First observation of the decay
$B^0 \rightarrow D^0 \bar{D}^0 K^+ \pi^-$

LHCb collaboration†

Abstract

The first observation of the decay $B^0 \rightarrow D^0 \bar{D}^0 K^+ \pi^-$ is reported using proton-proton collision data corresponding to an integrated luminosity of 4.7 fb$^{-1}$ collected by the LHCb experiment in 2011, 2012 and 2016. The measurement is performed in the full kinematically allowed range of the decay outside of the $D^{*-}$ region. The ratio of the branching fraction relative to that of the control channel $B^0 \rightarrow D^{*-}D^0 K^+$ is measured to be $R = (14.2 \pm 1.1 \pm 1.0)\%$, where the first uncertainty is statistical and the second is systematic. The absolute branching fraction of $B^0 \rightarrow D^0 \bar{D}^0 K^+ \pi^-$ decays is thus determined to be $B(B^0 \rightarrow D^0 \bar{D}^0 K^+ \pi^-) = (3.50 \pm 0.27 \pm 0.26 \pm 0.30) \times 10^{-4}$, where the third uncertainty is due to the branching fraction of the control channel.

Submitted to Phys. Rev. Lett.

© 2020 CERN for the benefit of the LHCb collaboration. CC BY 4.0 licence

†Authors are listed at the end of this paper.
The family of $B \to D^{(*)} \bar{D}^{(*)} K$ and $B \to D^{(*)} \bar{D}^{(*)} K \pi$ decays, each with two charm hadrons and a kaon in the final state, proceed at quark level through Cabibbo-Kobayashi-Maskawa favoured $b \to c \bar{s}s$ transitions. These transitions occur with either an external or internal $W$ emission process, as shown in Fig. 1, offering the opportunity to search for new $c\bar{s}$ or $c\bar{c}$ states. In addition, measurements of the amplitude structure of the $D^{(*)}\bar{D}^{(*)}$ system in these processes can provide important information to calculations of the $c\bar{c}$ contribution above the open-charm threshold in $b \to s\ell^+\ell^-$ decays \[1\]. There is considerable debate whether the theoretical uncertainties associated with these long-distance contributions \[2–5\] could alleviate the tensions in a wide range of measurements involving $b \to s\ell^+\ell^-$ transitions \[6 –16\] with Standard Model predictions. Therefore, measurements that can provide input to these calculations are of the utmost importance.

Although measurements involving $B \to D^{(*)} \bar{D}^{(*)} K$ decays have been performed by the ALEPH, BaBar, Belle and LHCb collaborations \[17 –22\], no measurements involving $B \to D^{(*)} \bar{D}^{(*)} K \pi$ transitions have been performed to date. The $B^0 \to D^0 \bar{D}^0 K^+\pi^-$ branching fraction, based on considerations of similar decay modes, is expected to be $\mathcal{O}(10^{-4})$, but the product of the branching fractions including the $D^0 \to K^-\pi^+$ charm meson decays is much smaller, at the level of $\mathcal{O}(10^{-7})$.

This Letter presents the first observation of the $B^0 \to D^0 \bar{D}^0 K^+\pi^-$ decay, excluding contributions from $B^0 \to D^{\ast-} D^0 K^+$ transitions, with $D^{\ast-} \to \bar{D}^0 \pi^-$ decays.\[1\] The branching fraction of this decay is measured in the full kinematically allowed range of the decay outside of the $D^{\ast-}$ region, relative to the control mode $B^0 \to D^{\ast-} D^0 K^+$. After the decay of the $D^{\ast-}$ meson via the strong interaction, signal and control modes present the same final-state particles $D^0 \bar{D}^0 K^+\pi^-$. The measurement is performed using data collected with the LHCb detector in proton-proton collisions at centre-of-mass energies of 7 TeV and 8 TeV during Run 1, and 13 TeV during 2016. The corresponding integrated luminosities for the years 2011, 2012 and 2016 are 1.0, 2.0 and 1.7fb$^{-1}$, respectively.

The LHCb detector \[23,24\] is a single-arm forward spectrometer covering the pseudo-rapidity range $2 < \eta < 5$, designed for the study of particles containing $b$ or $c$ quarks. The detector elements that are particularly relevant to this analysis are: a silicon-strip vertex detector surrounding the $pp$ interaction region \[25\] that allows $c$ and $b$ hadrons to be identified from their characteristically long flight distance; a tracking system that provides a measurement of the momentum, $p$, of charged particles \[26,27\]; and two ring-imaging Cherenkov detectors that are able to discriminate between different species of charged hadrons \[28\]. Photons, electrons and hadrons are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic and a hadronic

\[\begin{align*}
B^0 & \xrightarrow{\bar{b}} \bar{c} & D^+_s \quad D^{+}_s \quad \bar{s} \\
B^0 & \xrightarrow{d} \bar{u} & \bar{D}^0 \quad \bar{c} & \quad W^+ \end{align*}\]

\[\begin{align*}
B^0 & \xrightarrow{d} \bar{D}^0 \quad \bar{s} & \quad W^+ \quad \bar{c} & \quad K^{*0} \\
B^0 & \xrightarrow{d} \bar{D}^0 \quad \bar{s} & \quad W^+ \quad \bar{c} & \quad K^{*0}
\end{align*}\]

Figure 1: Feynman diagrams of the external (left) and internal (right) $W$ emission contributing to $B^0 \to D^0 \bar{D}^0 K^+\pi^-$ decays.

\[\text{\[1\]The inclusion of charge-conjugate processes is implied throughout this paper unless otherwise noted.}\]
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. Events retained following the hardware trigger decision are split into two independent categories, those with a positive decision based on activity in the hadronic calorimeter associated with the signal candidate decay and those based on signatures from other particles in the event. The data are further split into two data taking categories for the Run 1 and 2016 samples. The software trigger stage requires a two-, three- or four-track secondary vertex with a significant displacement from any primary $pp$ interaction vertex (PV).

Simulation is required to model the effects of the detector acceptance and the imposed selection requirements. It is also used to train multivariate classifiers for background suppression, and to obtain the shape of the invariant-mass distribution for candidate $B^0$ hadrons. In the simulation, $pp$ collisions are generated using PYTHIA [29] with a specific LHCb configuration [30]. Decays of unstable particles are described by EVTGEN [31,32], in which final-state radiation is generated using PHOTOS [33]. The interaction of the generated particles with the detector, and its response, are implemented using the GEANT4 toolkit [34] as described in Ref. [35].

The simulated samples of the signal- and control-mode decays are corrected to improve agreement with the data. A fit to the $B^0$ candidate invariant-mass distribution of the $B^0 \rightarrow D^{*-}D^0K^+$ sample is performed using the $s$Plot technique [36] to calculate weights that statistically remove background contributions. Subsequently, a correction to the simulation is derived as a function of event track multiplicity and impact parameter significance of the $B^0$ candidate with respect to the associated PV, by comparing $B^0 \rightarrow D^{*-}D^0K^+$ candidates in simulation and background-subtracted data. In addition, the particle identification (PID) variables in the simulation are corrected using control data samples with the MEERKAT software package [37,38].

The $D^0$ ($\bar{D}^0$) candidates are reconstructed in the $K^-\pi^+ (K^+\pi^-)$ final state, in a $\pm 30$ MeV/c$^2$ window around the known mass [39]. The $K^+\pi^-$ candidates originating directly from the $B^0$ decay are required to have an invariant mass below 1600 MeV/c$^2$ and are subsequently combined with the charm mesons to form the $B^0$ candidates.

The selection comprises two stages. First, a loose selection is applied that relies on PID criteria to correctly identify charged kaons and pions, and on the flight distance significance of the $D^0$ candidates to reject charmless backgrounds. The signal and control mode data samples are then split using the requirement $|m(D^0\pi^-) - m(D^0) - |m_0(D^{*-}) - m_0(D^0)|| < (4 \times 0.724)$ MeV/c$^2$ to select candidates consistent with the $B^0 \rightarrow D^{*-}D^0K^+$ hypothesis, where $m_0$ is the known mass of the particle [39] and 0.724 MeV/c$^2$ is the resolution of the $D^{*-}$ contribution. To improve the mass resolution a global kinematic fit is performed constraining the mass of the $D^0$ mesons to its known value [40]. In this kinematic fit the $B^0$ candidate is also constrained to originate from the associated PV.

The second selection stage relies on two neural networks: one to identify good-quality $D^0$ candidates from $B^0$ meson decays (NN$_D$); and another to reduce the combinatorial background, which consists of candidates constructed from one or two random tracks in place of the $K^+$ and $\pi^-$ from the $B^0$ meson decay (NN$_B$). A multilayer perceptron model is used, implemented using the KERAS library [41] in the TENSORFLOW [42] framework. These classifiers are trained separately for the Run 1 and 2016 data-taking periods and
for each trigger category. The training and testing is performed using the $k$-fold cross validation technique with $k = 10$ [43]. Simulated samples are used as a signal proxy and data from the sidebands of the $D^0$ or $B^0$ candidate invariant-mass distributions as the background proxy. Specifically, these are candidates outside of a $\pm 40\text{MeV}/c^2$ window around the known $D^0$-meson mass [39], for the $N_{D}$ classifier and candidates satisfying $m(D^0 \bar{D}^0 K^+ \pi^-) > m_0(B^0) + 100\text{MeV}/c^2$ for $N_{B}$.

The $N_{D}$ classifier is trained using fourteen variables including PID information as well as kinematic properties and the decay topology of the tracks and $D^0$ meson candidate. Fourteen variables are also used to train the $N_{B}$ classifier, including the output of the two $N_{D}$ classifiers and other observables describing the topology and kinematics of the $B^0$ meson decay. As the $N_{D}$ classifier is an input to the $N_{B}$ classifier, a requirement is only placed on the output of the $N_{B}$ classifier. This threshold is optimised by maximising the figure of merit $\frac{N_{S}}{\sqrt{N_{S} N_{B}}}$. Here $N_{S}$ is the expected signal yield calculated using the signal efficiency from the simulation and the estimated branching fraction based on branching fraction ratios of similar decays and the known branching fraction $B(B^+ \to D^0 \bar{D}^0 K^+)$ [39]. The background yield $N_{B}$ is extrapolated from fits to the sidebands of the $D^0$ candidate invariant-mass distribution. The classifiers are found to be independent of the $m(D^0 \bar{D}^0 K^+ \pi^-)$ distribution.

The family of decays $H_b \to D^{0(*)} \bar{D}^{0(*)} H^{(*)}$, where $H_b$ is a beauty hadron and $H^{(*)}$ any one- or two-body collection of light or strange hadrons, are examined to search for possible background contributions. These are referred to as peaking backgrounds. Of these, four decay modes ($B^+ \to D^0 \bar{D}^0 K^+$, $B^+ \to D^0 D^{0*} K^+$, $B_s^0 \to D^0 \bar{D}^0 \phi$ and $\bar{B}_s^0 \to D^0 \bar{D}^0 \bar{p} K^+$) are found to have substantial contributions to the signal channel. The $B^+ \to D^0 \bar{D}^0 K^+$ decays are removed using requirements on the three- and four-body invariant masses $5220 < m(D^0 \bar{D}^0 K^+) < 5340\text{ MeV}/c^2$ for candidates with $m(D^0 \bar{D}^0 K^+ \pi^-) > 5380\text{ MeV}/c^2$. The corresponding partially reconstructed decay $B^+ \to D^0 D^{0*} K^+$ is similarly removed with the requirement $5050 < m(D^0 \bar{D}^0 K^+) < 5200\text{ MeV}/c^2$. Contributions from $B_s^0 \to D^0 \bar{D}^0 \phi$ decays are suppressed using tighter PID requirements in the invariant-mass window $5321 < m(D^0 \bar{D}^0 K^+ K^-) < 5411\text{ MeV}/c^2$, where the $\pi^-$ candidate is reconstructed under the $K^-$ mass hypothesis. Similarly, $\bar{B}_s^0 \to D^0 \bar{D}^0 \bar{p} K^+$ candidates are removed using PID requirements for candidates satisfying $5575 < m(D^0 \bar{D}^0 \bar{p}) < 5665\text{ MeV}/c^2$, with the $\pi^+$ candidate reconstructed using the $\bar{p}$ mass hypothesis. All of these backgrounds are reduced to negligible levels, and only the $B^+ \to D^0 D^{0*} K^+$ veto induces a sizeable signal loss with an efficiency of 93%.

A particularly challenging source of background are the modes $B^0 \to D^0 K^+ \pi^- K^+ \pi^-$, $B^0 \to \bar{D}^0 K^- \pi^+ K^- \pi^+$ and $B^0 \to K^- \pi^+ K^+ \pi^- K^+ \pi^-$, so called single-charm and charmless backgrounds, respectively. Contributions from these decays are reduced by the flight distance criterion for the $D^0$ mesons, but must be estimated carefully because they peak at the known $B^0$ meson mass. The residual backgrounds are estimated from the sidebands of the $D^0$ invariant-mass distributions to be $10 \pm 7$ candidates. These candidates are subtracted from the yields during the fitting procedure described below.

The efficiency of the selections applied to the signal and control modes are calculated from simulated samples. The selection efficiencies include the geometrical acceptance of the LHCb detector, the online trigger and event reconstruction, offline selections and the neural network classifiers. For the signal mode, a single total efficiency is calculated and the resulting dependence on this efficiency model is considered as a systematic uncertainty.
For the control mode, efficiency variations are seen over the phase space. Therefore, an efficiency is calculated for each candidate that depends on the two-dimensional Dalitz plot of the control mode decay.

Extended unbinned maximum-likelihood fits are performed to the $B^0$ candidate invariant-mass distributions of the signal and control channels in the range $5235 < m(D^0\bar{D}^0 K^+\pi^-) < 5600 \text{MeV}/c^2$. The resolution of the $m(D^0\bar{D}^0 K^+\pi^-)$ distribution means that the contribution from partially reconstructed $B \to D^{*0}\bar{D}^{0}K^+\pi^-$ and $B \to D^{*0}\bar{D}^{0}K^+\pi^-$ decays is negligible in this fit range \[44\].

The fit to the control mode is performed separately in the four data samples, corresponding to the two trigger categories and two data taking periods. The fit to the signal channel is performed simultaneously to these four categories. The invariant-mass distributions for signal and control mode are modelled with a double-sided Crystal Ball function \[45\]. The parameters describing the tails of these distributions are fixed from fits to simulation separately for each of the four data samples. For the control mode, the mean and width of the mass distribution are determined directly from fits to the data subsamples. The resulting values are compared to those obtained on a fit to simulation to derive correction factors, which are subsequently used in the fits to the mass distribution of the signal channel. For the signal and control mode fits, the combinatorial background in each data sample is modelled with an exponential function with a slope allowed to vary in the fit. In the signal mode, the selections against the peaking backgrounds smoothly modify the shape of the mass distribution of the combinatorial background. This is accounted for by modulating the exponential function by an empirical correction from simulation. In the subsequent fits to the mass distribution of the signal candidates the ratio of branching fractions between the signal and control modes, $\mathcal{R} = \frac{\mathcal{B}(D^{*0} \to D^{*0}\bar{D}^{0}K^+\pi^-)}{\mathcal{B}(D^{*0} \to D^{*0}\bar{D}^{0}K^+\pi^-)}$, is expressed in terms of the signal yield in each of the four data samples as

$$\mathcal{R} = B(D^{*0} \to \bar{D}^{0}\pi^-) \times \left( \frac{N_{\text{sig}}}{N_{\text{con}}} \right) \times \left( \frac{\varepsilon_{\text{sig}}}{\varepsilon_{\text{con}}} \right),$$

where $N_{\text{sig}}$ and $N_{\text{con}}$ are the yields of the signal and control modes, respectively, and $\varepsilon_{\text{sig}}$, $\varepsilon_{\text{con}}$ are the corresponding efficiencies. The $\mathcal{R}$ parameter is determined from the simultaneous fit to the four data samples. The yield $N_{\text{con}}$ and its uncertainty are propagated from the fit to the control mode with a Gaussian constraint.

Invariant-mass distributions and fit projections of the $B^0$ candidates, summed over the trigger and data taking period subsamples, are shown in Fig.\[2\] In total 297 ± 14 signal and 1697 ± 42 control mode decays are found with a ratio of branching fractions $\mathcal{R} = (14.2 \pm 1.1)\%$, where the uncertainties are statistical only.

Figure \[3\] shows the background-subtracted invariant-mass distributions of $m(D^0\bar{D}^0)$, $m(D^0K^+)$ and $m(K^+\pi^-)$ overlaid with a simple phase-space distribution, including efficiency effects derived from simulation. There are hints of structures visible at the masses of the $\psi(3770)$, $D^*_{s(1,3)}(2573)^+$ and $D^*_{s(1,3)}(2860)^+$, and $K^+(892)^0$ states in the $m(D^0\bar{D}^0)$, $m(D^0K^+)$ and $m(K^+\pi^-)$ distributions, respectively. Care should be taken with any interpretation of these projections because structures may be caused by reflections. Further analysis of these structures is left for future studies.

Several sources of systematic uncertainty are taken into account. The impact of using an averaged efficiency in the signal mode is considered by comparing the results using samples of $B^0 \to D^0\bar{D}^0 K^{*0}$ simulated events. An event-by-event correction to the
Figure 2: Invariant-mass distributions and fit projections for $B^0$ candidates in (left) the signal and (right) control mode for all subsamples combined. The data are shown as black points with error bars and the fit components are as described in the legends. The small single-charm and charmless background is included in the signal component.

Figure 3: Projections of background-subtracted data (black points) in (left) $m(D^0 \bar{D}^0)$, (centre) $m(D^0 K^+)$ and (right) $m(K^+ \pi^-)$ with the phase-space only distribution (orange dashed line) superimposed for reference. The data contain a few single-charm and charmless background candidates.

Efficiency is also considered, based on various three-dimensional parameterisations of the full five-dimensional phase space. The fit model uncertainty is calculated by comparing the nominal background model to a polynomial form, and varying the signal shape parameters by sampling multivariate Gaussian distributions to account for the variance in the fit to simulation. The overall fit procedure is tested using pseudoexperiments and no bias is observed. The limited simulation sample size introduces a systematic uncertainty related to the spread in results obtained by varying the overall selection efficiencies within statistical uncertainties. Additionally, the weighting algorithm used to correct the simulation, as well as the data-driven method correcting the PID variables, introduce an associated statistical uncertainty. An uncertainty is also assigned to the estimation of single-charm and charmless background yields, by varying this contribution during the simultaneous fit to data. A correction is applied to the NN$_B$ neural network classifier to account for possible mismodelling between data and simulation, and this uncertainty is calculated from the resulting difference in selection efficiencies. A small uncertainty is introduced
Table 1: Systematic uncertainties expressed as a percentage of the branching fraction ratio $R$. The statistical uncertainty is included for comparison. The single-charm and charmless backgrounds are considered together.

| Source                  | Uncertainty (%) |
|-------------------------|-----------------|
| Signal model            | 5.0             |
| Background model        | 2.0             |
| Fixed fit parameters    | 2.0             |
| Simulation sample size  | 2.5             |
| Simulation weighting    | 2.0             |
| PID weighting           | 1.2             |
| Charmless backgrounds   | 2.0             |
| Classifier modelling    | 2.0             |
| Selection efficiency    | 0.6             |
| Sum in quadrature       | 7.3             |
| Statistical             | 7.7              |

due to the difference in the efficiency of selections applied to reconstruct candidates in signal and control modes. The systematic uncertainties are summarised in Table 1; they are summed in quadrature to give an overall relative systematic uncertainty on the ratio of branching fractions of 7.3%.

In summary, the decay $B^0 \to D^0\bar{D}^0 K^+\pi^-$ is observed for the first time, and its branching ratio relative to $B^0 \to D^{*-}D^0 K^+$ is measured to be

$$R = (14.2 \pm 1.1 \pm 1.0\%),$$

where the first uncertainty is statistical, and the second systematic. This measurement uses the full kinematically-allowed range of $B^0 \to D^0\bar{D}^0 K^+\pi^-$ outside of the $D^{*-}$ region, including the entire $K^+\pi^-$ mass range, encompassing the $K^*(892)^0$ resonance and the broad $K^+\pi^-$ $S$-wave. The most precise measurement of the branching fraction of $B^0 \to D^{*-}D^0 K^+$ decays, performed by the BaBar collaboration [21], is

$$B(B^0 \to D^{*-}D^0 K^+) = (2.47 \pm 0.21) \times 10^{-3} \ [21].$$

Substituting in this value gives

$$B(B^0 \to D^0\bar{D}^0 K^+\pi^-) = (3.50 \pm 0.27 \pm 0.26 \pm 0.30) \times 10^{-4},$$

where the third uncertainty comes from the uncertainty on the branching fraction $B(B^0 \to D^{*-}D^0 K^+)$. Recently, the LHCb collaboration performed a measurement of the ratio of branching fractions $B(B^0 \to D^{*-}D^0 K^+)$ [22]. However, the current precision on the branching fraction of the decay $B^0 \to D^0 D^+ K^+$ [39] does not yet allow for a more precise measurement of the decay rate $B(B^0 \to D^{*-}D^0 K^+)$. The results in this Letter provide a crucial first step towards studying the rich resonant structure of these decays. An amplitude analysis will provide insights to both the spectroscopy of $c\bar{s}$ and $c\bar{c}$ states, and charm-loop contributions to $b \to s\ell^+\ell^-$ decays.

**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); 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); MinECo (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United Kingdom); DOE NP and NSF (USA). 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 OSC (USA). 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 AvH Foundation (Germany); EPLANET, Marie Skłodowska-Curie Actions and ERC (European Union); A*MIDEX, ANR, Labex P2I0 and OCEVU, and Région Auvergne-Rhône-Alpes (France); Key Research Program of Frontier Sciences of CAS, CAS PIFI, and the Thousand Talents Program (China); RFBR, RSF and Yandex LLC (Russia); GVA, XuntaGal and GENCAT (Spain); the Royal Society and the Leverhulme Trust (United Kingdom).
Supplemental Material

This supplemental material includes additional information to that already provided in the main Letter. Section 1 details the correction applied to the selection efficiency of the control mode. Section 2 shows the invariant-mass distributions and fit projections separately for the data subsamples.

1 Efficiency correction

To account for variations across the Dalitz plot, \( m^2(D^{*-}K^+) \) and \( m^2(D^0K^+) \), in the control mode, the efficiency is calculated, in each category, as

\[
\varepsilon_{\text{con}}(m^2(D^{*-}K^+), m^2(D^0K^+)) = \frac{\sum_i w_i \varepsilon_i(m^2(D^{*-}K^+), m^2(D^0K^+))}{\sum_i w_i}. \tag{S1}
\]

Here \( w_i \) is the weight derived from the \( s\text{Plot} \) technique, with the \( B^0 \) candidate invariant mass distribution as the discriminating variable. The symbol \( \varepsilon_i \) refers to the efficiency as a function of the Dalitz plot position for candidate \( i \) in the control mode data sample. The efficiency variation across the Dalitz plot is displayed in Fig. S1. The average uncertainty across the bins of the Dalitz plane ranges from 1–8\% for the four data samples. The two exclusive trigger categories are henceforth referred to as Trigger on Signal (TOS), for events triggered by the signal decay, and Trigger Independent of Signal (TIS), for events triggered by other particles in the event and that do not fall into the TOS category.
Figure S1: The efficiency $\varepsilon_{\text{con}}$, as a function of position in the Dalitz plot of $B^0 \rightarrow D^{*-}D^0K^+$ decays obtained from simulated samples. The top row shows Run 1 samples for (left) TOS and (right) TIS. The bottom row shows the same trigger categories for 2016 samples.
2 Mass fits

The invariant-mass distributions and fit projections of the fit model results for each run period and trigger category are shown in Fig. S2 and Fig. S3 for the signal and control mode, respectively. Figure S4 shows the invariant-mass distributions for each category in an extended range to include the partially reconstructed backgrounds for illustrative purposes. The partially reconstructed peaks are modelled with Crystal Ball functions [45]. Further systematic uncertainties would be required to account for the effects of modelling these additional peaks, motivating the nominal fit range $5235 < m(D^0\bar{D}^0 K^+\pi^-) < 5600 \text{ MeV}/c^2$, where the partially reconstructed contributions are negligible. As a cross check, performing the fit in this extended range results in a branching fraction ratio $R$ that is in good agreement with the measurement from the nominal fit.

![Invariant-mass distributions and fit projections for $B^0 \rightarrow D^0\bar{D}^0 K^+\pi^-$ candidates, shown separately for each run period, (top) Run 1 and (bottom) 2016, and trigger category, (left) TOS and (right) TIS. The data are shown as black points with error bars and the fit components are as described in the legends. The data contain a few single-charm and charmless background candidates.](image)
Figure S3: Invariant-mass distributions and fit projections for $B^0 \to D^{*-} D^0 K^+$ candidates, shown separately for each run period, (top) Run 1 and (bottom) 2016, and trigger category, (left) TOS and (right) TIS. The data are shown as black points with error bars and the fit components are as described in the legends.
Figure S4: Invariant-mass distributions and fit projections for $B^0 \rightarrow D^0 D^0 K^+ \pi^-$ candidates over an extended range, shown separately for each run period, (top) Run 1 and (bottom) 2016, and trigger category, (left) TOS and (right) TIS. The data are shown as black points with error bars and the fit components are as described in the legends, where the green shaded distribution describes partially reconstructed backgrounds. The data contain a few single-charm and charmless background candidates.
References

[1] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y.-M. Wang, Charm-loop effect in $B \to K^{(*)}\ell^+\ell^-$ and $B \to K^{\gamma}$, *JHEP* 09 (2010) 089, arXiv:1006.4945.

[2] J. Lyon and R. Zwicky, Resonances gone topsy turvy - the charm of QCD or new physics in $b \to s \ell^+\ell^-$?, arXiv:1406.0566.

[3] M. Ciuchini et al., $B \to K^{*}\ell^+\ell^-$ decays at large recoil in the Standard Model: a theoretical reappraisal, *JHEP* 06 (2016) 116, arXiv:1512.07157.

[4] C. Bobeth, G. Hiller, and G. Piranishvili, CP Asymmetries in $B \to K^{*}(\to K\pi)\bar{\ell}\ell$ and Untagged $\bar{B}_s \to \phi(\to K^+K^-)\bar{\ell}\ell$ Decays at NLO, *JHEP* 07 (2008) 106, arXiv:0805.2525.

[5] T. Blake et al., An empirical model to determine the hadronic resonance contributions to $B^0 \to K^{*0}\mu^+\mu^-$ transitions, *Eur. Phys. J.* C 78 (2018) 453, arXiv:1709.03921.

[6] LHCb collaboration, R. Aaij et al., Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay, *Phys. Rev. Lett.* 125 (2020) 011802, arXiv:2003.04831.

[7] CMS collaboration, A. M. Sirunyan et al., Measurement of angular parameters from the decay $B^0 \to K^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV, *Phys. Lett.* B781 (2018) 517, arXiv:1710.02846.

[8] ATLAS collaboration, M. Aaboud et al., Angular analysis of $B^0_d \to K^{*}\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, *JHEP* 10 (2018) 047, arXiv:1805.04000.

[9] Belle collaboration, S. Wehle et al., Lepton-flavor-dependent angular analysis of $B \to K^{*}\ell^+\ell^-$, *Phys. Rev. Lett.* 118 (2017) 111801, arXiv:1612.05014.

[10] BaBar collaboration, B. Aubert et al., Measurements of branching fractions, rate asymmetries, and angular distributions in the rare decays $B \to K\ell^+\ell^-$ and $B \to K^*\ell^+\ell^-$, *Phys. Rev.* D73 (2006) 092001, arXiv:hep-ex/0604007.

[11] CDF collaboration, T. Altonen et al., Measurements of the angular distributions in the decays $B \to K^{(*)}\mu^+\mu^-$ at CDF, *Phys. Rev. Lett.* 108 (2012) 081807, arXiv:1108.0695.

[12] LHCb collaboration, R. Aaij et al., Angular moments of the decay $A_0^b \to \Lambda\mu^+\mu^-$ at low hadronic recoil, *JHEP* 09 (2018) 146, arXiv:1808.00264.

[13] LHCb collaboration, R. Aaij et al., Angular analysis and differential branching fraction of the decay $B^0_s \to \phi\mu^+\mu^-$, *JHEP* 09 (2015) 179, arXiv:1506.08777.

[14] LHCb collaboration, R. Aaij et al., Measurements of the S-wave fraction in $B^0 \to K^{*}\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^{*}(892)^0\mu^+\mu^-$ differential branching fraction, *JHEP* 11 (2016) 047, Erratum ibid. 04 (2017) 142, arXiv:1606.04731.
[15] LHCb collaboration, R. Aaij et al., Differential branching fraction and angular analysis of $Λ^0_\theta \to Λμ^+μ^-$ decays. JHEP 06 (2015) 115. Erratum ibid. 09 (2018) 145, arXiv:1503.07138.

[16] LHCb collaboration, R. Aaij et al., Differential branching fractions and isospin asymmetries of $B \to K^{(*)}μ^+μ^-$ decays. JHEP 06 (2014) 133, arXiv:1403.8044.

[17] ALEPH collaboration, R. Barate et al., Observation of doubly charmed $B$ decays at LEP. Eur. Phys. J. C 4 (1998) 387.

[18] BaBar collaboration, B. Aubert et al., Measurement of the branching fractions for the exclusive decays of $B_0$ and $B^+$ to $\bar{D}(s)D(s)K$, Phys. Rev. D 68 (2003) 092001, arXiv:hep-ex/0305003.

[19] Belle collaboration, G. Gokhroo et al., Observation of a Near-threshold $D^0\bar{D}^0π^0$ enhancement in $B \to D^0\bar{D}^0π^0K$ decay. Phys. Rev. Lett. 97 (2006) 162002, arXiv:hep-ex/0606055.

[20] Belle collaboration, J. Brodzicka et al., Observation of a new $D_{sJ}$ meson in $B^+ \to D^0\bar{D}^0K$ decays, Phys. Rev. Lett. 100 (2008) 092001, arXiv:0707.3941.

[21] BaBar collaboration, P. del Amo Sanchez et al., Measurement of the $B \to D^{(*)}D^{(*)}K$ branching fractions, Phys. Rev. D 83 (2011) 032004, arXiv:1011.3929.

[22] LHCb collaboration, R. Aaij et al., Measurement of the branching fractions for $B^+ \to D^{(*)}D^{(*)}K^+$, $B^+ \to D^{*-}D^+K^+$, and $B^0 \to D^{*-}D^0K^+$ decays. arXiv:2005.10264, submitted to JHEP.

[23] LHCb collaboration, A. A. Alves Jr. et al., The LHCb detector at the LHC. JINST 3 (2008) S08005.

[24] LHCb collaboration, R. Aaij et al., LHCb detector performance, Int. J. Mod. Phys. A30 (2015) 1530022 arXiv:1412.6352.

[25] R. Aaij et al., Performance of the LHCb Vertex Locator. JINST 9 (2014) P09007, arXiv:1405.7808.

[26] R. Arink et al., Performance of the LHCb Outer Tracker. JINST 9 (2014) P01002, arXiv:1311.3893.

[27] P. d’Argent et al., Improved performance of the LHCb Outer Tracker in LHC Run 2. JINST 12 (2017) P11016, arXiv:1708.00819.

[28] M. Adinolfi et al., Performance of the LHCb RICH detector at the LHC, Eur. Phys. J. C73 (2013) 2431, arXiv:1211.6759.

[29] T. Sjöstrand, S. Mrenna, and P. Skands, A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. 178 (2008) 852, arXiv:0710.3820.

[30] I. Belyaev et al., Handling of the generation of primary events in Gauss, the LHCb simulation framework. J. Phys. Conf. Ser. 331 (2011) 032047.
[31] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth A462 (2001) 152.

[32] D. Müller, M. Clemencic, G. Corti, and M. Gersabeck, *ReDecay: A novel approach to speed up the simulation at LHCb*, Eur. Phys. J. C78 (2018) 1009, arXiv:1810.10362.

[33] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays*, Eur. Phys. J. C45 (2006) 97, arXiv:hep-ph/0506026.

[34] Geant4 collaboration, J. Allison et al., *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. 53 (2006) 270; Geant4 collaboration, S. Agostinelli et al., *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. A506 (2003) 250.

[35] M. Clemencic et al., *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. 331 (2011) 032023.

[36] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. A555 (2005) 356, arXiv:physics/0402083.

[37] A. Poluektov, *Kernel density estimation of a multidimensional efficiency profile*, JINST 10 (2015) P02011, arXiv:1411.5528.

[38] R. Aaij et al., *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, Eur. Phys. J. Tech. Instr. 6 (2018) 1, arXiv:1803.00824.

[39] Particle Data Group, M. Tanabashi et al., *Review of particle physics*, Phys. Rev. D98 (2018) 030001, and 2019 update.

[40] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth A552 (2005) 566, arXiv:physics/0503191.

[41] F. Chollet et al., *Keras*, 2015. Software available from https://keras.io.

[42] M. Abadi et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015. Software available from https://tensorflow.org.

[43] A. Blum, A. Kalai, and J. Langford, *Beating the hold-out: bounds for k-fold and progressive cross-validation*, in *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, COLT ’99, (New York, NY, USA), 203, ACM, 1999.

[44] See Supplemental Material for additional details on the mass fits.

[45] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.
Institute of Theoretical and Experimental Physics NRC Kurchatov Institute (ITEP NRC KI), Moscow, Russia
Institute of Nuclear Physics, Moscow State University (SINP MSU), Moscow, Russia
Institute for Nuclear Research of the Russian Academy of Sciences (INR RAS), Moscow, Russia
Yandex School of Data Analysis, Moscow, Russia
Budker Institute of Nuclear Physics (SB RAS), Novosibirsk, Russia
Institute for High Energy Physics NRC Kurchatov Institute (IHEP NRC KI), Protvino, Russia, Protvino, Russia
ICCUB, Universitat de Barcelona, Barcelona, Spain
Instituto Galego de Física de Altas Enerxías (IGFAE), Universidade de Santiago de Compostela, Santiago de Compostela, Spain
Instituto de Física Corpuscular, Centro Mixto Universidad de Valencia - CSIC, Valencia, Spain
European Organization for Nuclear Research (CERN), Geneva, Switzerland
Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Physik-Institut, Universität Zürich, Zürich, Switzerland
NSC Kharkiv Institute of Physics and Technology (NSC KIPT), Kharkiv, Ukraine
Institute for Nuclear Research of the National Academy of Sciences (KINR), Kyiv, Ukraine
University of Birmingham, Birmingham, United Kingdom
H.H. Wills Physics Laboratory, University of Bristol, Bristol, United Kingdom
 Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
Department of Physics, University of Warwick, Coventry, United Kingdom
STFC Rutherford Appleton Laboratory, Didcot, United Kingdom
School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom
Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom
Imperial College London, London, United Kingdom
Department of Physics and Astronomy, University of Manchester, Manchester, United Kingdom
Department of Physics, University of Oxford, Oxford, United Kingdom
Massachusetts Institute of Technology, Cambridge, MA, United States
University of Cincinnati, Cincinnati, OH, United States
University of Maryland, College Park, MD, United States
Los Alamos National Laboratory (LANL), Los Alamos, United States
Syracuse University, Syracuse, NY, United States
Laboratory of Mathematical and Subatomic Physics, Constantine, Algeria, associated to 2
Pontificia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil, associated to 2
Guangdong Provencial Key Laboratory of Nuclear Science, Institute of Quantum Matter, South China Normal University, Guangzhou, China, associated to 3
School of Physics and Technology, Wuhan University, Wuhan, China, associated to 3
Departamento de Física, Universidad Nacional de Colombia, Bogotá, Colombia, associated to 12
Universität Bonn - Helmholtz-Institut für Strahlen und Kernphysik, Bonn, Germany, associated to 16
Institut für Physik, Universität Rostock, Rostock, Germany, associated to 16
Van Swinderen Institute, University of Groningen, Groningen, Netherlands, associated to 31
Universiteit Maastricht, Maastricht, Netherlands, associated to 31
National Research Centre Kurchatov Institute, Moscow, Russia, associated to 38
National University of Science and Technology “MISIS”, Moscow, Russia, associated to 38
National Research University Higher School of Economics, Moscow, Russia, associated to 41
National Research Tomsk Polytechnic University, Tomsk, Russia, associated to 38
University of Michigan, Ann Arbor, United States, associated to 67
Laboratoire Leprince-Ringuet, Palaiseau, France
AGH - University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Kraków, Poland
Universidade Federal do Triângulo Mineiro (UFTM), Uberaba-MG, Brazil
Laboratoire Leprince-Ringuet, Palaiseau, France
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 Genova, Genova, Italy
Università di Milano Bicocca, Milano, Italy
Università di Roma Tor Vergata, Roma, Italy
AGH - University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Kraków, Poland
DS4DS, La Salle, Universitat Ramon Llull, Barcelona, Spain
Hanoi University of Science, Hanoi, Vietnam
Università di Padova, Padova, Italy
Università di Pisa, Pisa, Italy
Università degli Studi di Milano, Milano, Italy
Università di Urbino, Urbino, Italy
Università della Basilicata, Potenza, Italy
Scuola Normale Superiore, Pisa, Italy
Università di Modena e Reggio Emilia, Modena, Italy
Università di Siena, Siena, Italy
MSU - Iligan Institute of Technology (MSU-IIT), Iligan, Philippines
Novosibirsk State University, Novosibirsk, Russia
INFN Sezione di Trieste, Trieste, Italy
Universidad Nacional Autonoma de Honduras, Tegucigalpa, Honduras