



# Search for rare decays of $D^0$ mesons into two muons

LHCb collaboration

## Abstract

A search for the very rare  $D^0 \rightarrow \mu^+ \mu^-$  decay is performed using data collected by the LHCb experiment in proton-proton collisions at  $\sqrt{s} = 7, 8$  and  $13$  TeV, corresponding to an integrated luminosity of  $9 \text{ fb}^{-1}$ . The search is optimised for  $D^0$  mesons from  $D^{*+} \rightarrow D^0 \pi^+$  decays but is also sensitive to  $D^0$  mesons from other sources. No evidence for an excess of events over the expected background is observed. An upper limit on the branching fraction of this decay is set at  $\mathcal{B}(D^0 \rightarrow \mu^+ \mu^-) < 3.1 \times 10^{-9}$  at a 90% CL. This represents the world's most stringent limit, constraining models of physics beyond the Standard Model.

Published in Phys. Rev. Lett. 131 (2023) 041804



Processes with a change in quark flavour without a change in electric charge are forbidden at the lowest order in the Standard Model (SM) of particle physics. Flavour Changing Neutral Currents (FCNC) are additionally suppressed by the Glashow-Iliopoulos-Maiani (GIM) mechanism [1]. FCNC have been extensively studied in strange- and beauty-quark hadrons. In the charm sector the GIM suppression is stronger because the mass differences between down-type quarks are smaller than the ones between up-type quarks. These processes can be enhanced by several orders of magnitude in new physics (NP) scenarios when compared to the SM.

The  $D^0 \rightarrow \mu^+ \mu^-$  decay is among the most interesting charm-hadrons decays,<sup>1</sup> being fully leptonic and additionally suppressed by helicity reasons. Its SM short-distance contribution is extremely suppressed, yielding a branching fraction on the order of  $10^{-18}$  [2]. Long-distance contributions dominate through an intermediate two-photon state, and can be estimated to be  $\mathcal{B}(D^0 \rightarrow \mu^+ \mu^-) \simeq 2.7 \times 10^{-5} \mathcal{B}(D^0 \rightarrow \gamma\gamma)$ , leading to a  $D^0 \rightarrow \mu^+ \mu^-$  branching fraction of at least  $3 \times 10^{-13}$  [2]. The best upper limit on the  $D^0 \rightarrow \gamma\gamma$  decay rate was set by the Belle collaboration to be  $8.5 \times 10^{-7}$  at a 90% C.L. [3]; using the same relation this turns into an upper limit on the long-distance contribution to the  $D^0 \rightarrow \mu^+ \mu^-$  branching fraction of  $2.3 \times 10^{-11}$ . The  $D^0 \rightarrow \mu^+ \mu^-$  decay rate can be enhanced in many NP models [4]. Being one of the most sensitive FCNC processes in the up-quark sector, its branching fraction is used as a primary building block of different models, constraining the relevant couplings saturating the branching fraction limit [24, 27, 29–32, 35–37, 41, 47, 49]. Most significantly, model-independent bounds on the Wilson coefficients related to charm physics and in particular to  $D^0 \rightarrow \mu^+ \mu^-$  decays have been set in Ref. [48, 49]. Furthermore, the  $D^0 \rightarrow \mu^+ \mu^-$  rate is correlated to the rate of  $D^0 - \bar{D}^0$  mixing in many NP models [4]. This is of uttermost importance given the recent first observation of the mass difference between neutral charm-meson eigenstates [50]. Concerning specific models, in the Minimal Supersymmetric Standard Model no sizeable contribution would enhance the  $D^0 \rightarrow \mu^+ \mu^-$  rate [4]. Conversely, in some Supersymmetric models with  $R$ -parity symmetry violation tree level contributions would be allowed [2, 39]. Recent discussions of predictions for  $D^0 \rightarrow \mu^+ \mu^-$  decays in such models can be found in Refs. [40, 41]. In addition, it is interesting to note the importance of  $D^0 \rightarrow \mu^+ \mu^-$  decays as testing ground for models with leptoquarks proposed to explain deviations from the SM observed in  $B$  physics measurements [?, 5–23]. In some of these models rare  $B$  decays receive contributions at the loop level, while new particles could mediate the  $D^0 \rightarrow \mu^+ \mu^-$  decay at tree level. This phenomenon has been extensively discussed in the literature [24–38]. In those where additional vector bosons ( $Z'$ ) are introduced, the  $D^0 \rightarrow \mu^+ \mu^-$  decay usually does not give strong constraints [42, 43]. Similarly, in other explanatory models [44–46] the bound from  $D^0 \rightarrow \mu^+ \mu^-$  decays can be avoided. Instead in some models with vector-like fermions [47], the  $D^0 \rightarrow \mu^+ \mu^-$  decay gives the strongest constraint.

The current world's best limit on this decay is  $\mathcal{B}(D^0 \rightarrow \mu^+ \mu^-) < 6.2 \times 10^{-9} (7.6 \times 10^{-9})$  at 90% (95%) CL, and was obtained by the LHCb experiment exploiting about  $0.9 \text{ fb}^{-1}$  of 2011 data [51]. The data used in Ref. [51] are also used in this analysis and those results are superseded by this Letter.

This Letter presents a search for the  $D^0 \rightarrow \mu^+ \mu^-$  decay based on data collected by the LHCb experiment in  $pp$  collisions corresponding to  $9 \text{ fb}^{-1}$  of integrated luminosity. The

---

<sup>1</sup>Charge conjugate processes are implied throughout.

data have been collected in 2011, 2012 (Run 1) and 2015–2018 (Run 2) at  $\sqrt{s} = 7, 8$  and 13 TeV, respectively. Compared to the previous publication [51], the present work benefits from various improvements in the analysis, such as refined multivariate algorithms against combinatorial and misidentified background as well as an improved trigger [?], described throughout the paper. These allow to mitigate the impact of the harsher experimental conditions in Run 2. Both the higher energy and instantaneous luminosity produce a higher track multiplicity, which increases the combinatorial background and worsens the particle identification performance. The  $D^0 \rightarrow \mu^+\mu^-$  decay is searched for using  $D^{*+} \rightarrow D^0\pi^+$  decays, as this improves the background rejection and allows the yield of the decay to be obtained from a two-dimensional fit to the dimuon invariant mass,  $m(\mu^+\mu^-)$ , and the difference between the  $D^{*+}$  and  $D^0$  candidate masses,  $\Delta m$ . The yield is converted to the decay branching fraction by normalising to two hadronic decays,  $D^0 \rightarrow K^-\pi^+$  and  $D^0 \rightarrow \pi^+\pi^-$ , selected concurrently to the signal (collectively referred to as  $D^0 \rightarrow h^+h^-$ ).

The LHCb detector is a single-arm forward spectrometer covering the pseudorapidity range  $2 < \eta < 5$ , described in detail in Refs. [52, 53]. The simulated events used in this analysis are produced with the software described in Refs. [54–60].

Events are selected online by a trigger that consists of a hardware stage, which is based on information from the calorimeter and muon systems, followed by two software stages. At the hardware trigger stage, events are required to have a muon candidate with high transverse momentum,  $p_T$ , or a hadron, photon or electron candidate with high transverse energy in the calorimeters. A first stage of the software trigger selects events with a muon candidate, or a high  $p_T$  charged particle, or a combination of two tracks, each of these displaced from the primary  $pp$  collision vertex (PV). In the second stage of the software trigger, dedicated algorithms select candidate  $D^0 \rightarrow \mu^+\mu^-$ ,  $D^0 \rightarrow K^-\pi^+$  and  $D^0 \rightarrow \pi^+\pi^-$  decays, combining two oppositely charged tracks with loose particle identification (PID) requirements that form a secondary vertex separated from any PV. The invariant mass of the  $D^0$  candidate must lie in an interval of  $\pm 300$  ( $\pm 70$ ) MeV/ $c^2$  centred on the known  $D^0$  mass for signal (normalisation) channel candidates. To keep the same trigger selection for the signal and normalisation channels, aside from PID, a scale factor of order 0.2–3.0% depending on the data taking period is applied to the normalisation channels trigger selections to limit their rate, keeping randomly a fraction of the events. In the offline selection, only candidates associated with a muon hardware trigger, or those where the rest of the event contained a high transverse energy hadron or electron are kept for the signal channel. For the normalisation channels, only candidates associated with a high transverse energy hadron are kept.

In the offline analysis,  $D^0$  candidates satisfying the trigger requirements are formed with similar but more stringent criteria than the second software-trigger stage. These  $D^0$  candidates are then combined with a charged particle originating from the same PV and having  $p_T > 110$  MeV/ $c$  to form  $D^{*+} \rightarrow D^0\pi^+$  candidates. To improve the mass resolution, the  $D^{*+}$  meson decay vertex is constrained to coincide with the PV [61]. The candidate  $\Delta m$  is required to be in the range 139.6–151.6 MeV/ $c^2$ . A multivariate selection based on a boosted decision tree (BDT) algorithm [62–64] is used to suppress background from random combinations of charged particles, using as input: the  $p_T$  of the pion from the  $D^{*+}$  decay, the smallest  $p_T$  and impact parameter significance with respect to the PV of the  $D^0$  decay products, the angle between the  $D^0$  momentum and the vector connecting the primary and secondary vertices, and the quality of the  $D^0$  vertex. The BDT is trained separately for each Run of data taking, using simulated decays as signal and data

candidates from the dimuon sample with  $m(\mu^+\mu^-) > 1894 \text{ MeV}/c^2$  as background. The  $k$ -folding technique, with  $k = 9$ , is applied [65]. The BDT output ranges from 0 to 1, from background-like candidates to more signal-like candidates. The BDT output is used to define three search regions:  $\text{BDT} \in [0.15, 0.33]$ ,  $[0.33, 0.66]$ ,  $[0.66, 1.]$ . The output of the same BDT algorithm is computed also for the  $D^0 \rightarrow \pi^+\pi^-$  candidates for calibration purposes, but is not required in the selection.

A second source of background is due to two- and three-body  $D^0$  decays, with one or two hadrons misidentified as muons (*e.g.*  $D^0 \rightarrow h^+h^-$  or semileptonic decays). The misidentification occurs mainly for hadrons that decay into a muon before the muon sub-detector. Although this process is relatively rare, the large branching fractions of these modes produce a background peaking in the signal region of the  $m(\mu^+\mu^-)$  distribution that is partially suppressed by a multivariate muon identification discriminant combining information from the Cherenkov detectors, the calorimeters, and the muon sub-detector [66]. In addition, the muon candidates are required to have associated muon chamber hits that are not shared with any other track in the event. Signal and background sources can originate both from  $D^{*+} \rightarrow D^0\pi^+$  decays or from other sources (the PV or  $B$  decays) combined with an unrelated pion (untagged); both are taken into account in the yield estimation. A requirement on the output of the multivariate muon identification discriminant is simultaneously chosen for the three BDT regions, by optimising the sensitivity to the minimum visible cross-section, as defined by an extension of the figure of merit defined in Ref. [67]. Roughly one percent of events contain more than one signal candidate after all selection requirements, all of which are retained.

The signal yield is converted to the decay branching fraction by normalising to the hadronic decays  $D^0 \rightarrow \pi^+\pi^-$  and  $D^0 \rightarrow K^-\pi^+$ , with branching fractions of  $(1.490 \pm 0.027) \times 10^{-3}$  and  $(3.999 \pm 0.045) \times 10^{-2}$ , respectively [68], as

$$\mathcal{B}(D^0 \rightarrow \mu^+\mu^-) = \frac{N_{D^0 \rightarrow \mu^+\mu^-}}{N_{D^0 \rightarrow h^+h^-}} \cdot \frac{\varepsilon_{h^+h^-}}{\varepsilon_{\mu^+\mu^-}} \cdot s \cdot \mathcal{B}(D^0 \rightarrow h^+h^-) \equiv \alpha N_{D^0 \rightarrow \mu^+\mu^-} \quad (1)$$

where  $\varepsilon$  is the efficiency and  $N$  is the yield of the given channel,  $s$  is the scale factor of the normalisation channel and  $\alpha$  is defined as the single event sensitivity.

The efficiencies in Eq. (1) are factorised into different steps for ease of estimation and evaluated with respect to the previous steps: detector acceptance, reconstruction and selection, PID, and trigger.

The reconstruction and selection efficiencies are obtained from simulated samples. The simulated candidates are assigned weights with an iterative procedure that improves the agreement with data using the following variables: pseudorapidity of the  $D^0$  meson, transverse momentum of the  $D^0$  meson and number of tracks in the event. It is verified that after weighting, all variables used in the selection agree well between data and simulation. The weights obtained from the  $D^0 \rightarrow \pi^+\pi^-$  candidates are used to correct also the signal simulation.

Possible residual differences between data and simulation in the tracking efficiencies are determined using control channels in data [69]. The PID efficiencies are determined from data using samples of kinematically identified charged particles from  $B^+ \rightarrow J/\psi K^+$  and  $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$  decays [66], weighted to match the kinematic properties of the signal and the normalisation channels, respectively. The efficiencies are determined in bins of the  $p$  and  $p_T$  of the tracks. A total systematic uncertainty of 1–3% is associated to the binning scheme and background determination in the calibration samples.

The efficiency of the second level of the software trigger is unity with respect to the offline-selected candidates by construction, as the selection is tighter in every requirement. The hardware and first level software trigger efficiencies are evaluated with the TISTOS method [70] in data. For the signal channel, the  $B^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^+$  decay is used as the calibration channel, selected with the same requirements as those used for the analysis of  $B$  decays into two muons [71, 72]. The calibration is performed in intervals of the  $J/\psi$   $p_T$  and pseudorapidity. For each interval, a scaling factor between data and simulation is obtained and applied to the  $D^0 \rightarrow \mu^+\mu^-$  simulation. Compatible results are obtained repeating the calibration in intervals of  $p_T$  and the maximum  $\text{IP}\chi^2$  of the muons, where  $\text{IP}\chi^2$  is defined as the difference in the vertex-fit  $\chi^2$  of a given PV reconstructed with and without the track under consideration. The typical scaling between data and simulation deviates from unity by 2–6%. The normalisation channels, given their high yields, are self-calibrated. The TISTOS method is applied to the  $D^0 \rightarrow K^-\pi^+$  and  $D^0 \rightarrow \pi^+\pi^-$  channels and trigger efficiencies are obtained. To minimise cross-correlation biases, only candidates in events that satisfy a muon trigger independently of the candidate are used as calibration sample. The calibration of the hadronic hardware trigger is also validated with independent estimates based on control samples in data, obtained with similar methods as in Ref. [73], from which a 15% relative systematic uncertainty is assigned to the hadronic trigger efficiency calibration.

The efficiency of the BDT requirement, and the signal fraction in the BDT intervals, are calibrated in data by applying the same estimator to the  $D^0 \rightarrow \pi^+\pi^-$  decay, which is topologically very similar to the signal. The distribution of the BDT output is obtained in background subtracted  $D^0 \rightarrow \pi^+\pi^-$  decays in data and simulation, and found to be compatible, as shown in the Supplemental Material [74]. A small correction is determined and applied to the signal; its uncertainty is assigned as systematic uncertainty to the signal efficiency.

The yields of the normalisation channels are obtained through a fit to the  $\Delta m$  distribution (Fig. 1), requiring the reconstructed  $D^0$  mass to be within  $\pm 10 \text{ MeV}/c^2$  of the known  $D^0$  mass. The signal probability distribution function is composed of a sum of a Gaussian and a Crystal-Ball function [75] with power-law tails on both sides. The background is described with a threshold function, as defined in Ref. [51]. The parameters of the Crystal-Ball function are estimated with simulation: the power of the tail is fixed, while the position where the power tails start may vary freely in the fit. In addition, the signal width and all background parameters are left free in the fit.

Using Eq. (1), values of  $\alpha$  for both normalisation channels are obtained, and found to be in good agreement with each other for each data taking run and for the full sample. As an additional cross-check, the ratio of the efficiency corrected yields of the two normalisation channels is obtained and compared to the ratio of their branching fractions. The value is stable across the data taking years and compatible with the world average [68]. The average single event sensitivity is found to be  $\alpha = (2.15 \pm 0.34) \times 10^{-11}$ , corresponding to at most one expected signal  $D^0 \rightarrow \mu^+\mu^-$  decay under the SM hypothesis.

The signal yield is obtained through an unbinned maximum-likelihood fit to the two-dimensional distribution of  $m(\mu^+\mu^-)$  and  $\Delta m$ , performed simultaneously in the three BDT intervals and in the two data taking Runs. The distributions projected onto the two variables are shown in Fig. 2. Each of the two projections is selected using only candidates in the signal region of the other variable, where the signal regions are defined as  $m(D^0) \in [1840, 1885] \text{ MeV}/c^2$  and  $\Delta m \in [144.9, 146.1] \text{ MeV}/c^2$ , respectively.

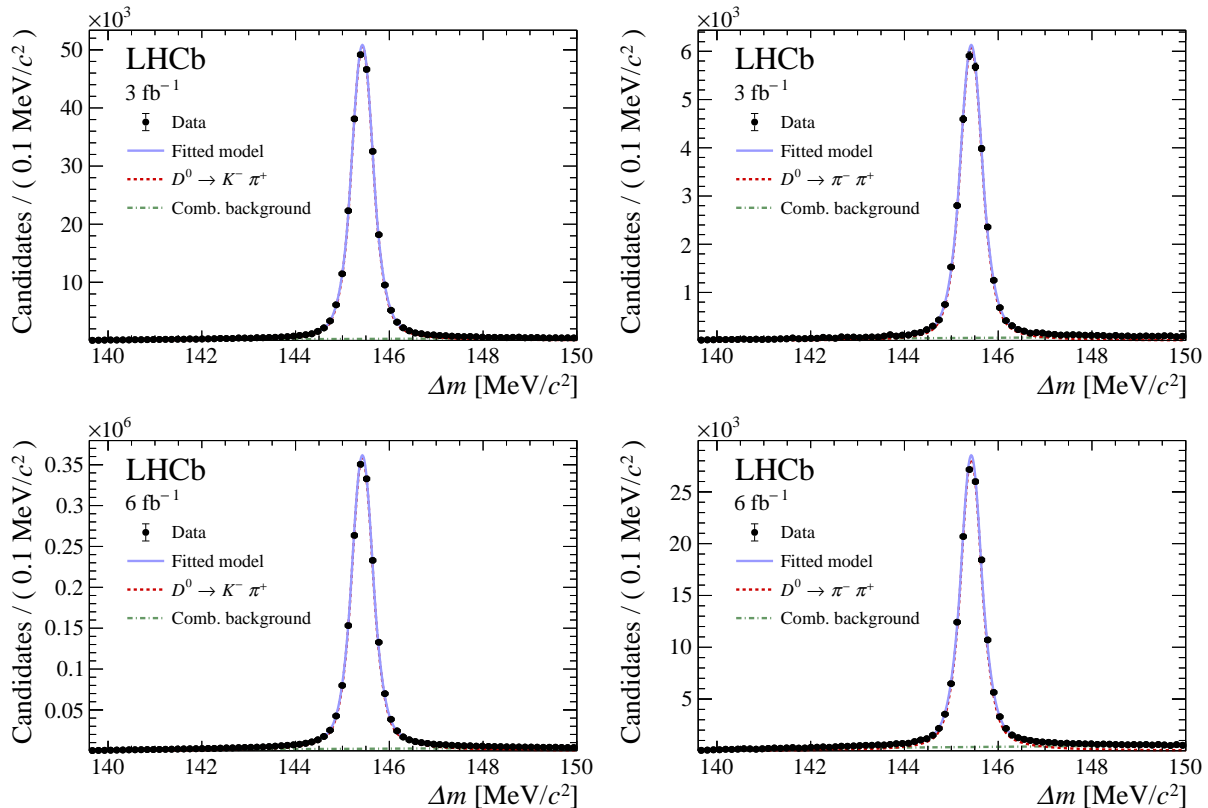


Figure 1: Distributions of  $\Delta m$  for (left)  $D^0 \rightarrow K^- \pi^+$  and (right)  $D^0 \rightarrow \pi^+ \pi^-$  normalisation channels candidates for (top) Run 1 and (bottom) Run 2 data. The distributions are superimposed with the fit.

The full distributions can be seen in the Supplemental Material [74]. The correlation between the two variables is found to be negligible for all contributing decay modes, thus they are treated as uncorrelated. After the full selection, only combinatorial and misidentified hadronic  $D^0$  decays are found to contribute to the background. Background from semileptonic decays is found to be negligible, and any remaining background from other sources is well modelled as part of the combinatorial background component.

The shape of the signal and misidentified background ( $D^0 \rightarrow \pi^+ \pi^-$ ,  $D^0 \rightarrow K^- \pi^+$ ) distributions in the two fit variables is obtained from simulation, reconstructed as  $D^0 \rightarrow \mu^+ \mu^-$  decays. The model parameters are determined separately for Run 1 and Run 2; the resulting PDF describes the distribution in each data taking year and BDT interval well. For signal, the  $m(\mu^+ \mu^-)$  and  $\Delta m$  distributions are both parametrised by Crystal-Ball [75] functions with power tails on both sides. For  $D^0 \rightarrow \pi^+ \pi^-$  decays, a single Crystal-ball function in  $m(\mu^+ \mu^-)$  and the sum of a Johnson [76] and Gaussian function in  $\Delta m$  are employed. For  $D^0 \rightarrow K^- \pi^+$  decays, a Johnson function is used for the  $m(\mu^+ \mu^-)$  distribution, while the  $\Delta m$  distribution is described by three Gaussian functions. The combinatorial background is described by an exponential function in  $m(\mu^+ \mu^-)$  and a threshold function in  $\Delta m$  [51]. Untagged signal and  $D^0 \rightarrow \pi^+ \pi^-$  components are included and parametrised as their respective tagged component in  $m(\mu^+ \mu^-)$  and with the same threshold function of the combinatorial background in  $\Delta m$ . The fraction of this component is fixed to the value determined in each BDT interval from a fit to  $D^0 \rightarrow \pi^+ \pi^-$  data. The shape parameters

obtained from the simulated samples are fixed in the data fit, while the slope of the exponential of the combinatorial background is left free to vary in each BDT interval.

A constraint on the expected number of misidentified  $D^0 \rightarrow \pi^+\pi^-$  decays is determined from a dedicated, high-statistics, simulation sample with the trigger and offline selection applied. The most critical part of the simulated sample is the PID efficiency due to the presence of a large fraction of  $\pi \rightarrow \mu$  decays that mimic the signal and are not considered in the standard calibration tools. The PID efficiency is obtained from simulation but it is cross-checked using  $D^+ \rightarrow \pi^+\pi^-\pi^+$  and  $D_s^+ \rightarrow \pi^+\pi^-\pi^+$  control samples in data where same-sign pions are weighted to match the kinematics of  $D^0 \rightarrow \pi^+\pi^-$  decays. The agreement between the PID efficiency determined with both methods is satisfactory over the full range of the muon identification discriminant variable [74]. Therefore, no systematic uncertainty is assigned on this estimate. The uncertainty on the expected  $D^0 \rightarrow \pi^+\pi^-$  yield is propagated through a Gaussian constraint on the relevant parameter in the final fit.

The yield of the misidentified  $D^0 \rightarrow K^-\pi^+$  decays is constrained from an auxiliary fit to the  $m(\mu^+\mu^-)$  sideband data, recomputed with the correct mass hypothesis. The fit is performed using the  $\Delta m$  distribution within a  $\pm 10 \text{ MeV}/c^2$  region around the  $D^0$  mass in the  $K^-\pi^+$  mass hypothesis. A correction is applied to take into account this mass requirement. The correlation between this estimate and the yield in the final fit is found not to influence the estimate of the signal branching fraction.

The systematic uncertainties related to both the normalisation, through  $\alpha$ , and the background shapes and yields, are included in the fit as Gaussian constraints on the relevant parameters. The dominant systematic uncertainty comes from the calibration of the hadronic trigger efficiency, which is shared through auxiliary parameters among the normalisation channels, and also with the misidentified  $D^0 \rightarrow \pi^+\pi^-$  yields that depend on the same estimate. The fit procedure is tested with pseudoexperiments. The values of the floating shape parameters are obtained from the data fit. Unbiased estimates of the branching fraction with correct coverage are obtained.

The  $m(\mu^+\mu^-)$  and  $\Delta m$  distributions in data are shown for the most sensitive BDT interval in Fig. 2 and for all intervals in Ref. [74], overlaid with the result of the fit. The data are consistent with the expected background. The value obtained for the  $D^0 \rightarrow \mu^+\mu^-$  branching fraction is  $\mathcal{B}(D^0 \rightarrow \mu^+\mu^-) = (1.7 \pm 1.0) \times 10^{-9}$ , corresponding to  $79 \pm 45$  signal decays. The significance of this signal is estimated comparing the test statistics in data with the distribution of the test statistics in background-only pseudoexperiments, and is found to have a  $p$ -value of 0.068, corresponding to a significance of  $1.5\sigma$  (see also Ref. [74]). An upper limit on the branching fraction is derived using the frequentist  $\text{CL}_s$  method [77] as implemented in the GAMMACOMBO framework [78, 79]. This yields

$$\mathcal{B}(D^0 \rightarrow \mu^+\mu^-) < 3.1 (3.5) \times 10^{-9} \text{ at } 90 (95)\% \text{ CL} .$$

The observed limit is larger than the one expected from background-only pseudoexperiments,  $\mathcal{B}(D^0 \rightarrow \mu^+\mu^-) < 1.9 (2.3) \times 10^{-9}$  at a 90 (95)% CL, coherently with the central value for the signal branching fraction.

The fit is repeated with different configurations: allowing the resolution of the misidentified  $D^0 \rightarrow \pi^+\pi^-$  background to vary, using a double exponential function in place of a single one for the combinatorial background, and reducing the range in the  $\Delta m$  variable. No significant change was found in the signal branching fraction with any configuration.



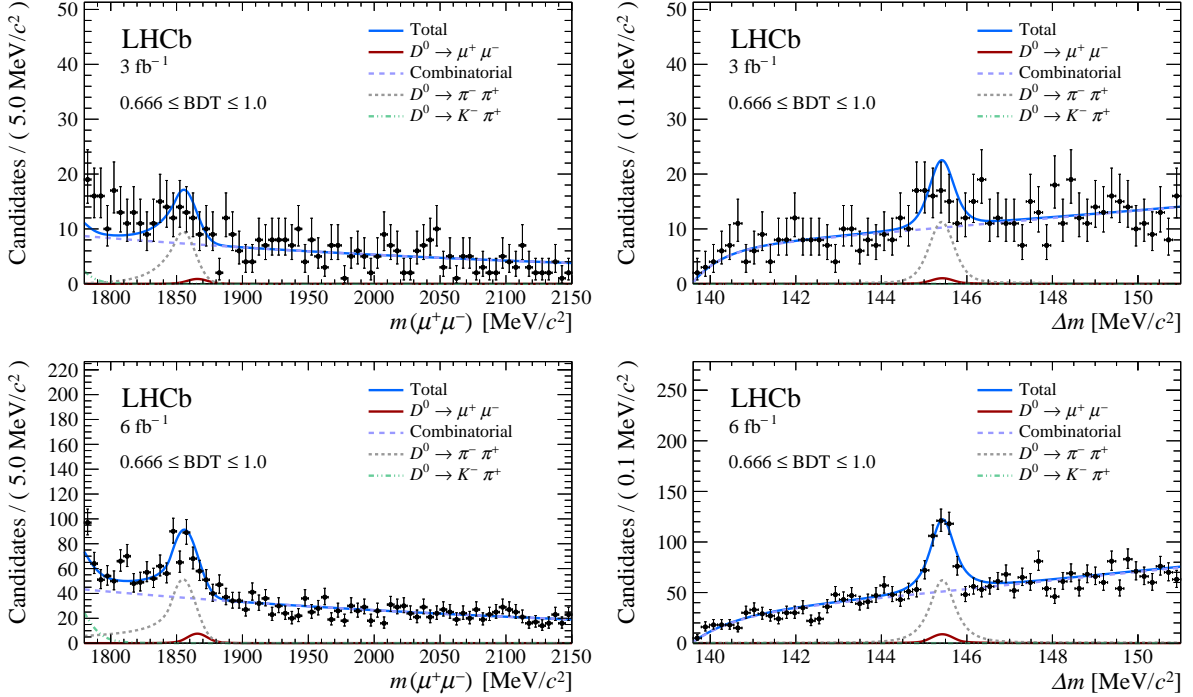


Figure 2: Distribution of (left)  $m(\mu^+\mu^-)$  and (right)  $\Delta m$  for the  $D^0 \rightarrow \mu^+\mu^-$  candidates in data from (top) Run 1 and (bottom) Run 2, for the most sensitive BDT interval. The distribution is superimposed with the fit to data. Each of the two distributions is in the signal region of the other variable, see text for details. Untagged and tagged decays are included in a single component for signal and  $D^0 \rightarrow \pi^+\pi^-$  background.

In summary, a search for the  $D^0 \rightarrow \mu^+\mu^-$  decay in data corresponding to  $9 \text{ fb}^{-1}$  of  $pp$  collision data collected by the LHCb experiment is performed. No excess with respect to the background expectation has been found and an upper limit of  $\mathcal{B}(D^0 \rightarrow \mu^+\mu^-) < 3.1 \times 10^{-9}$  at 90% CL has been set. This result represents an improvement of more than a factor two with respect to the previous LHCb result. This measurement constitutes the most stringent limit on the relevant FCNC couplings in the charm sector, allowing to set additional constraints on physics models beyond the SM which predict the branching fractions of  $D^0 \rightarrow \mu^+\mu^-$  and describe results from  $B$  physics measurements.

## References

- [1] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, Phys. Rev. **D2** (1970) 1285.
- [2] G. Burdman, E. Golowich, J. L. Hewett, and S. Pakvasa, *Rare charm decays in the Standard Model and beyond*, Phys. Rev. **D66** (2002) 014009, arXiv:hep-ph/0112235.
- [3] Belle collaboration, N. K. Nisar *et al.*, *Search for the rare decay  $D^0 \rightarrow \gamma\gamma$  at Belle*, Phys. Rev. **D93** (2016) 051102, arXiv:1512.02992.
- [4] E. Golowich, J. Hewett, S. Pakvasa, and A. A. Petrov, *Relating  $D^0 - \bar{D}^0$  mixing and  $D^0 \rightarrow l^+l^-$  with new physics*, Phys. Rev. **D79** (2009) 114030, arXiv:0903.2830.

- [5] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with  $B^0 \rightarrow K^{*0} \ell^+ \ell^-$  decays*, JHEP **08** (2017) 055, [arXiv:1705.05802](#).
- [6] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, [arXiv:2103.11769](#), to appear in Nature Physics.
- [7] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality using  $\Lambda_b^0 \rightarrow p K^- \ell^+ \ell^-$  decays*, JHEP **05** (2020) 040, [arXiv:1912.08139](#).
- [8] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the  $B^+ \rightarrow K^{*+} \mu^+ \mu^-$  decay*, Phys. Rev. Lett. **126** (2021) 161802, [arXiv:2012.13241](#).
- [9] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP-averaged observables in the  $B^0 \rightarrow K^{*0} \mu^+ \mu^-$  decay*, Phys. Rev. Lett. **125** (2020) 011802, [arXiv:2003.04831](#).
- [10] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the  $B^0 \rightarrow K^{*0} \mu^+ \mu^-$  decay using  $3 \text{ fb}^{-1}$  of integrated luminosity*, JHEP **02** (2016) 104, [arXiv:1512.04442](#).
- [11] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the rare decay  $B_s^0 \rightarrow \phi \mu^+ \mu^-$* , JHEP **11** (2021) 043, [arXiv:2107.13428](#).
- [12] ATLAS collaboration, M. Aaboud *et al.*, *Angular analysis of  $B_d^0 \rightarrow K^{*0} \mu^+ \mu^-$  decays in pp collisions at  $\sqrt{s} = 8 \text{ TeV}$  with the ATLAS detector*, JHEP **10** (2018) 047, [arXiv:1805.04000](#).
- [13] BaBar collaboration, B. Aubert *et al.*, *Measurements of branching fractions, rate asymmetries, and angular distributions in the rare decays  $B \rightarrow K \ell^+ \ell^-$  and  $B \rightarrow K^* \ell^+ \ell^-$* , Phys. Rev. **D73** (2006) 092001, [arXiv:hep-ex/0604007](#).
- [14] BaBar collaboration, J. P. Lees *et al.*, *Measurement of angular asymmetries in the decays  $B \rightarrow K^* \ell^+ \ell^-$* , Phys. Rev. **D93** (2016) 052015, [arXiv:1508.07960](#).
- [15] Belle collaboration, J.-T. Wei *et al.*, *Measurement of the differential branching Fraction and forward-backward Asymmetry for  $B \rightarrow K^{(*)} \ell^+ \ell^-$* , Phys. Rev. Lett. **103** (2009) 171801, [arXiv:0904.0770](#).
- [16] Belle collaboration, S. Wehle *et al.*, *Lepton-flavor-dependent angular analysis of  $B \rightarrow K^* \ell^+ \ell^-$* , Phys. Rev. Lett. **118** (2017) 111801, [arXiv:1612.05014](#).
- [17] CDF collaboration, T. Aaltonen *et al.*, *Measurements of the angular distributions in the decays  $B \rightarrow K^{(*)} \mu^+ \mu^-$  at CDF*, Phys. Rev. Lett. **108** (2012) 081807, [arXiv:1108.0695](#).
- [18] CMS collaboration, V. Khachatryan *et al.*, *Angular analysis of the decay  $B^0 \rightarrow K^{*0} \mu^+ \mu^-$  from pp collisions at  $\sqrt{s} = 8 \text{ TeV}$* , Phys. Lett. **B753** (2016) 424, [arXiv:1507.08126](#).
- [19] CMS collaboration, A. M. Sirunyan *et al.*, *Measurement of angular parameters from the decay  $B^0 \rightarrow K^{*0} \mu^+ \mu^-$  in proton-proton collisions at  $\sqrt{s} = 8 \text{ TeV}$* , Phys. Lett. **B781** (2018) 517, [arXiv:1710.02846](#).

- [20] LHCb collaboration, R. Aaij *et al.*, *Measurements of the S-wave fraction in  $B^0 \rightarrow K^+\pi^-\mu^+\mu^-$  decays and the  $B^0 \rightarrow K^*(892)^0\mu^+\mu^-$  differential branching fraction*, JHEP **11** (2016) 047, Erratum *ibid.* **04** (2017) 142, [arXiv:1606.04731](#).
- [21] LHCb collaboration, R. Aaij *et al.*, *Branching fraction measurements of the rare  $B_s^0 \rightarrow \phi\mu^+\mu^-$  and  $B_s^0 \rightarrow f_2'(1525)\mu^+\mu^-$  decays*, Phys. Rev. Lett. **127** (2021) 151801, [arXiv:2105.14007](#).
- [22] LHCb collaboration, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of  $B \rightarrow K^{(*)}\mu^+\mu^-$  decays*, JHEP **06** (2014) 133, [arXiv:1403.8044](#).
- [23] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of  $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$  decays*, JHEP **06** (2015) 115, Erratum *ibid.* **09** (2018) 145, [arXiv:1503.07138](#).
- [24] M. Bauer and M. Neubert, *Minimal leptoquark explanation for the  $R_{D^{(*)}}$ ,  $R_K$ , and  $(g-2)_g$  anomalies*, Phys. Rev. Lett. **116** (2016) 141802, [arXiv:1511.01900](#).
- [25] M. Freytsis, Z. Ligeti, and J. T. Ruderman, *Flavor models for  $\bar{B} \rightarrow D^{(*)}\tau\bar{\nu}$* , Phys. Rev. **D92** (2015) 054018, [arXiv:1506.08896](#).
- [26] R. Benbrik and C.-H. Chen, *Leptoquark on  $P \rightarrow \ell^+\nu$ , FCNC and LFV*, Phys. Lett. **B672** (2009) 172, [arXiv:0807.2373](#).
- [27] S. Fajfer and N. Košnik, *Leptoquarks in flavor changing neutral current charm decays*, Phys. Rev. **D79** (2009) 017502.
- [28] S. de Boer and G. Hiller, *Flavor and new physics opportunities with rare charm decays into leptons*, Phys. Rev. **D93** (2016) 074001, [arXiv:1510.00311](#).
- [29] D. Bečirević, N. Košnik, O. Sumensari, and R. Zukanovich Funchal, *Palatable leptoquark scenarios for lepton flavor violation in exclusive  $b \rightarrow s\ell_1\ell_2$  modes*, JHEP **11** (2016) 035, [arXiv:1608.07583](#).
- [30] Y. Cai, J. Gargalionis, M. A. Schmidt, and R. R. Volkas, *Reconsidering the One Leptoquark solution: flavor anomalies and neutrino mass*, JHEP **10** (2017) 047, [arXiv:1704.05849](#).
- [31] K. Kowalska, E. M. Sessolo, and Y. Yamamoto, *Constraints on charmphilic solutions to the muon  $g-2$  with leptoquarks*, Phys. Rev. **D99** (2019) 055007, [arXiv:1812.06851](#).
- [32] I. Bigaran, J. Gargalionis, and R. R. Volkas, *A near-minimal leptoquark model for reconciling flavour anomalies and generating radiative neutrino masses*, JHEP **10** (2019) 106, [arXiv:1906.01870](#).
- [33] I. Doršner *et al.*, *Physics of leptoquarks in precision experiments and at particle colliders*, Phys. Rept. **641** (2016) 1, [arXiv:1603.04993](#).
- [34] I. Doršner, S. Fajfer, and M. Patra, *A comparative study of the  $S_1$  and  $U_1$  leptoquark effects in the light quark regime*, Eur. Phys. J. **C80** (2020) 204, [arXiv:1906.05660](#).

- [35] I. Doršner, S. Fajfer, and O. Sumensari, *Muon  $g - 2$  and scalar leptoquark mixing*, JHEP **06** (2020) 089, arXiv:1910.03877.
- [36] K. Kowalska, E. M. Sessolo, and Y. Yamamoto, *Flavor anomalies from asymptotically safe gravity*, Eur. Phys. J. **C81** (2021) 272, arXiv:2007.03567.
- [37] M. Bordone, O. Catà, T. Feldmann, and R. Mandal, *Constraining flavour patterns of scalar leptoquarks in the effective field theory*, JHEP **03** (2021) 122, arXiv:2010.03297.
- [38] F. Abudinén, T. Blake, U. Egede, and T. Gershon, *Prospects for studies of  $D^{*0} \rightarrow \mu^+ \mu^-$  and  $B_{(s)}^{*0} \rightarrow \mu^+ \mu^-$  decays*, Eur. Phys. J. C **82** (2022) 459, arXiv:2202.03916.
- [39] G. Burdman and I. Shipsey,  *$D^0 - \bar{D}^0$  mixing and rare charm decays*, Ann. Rev. Nucl. Part. Sci. **53** (2003) 431, arXiv:hep-ph/0310076.
- [40] R.-M. Wang *et al.*, *Decays  $D_{(s)}^+ \rightarrow \pi(K)^+ \ell^+ \ell^-$  and  $D^0 \rightarrow \ell^+ \ell^-$  in the MSSM with and without  $R$ -parity*, Int. J. Mod. Phys. **A30** (2015) 1550063, arXiv:1409.0181.
- [41] W. Altmannshofer, P. S. B. Dev, A. Soni, and Y. Sui, *Addressing  $R_{D^{(*)}}$ ,  $R_{K^{(*)}}$ , muon  $g - 2$  and ANITA anomalies in a minimal  $R$ -parity violating supersymmetric framework*, Phys. Rev. **D102** (2020) 015031, arXiv:2002.12910.
- [42] Z. Kang and Y. Shigekami,  *$(g - 2)_\mu$  versus flavor changing neutral current induced by the light  $(B - L)_{\mu\tau}$  boson*, JHEP **11** (2019) 049, arXiv:1905.11018.
- [43] A. K. Alok, A. Dighe, S. Gangal, and D. Kumar, *Predictions for  $B_s \rightarrow \bar{K}^* \ell \ell$  in non-universal  $Z'$  models*, Eur. Phys. J. **C80** (2020) 682, arXiv:1912.02052.
- [44] D. Guadagnoli, M. Reboud, and O. Sumensari, *A gauged horizontal  $SU(2)$  symmetry and  $R_{K^{(*)}}$* , JHEP **11** (2018) 163, arXiv:1807.03285.
- [45] R. Alonso *et al.*, *A clockwork solution to the flavor puzzle*, JHEP **10** (2018) 099, arXiv:1807.09792.
- [46] A. K. Alok *et al.*, *New-physics signals of a model with a vector-singlet up-type quark*, Phys. Rev. **D92** (2015) 013002, arXiv:1504.00517.
- [47] K. Ishiwata, Z. Ligeti, and M. B. Wise, *New vector-like fermions and flavor physics*, JHEP **10** (2015) 027, arXiv:1506.03484.
- [48] S. Fajfer and N. Kosnik, *Prospects of discovering new physics in rare charm decays*, Eur. Phys. J. **C75** (2015) 567, arXiv:1510.00965.
- [49] A. Bharucha, D. Boito, and C. Méaux, *Disentangling QCD and new physics in  $D^+ \rightarrow \pi^+ \ell^+ \ell^-$* , JHEP **04** (2021) 158, arXiv:2011.12856.
- [50] LHCb collaboration, R. Aaij *et al.*, *Observation of the mass difference between neutral charm-meson eigenstates*, Phys. Rev. Lett. **127** (2021) 111801, arXiv:2106.03744.
- [51] LHCb collaboration, R. Aaij *et al.*, *Search for the rare decay  $D^0 \rightarrow \mu^+ \mu^-$* , Phys. Lett. **B725** (2013) 15, arXiv:1305.5059.

- [52] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.
- [53] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.
- [54] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820; T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.
- [55] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.
- [56] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.
- [57] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.
- [58] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.
- [59] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.
- [60] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and physics documentation*, Comp. Phys. Comm. **199** (2016) 86, arXiv:1011.0937.
- [61] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, arXiv:physics/0503191.
- [62] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.
- [63] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.
- [64] H. Voss, A. Hoecker, J. Stelzer, and F. Tegenfeldt, *TMVA - Toolkit for Multivariate Data Analysis with ROOT*, PoS **ACAT** (2007) 040.
- [65] 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.
- [66] 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** (2019) 1, arXiv:1803.00824.
- [67] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, arXiv:physics/0308063.

- [68] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of  $b$ -hadron,  $c$ -hadron, and  $\tau$ -lepton properties as of 2018*, Eur. Phys. J. **C81** (2021) 226, [arXiv:1909.12524](https://arxiv.org/abs/1909.12524), updated results and plots available at <https://hflav.web.cern.ch>.
- [69] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, [arXiv:1408.1251](https://arxiv.org/abs/1408.1251).
- [70] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039, CERN, Geneva, 2014.
- [71] LHCb collaboration, R. Aaij *et al.*, *Analysis of neutral  $B$ -meson decays into two muons*, Phys. Rev. Lett. **128** (2022) 041801, [arXiv:2108.09284](https://arxiv.org/abs/2108.09284).
- [72] LHCb collaboration, R. Aaij *et al.*, *Measurement of the  $B_s^0 \rightarrow \mu^+\mu^-$  decay properties and search for the  $B^0 \rightarrow \mu^+\mu^-$  and  $B_s^0 \rightarrow \mu^+\mu^-\gamma$  decays*, Phys. Rev. **D105** (2022) 012010, [arXiv:2108.09283](https://arxiv.org/abs/2108.09283).
- [73] A. Martin Sanchez, P. Robbe, and M.-H. Schune, *Performances of the LHCb L0 Calorimeter Trigger*, LHCb-PUB-2011-026, CERN-LHCb-PUB-2011-026, CERN, Geneva, 2012.
- [74] See Supplemental Material at the end of this Letter for additional figures.
- [75] 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.
- [76] N. L. Johnson, *Systems of frequency curves generated by methods of translation*, Biometrika **36** (1949) 149.
- [77] A. L. Read, *Presentation of search results: The  $CL_S$  technique*, J. Phys. **G28** (2002) 2693.
- [78] M. Kenzie *et al.*, *GammaCombo: A statistical analysis framework for combining measurements, fitting datasets and producing confidence intervals*, doi: 10.5281/zenodo.3371421.
- [79] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CKM angle  $\gamma$  from a combination of LHCb results*, JHEP **12** (2016) 087, [arXiv:1611.03076](https://arxiv.org/abs/1611.03076).

# Supplemental Material

This section contains the additional figures mentioned in the main text. Figures 3 and 4 present the same data shown in Figure 1 in the main body, but for all the BDT intervals. Figures 5 and 6 represent the same distributions and subdivisions of the previous plots, but without restricting each to the signal region of the other variable, *i.e.* they contain the full data used for the signal search. Figure 7 shows the data in the two-dimensional plane of the  $m(\mu^+\mu^-)$  and  $\Delta m$  variables, as well as the signal regions in each variable. Figure 8 displays the result of the calibration of the BDT output described in the text. Figure 9 shows the test on the particle identification variable mentioned in the main body. Finally, Figure 10 shows the value of the  $CL_s$  estimator used to compute the upper limit on the  $D^0 \rightarrow \mu^+\mu^-$  branching fraction, as a function of the branching fraction itself.

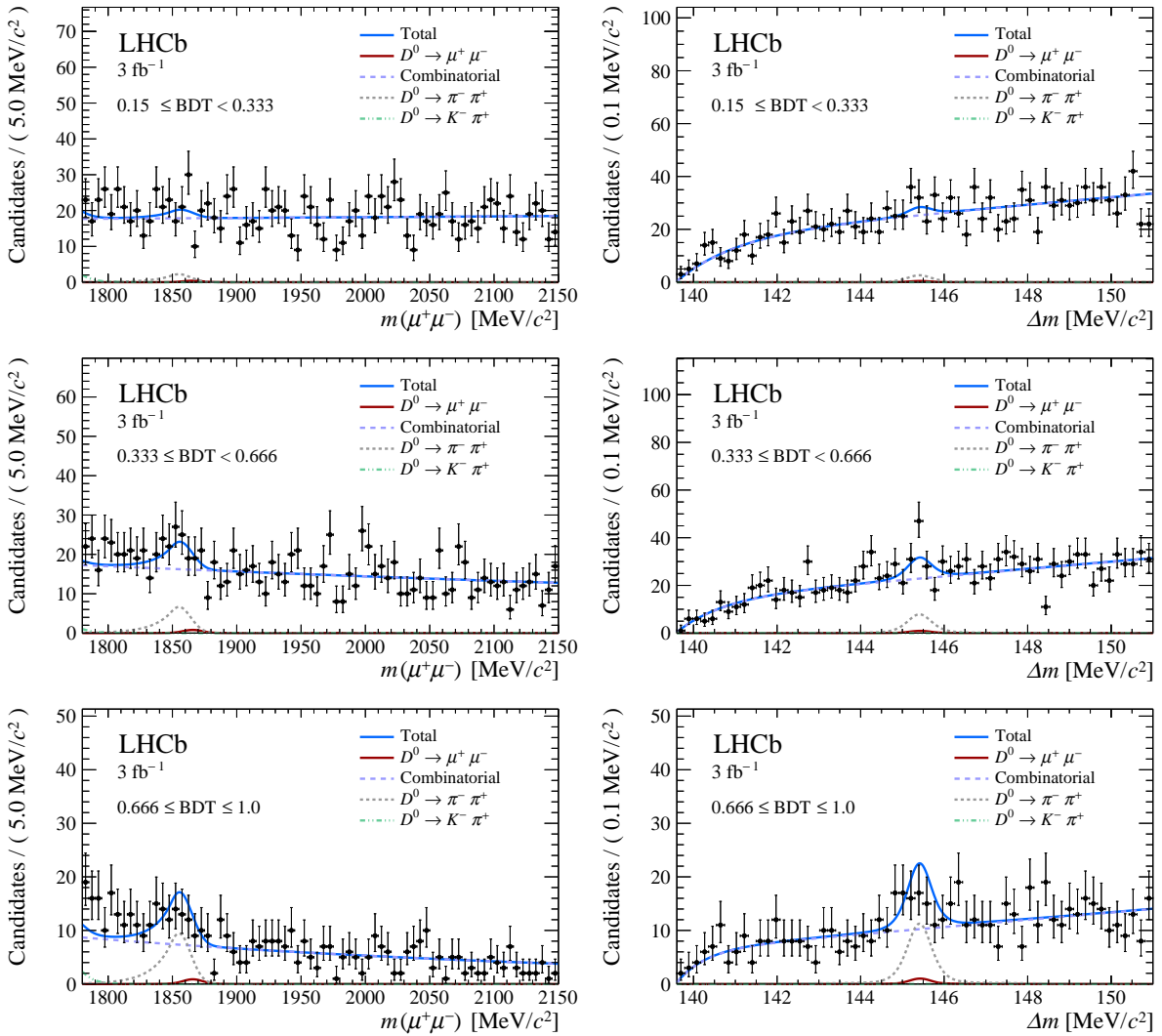


Figure 3: Distributions of (left)  $m(\mu^+\mu^-)$  and (right)  $\Delta m$  for the  $D^0 \rightarrow \mu^+\mu^-$  candidates for Run 1 data in, top to bottom, the three BDT intervals. The distributions are superimposed with the fit to data. Each of the two distributions is in the signal region of the other variable, see text for details. Untagged and tagged decays are included in a single component for signal and  $D^0 \rightarrow \pi^+\pi^-$  background.

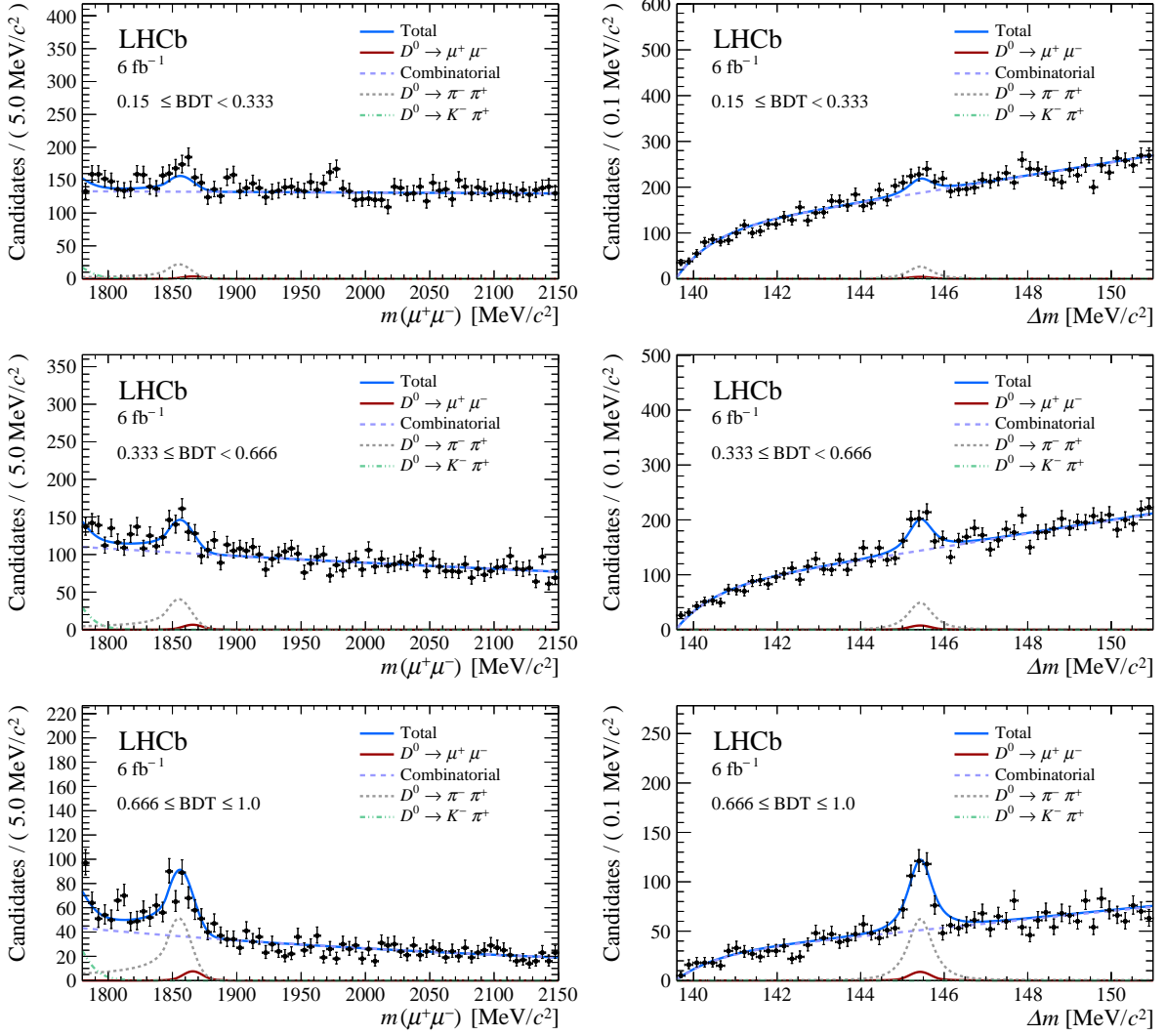


Figure 4: Distributions of (left)  $m(\mu^+\mu^-)$  and (right)  $\Delta m$  for the  $D^0 \rightarrow \mu^+\mu^-$  candidates for Run 2 data in, top to bottom, the three BDT bins. The distributions are superimposed with the fit to data. Each of the two distributions is in the signal region of the other variable, see text for details. Untagged and tagged decays are included in a single component for signal and  $D^0 \rightarrow \pi^+\pi^-$  background.



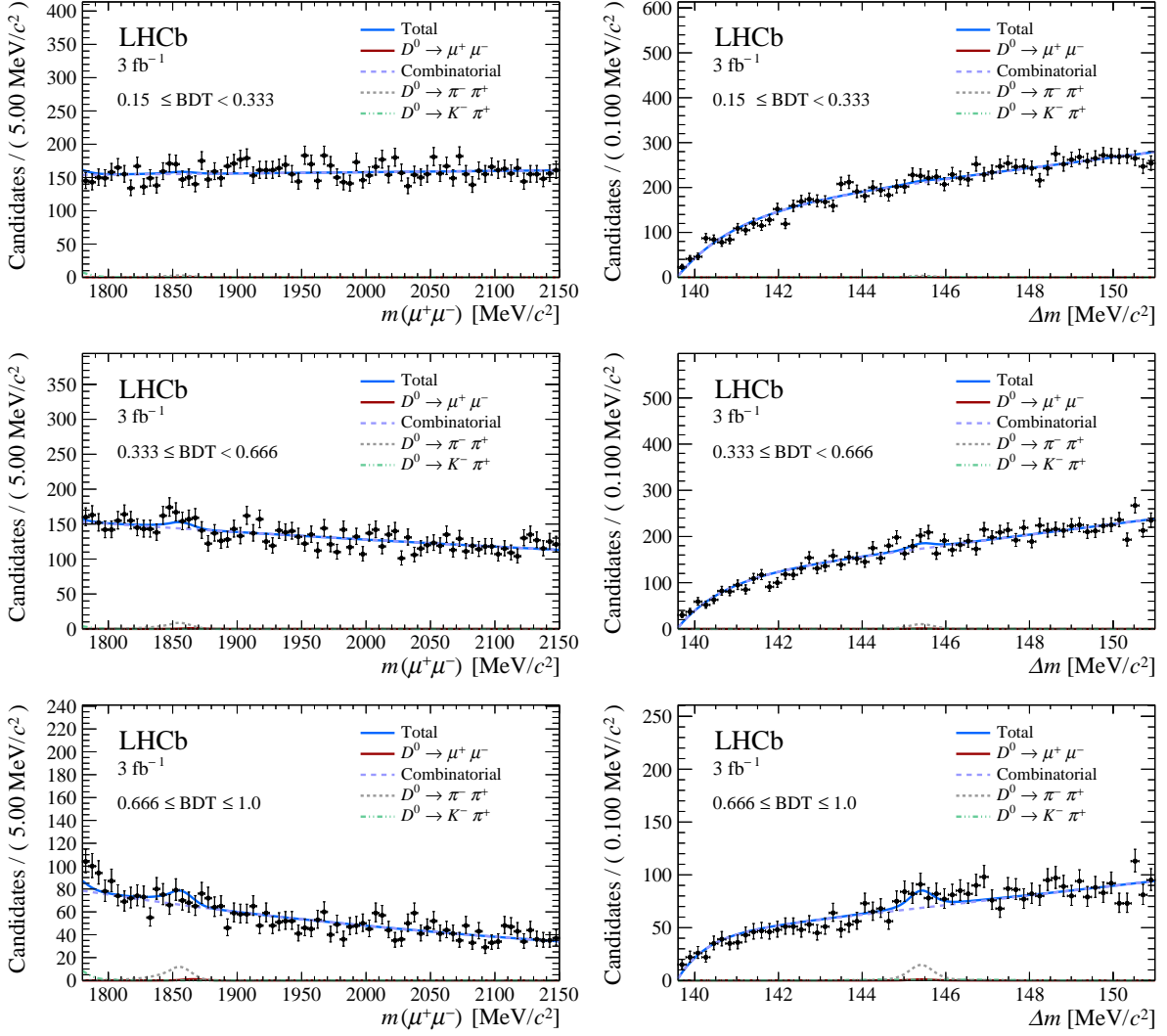


Figure 5: Distributions of (left)  $m(\mu^+\mu^-)$  and (right)  $\Delta m$  for the  $D^0 \rightarrow \mu^+\mu^-$  candidates for Run 1 data in, top to bottom, the three BDT bins. The distributions are superimposed with the fit to data. Untagged and tagged decays are included in a single component for signal and  $D^0 \rightarrow \pi^+\pi^-$  background. Unlike the correspondent figures in the main body of the Letter, here all events are shown.

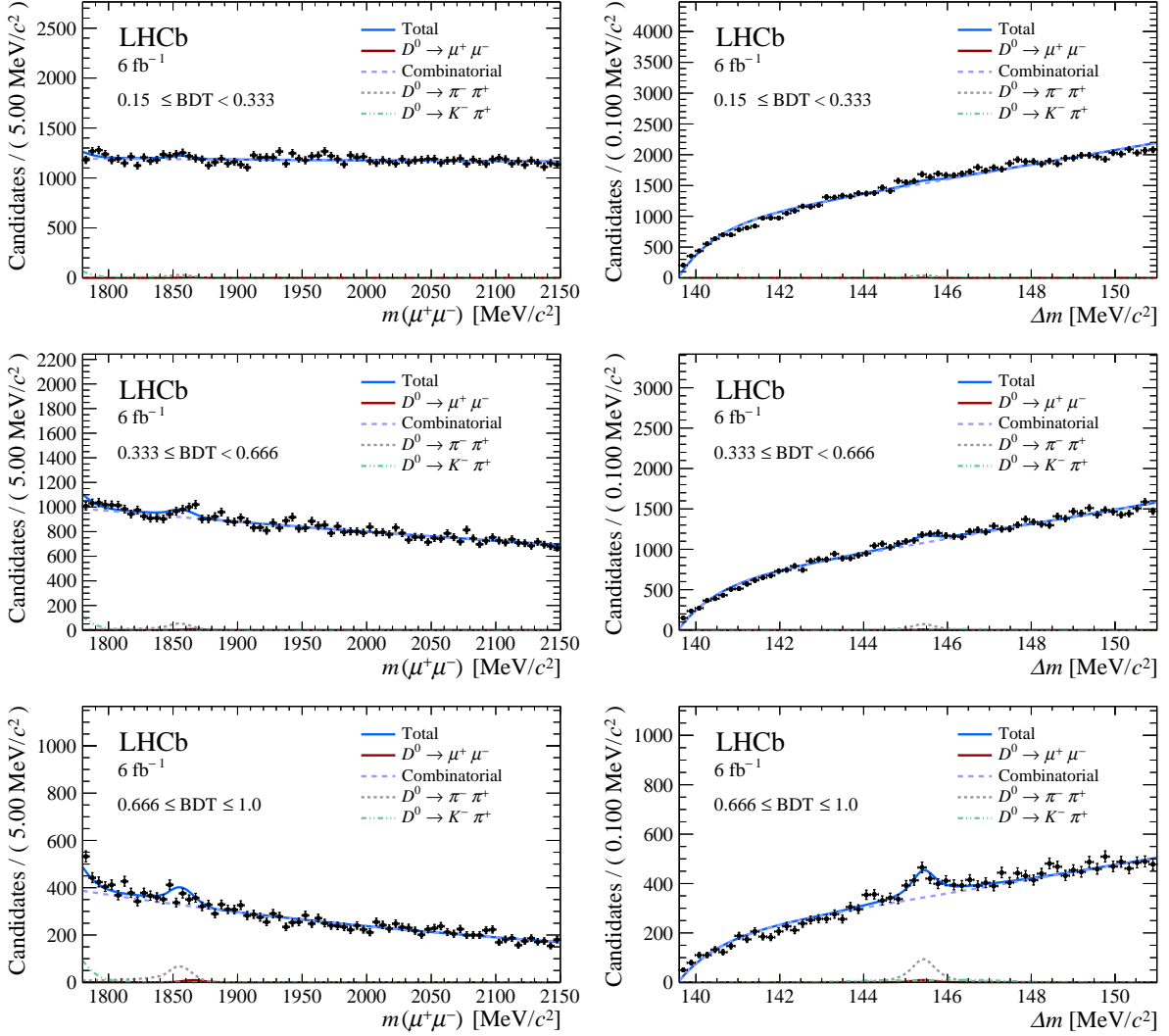


Figure 6: Distributions of (left)  $m(\mu^+\mu^-)$  and (right)  $\Delta m$  for the  $D^0 \rightarrow \mu^+\mu^-$  candidates for Run 2 data in, top to bottom, the three BDT bins. The distributions are superimposed with the fit to data. Untagged and tagged decays are included in a single component for signal and  $D^0 \rightarrow \pi^+\pi^-$  background. Unlike the correspondent figures in the main body of the Letter, here all events are shown.

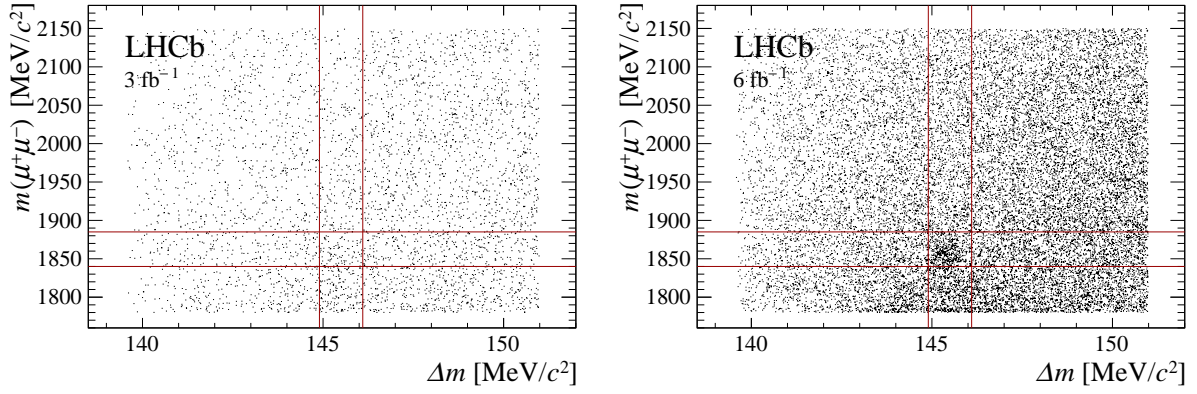


Figure 7: Two-dimensional distribution of  $m(D^0)$  versus  $\Delta m$  for the most sensitive BDT interval in the (left) Run 1 and (right) Run 2 data. Red lines represent the signal region of each variable as defined in the text.

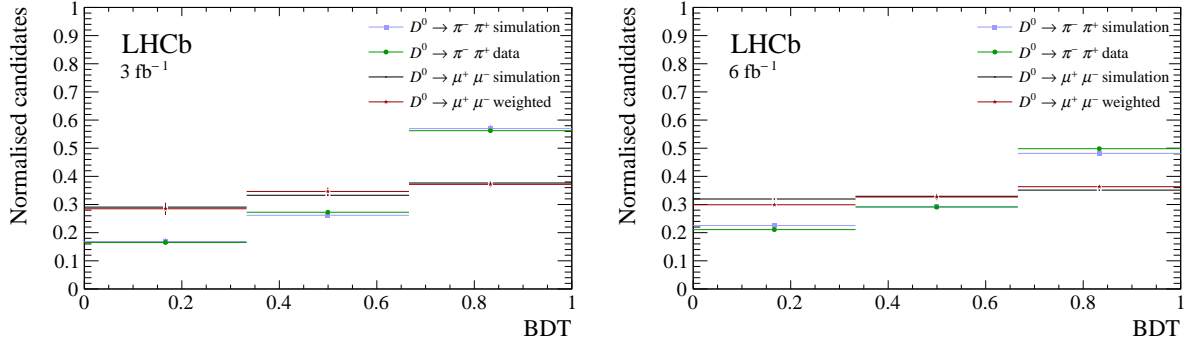


Figure 8: Distribution of the BDT output. The  $D^0 \rightarrow \pi^+ \pi^-$  candidates in data and simulation are shown and used to weight the simulated  $D^0 \rightarrow \mu^+ \mu^-$  decays distribution used for the search.

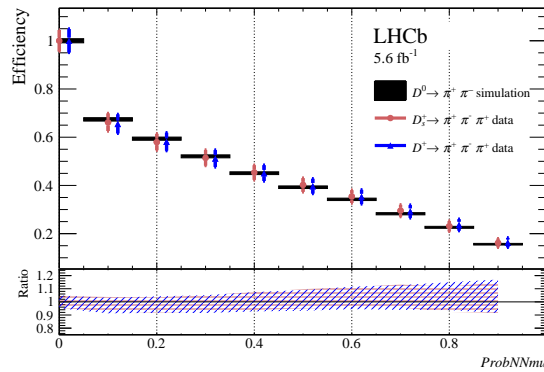


Figure 9: Efficiency for two pions to pass a requirement on the  $ProbNNmu$  variable (after additional requirements on its identification as muon) in Run 2, as measured from  $D^+ \rightarrow \pi^+ \pi^- \pi^+$  and  $D_s^+ \rightarrow \pi^+ \pi^- \pi^+$  decays in data and  $D^0 \rightarrow \pi^+ \pi^-$  decays in simulation. Same sign pions are used in data to avoid contamination for hadronic resonances decaying to two real muons.

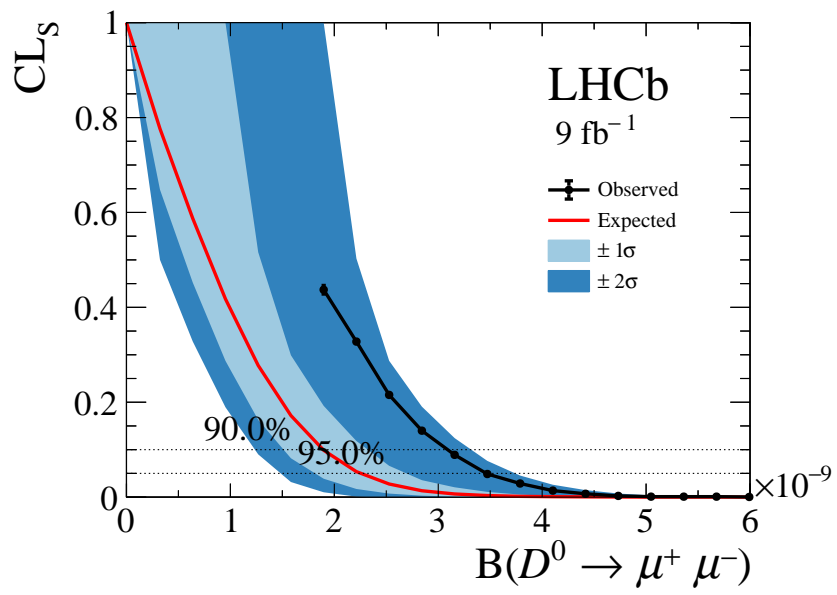


Figure 10: Value of  $CL_s$  as a function of the  $D^0 \rightarrow \mu^+ \mu^-$  branching fraction.