CMS Physics Analysis Summary

Contact: cms-pag-conveners-bphysics@cern.ch 2024/07/23

Search for Rare Charm Decays Into Two Muons

The CMS Collaboration

Abstract

A search for the rare decay $D^0 \to \mu^+\mu^-$ was conducted using $\sqrt{s} = 13.6$ TeV proton-proton collision data collected by the CMS experiment during the 2022-2023 operation of the CERN LHC, with an integrated luminosity of 64.5 fb^{-1}. The search exploits a new high-rate trigger that records events with two muons with transverse momenta as low as 4 and 3 GeV. Rare decays of the charm quark, which are less explored than rare bottom quark decays, offer a unique way to probe for new physics effects beyond the reach of modern colliders. No significant excess above the expected background was observed, leading to an upper limit of $\mathcal{B}(D^0 \to \mu^+\mu^-) < 2.6 \times 10^{-9}$ at the 95% confidence level. This is the most sensitive measurement to date, imposing additional constraints on new physics models.

1 Introduction

Studies of rare decays of hadrons have long been considered one of the most promising methods for searching for new physics. Small contributions from new physics effects are more easily detected against the backdrop of smaller Standard Model (SM) contributions. Decays mediated by flavor-changing neutral currents, which are forbidden at leading order, have been studied in various experiments [1–19]. These studies have primarily focused on rare decays of bottom and strange hadrons involving $b \to s$ and $s \to d$ currents. In contrast, rare decays of charmed hadrons that are mediated by the $c \to u$ transition, which tend to have smaller SM predictions, have been less popular and consequently less studied. In this note, charge-conjugated decay modes are implied when decays are mentioned.

The primary distinction between rare decays of charmed and those of bottom and strange hadrons is that the loop contributions in charm decays are mediated by lighter quarks. This leads to substantial long-distance contributions, which are challenging to predict analytically. As a result, the SM predictions become less reliable, limiting the interesting regions for new physics searches mainly to contributions that substantially exceed those predicted by the SM.

The rare decay of the D^0 meson into two muons is one such interesting case. Various new physics models [20–22], including those involving leptoquarks, R-parity violating supersymmetry, and extra fermions or gauge bosons, predict enhancements in this decay rate. The SM prediction for the branching fraction is of order of 3×10^{-13} [22]. The most sensitive experimental search to date, which was conducted by the LHCb collaboration, set an upper limit for $\mathcal{B}(D^0 \to \mu^+\mu^-)$ at 3.5×10^{-9} at 95% confidence level (CL) [23]. This value is four orders of magnitude larger than the SM prediction, indicating a substantial unexplored territory.

In this note, we report on the measurement of the $D^0 \to \mu^+\mu^-$ branching fraction based on proton-proton collision data collected by the CMS experiment in 2022–2023 at a center-of-mass energy of 13.6 TeV, corresponding to an integrated luminosity of $64.5\,\text{fb}^{-1}$. This is the first publication utilizing the high-rate double muon trigger [24] developed for CMS Run 3 and the first public result of the $D^0 \to \mu^+\mu^-$ branching fraction measurement at CMS.

2 Analysis strategy

One of the main challenges in the search for $D^0 \to \mu^+\mu^-$ decays is identifying a subtle signal among a large number of background events. We look for the signal in the cascade decays of the $D^*(2010)^\pm$ mesons, i.e., $D^{*+} \to D^0\pi^+$. Though this approach reduces the event yield by about a factor of four [25] relative to inclusive D^0 production, it reduces the dominant combinatorial background by orders of magnitude. It also allows the signal extraction using two-dimensional unbinned maximum likelihood (UML) fit on the D^0 candidate mass $m(D^0)$ and the mass difference between the D^{*+} and the daughter D^0 meson $\Delta m = m(D^{*+}) - m(D^0)$. The mass difference Δm is a powerful observable for suppressing the background as it yields a narrow distribution by canceling out the muon resolution effect. In addition to the combinatorial background, signal-like signatures can be caused by misidentified muons originating from the decay in flight of pions and kaons. While their contributions tend to be negligible compared to the dominant background, they require a reliable estimation of the misidentification rates to avoid biasing the signal extraction.

Given the limited precision of the D^* production cross section at the LHC, a direct extraction of the branching fractions of the $D^0 \to \mu^+\mu^-$ decay would be affected by a large uncertainty. Therefore, the branching fraction is estimated with respect to the $D^{*+} \to D^0\pi^+$, $D^0 \to \pi^+\pi^-$

decays (normalization channel), which are collected with the "zero bias" triggers, where the only requirement is a proton-proton bunch crossing. We determine the branching ratio for $D^0 \to \mu^+\mu^-$ using

$$\mathcal{B}(D^0 \to \mu^+\mu^-) = \mathcal{B}(D^0 \to \pi^+\pi^-)\frac{N_{D^0\to\mu^+\mu^-}}{N_{D^0\to\pi^+\pi^-}}\frac{\varepsilon_{D^0\to\pi^+\pi^-}}{\varepsilon_{D^0\to\mu^+\mu^-}} \quad (1)$$

where N_x is the number of fitted candidates for the decay x, and ε_x is the corresponding full selection efficiency.

The $D^0 \to K^-\pi^+$ process is not used as it has a larger kinematic difference from the signal and is only employed for cross-checking the $D^0 \to \pi^+\pi^-$ normalization.

Another advantage of measuring the branching fractions with respect to the normalization channel is that it allows for a cancellation of many systematic uncertainties in the selection and reconstruction efficiencies of the signal. This becomes particularly valuable in addressing uncertainties caused by the mismodeling of the D meson production contribution from $c\bar{c}$ and $b\bar{b}$. The selection efficiencies are measured using control samples in data, while MC simulations are used to take into account differences between the final states.

The displaced D^0 vertex is one of the most powerful discriminators for separating signal decays from the combinatorial background. Although the mean lifetime of the D^0 is relatively short ($c\tau = 0.123$ mm) when compared to the vertex resolution, $D^{*+} \to D^0\pi^+$ decays originating from B-hadrons tend to have a more displaced D^0 vertex, thanks to the significantly larger mean lifetime of B-hadrons ($c\tau$ ranges from 0.455 to 0.491 mm). On the other hand, prompt $D^{*\pm}$ reconstruction can benefit from vertexing the soft pion originating from the $D^{*\pm}$ meson with the primary vertex (PV), which improves the Δm resolution and reduces some of the combinatorial background. The analysis is optimized by considering both production mechanisms, focusing on achieving the best sensitivity for $D^0 \to \mu^+\mu^-$ signal extraction.

To reduce unintentional bias, the analysis employs a data blinding strategy consisting of hiding the branching fraction results in the fit and using 12.5% of unblinded data for the analysis development.

3 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter, each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity η coverage provided by the barrel and endcap detectors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [26].

The silicon strip tracker in combination with the pixel detector measured charged particles up to $|\eta| = 3.0$ with typical resolutions of 1.5% in p_T and 20–75 μm in the transverse impact parameter [27] for nonisolated particles with $1 < p_T < 10$ GeV. Muons are measured in the range $|\eta| < 2.4$, with detection planes made using three technologies: drift tubes (DTs), cathode strip chambers (CSCs), and resistive-plate chambers. Matching muons to tracks measured in

the silicon tracker results in a relative p_T resolution, for muons with p_T up to 100 GeV, of 1% in the barrel and 3% in the endcaps [28].

Events of interest are selected using a two-tiered trigger system. The first level (L1), composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of about 4 μs [29]. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to the order of kHz before data storage [30]. In CMS Run 3 a new double muon trigger was designed using the data parking technique [24], operating at high rates with low trigger thresholds.

4 Data and Monte Carlo simulation

The analysis is performed using two primary HLTs. Signal events are collected with the low-mass double muon trigger, requiring the minimum transverse muon momentum to be 4 GeV and 3 GeV for the leading and trailing muons, respectively. To select the reference $D^{*+} \to D^0\pi^+$, $D^0 \to \pi^+\pi^-$ events, we used a trigger that collected events from all colliding bunches, with an effective reduced rate (prescale) of the order of $(1.3 - 1.5) \times 10^6$.

We used multiple samples of MC simulated events to evaluate the signal efficiency, the detector response, and the background yields. The simulated event samples are generated with PYTHIA 8.212 [31] using the CP5 underlying event tune [32] and propagated through the CMS detector model using the GEANT4 [33] package. The decays of B hadrons are simulated using the EVTGEN 1.3.0 [34] program and final-state photon radiation is described using PHOTOS 3.56 [35].

Multiple interactions within the same or nearby bunch crossings (pileup) are simulated for all samples by overlapping the hard-scattering event with several minimum bias events, with a multiplicity similar to the one observed in data (averaging to 46 for 2022 and 52 for 2023).

5 Event reconstruction and selection

The candidate selection for $D^{*+} \to D^0\pi^+$ decays consists of two major parts: preselection, and final selection. The preselection includes efficient background rejection requirements that are largely independent of each other and constrained by trigger requirements. The final selection employs a multivariate analysis (MVA) to achieve optimal background rejection using several observables, except those used in the final fit.

The preselection begins with the reconstruction of $D^0 \to \mu^+\mu^-$ and $D^0 \to \pi^+\pi^-$ candidates. The selection criteria are kept as consistent as possible between the two channels to minimize the systematic uncertainty on the acceptance and efficiency for the branching fraction measurement. The D^0 trajectory is extrapolated to the beam line. The absolute distance between the closest point on this trajectory and the PV in 3D space is defined as the impact parameter. The PV with the smallest impact parameter is selected as the most suitable PV for the $D^{*+} \to D^0\pi^+$ decay candidate. Subsequently, a soft pion candidate is selected from the list of available charged tracks, and its compatibility with the PV is checked. No additional requirement is applied on the soft pion.

Each muon candidate must be reconstructed using both the "tracker muon" algorithm, which extrapolates a reconstructed track in the inner tracker to the muon system with loose matching

to DT or CSC segments, and the "global muon" algorithm, which combines compatible tracks from the muon system with those from the inner tracker. These candidates are required to pass the loose muon identification [28] and have a high-quality inner track [36]. This combination of requirements effectively rejects the majority of poorly reconstructed muon candidates. Using special MC simulation samples with simulated hit information available for all interactions, we estimate that the fraction of muon candidates not originating from heavy flavor decays or decays in flight is of the order of 0.2% and can be neglected.

The tracks of the D^0 candidates are refitted to a common vertex using the standard CMS kinematic vertex fitting package [37]. The PV is refitted using only those tracks that are tightly associated with the PV and the soft pion. The refitted PV corresponds to the D^* candidate vertex and is used to extract the $D^{*\pm}$ vertex probability. The refitted soft pion momentum is used for D^* candidate reconstruction.

A complete list of preselection requirements for $D^{*+} \to D^0\pi^+$ candidate selection is:

- $p_T > 4\,\text{GeV}$ and $|\eta| < 2.4$ for decay products of D^0
- D^0 reconstructed mass is in [1.75, 1.95] GeV
- D^0 refitted mass is in [1.81, 1.94] GeV
- reconstructed Δm is in [0.135, 0.160] GeV
- refitted Δm is in [0.140, 0.150] GeV
- refitted D^0 pointing angle less than 0.1
- D^0 vertex probability > 0.01 to match the trigger requirement, which is 0.005
- $D^{*\pm}$ vertex probability > 0.1
- D^0 flight length significance in 3D greater than 3

There can be multiple candidates reconstructed for the same signal event. Most of them have the wrong soft pion, which leads to an incorrect Δm estimation that looks like background. Such candidates are included in background modeling and will not bias signal extraction in the fit. The absolute efficiency after the preselection is 6.35×10^{-4} (7.30×10^{-4}) for simulations of the $D^{*+} \to D^0\pi^+, D^0 \to \mu^+\mu^-$ ($\pi^+\pi^-$) process.

6 Multivariate analysis

After applying the preselection requirements, some of the remaining signal-background discriminating observables show significant correlations. To achieve optimal performance, we employ a multivariate analysis to construct a single, more powerful discriminator (MVA_D). For this purpose, we use the XGBOOST library [38], which implements an advanced gradient boosting decision tree algorithm.

The training is performed on the simulated $D^0 \to \mu^+\mu^-$ signal events and the right side bands of the Δm distribution, $\Delta m \in [0.150, 0.155]$ GeV, in data, with mass window $m(D^0) \in [1.81, 2.45]$ GeV, as the background events. The contribution of true D^0 decays to the background is negligible and does not impact the training. The samples are divided into five groups and trained using k-fold cross-validation. This method ensures that we can classify all events in the data set and guarantees that no event is predicted by a classifier that was trained on the event itself.

The list of variables used as input to the MVA_D includes:

- leading muon p_T

- trailing muon p_{T}
- p_{T} of the soft pion
- D^0 pointing angle
- D^0 flight length significance in 3D
- D^0 vertex probability
- $\mathrm{D}^{*\pm}$ vertex probability
- the 3D impact parameter of the D^0 candidate with respect to the selected PV
- the significance of the 3D impact parameter of the D^0
- D^0 mass resolution over D^0 mass

Since our training is partially conducted on simulated signal events, with the potential mis-modeling of detector effects and D^* production, it cannot fully represent the true performance. To address this, we use $\mathrm{D}^0 \to \pi^+\pi^-$ zero bias data to evaluate the actual MVA_D performance. We apply MVA_D selection requirements in both the signal and normalization channels so that the major systematic uncertainties and mis-modeling effects are canceled out in the ratio. The MVA_D selection efficiencies are similar between the $\mathrm{D}^0 \to \mu^+\mu^-$ and $\mathrm{D}^0 \to \pi^+\pi^-$ processes and the residual difference is considered in the efficiency correction. The MVA_D selection has a 72% signal efficiency while rejecting 80% of the background events.

7 Maximum likelihood fits

In order to extract the results, we perform UML fits on both normalization and signal channels.

7.1 Normalization channel

The normalization channel $\mathrm{D}^{*+} \to \mathrm{D}^0\pi^+$, $\mathrm{D}^0 \to \pi^+\pi^-$ serves as an input to measure the signal branching fraction, as described by Eq. 1. The events are collected using the zero bias trigger, which has a fairly stable prescale at L1 and HLT making the conditions representative of the signal channel.

The yield of $\mathrm{D}^{*+} \to \mathrm{D}^0\pi^+$, $\mathrm{D}^0 \to \pi^+\pi^-$ is determined with a two-dimensional (2D) UML fit to the dipion invariant mass (denoted as $m_{\pi\pi}$) and the D^*–D^0 mass difference (denoted as $\Delta m = m_{(\pi\pi)\pi} - m_{\pi\pi}$).

The fit is implemented with `RooFit` tools [39] with the following components:

- signal $\mathrm{D}^{*+} \to \mathrm{D}^0\pi^+, \mathrm{D}^0 \to \pi^+\pi^-$ decay, denoted as "$\mathrm{D}^*, \mathrm{D}^0 \to \pi\pi$ ";
- background from $\mathrm{D}^{*+} \to \mathrm{D}^0\pi^+, \mathrm{D}^0 \to \mathrm{K}^-\pi^+$ process, denoted as "$\mathrm{D}^*, \mathrm{D}^0 \to \mathrm{K}\pi$ ";
- combinatorial background events;
- other non-D^* $\mathrm{D}^0 \to \pi^+\pi^-$ contributions, denoted as "non$-\mathrm{D}^*, \mathrm{D}^0 \to \pi\pi$ ";
- other non-D^* $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ contributions, denoted as "non$-\mathrm{D}^*, \mathrm{D}^0 \to \mathrm{K}\pi$ ";

The individual probability density functions (PDFs) for each component are described as follows:

- **Model for signal $\mathrm{D}^*, \mathrm{D}^0 \to \pi\pi$ events**: the shape for $m_{\pi\pi}$ is modeled with a sum of two Gaussians with a common mean, and the shape for Δm is a sum of three Gaussians, also with a common mean value. The PDF parameters are obtained from a fit to a large control sample of $\mathrm{D}^*, \mathrm{D}^0 \to \pi\pi$ events.

- **Background $D^*, D^0 \to K\pi$ model**: the $D^0 \to K^-\pi^+$ peak in $m_{\pi\pi}$ is shifted to the left-hand side due to an incorrect mass assignment in the $D^0 \to \pi^+\pi^-$ reconstruction. An exponential function and a Gaussian function are used to model the shape for $m_{\pi\pi}$ and Δm, respectively. The parameters are obtained from the fit to the simulated events and the exponential parameters are floated in the data fitting.
- **Combinatorial background model**: combinatorial background events are modeled with an exponential function in $m_{\pi\pi}$, with a reduced `RooDstD0BG` function [40]:

$$P(m \mid m_0, A, B, C) = \left(1 - \exp\left(-\frac{m - m_0}{C}\right)\right) \cdot \left(\frac{m}{m_0}\right)^A, \tag{2}$$

  where the linear term is dropped to improve the stability. The model parameter m_0 is the kinematic threshold and is fixed to the pion mass, and the shape parameters A and C are determined from fits to data.
- **Model for non$-D^*, D^0 \to \pi\pi$ events**: the model for $m_{\pi\pi}$ is in common with signal $D^*, D^0 \to \pi\pi$ events, but the shape for Δm is assumed to be the same as in the combinatorial events.
- **Model for non$-D^*, D^0 \to K\pi$ events**: similar to the non$-D^*, D^0 \to \pi\pi$, the $m_{\pi\pi}$ model is shared with $D^*, D^0 \to K\pi$, but Δm is assumed to be the same as the combinatorial component as well.

To determine the model parameters, we fit $D^*, D^0 \to \pi\pi$ events in a control sample collected with the double muon trigger in 2022 and 2023. The shape is then corrected based on the shape difference between the $D^*, D^0 \to K\pi$ control samples collected with the double muon trigger and the zero bias trigger. Since the non$-D^*, D^0 \to K\pi$ component turns out to be negligible, it is not included in the fitting configuration for the sake of stability. The $D^0 \to KK$ process is not considered as its reconstructed mass falls below 1.81 GeV under the dipion mass hypothesis.

Fitted event yields and graphs after the MVA_D selection are shown in Table 1 and Figure 1 respectively.

Table 1: Fitted yields with zero bias events after MVA_D selection.

Parameter	Fitted mean	Error
$D^*, D^0 \to \pi\pi$ yield	195	$-17/+18$
$D^*, D^0 \to K\pi$ yield	74	$-19/+23$
Combinatorial yield	140	$-19/+21$
non$-D^*, D^0 \to \pi\pi$ yield	0	Limit$/+8$

7.2 Signal channel

The decay branching fraction (and the associated upper limit) of $D^0 \to \mu^+\mu^-$ is extracted using a 2D UML fit on the invariant mass of D^0 meson candidates (denoted as $m_{\mu\mu}$) and D^*–D^0 mass (denoted as $\Delta m = m_{(\mu\mu)\pi} - m_{\mu\mu}$).

The fitting model is constructed using `RooFit` interfaced with the `Combine` tool [41], which is responsible for performing UML fit and upper limit evaluation. In the fit the following components are included:

- signal $D^{*+} \to D^0\pi^+, D^0 \to \mu^+\mu^-$ decay, denoted as "$D^*, D^0 \to \mu\mu$ ";
- peaking background from $D^{*+} \to D^0\pi^+, D^0 \to \pi^+\pi^- \to \mu^+\nu_\mu\mu^-\overline{\nu}_\mu$ process, denoted as "$D^*, D^0 \to \pi\pi \to \mu\mu$ ";

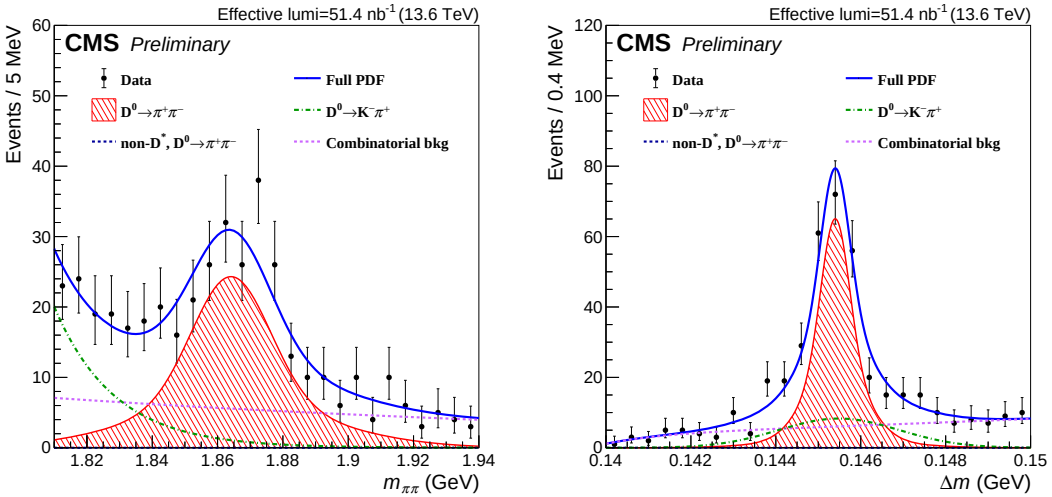

Figure 1: Dipion invariant mass $m_{\pi\pi}$ (left) and D*–D^0 mass difference Δm (right) from the reconstructed candidates from zero bias samples after MVA$_D$ selection.

- semileptonic background from $\mathrm{D}^{*+} \to \mathrm{D}^0\pi^+, \mathrm{D}^0 \to \pi^-\mu^+\nu$ process when the charged pion is misidentified as a muon, denoted as "$\mathrm{D}^*, \mathrm{D}^0 \to \pi\mu\nu$ ";
- combinatorial background events;

The semileptonic background $\mathrm{D}^{*+} \to \mathrm{D}^0\pi^+, \mathrm{D}^0 \to \pi^-\mu^+\nu$ is included since it exhibits a peaking distribution in Δm when the dimuon mass is close to the D^0 mass, indicating that the neutrino carries away less energy. It has a 2–3% impact on the signal extraction. There is a correlation between the $m_{\mu\mu}$ and Δm distribution, such that the closer the $m_{\mu\mu}$ is to the D^0 mass, the narrower the Δm peak. We use an uncorrelated 2D PDF for its modeling, and the correlation effect is considered in the fit bias uncertainty.

The non-D* $\mathrm{D}^0 \to \mu^+\mu^-$, $\mathrm{D}^0 \to \pi^+\pi^- \to \mu^+\nu_\mu\mu^-\overline{\nu}_\mu$, and $\mathrm{D}^0 \to \pi^-\mu^+\nu$ contributions are not expected to make any impact on the measurement, as they are mostly rejected by the Δm requirements. Their contributions have been considered in the fit and found to be negligible. The parameter of interest is the $\mathrm{D}^*, \mathrm{D}^0 \to \mu\mu$ signal strength with respect to the expected yield. Normalizations of the peaking components, including signal $\mathrm{D}^*, \mathrm{D}^0 \to \mu\mu$, $\mathrm{D}^*, \mathrm{D}^0 \to \pi\pi \to \mu\mu$, and $\mathrm{D}^*, \mathrm{D}^0 \to \pi\mu\nu$, are derived with the yield of normalization channel $\mathrm{D}^{*+} \to \mathrm{D}^0\pi^+, \mathrm{D}^0 \to \pi^+\pi^-$ decays (denoted as "$\mathrm{D}^*, \mathrm{D}^0 \to \pi\pi$ ") measured from the zero bias triggered events:

$$N_{\mathrm{D}^*,\mathrm{D}^0\to\mu\mu} = N_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi,\mathrm{MVA}_D} \times S_{\mathrm{ZB}} \times \frac{\varepsilon_{\mathrm{D}^*,\mathrm{D}^0\to\mu\mu}}{\varepsilon_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi}} \times \frac{\mathcal{B}_{\mathrm{D}^*,\mathrm{D}^0\to\mu\mu}}{\mathcal{B}_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi}} \times \mathrm{MVA}_{D,cor} \times T_{cor} \quad (3)$$

$$N_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi\to\mu\mu} = N_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi,\mathrm{MVA}_D} \times S_{\mathrm{ZB}} \times \frac{\varepsilon_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi\to\mu\mu}}{\varepsilon_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi}} \times \left(f_{\pi\to\mu}\right)^2 \times \mathrm{MVA}_{D,cor} \times T_{cor} \quad (4)$$

$$N_{\mathrm{D}^*,\mathrm{D}^0\to\pi\mu\nu} = N_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi,\mathrm{MVA}_D} \times S_{\mathrm{ZB}} \times \frac{\varepsilon_{\mathrm{D}^*,\mathrm{D}^0\to\pi\mu\nu}}{\varepsilon_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi}} \times \frac{\mathcal{B}_{\mathrm{D}^*,\mathrm{D}^0\to\pi\mu\nu}}{\mathcal{B}_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi}} \times \left(f_{\pi\to\mu}\right) \times T_{cor} \quad (5)$$

where S_{ZB} is the effective prescaling factor for the zero bias data sets; N_X, ε_X, and $\mathcal{B}_X$ are the yield, efficiency, and decay branching fraction for process X, respectively; $N_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi,\mathrm{MVA}_D}$ means $\mathrm{D}^*, \mathrm{D}^0 \to \pi\pi$ yield after MVA$_D$ selection. $f_{\pi\to\mu}$ is the pion to muon misidentification ratio between data and MC; T_{cor} is the trigger efficiency correction. MVA$_{D,cor}$ is the MVA$_D$

correction used to account for the small difference of MVA_D efficiency between $\mathrm{D}^0 \to \mu^+\mu^-$ ($\mathrm{D}^0 \to \pi^+\pi^- \to \mu^+\nu_\mu\mu^-\overline{\nu}_\mu$) and $\mathrm{D}^0 \to \pi^+\pi^-$. The $\mathrm{MVA}_{D,cor}$ is not used in the $\mathrm{D}^*, \mathrm{D}^0 \to \pi\mu\nu$ process since it is negligible compared to its efficiency uncertainty. All the efficiencies (ε_X), decay branching fraction ($\mathcal{B}_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi}$), the yield for the normalization channel ($N_{\mathrm{D}^*,\mathrm{D}^0\to\pi\pi}$), the misidentification ratio ($f_{\pi\to\mu}$) and T_{cor} are implemented as Gaussian-constrained nuisance parameters.

The models introduced for each component are described as follows:

- **Signal $\mathbf{D^*, D^0 \to \mu\mu}$ model**: the signal model is a 2D PDF consisting of a sum of two Gaussians for the dimuon invariant mass $m_{\mu\mu}$, with a common mean and a sum of three Gaussians for the mass difference Δm, also with a common mean. All model parameters, including the mean and the widths of the individual Gaussians for both dimensions, are determined from the fits to the simulated signal events.
- **Peaking background $\mathbf{D^*, D^0 \to \pi\pi \to \mu\mu}$ model**: this process is also modeled with a sum of two Gaussians for $m_{\mu\mu}$, with different means, and a sum of three Gaussians for Δm, with a common mean. The model parameters, including the mean and width of the individual Gaussians for both dimensions, are determined from the fits to the simulated $\mathrm{D}^{*+} \to \mathrm{D}^0\pi^+, \mathrm{D}^0 \to \pi^+\pi^- \to \mu^+\nu_\mu\mu^-\overline{\nu}_\mu$ events.
- **Semi-Peaking background $\mathbf{D^*, D^0 \to \pi\mu\nu}$ model**: this process is also modeled with a sum of two Gaussians for $m_{\mu\mu}$, and a sum of three Gaussians for Δm. The model parameters are determined from the fits to the simulation.
- **Model for combinatorial background**: the combinatorial events are also described with a 2D PDF, with the $m_{\mu\mu}$ distribution modeled using a discrete profiling method [42] that freely floats a first-order Bernstein polynomial, a power law, and an exponential function, while the Δm distribution is modeled by a specialized shape commonly used for background modeling:

$$\mathrm{P}(m \mid m_0, A, B, C) = \left(1 - \exp\left(-\frac{m - m_0}{C}\right)\right) \cdot \left(\frac{m}{m_0}\right)^A + B \cdot \left(\frac{m}{m_0} - 1\right), \quad (6)$$

  where m_0 is the kinematic threshold and is fixed to pion mass, and parameters A, B, and C are determined from fits to data.

The mean and width of the Gaussian functions describing the signal component are corrected based on the shape difference between data and MC simulation derived from the control samples $\mathrm{D}^0 \to \pi^+\pi^-$ and $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$.

8 Systematic effects

Most of the systematic effects encountered in the data analysis are canceled out in the ratio of the signal and reference event yield measurements. This section deals with the residual effects and remaining differences.

The double muon trigger is only applied in the signal channel, which requires dedicated correction. The trigger efficiency is estimated from MC simulated events and is corrected in data using dedicated utility triggers. The L1 trigger efficiency correction is derived by measuring the efficiency of the double muon L1 trigger with respect to the single muon triggers. The ratio of Data to MC efficiencies is found to be $(99.0 \pm 0.5)\%$. The HLT efficiency correction is derived by measuring the efficiency of the double muon HLT with respect to the double muon L1 trigger. The ratio of Data to MC efficiencies is found to be $(95.3 \pm 0.5)\%$. The total correction is

estimated as the product of the L1 and HLT corrections and is found to be $(94.3 \pm 0.7)\%$. The uncertainty includes the statistical error from the measured samples and efficiency variations across different data periods and utility triggers.

The misidentified muons from pion decays are reasonably well modeled in MC simulations. However, an accurate simulation of the reconstruction and identification of such muons is more challenging. To derive a correction, we measured the relative rate of $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \mu\pi$ (where μ is the misidentified muon from pion) decays in data and MC simulation. Good agreement is observed with an average correction of (1.06 ± 0.15) for the 2022 and 2023 data.

Muon reconstruction and identification efficiency corrections are measured using a tag-and-probe method on $J/\psi \to \mu\mu$ decays. The combined effect is found to be less than 1%. The charged track reconstruction efficiency was assessed using the D^* method [43]. This method involves computing the ratio of $D^0 \to K^-\pi^+$ and $D^0 \to K\pi\pi\pi$ event yields coming from $D^{*+} \to D^0\pi^+$ decays and comparing this ratio to the world average value. Based on this study, we assign a conservative systematic uncertainty of 2.3% per charged track to the tracking efficiency.

To evaluate the impact of the zero bias trigger prescale non-uniformity on the normalization channel selection efficiency, we compared a number of primary vertex distributions for the signal and normalization channels and calculated the reweighted efficiency using MC simulations. The effect was found to be less than 1%.

We have examined the fit bias in the branching fraction extraction with ML fits using pseudo-experiments, with an injected $D^0 \to \mu^+\mu^-$ signal branching fraction near our expected sensitivity limit. The difference between the median expected branching fraction and the simulated branching fraction is less than 1%. The potential correlation effect of $D^0 \to \pi^-\mu^+\nu$ background modeling has been examined, resulting in a fit bias of 1.8%. The bias from using alternative modeling for signal and background components is also examined and found to be negligible. Therefore, the total fit bias uncertainty is 2%.

Table 2 presents a summary of all systematic uncertainties. The nuisance parameters, including tracking efficiency, muon efficiency, zero bias prescale non-uniformity, fit bias, and MVA_D correction, are constrained using a log-normal PDF, while the remaining parameters are constrained using a Gaussian PDF.

Table 2: Summary of systematic uncertainties for the $D^0 \to \mu^+\mu^-$ branching ratio measurement.

Effect	$D^0 \to \mu^+\mu^-$	$D^0 \to \pi\pi \to 2\mu 2\nu$	$D^0 \to \pi\mu\nu$
Trigger efficiency correction		0.7%	
Zero bias prescale non-uniformity		1 %	
Tracking efficiency (per track)		2.3%	
Muon efficiency (per muon)		1%	
$D^0 \to \pi^+\pi^-$ normalization		8.7%	
Preselection efficiency	0.2%	0.6%	12%
MVA_D correction for $D^0 \to \mu^+\mu^-$	1.2%	2.0%	—
$\mathcal{B}(D^0 \to \pi^+\pi^-)$	1.7%	—	1.7%
$\mathcal{B}(D^0 \to \pi^-\mu^+\nu)$	—	—	4.5%
Fit bias	2 %	—	—
Misidentification rate (per muon)	—	14%	14%

9 Results

The $m_{\mu\mu}$ and Δm projections in the full range and sub-range in $m_{\mu\mu}$ and Δm are shown in Figs. 2 and 3, respectively. No obvious excess of events with respect to the background expectation is observed in the data. The exclusion upper limit is computed using `Combine` tool [41]. A profile likelihood ratio test statistic is built including the systematic uncertainties as nuisance parameters. A log-normal or Gaussian PDF is used to describe each constrained nuisance parameter.

The upper limit at 95% confidence level (CL) based on the Asymptotic CLs method is:

$$\mathcal{B}(\mathrm{D}^0 \to \mu^+\mu^-) < 2.6 \times 10^{-9} \text{ at } 95\% \text{ CL}. \tag{7}$$

The best-fit value of the branching fraction of the $\mathrm{D}^0 \to \mu^+\mu^-$ is found to be:

$$\mathcal{B}(\mathrm{D}^0 \to \mu^+\mu^-) = (1.0 \pm 0.9) \times 10^{-9}, \tag{8}$$

which corresponds to a signal yield of 139 ± 123. The observed significance evaluated from the log-likelihood difference is 1.1σ. The event yields for each component of the fit are summarized in Table 3.

Table 3: The expected event yields for signal, the combinatorial background, the peaking background, and the semileptonic background are summarized (post-fit). The total expected and observed event yields are given in Total and Data column, respectively. The subrange is in one dimension with a full range in the other dimension.

Range	Signal	Comb	Peak	Semi	Total	Data
Full range	139 ± 123	126185 ± 366	220 ± 58	207 ± 40	126751 ± 355	126752
$0.145 < \Delta m < 0.146$ GeV	92 ± 81	14044 ± 63	141 ± 37	91 ± 17	14367 ± 81	14412
$1.84 < m_{\mu\mu} < 1.89$ GeV	120 ± 106	48553 ± 204	123 ± 33	55 ± 11	48851 ± 211	48798

Compared to the current world best measurement by LHCb [23], $\mathcal{B}(\mathrm{D}^0 \to \mu^+\mu^-) < 3.5 \times 10^{-9}$ at 95% CL, the result represents an improvement of about 35%. The observed limit is larger than the expected limit from background-only pseudo-experiments, $\mathcal{B}(\mathrm{D}^0 \to \mu^+\mu^-) < 1.8 \times 10^{-9}$ at 95% CL, which is consistent with the central value of the signal branching fraction.

The profile likelihood as a function of the $\mathcal{B}(\mathrm{D}^0 \to \mu^+\mu^-)$ branching fractions is shown in Fig. 4. The fit and likelihood scan is repeated using different configurations of the individual function in the discrete profiling method that describes the combinatorial background, leading to the same result. The result using the alternative normalization channel $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ is consistent with the nominal result.

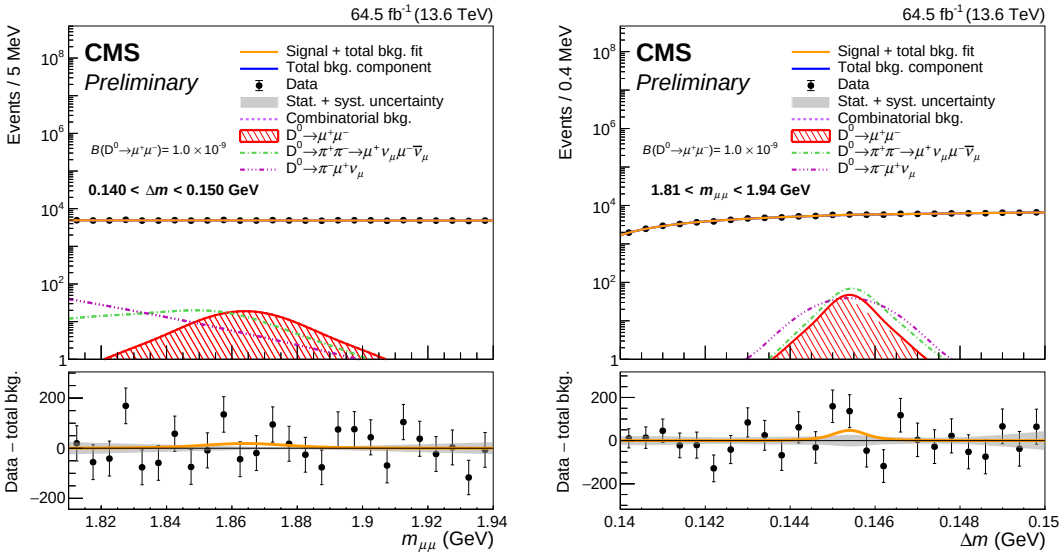

Figure 2: The projection of $m_{\mu\mu}$ (left) and Δm (right) of the fit in the full $m_{\mu\mu}$ and Δm range. The bottom panel shows the data and the fit result after subtraction of the total background component. The grey error band represents the statistical and systematic uncertainties of the total background component.

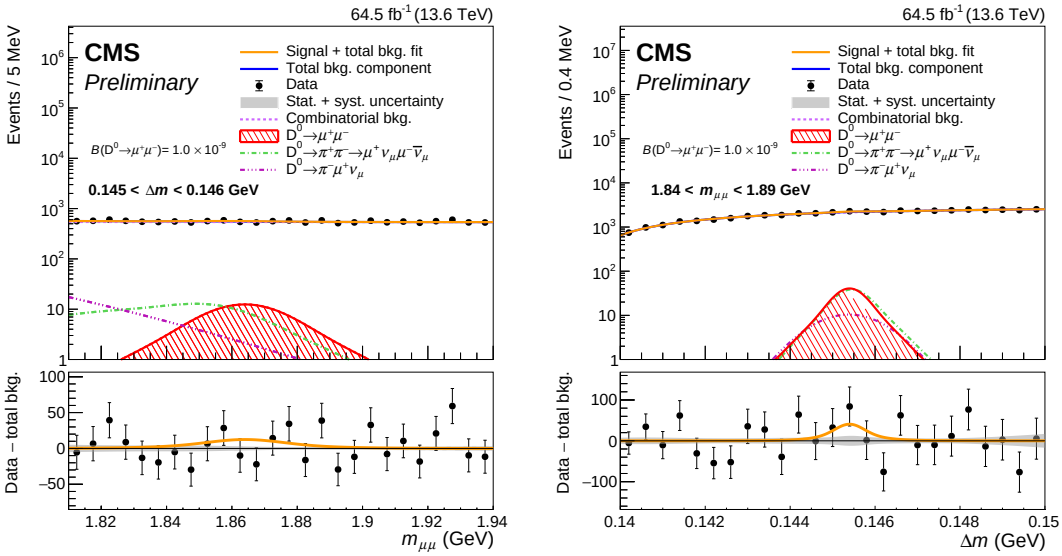

Figure 3: The projection of $m_{\mu\mu}$ (left) and Δm (right) of the fit with requirements $0.145 < \Delta m < 0.146$ GeV and $1.84 < m_{\mu\mu} < 1.89$ GeV, respectively. The bottom panel shows the data and the fit result after subtraction of the total background component. The grey error band represents the statistical and systematic uncertainties of the total background component.

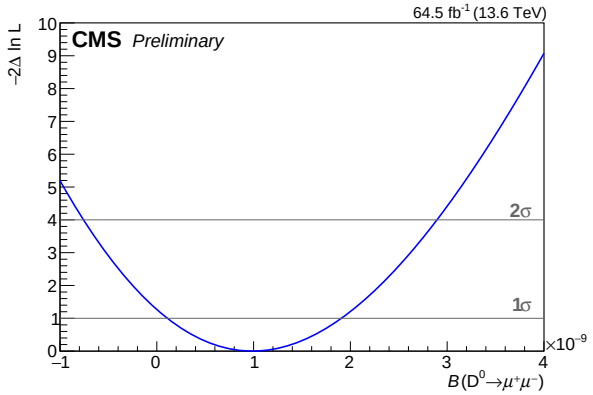

Figure 4: The profile likelihood scan as a function of $D^0 \to \mu^+\mu^-$ decay branching fraction.

10 Summary

The measurement of the branching fraction ($\mathcal{B}$) of the $D^0 \to \mu^+\mu^-$ decay, based on a data set of proton-proton collisions at $\sqrt{s} = 13.6$ TeV collected by the CMS experiment corresponding to an integrated luminosity of $64.5\,\text{fb}^{-1}$, has been presented. The branching fraction is measured using the cascade decay $D^{*+} \to D^0\pi^+$, $D^0 \to \mu^+\mu^-$ with the $D^0 \to \pi^+\pi^-$ decay mode as the normalization channel, considering both prompt and non-prompt D^0 meson production.

No significant excess above the background expectation was observed and an upper limit on the branching fraction has been set to be:

$$\mathcal{B}(D^0 \to \mu^+\mu^-) < 2.6 \times 10^{-9} \text{ at } 95\% \text{ CL}. \tag{9}$$

The obtained value of $D^0 \to \mu^+\mu^-$ branching fraction is found to be $\mathcal{B}(D^0 \to \mu^+\mu^-) = (1.0 \pm 0.9) \times 10^{-9}$.

The measurement is the most sensitive to date, representing a 35% improvement over the current best measurement [23]. This measurement provides the most stringent limit on flavor-changing neutral currents in the charm sector, setting additional constraints on new physics models that modify the decay branching fraction of $D^0 \to \mu^+\mu^-$.

The analysis is the first measurement at CMS using the newly developed low-mass double muon parking trigger [24]. It also marks the first publication on the $D^0 \to \mu^+\mu^-$ branching fraction measurement at CMS, made possible by the enriched dimuon events collected by this innovative trigger.

References

[1] CMS Collaboration, "Angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from pp collisions at $\sqrt{s} = 8$ TeV", *Physics Letters B* **753** (2016) 424, doi:10.1016/j.physletb.2015.12.020, arXiv:1507.08126.

[2] CMS Collaboration, "Measurement of angular parameters from the decay $B^0 \to K^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Physics Letters B* **781** (2018) 517, doi:10.1016/j.physletb.2018.04.030, arXiv:1710.02846.

[3] CMS Collaboration, "Measurement of properties of $B_s^0 \to \mu^+\mu^-$ decays and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment", *JHEP* **04** (2020) 188, doi:10.1007/JHEP04(2020)188, arXiv:1910.12127.

[4] ATLAS, CMS and LHCb Collaborations, "Combination of the ATLAS, CMS and LHCb results on the $B_s^0 \to \mu^+\mu^-$ decays", Physics Analysis Summary CMS-PAS-BPH-20-003, LHCb-CONF-2020-002, ATLAS-CONF-2020-049, 2020.

[5] CMS Collaboration, "Measurement of the $B_s^0 \to \mu^+\mu^-$ decay properties and search for the $B^0 \to \mu^+\mu^-$ decay in proton-proton collisions at $\sqrt{s}$ = 13 TeV", *Phys. Lett. B* **842** (2023) 137955, doi:10.1016/j.physletb.2023.137955, arXiv:2212.10311.

[6] LHCb Collaboration, "Angular analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ Decay", *Phys. Rev. Lett.* **126** (2021) 161802, doi:10.1103/PhysRevLett.126.161802, arXiv:2012.13241.

[7] LHCb Collaboration, "Measurement of *CP*-Averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay", *Phys. Rev. Lett.* **125** (2020) 011802, doi:10.1103/PhysRevLett.125.011802, arXiv:2003.04831.

[8] LHCb Collaboration, "Angular analysis of the rare decay $B_s^0 \to \phi\mu^+\mu^-$", *JHEP* **11** (2021) 043, doi:10.1007/JHEP11(2021)043, arXiv:2107.13428.

[9] LHCb Collaboration, "Branching Fraction Measurements of the Rare $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ Decays", *Phys. Rev. Lett.* **127** (2021) 151801, doi:10.1103/PhysRevLett.127.151801, arXiv:2105.14007.

[10] LHCb Collaboration, "Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays", *Phys. Rev. Lett.* **118** (2017) 191801, doi:10.1103/PhysRevLett.118.191801, arXiv:1703.05747.

[11] LHCb Collaboration, "Measurement of the $B_s^0 \to \mu^+\mu^-$ decay properties and search for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays", *Phys. Rev. D* **105** (2022) 012010, doi:10.1103/PhysRevD.105.012010, arXiv:2108.09283.

[12] LHCb Collaboration, "Analysis of neutral B-meson decays into two muons", *Phys. Rev. Lett.* **128** (2022) 041801, doi:10.1103/PhysRevLett.128.041801, arXiv:2108.09284.

[13] ATLAS Collaboration, "Angular analysis of $B^0 \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$TeV with the ATLAS detector", *JHEP* **10** (2018) 47, doi:10.1007/JHEP10(2018)047, arXiv:1805.04000.

[14] ATLAS Collaboration, "Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector", *JHEP* **04** (2019) 098, doi:10.1007/JHEP04(2019)098, arXiv:1812.03017.

[15] CMS and LHCb Collaboration, "Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data", *Nature* **522** (2015) 68, doi:10.1038/nature14474, arXiv:1411.4413.

[16] BABAR Collaboration, "Measurement of angular asymmetries in the decays $B \to K^*\ell^+\ell^-$", *Phys. Rev. D* **93** (2016) 052015, doi:10.1103/PhysRevD.93.052015, arXiv:1508.07960.

[17] Belle Collaboration, "Measurement of the Differential Branching Fraction and Forward-Backward Asymmetry for $B \to K^* \ell^+ \ell^-$", *Phys. Rev. Lett.* **103** (2009) 171801, doi:10.1103/PhysRevLett.103.171801, arXiv:0904.0770.

[18] Belle Collaboration, "Lepton-Flavor-Dependent Angular Analysis of $B \to K^* \ell^+ \ell^-$", *Phys. Rev. Lett.* **118** (2017) 111801, doi:10.1103/PhysRevLett.118.111801, arXiv:1612.05014.

[19] CDF Collaboration, "Measurements of the Angular Distributions in the Decays $B \to K^{(*)} \mu^+ \mu^-$ at CDF", *Phys. Rev. Lett.* **108** (2012) 081807, doi:10.1103/PhysRevLett.108.081807, arXiv:1108.0695.

[20] E. Golowich, J. Hewett, S. Pakvasa, and A. A. Petrov, "Relating D^0 - $\overline{D}^0$ mixing and $D^0 \to \mu^+ \mu^-$ with new physics", *Phys. Rev. D* **79** (2009) 114030, doi:10.1103/PhysRevD.79.114030, arXiv:0903.2830.

[21] S. Fajfer and N. Košnik, "Leptoquarks in flavor changing neutral current charm decays", *Phys. Rev. D* **79** (2009) 017502, doi:10.1103/PhysRevD.79.017502, arXiv:0810.4858.

[22] G. Burdman, E. Golowich, J. L. Hewett, and S. Pakvasa, "Rare charm decays in the standard model and beyond", *Phys. Rev. D* **66** (2002) 014009, doi:10.1103/PhysRevD.66.014009, arXiv:hep-ph/0112235.

[23] LHCb Collaboration, "Search for rare decays of D^0 mesons into two muons", *Phys. Rev. Lett.* **131** (2023) 041804, doi:10.1103/PhysRevLett.131.041804, arXiv:2212.11203.

[24] CMS Collaboration, "Enriching the physics program of the CMS experiment via data scouting and data parking", doi:10.48550/arXiv.2403.16134, arXiv:2403.16134.

[25] LHCb Collaboration, "Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV", *JHEP* **03** (2016) 159, doi:10.1007/JHEP03(2016)159, arXiv:1510.01707. [Erratum: JHEP 09, 013 (2016), Erratum: JHEP 05, 074 (2017)].

[26] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[27] CMS Collaboration, "Track impact parameter resolution for the full pseudorapidity coverage in the 2017 dataset with the CMS Phase-1 pixel detector", CMS Detector Performance Report CMS-DP-2020-049, 2020.

[28] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018) P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.

[29] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.

[30] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[31] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[32] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020) 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[33] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[34] D. J. Lange, "The EvtGen particle decay simulation package", *Nucl. Instrum. Meth. A* **462** (2001) 152, doi:10.1016/S0168-9002(01)00089-4.

[35] N. Davidson, T. Przedzinski, and Z. Was, "PHOTOS interface in C++: technical and physics documentation", *Comput. Phys. Commun.* **199** (2016) 86, doi:10.1016/j.cpc.2015.09.013, arXiv:1011.0937.

[36] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[37] K. Prokofiev and T. Speer, "A kinematic and a decay chain reconstruction library",. Prepared for Computing in High-Energy Physics (CHEP '04), Interlaken, Switzerland, 27 Sep - 1 Oct 2004.

[38] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, p. 785. ACM, New York, NY, USA, 2016. doi:10.1145/2939672.2939785.

[39] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling", *eConf* **C0303241** (2003) MOLT007, doi:10.1142/9781860948985_0039, arXiv:physics/0306116.

[40] LHCb Collaboration, "Search for the rare decay $D^0 \to \mu^+\mu^-$", *Physics Letters B* **725** (2013) 15, doi:10.1016/j.physletb.2013.06.037, arXiv:2106.03744.

[41] CMS Collaboration, "The CMS statistical analysis and combination tool: COMBINE", doi:10.48550/arXiv.2404.06614, arXiv:2404.06614.

[42] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies, "Handling uncertainties in background shapes: the discrete profiling method", *JINST* **10** (2015) P04015, doi:10.1088/1748-0221/10/04/P04015, arXiv:1408.6865.

[43] CMS Collaboration, "Tracking performances for charged pions with Run2 legacy data", CMS Detector Performance Report CMS-DP-2022-012, 2022.