, as well as the efficiencies obtained in simulated samples, are compared to evaluate the corresponding systematic uncertainty. The same approach is used to evaluate the efficiency of the kaon identification requirement. In this case, the calibration kaons are provided by $D^0 \rightarrow K^-\pi^+(\rightarrow K^\pi\pi\pi)\pi^\pm$ decays in data.

In the fit outlined in Sect. 4, the parameters of the function that describe the lower-mass tail of the $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ peak are fixed to values obtained from simulation. The corresponding systematic uncertainty is determined by repeating the fit using the values obtained by a fit to the signal TOS control sample. A similar difference is observed when the corresponding test is performed for $D^0 \rightarrow K^-\pi^+\pi^-\pi^-$ candidates.

The systematic uncertainty related to the description of the peaking background is determined by the change observed in $B(D^0 \rightarrow K^-\pi^+\mu^+\mu^-)$ when the components due to the decay of one or two pions in flight are neglected, and when their yields relative to the rest of the peaking background are enhanced by twice their uncertainty.

Two other systematic uncertainties have been evaluated. To estimate the impact of the signal PDF employed, the fit is repeated using the Cudiff function [32] instead. Potential effects arising from non-peaking backgrounds are assessed by repeating the fits with the non-peaking backgrounds assumed to be linear in $m(K^-\pi^+\mu^+\mu^-)$. The values of the systematic uncertainties associated with the choice of fit model and its parameters were also further validated using pseudoexperiments.

The impact on the fit of the similarities between the shapes of the signal and background components was further controlled in two ways. First, fixing the background yields decreases the relative uncertainty on $B(D^0 \rightarrow K^-\pi^+\mu^+\mu^-)$ from 2.9% to 2.4%. This variation is far lower than the total systematic uncertainty due to the yield determination (2.8%). Moreover, another study is performed based on pseudoexperiments, generated with realistic values of the yields and PDFs shape parameters. The fit proved able to return unbiased measurements of the generated value of $B(D^0 \rightarrow K^-\pi^+\mu^+\mu^-)$ and an accurate estimation of the statistical uncertainty, consistent with the uncertainty obtained in data.
As can be seen in Table 3, the systematic uncertainties are dominated by the uncertainty on the $D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-$ to $D^0 \rightarrow K^- \pi^+ \pi^- \pi^-$ efficiency ratio, which is larger than the 2.9% statistical uncertainty on $B(D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-)$. As expected, this systematic uncertainty is primarily due to the different final state particles of the two decays. The trigger efficiencies, and the muon identification and track reconstruction efficiencies, are responsible for about 90% of this uncertainty. The uncertainties due to the yield determination and the knowledge of $B(D^0 \rightarrow K^- \pi^+ \pi^- \pi^-)$ represent secondary contributions.

6. Conclusions

The decay $D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-$ is studied using proton–proton collision data corresponding to an integrated luminosity of 2.0 fb$^{-1}$ collected in 2012 by the LHCb detector at a centre-of-mass energy of 8 TeV. The branching fraction of the decay $D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-$ in the dimuon mass range 675–875 MeV/$c^2$ is measured to be

\[
B(D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-) = (4.17 \pm 0.12{\text{(stat)}} \pm 0.40{\text{(syst)}}) \times 10^{-6}.
\]

This branching fraction can be compared to the Standard Model value calculated in Ref. [4], $B(D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-) = 6.7 \times 10^{-6}$, in the full dimuon mass range. This is the first observation of this decay. The branching fraction is measured with an overall precision of 10% and is one order of magnitude lower than the previous most stringent upper limit. Precise measurements of the $D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-$ and $D^0 \rightarrow K^- K^- \mu^+ \mu^-$ decays are now possible in all regions of the dimuon invariant mass since they can be compared with a normalisation mode that has similar features and a precisely known branching fraction. This will allow more stringent constraints on new physics to be obtained using data already collected by the LHCb detector, and the sensitivity of future experiments to angular asymmetries to be assessed.

The distributions of the $K^- \pi^+$ and $\mu^+ \mu^-$ invariant masses in $D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-$ decays are shown in Fig. 3, where the background contribution is removed using the sPlot technique [30], taking the $m(K^- \pi^+ \mu^+ \mu^-)$ invariant mass as the discriminating variable. An amplitude analysis would be required for a full understanding of the decay dynamics. The distributions in Fig. 3 suggest the presence of additional contributions, including the $\omega$ resonance, beyond the $F_{\pi}^0 \varphi_{\pi}^0$ intermediate state that, according to Ref. [4], should strongly dominate the decay amplitude.

Acknowledgements

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 (Brazil); NSFC (China); CNRS/IN2P3 (France); BMBF, DFG and MPG (Germany); INFN (Italy); FOM and NWO (The Netherlands); MNiSW and NCN (Poland); MEN/IFA (Romania); MinES and FANO (Russia); MiniEco (Spain); SNSF and SER (Switzerland); STFC (United Kingdom); NSF (USA). We acknowledge the computing resources that are provided by CERN, IN2P3 (France), KIT and DESY (Germany), INFN (Italy), SURF (The Netherlands), PIC (Spain), GridPP (United Kingdom), RRCCKI (Russia), CSCS (Switzerland), IFIN–HH (Romania), CBPF (Brazil), PL–GRID (Poland) and OSC (USA). We are indebted to the communities behind the multiple open source software packages on which we depend. We are also thankful for the computing resources and the access to software R&D tools provided by Yandex LLC (Russia). Individual groups or members

References

LHCb Collaboration

35 Budker Institute of Nuclear Physics (SB RAS), Russia
36 Institute for High Energy Physics (BHEP), Protvino, Russia
37 Universitat de Barcelona, Barcelona, Spain
38 Universidad de Santiago de Compostela, Santiago de Compostela, Spain
39 European Organization for Nuclear Research (CERN), Geneva, Switzerland
40 Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
41 Physik-Institut, Universität Zürich, Zürich, Switzerland
42 Nikhef National Institute for Subatomic Physics, Amsterdam, The Netherlands
43 Nikhef National Institute for Subatomic Physics and VU University Amsterdam, Amsterdam, The Netherlands
44 NSC Kharkiv Institute of Physics and Technology (NSC KIPT), Kharkiv, Ukraine
45 Institute for Nuclear Research of the National Academy of Sciences (KINR), Kyiv, Ukraine
46 University of Birmingham, Birmingham, United Kingdom
47 H.H. Wills Physics Laboratory, University of Bristol, Bristol, United Kingdom
48 Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
49 Department of Physics, University of Warwick, Coventry, United Kingdom
50 STFC Rutherford Appleton Laboratory, Didcot, United Kingdom
51 School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
52 School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom
53 Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom
54 Imperial College London, London, United Kingdom
55 School of Physics and Astronomy, University of Manchester, Manchester, United Kingdom
56 Department of Physics, University of Oxford, Oxford, United Kingdom
57 Massachusetts Institute of Technology, Cambridge, MA, United States
58 University of Cincinnati, Cincinnati, OH, United States
59 University of Maryland, College Park, MD, United States
60 Syracuse University, Syracuse, NY, United States
61 Pontificia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil
62 Institute of Particle Physics, Central China Normal University, Wuhan, Hubei, China
63 Departamento de Física, Universidad Nacional de Colombia, Bogotá, Colombia
64 Institut für Physik, Universität Rostock, Rostock, Germany
65 National Research Centre Kurchatov Institute, Moscow, Russia
66 Yandex School of Data Analysis, Moscow, Russia
67 Instituto de Física Corpuscular (IFIC), Universitat de València-CSIC, Valencia, Spain
68 Van Swinderen Institute, University of Groningen, Groningen, The Netherlands
69 Novosibirsk State University, Novosibirsk, Russia

* Corresponding author. Tel.: +331 64 46 8319.
E-mail address: viaud@lal.in2p3.fr (B. Viaud).

 Universidad Federal do Triângulo Mineiro (UFTM), Uberaba-MG, Brazil.
† P.N. Lebedev Physical Institute, Russian Academy of Science (LPI RAS), Moscow, Russia.

Università di Bari, Bari, Italy.

Università di Bologna, Bologna, Italy.

Università di Cagliari, Cagliari, Italy.

Università di Ferrara, Ferrara, Italy.

Università di Urbino, Urbino, Italy.

Università di Modena e Reggio Emilia, Modena, Italy.

Università di Genova, Genova, Italy.

Università di Milano Bicocca, Milano, Italy.

Università di Roma Tor Vergata, Roma, Italy.

Università di Roma La Sapienza, Roma, Italy.

Università della Basilicata, Potenza, Italy.

AGH – University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Kraków, Poland.

LIFAELS, La Salle, Universitat Ramon Llull, Barcelona, Spain.

Hanoi University of Science, Hanoi, Viet Nam.

Università di Padova, Padova, Italy.

Università di Pisa, Pisa, Italy.

Scuola Normale Superiore, Pisa, Italy.

Università degli Studi di Milano, Milano, Italy.

Associated to Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

Associated to Center for High Energy Physics, Tsinghua University, Beijing, China.

Associated to LPNHE, Université Pierre et Marie Curie, Université Paris Diderot, CNRS/IN2P3, Paris, France.

Associated to Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany.

Associated to Institute of Theoretical and Experimental Physics (ITEP), Moscow, Russia.

Associated to Universitat de Barcelona, Barcelona, Spain.

Associated to Nikhef National Institute for Subatomic Physics, Amsterdam, The Netherlands.

Deceased.