CMS Physics Analysis Summary

Contact: cms-pag-conveners-bphysics@cern.ch 2022/07/07

Measurement of $B_s^0 \to \mu^+\mu^-$ decay properties and search for the $B^0 \to \mu\mu$ decay in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

We report a measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and its effective lifetime, as well as results of a search for the $B^0 \to \mu^+\mu^-$ decay in proton-proton collisions at $\sqrt{s} = 13$ TeV at the LHC. The analysis is based on data collected with the CMS detector in 2016-2018 corresponding to an integrated luminosity of 140 fb^{-1}. The measured branching fraction of the $B_s^0 \to \mu^+\mu^-$ decay and the effective B_s^0 lifetime are the most precise measurements to date. No evidence for the $B^0 \to \mu^+\mu^-$ decay has been found. All results are found to be consistent with the standard model predictions.

1 Introduction

Rare B meson decays provide a sensitive probe for beyond-the-standard-model (BSM) effects and can explore energy scales much higher than the ones directly accessible at the Large Hadron Collider (LHC) at CERN. A key factor in the success of these studies is the availability of precise theoretical predictions for experimentally accessible processes. The $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays represent a rare case where theoretically clean predictions are matched with a clear experimental signature. (In what follows, charge-conjugated decays are implied.) These rare decays are examples of flavor changing neutral current processes, which are strongly suppressed in the standard model (SM), making them sensitive to BSM physics contributions.

The $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays proceed through penguin and box diagrams that involve Z or W boson exchange and are furthermore helicity suppressed by a factor m_μ^2/m_B^2, where m_μ and m_B denote the masses of the muon and either the B_s^0 or B^0 meson, respectively. Moreover, the $B^0 \to \mu^+\mu^-$ decay is also Cabibbo suppressed. As a result, in the SM, the average time-integrated branching fractions for these decays are very small [1]:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.66 \pm 0.14) \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.03 \pm 0.05) \times 10^{-10}.$$

These predictions include next-to-leading order (NLO) corrections of electroweak origin and next-to-next-to-leading order (NNLO) QCD corrections. The largest contribution to the theoretical error is due to the uncertainties in the CKM matrix element values.

A number of experiments at e^+e^- and hadron colliders have searched for these decays, but only recently the first observation of the $B_s^0 \to \mu^+\mu^-$ decay was reported in a combined analysis of data taken by the LHCb and CMS Collaborations [2], which was later confirmed by the ATLAS [3], CMS [4], and LHCb [5, 6] experiments individually. Currently, the most precise measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction is achieved in a combined analysis of data of the three experiments [7]. They show a small deviation from the SM prediction at the level of 2.1σ. No significant indication of the $B^0 \to \mu^+\mu^-$ decay has been reported so far.

A few recent measurements of $b \to s\ell^+\ell^-$ processes have reported disagreements at the level of 2–3 standard deviations from the SM predictions in the branching fraction of the $B \to K^*\mu^+\mu^-$ and $B \to \phi\mu^+\mu^-$ decays [8–11], angular observables in the $B \to K^*\mu^+\mu^-$ decays [12, 13] and in searches for lepton flavor universality violation in measurements of the R_K and R_{K^*} ratios [14–17]. While not all individual measurements confirm these observations [18, 19], the global fits of rare B decay data show a strong preference of the BSM physics scenario over the SM by more than 4 standard deviations [20].

In the framework of an effective field theory, $b \to s\ell^+\ell^-$ decays are dominated by the semileptonic operators $O_9 = (\bar{s}_L\gamma_\mu b_L)(\bar{\ell}\gamma^\mu\ell)$ and $O_{10} = (\bar{s}_L\gamma_\mu b_L)(\bar{\ell}\gamma^\mu\gamma_5\ell)$. As the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays are dominated by the O_{10} operator, they can be sensitive to the same effects. The BSM physics contributions could be observed as deviations of the corresponding Wilson coefficients (C_9 and C_{10}) from their SM values. What differentiates the fully leptonic B meson decays from the semileptonic ones is that all non-perturbative hadronic contributions enter through the decay constants, which are known precisely from the Lattice QCD calculations. Therefore a precise measurement of the $B_s^0 \to \mu^+\mu^-$ decay properties may have a big impact on interpretations of these anomalies.

The effective lifetime of the B_s^0 meson measured in the $B_s^0 \to \mu^+\mu^-$ decay is an independent theoretically clean probe for BSM physics [21]. In the SM, only the heavy $B_{s,H}^0$ mass eigenstate

can decay to the $\mu^+\mu^-$ final state. This is because, in the absence of CP violation, the B_s^0 mass eigenstates are also CP eigenstates, with the heavier one being CP odd, and $\mu^+\mu^-$ is also a CP-odd final state. Therefore, any significant deviation of the measurement from the SM prediction for the $B_{s,H}^0$ lifetime would indicate a BSM physics contribution. Currently, the most precise measurement of the B_s^0 meson lifetime in $B_s^0 \to \mu^+\mu^-$ decays of $\tau(B_s^0 \to \mu^+\mu^-) = 2.07 \pm 0.29$ ps comes from the LHCb experiment [6].

In this note, we report on a new measurement of the $B_s^0 \to \mu^+\mu^-$ decay and a search for the $B^0 \to \mu^+\mu^-$ decay based on the proton-proton (pp) collision data at a center-of-mass energy of 13 TeV collected by the CMS experiment in 2016–2018, and corresponding to an integrated luminosity of 140 fb^{-1}. The new analysis uses improved techniques compared to our earlier publication based on 2011–2012 and 2016 data [4]; consequently the 2016 data sample has been reanalyzed, and the new results supersede the ones from Ref. [4]. Given that the sensitivity of the 2016–2018 data significantly exceeds that of the 2011–2012 sample, no attempt is made to combine the new results with the 2011–2012 data.

2 Data analysis overview

The data analysis strategy employed in this measurement is based on the previous CMS studies of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays. We made a few modifications and improvements to increase the analysis sensitivity, benefiting from the large amount of data collected in 2016–2018. With the increased sensitivity, we had to develop new methods to achieve better understanding of various systematic effects.

We reconstruct leptonic B meson decays by combining two oppositely charged muons, performing a common vertex fit, and imposing selection criteria to separate small signals from large backgrounds. (Here and in what follows, we use the notation B to denote either the B^0 or the B_s^0 meson.) The dominant background sources are the combinatorial background where the two muons originate from two different heavy quarks, the partially reconstructed semileptonic decays where both muons originate from the same B meson, and the peaking background coming from the charmless two-body hadronic decays of B mesons.

The combinatorial and partially reconstructed backgrounds are the main limiting factors in the analysis sensitivity. Despite being reducible backgrounds with a number of distinct features, they are copious, which makes it difficult to reject them completely without losing a significant fraction of signal events. In order to maximize the analysis sensitivity, we perform a multivariate analysis (MVA_B) combining multiple discriminating observables in a single powerful discriminator. Training and performance evaluation of the MVA_B is described in Section 6.

The charmless two-body decays, such as $B^0 \to K^+\pi^-$, $B_s^0 \to K^+K^-$ etc., may look like signal when both charged hadrons are misidentified as muons. We measure the misidentification probabilities in data using $K_S^0 \to \pi^+\pi^-$, $\phi \to K^+K^-$, and $\Lambda \to p\pi^-$ decays. We find a reasonable agreement between the data and Monte Carlo (MC) simulation for pions and kaons, where the misidentification probability is dominated by the pion and kaon decays in flight to a muon and a muon neutrino. The probability to misidentify protons as muons is an order of magnitude smaller according to simulation, and found to be even smaller in data making contributions from associated processes unimportant. With a tight muon identification based on a multivariate analysis (MVA_μ), we reduce the charmless two-body backgrounds to a negligible level.

We split events into 16 categories: four data-taking periods, two pseudorapidity regions, and

two signal regions with different signal purity defined by the analysis $\mathrm{MVA_B}$ discriminator value. The pseudorapidity regions are $0 < |\eta_\mathrm{F}| < 0.7$ and $0.7 < |\eta_\mathrm{F}| < 1.4$, where η_F is the pseudorapidity of the most forward muon. The absolute pseudorapidity is limited to 1.4 because of the trigger requirement.

The results are extracted using simultaneous unbinned maximum likelihood (UML) fits. For the branching fraction measurements, we perform a two-dimensional (2D) fit using the dimuon invariant mass and its uncertainty as the observables. For the lifetime extraction, we perform a three-dimensional (3D) fit using the dimuon mass, the decay time, and its uncertainty as the observables.

Given the poor precision in the knowledge of the $\mathrm{b\bar{b}}$ cross section at the LHC, a direct extraction of the branching fractions of the $\mathrm{B_s^0} \to \mu^+\mu^-$ and $\mathrm{B^0} \to \mu^+\mu^-$ decays would be affected by a large uncertainty. As commonly done in B physics analyses, the signal branching fraction is instead calculated by normalizing it to another decay channel of a B meson, for which the branching fraction is well known and whose characteristics allow for a precise reconstruction with low backgrounds. The $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ with $\mathrm{J/\psi} \to \mu\mu$ is the best candidate in our case. We also consider $\mathrm{B_s^0} \to \mathrm{J/\psi\phi}$ decays with $\phi(1020) \to \mathrm{KK}$ as an alternative. The signal branching fractions $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$ and $\mathcal{B}(\mathrm{B^0} \to \mu^+\mu^-)$ can be then extracted as

$$\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-) = \mathcal{B}(\mathrm{B^+} \to \mathrm{J/\psi K^+}) \frac{N_{\mathrm{B_s^0}\to\mu^+\mu^-}}{N_{\mathrm{B^+}\to\mathrm{J/\psi K^+}}} \frac{\varepsilon_{\mathrm{B^+}\to\mathrm{J/\psi K^+}}}{\varepsilon_{\mathrm{B_s^0}\to\mu^+\mu^-}} \frac{f_\mathrm{u}}{f_\mathrm{s}}, \tag{1}$$

$$\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-) = \mathcal{B}(\mathrm{B_s^0} \to \mathrm{J/\psi\phi}) \frac{N_{\mathrm{B_s^0}\to\mu^+\mu^-}}{N_{\mathrm{B_s^0}\to\mathrm{J/\psi\phi}}} \frac{\varepsilon_{\mathrm{B_s^0}\to\mathrm{J/\psi\phi}}}{\varepsilon_{\mathrm{B_s^0}\to\mu^+\mu^-}}, \tag{2}$$

$$\mathcal{B}(\mathrm{B^0} \to \mu^+\mu^-) = \mathcal{B}(\mathrm{B^+} \to \mathrm{J/\psi K^+}) \frac{N_{\mathrm{B^0}\to\mu^+\mu^-}}{N_{\mathrm{B^+}\to\mathrm{J/\psi K^+}}} \frac{\varepsilon_{\mathrm{B^+}\to\mathrm{J/\psi K^+}}}{\varepsilon_{\mathrm{B^0}\to\mu^+\mu^-}} \frac{f_\mathrm{u}}{f_\mathrm{d}}, \tag{3}$$

where N_X is the number of the decay X candidates, as extracted from the fit, and ε_X is the corresponding full selection efficiency. In addition, f_u, f_d and f_s are the B hadron production fractions for $\mathrm{B^+}$, $\mathrm{B^0}$ and $\mathrm{B_s^0}$ mesons, respectively. The ratio $f_\mathrm{u}/f_\mathrm{d}$ is expected to be 1 in the SM due to the isospin symmetry. The ratio $f_\mathrm{s}/f_\mathrm{u}$, as well as $\mathcal{B}(\mathrm{B^+} \to \mathrm{J/\psi K^+})$ and $\mathcal{B}(\mathrm{B_s^0} \to \mathrm{J/\psi\phi})$, are external inputs discussed later in Section 9.

A key advantage of measuring the branching fractions with respect to other decays is that it allows for a cancellation of many systematic uncertainties in the selection and reconstruction efficiencies of the signal and normalization channels.

In order to avoid unconscious biases, the analysis employed a "data blinding" technique. All optimization studies were performed on data excluding events with the dimuon mass from 5.15 to 5.50 GeV. Once the selection criteria and measurement procedure were fixed, the data were unblinded.

3 The CMS detector

The central feature of the CMS detector 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 (ECAL), and a brass and scintillator

hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity 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. [22].

The silicon tracker used in 2016 measured charged particles within the range $|\eta| < 2.5$. For nonisolated particles of $1 < p_T < 10\,\text{GeV}$ and $|\eta| < 1.4$, the track resolutions were typically 1.5% in p_T and 25–90 (45–150) μm in the transverse (longitudinal) impact parameter [23]. At the start of 2017, a new pixel detector was installed [24]; the upgraded tracker measured particles up to $|\eta| < 3.0$ with typical resolutions of 1.5% in p_T and 20–75 μm in the transverse impact parameter [25] for nonisolated particles of $1 < p_T < 10\,\text{GeV}$.

Muons are measured in the pseudorapidity range $|\eta| < 2.4$, with detection planes made using three technologies: drift tubes, cathode strip chambers, and resistive plate chambers. Matching muons to tracks measured in the silicon tracker results in a relative transverse momentum resolution, for muons with p_T up to 100 GeV, of 1% in the barrel and 3% in the endcaps [26].

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 [27]. 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 around 1 kHz before data storage [28].

4 Data and Monte Carlo simulations

We split data collected with the CMS detector in 2016–2018 into four distinct periods: 2016a, 2016b, 2017, and 2018. The integrated luminosities of the corresponding samples amount to 20.0, 16.6, 42.0, and 61.3 fb^{-1}, respectively [29–31]. Data from the 2016a period was affected by the strip tracker dynamic hit inefficiency. The problem was resolved in August 2016 and the data for the rest of the 2016 data-taking period is referred to as 2016b. During the winter break of 2016–2017, the pixel detector was upgraded. The new detector provides a higher acceptance and better resolution. During the 2017 data taking, the pixel detector had synchronization issues at the beginning of the run, followed by the DC-DC converter failures leading to about 11% of the detector being unresponsive. Most of these issues were resolved in 2018.

We use 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 [32] using tune CP5 [33] and propagated through the CMS detector model using the GEANT4 [34] package. The decay of B hadrons is described using the EVTGEN 1.3.0 [35] program and final-state photon radiation using the PHOTOS 3.56 [36] program.

Multiple interactions within the same or nearby bunch crossings (pileup) are simulated for all samples by overlapping simulated minimum bias events with the hard scattering, with the multiplicity matching that observed in data (averaging to 23 for the 2016 and 32 for 2017–2018 data).

5 Event reconstruction and selection

The events used in this analysis were collected with a set of exclusive dimuon triggers designed to select $B \to \mu^+\mu^-$, $B^+ \to J/\psi K^+$, and $B_s^0 \to J/\psi\phi$ events. Given the high rates of dimuon events, the triggers were designed with tight selection requirements to achieve acceptable data rates. The L1 hardware triggers required two high-quality [27] oppositely charged muons restricted in rapidity to $|\eta| < 1.5$. At the HLT, a high-quality secondary dimuon vertex [28] was required and the events were restricted to narrow mass windows around the B meson and J/ψ masses. The J/ψ triggers additionally required the secondary vertex (SV) to be displaced from the beam spot.

The candidate selection starts with the reconstruction of dimuon candidates, which are used to build different B candidates for the signal and control regions. The selection is kept as similar as possible among all the candidates to be able to benefit from the cancellation of systematic effects in the ratio of efficiencies used for the branching fraction normalization. Both muons were required to have a high-quality central track [23] with the transverse momentum $p_T > 4\,\text{GeV}$ and rapidity $|\eta| < 1.4$. In addition, both muon candidates were required to pass tight muon identification requirements [4, 26], which suppressed misidentified muons from pion and kaon decays in flight and other sources.

We fit the B meson candidates employing a kinematic fitter described in Ref. [37]. We apply different kinematic constraints depending on the final state. For the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays, we use a common vertex constraint, while for the $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ decays we also add a mass constraint for the J/ψ candidate.

From the candidate's decay vertex and its momentum we build a refitted trajectory representing the B candidate. Then, for each primary vertex (PV) in the event, the trajectory is extrapolated to the closest point to that vertex. The absolute distance between the closest point and the PV in 3D is defined as the impact parameter. The PV with the smallest impact parameter is selected as the best PV for the B candidate

Table 1 shows a summary of the B candidate selection criteria used as inputs to the MVA_B and the UML fits. The 3D SV displacement significance is defined as the 3D distance between the PV and the dimuon SV, divided by their combined uncertainty. The 2D $\mu\mu$ pointing angle α is defined as the angle between the dimuon momentum and the line connecting the PV and SV and is calculated in the transverse plane with respect to the beam direction. The pointing angle requirement is introduced to match the $\cos(\alpha) > 0.9$ requirement used in the J/ψ triggers.

Table 1: Selection summary before the analysis MVA_B.

Selection	$B \to \mu^+\mu^-$	$B^+ \to J/\psi K^+$	$B_s^0 \to J/\psi\phi$
B candidate mass [GeV]	[4.90, 5.90]	[4.90, 5.90]	[4.90, 5.90]
Blinding window [GeV]	[5.15, 5.50]		
$p_{T\mu}$ [GeV]	> 4	> 4	> 4
$\lvert\eta_\mu\rvert$	< 1.4	< 1.4	< 1.4
3D SV displacement significance	> 6	> 4	> 4
$p_{T\mu\mu}$ [GeV]	> 5	> 7	> 7
$\mu\mu$ SV probability	> 0.025	> 0.1	> 0.1
$\mu\mu$ invariant mass [GeV]	[4.9, 5.9]	[2.9, 3.3]	[2.9, 3.3]
Kaon p_T [GeV]		> 1	> 1
Mass-constrained fit probability		> 0.025	> 0.025
2D $\mu\mu$ pointing angle [rad]		< 0.4	< 0.4
ϕ candidate mass [GeV]			[1.01, 1.03]

6 Multivariate analysis

Most of the observables used to distinguish signal from background have rather weak discriminating power. Therefore we employ a multivariate analysis $\mathrm{MVA_B}$ to combine them into a single more powerful discriminator. Compared with the previous analysis [4], we have relaxed the preselection requirements, developed new discriminating observables, added significantly more data for the model training, and used a more advanced machine learning algorithm. This allowed us to significantly improve the analysis sensitivity achieving the same level as in the previous measurement with just $\sim 60\%$ of the previous data.

Inputs to the $\mathrm{MVA_B}$ can be split into three major classes. The first class includes pointing angles, which are defined as the angles between the B candidate momentum and the line connecting the PV and SV, either in 2D or 3D. We use both definitions since the 2D version benefits from a smaller uncertainty on the vertex position and the 3D version provides additional matching information along the beam line. These observables are effective at rejecting all types of backgrounds except for the ones originating from the two-body decays.

The second class of observables is related to the SV. The dimuon candidates from the combinatorial background tend to correspond to a low-quality SV fit. Therefore the SV probability is one of the most powerful discriminators. In addition to that, it is often possible to find another track that forms a better SV with one of the muons. Furthermore most of the misreconstructed SVs tend to be close to the PV and they can be rejected using the SV displacement information.

The last class of observables represents various isolation variables that are designed to detect nearby decay products present in the semileptonic B meson decays. We examine the isolation environment of each muon as well as the dimuon candidate as a whole. Besides the isolation variables we use additional observables designed to detect extra tracks or vertices compatible with the two muons, which often happen for the partially reconstructed and combinatorial backgrounds. In particular, we compute the number of tracks compatible with the $\mu\mu$ SV based on various track selection requirements. We also look for better-quality SVs formed by one of the muons and an arbitrary track.

The isolation I is determined from the B candidate transverse momentum and other charged-particle tracks in a cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.7$ around the B candidate momentum as follows:

$$I = \frac{p_{\mathrm{T}}(\mathrm{B})}{p_{\mathrm{T}}(\mathrm{B}) + \sum_{\mathrm{trk}} |p_{\mathrm{T}}|},$$

where the sum includes all charged-particle tracks with $p_{\mathrm{T}} > 0.9\,\mathrm{GeV}$ that do not belong to the B candidate and are associated with the same PV as the B candidate.

This selection has been optimized in previous studies to maximize the separation power between $\mathrm{B_s^0} \to \mu^+\mu^-$ signal and background and to achieve a reasonable agreement of the MC simulation with data for the $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ normalization channel.

In addition to the B candidate isolation, we also compute a single-muon isolation I_μ for each muon using charged-particle tracks in a cone with radius $\Delta R = 0.5$ around the muon direction:

$$I_\mu = \frac{|\vec{p}_\mu|}{|\vec{p}_\mu| + \sum_{\mathrm{trk}} |\vec{p}_{\mathrm{trk}}|},$$

where the sum includes all tracks with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ that do not belong to the B candidate and are associated with the same PV as the B candidate.

The parameters for the single-muon isolation have been optimized to achieve maximum separation between signal and background, while minimizing the dependence on the level of pileup.

For the $\mathrm{MVA_B}$ training, we employ the XGBOOST library [38], which implements an advanced gradient boosting algorithm. The training is performed on a mixture of simulated $\mathrm{B_s^0} \to \mu^+\mu^-$ signal events and background events in data selected using the sidebands of the dimuon mass distribution consisting of two regions: $[5.5, 5.9]\,\mathrm{GeV}$ populated by the combinatorial background and $[4.9, 5.1]\,\mathrm{GeV}$ representing a combination of the partially reconstructed and combinatorial backgrounds. The events are split into training and testing categories in a 2:1 proportion. To reuse all available events we trained three classifiers by assigning events to one of the categories based on their event number modulo 3. This allowed us to classify all events in data making sure that no event was evaluated by a classifier that was trained on the event itself.

In order to measure the actual $\mathrm{MVA_B}$ performance in data we use the $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ control sample with modified selection requirements to improve the matching between the $\mathrm{B} \to \mu^+\mu^-$ and $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ kinematic distributions. In addition to the standard selection requirements we require the kaon $p_\mathrm{T} < 1.5\,\mathrm{GeV}$, effectively requiring the kaon to be soft and carry just a small fraction of the original B meson momentum.

Even with the optimized selection, we still have some differences in the kinematic distributions for the two decays, which may have an impact on the analysis. The most important one is the difference in the invariant mass of the $\mathrm{B_s^0}$ and $\mathrm{J/\psi}$ mesons. It has a significant impact on the opening angle between the two muons, which leads to a larger uncertainty in the $\mu\mu$ vertex position along the dimuon momentum and therefore the SV significance that needs to be scaled by a factor ~ 1.6 to match the $\mathrm{B_s^0} \to \mu^+\mu^-$ distribution to the $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ distribution.

To achieve the best matching between the $\mathrm{MVA_B}$ distributions for the $\mathrm{B_s^0} \to \mu^+\mu^-$ and $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ decays we need to select the right input observables for $\mathrm{MVA_B}$ in the $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ channel. For the pointing angle and the impact parameter we use the $\mu\mu K$ final state observables since otherwise we would have a wrong B candidate momentum vector. We also ignore the kaon track in all the isolation calculations and extra track counting. For the rest of the inputs we rely on the $\mu\mu$ observables only.

Figure 1 shows a comparison of the $\mathrm{MVA_B}$ distributions for the $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ simulation and data. The data plots have background subtracted using the *sPlot* technique [39] applied to the UML fits of the $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ invariant mass distributions. We observe good agreement between MC simulation and data for the 2016a and 2016b samples. The agreement is worse for the 2017 and 2018 samples.

We derive corrections to the efficiency of the $\mathrm{MVA_B}$ selection requirements in two different ways. The first ("Ratio") method derives the corrections using the $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ sample and applies them to the $\mathrm{B_s^0} \to \mu^+\mu^-$ and $\mathrm{B^0} \to \mu^+\mu^-$ efficiencies. The second ("XGBOOST") method is based on the idea of reweighting the MC simulation samples to match the data. We were not able to find a single variable, which would allow us to compensate for the discrepancy, and therefore we developed an approach using the XGBOOST algorithm to train a classifier on the difference between the simulation and data in $\mathrm{B^+} \to \mathrm{J/\psi K^+}$ events and use it to reweight the simulated $\mathrm{B} \to \mu^+\mu^-$ events. We trained the XGBOOST classifier using the same inputs that we use for the analysis $\mathrm{MVA_B}$. The data have the backgrounds subtracted via the *sPlot* technique, as described above. The corrections from the two methods are summarized in Table 2. In general the two methods give compatible results within 1–2 standard deviations. We are using results from the XGBOOST method as default, and take the difference between the two

Table 2: Efficiency corrections for the $B_s^0 \to \mu^+\mu^-$ decays derived using two different methods: the efficiency ratio between data and simulation and XGBOOST reweighting in $B^+ \to J/\psi K^+$ events. The loose region is defined as $MVA_B > 0.9$ and the tight one as $MVA_B > 0.99$.

Method	Loose MVA_B selection			Tight MVA_B selection		
	2016	2017	2018	2016	2017	2018
Ratio	1.011 ± 0.013	0.939 ± 0.007	0.903 ± 0.008	1.058 ± 0.019	0.891 ± 0.008	0.885 ± 0.010
XGBOOST	0.991 ± 0.008	0.949 ± 0.003	0.917 ± 0.002	1.008 ± 0.011	0.905 ± 0.004	0.908 ± 0.002

methods as a systematic uncertainty.

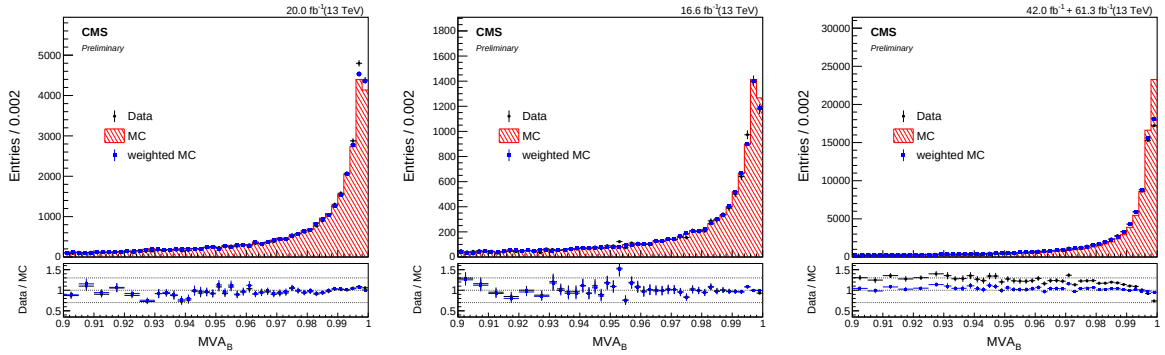

Figure 1: Distributions of the MVA_B output for 2016a (left), 2016b (center), and 2017–2018 (right) data and the corresponding simulations. The blue histograms represent reweighted MC simulations using the XGBOOST reweighting method. The MC distributions are normalized to the data integral.

7 Data analysis

We perform a set of UML fits: the $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ yield fits, simultaneous $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fraction fit, and $B_s^0 \to \mu^+\mu^-$ effective lifetime fit.

For the branching fraction extraction, we perform a 2D fit of the dimuon invariant mass and the relative mass resolution distributions within multiple event categories. The events are categorized using the following three independent criteria:

- data-taking period: 2016a, 2016b, 2017, or 2018;
- signal purity based on the analysis MVA_B: [0.90, 0.99] or [0.99, 1.00]; and
- absolute rapidity of the most forward muon: [0.0, 0.7] or [0.7, 1.4],

leading to 16 unique categories.

The parameters of interest are the branching fractions of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays, which are derived from the corresponding yields. The signal mass model includes the per-event mass resolution in the parameterization. The likelihood consists of five components: B_s^0 signal, B^0 signal, partially reconstructed semileptonic 3-body $B \to h\mu\nu$ and $B \to h\mu\mu$ background, peaking $B \to h^+h^-$ background, and the combinatorial background (h represents a hadron).

We estimated the expected performance of the branching fraction measurement via an ensemble of pseudo-experiments generated using the SM values for the branching fractions and the lifetime. The relative uncertainties in the $\mathcal{B}$ ($B_s^0 \to \mu^+\mu^-$) are expected to be $+11.1/-10.5\%$ with the expected signal significance of 13.1 standard deviations. For the $\mathcal{B}$ ($B^0 \to \mu^+\mu^-$), the expectations are $+67/-62\%$ and 1.67 standard deviations.

In order to extract the effective lifetime of the B_s^0 meson in the $B_s^0 \to \mu^+\mu^-$ decay, we perform an UML fit of the dimuon invariant mass, decay time, and decay time uncertainty. The signal component of the decay time is corrected for the reconstruction and selection efficiency measured in simulation and corrected for mismodeling using the $B^+ \to J/\psi K^+$ decays in data. In order to minimize the difference between the two channels, we used $B^+ \to J/\psi K^+$ events selected in as similar a manner as possible to the $B_s^0 \to \mu^+\mu^-$ decays including the analysis MVA_B, constructed in such a way that it matches MVA_B of $B_s^0 \to \mu^+\mu^-$ decays. The combinatorial background decay time distribution was obtained from the mass sideband in data. The decay time uncertainty is calculated for each event and used as a conditional observable in the fit.

Using pseudo-experiments generated with a complete $B_s^0 \to \mu^+\mu^-$ model, the expected uncertainty in the lifetime is found to be $(+0.18, -0.16)$ ps.

8 Systematic uncertainties

8.1 Branching fraction measurement

The branching fraction measurements have multiple sources of experimental and theoretical systematic uncertainties. The experimental uncertainties are dominated by the uncertainty in the $B \to \mu^+\mu^-$ signal efficiency corrections due to mismodeling of MVA_B in MC simulation, the kaon reconstruction and selection efficiency for the $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ normalization measurements, and the trigger efficiency difference between the signal and normalization channels. The uncertainties in the branching fractions of the $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ decays, as well as in f_s/f_u are considered to be external uncertainties, which are factorized out in the final results.

The signal efficiency corrections for mismodeling of the MVA_B distribution are estimated with two different methods described in Section 6. The two methods give results compatible with each other. Based on the difference between the two methods, we assign a 2 (3)% systematic uncertainty in the corrections for the $MVA_B > 0.9$ (0.99) $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ signal efficiency.

The tracking efficiency difference between data and simulation is assessed using $D^{*\pm}$ decays [40]. This method computes the ratio of the $D^0 \to K^-\pi^+$ to $D^0 \to K^-\pi^+\pi^+\pi^-$ event yields for D^0 mesons coming from $D^{*\pm} \to D^0\pi^\pm$ decays and compares it with the world average value [41]. Based on these results we assign an overall 2.3% systematic uncertainty due to the tracking efficiency for each kaon.

As for other systematic uncertainties, the fit bias is extracted from the difference between the measured branching fraction from the pseudo-experiments and the SM value. The shape uncertainty of the normalization channel is derived by using the different signal pdfs in the yield fits. The pileup uncertainty is extracted from the efficiency ratio difference derived using the pileup distribution of data and MC simulation. The normalization channels require a tighter SV probability than the signal channel due to the different triggers. Its uncertainty is evaluated by the efficiency difference of the tighter SV probability requirement (0.1 with respect to 0.025)

between the data and MC simulation in the $B^+ \to J/\psi K^+$ events.

Table 3 summarizes the systematic uncertainties for the branching fraction measurements using the $B^+ \to J/\psi K^+$ events for normalization. For the $B_s^0 \to \mu^+\mu^-$ branching fraction measurement with the $B_s^0 \to J/\psi\phi$ normalization the tracking efficiency systematics is doubled to 4.6% due to presence of two kaons in the final state and the shape uncertainty is found to be a bit larger, i.e. 1.5%. At the same time, this measurement is free from the B production fraction systematic uncertainty.

The lifetime of B_s^0 mesons has a significant impact on the signal efficiency for the $B \to \mu^+\mu^-$ decays. The branching fraction measurements are reported assuming the SM value for the lifetime. For an alternative hypothesis, the scale factor for the branching fraction is $1.577 - 0.358\tau$, where τ is the B meson lifetime in ps for the alternative hypothesis.

Table 3: Summary of the systematic uncertainties for the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fraction measurements.

Effect	$B_s^0 \to \mu^+\mu^-$	$B^0 \to \mu^+\mu^-$
Trigger efficiency	2 – 4%	
Pileup	1%	
Vertex quality requirement	1%	
MVA_B correction	2–3%	
Tracking efficiency (per kaon)	2.3%	
$B^+ \to J/\psi K^+$ shape uncertainty	1%	
Fit bias	2.2%	4.5%
f_s/f_u- ratio of the B meson production fractions	3.5%	-

8.2 Lifetime measurement

The dominant sources of systematic uncertainties in the lifetime measurement are associated with correlations of the MVA_B with the decay time and with mismodeling of these correlations and other effects in simulation.

The correlation of the decay time and MVA_B comes from the fact that one of the most discriminating variables, the pointing angle, strongly correlates with the decay time. The correlation enters via the decay distance: the larger the decay distance the better one knows the direction between the SV and PV. As the decay distance gets shorter, the uncertainty in the pointing angle increases, making such events harder to distinguish from the background. Mismodeling of these correlations in the simulation can have a significant impact on the decay time distribution.

The decay time is correlated with many selection requirements. Most of them are well simulated. We measure the lifetime bias in $B^+ \to J/\psi K^+$ events using a relaxed selection requirement $MVA_B > 0.9$ and compare it to the prediction from simulation. We find that the bias for the lifetime measurement to be 0.04–0.05 ps depending on the data-taking period.

For the final selection, we derive a correction as a ratio of the decay time distributions at $MVA_B > 0.99$ and $MVA_B > 0.9$ using $B^+ \to J/\psi K^+$ events in data and apply this correction to the $B_s^0 \to \mu^+\mu^-$ decay time distribution extracted from the simulation using $MVA_B > 0.9$ selection requirement. Repeating the procedure using simulated events, we find that the method may introduce up to 0.10 ps bias in 2016 data. The bias is found to be much smaller in 2017 and 2018 data. These effects are taken into account in the lifetime fit by introducing independent nuisance parameters in the fit model.

As for minor systematic uncertainties, the uncertainty of efficiency modeling is derived using the different efficiency functions of the decay time in the lifetime fit of the $B^+ \to J/\psi K^+$ events. We measure the lifetime of the MC samples generated with different lifetimes from the pseudo-experiments while sharing the same efficiency function. The difference between the measured lifetime and the input lifetime of the MC samples is assigned as the systematic uncertainty.

Table 4 summarizes the systematic uncertainties in the lifetime measurement. Most of the uncertainties are treated as uncorrelated between different data-taking periods unless a clear correlation is established.

Table 4: Summary of the systematic uncertainties in the $B_s^0 \to \mu^+\mu^-$ effective lifetime measurement (ps).

Effect	2016a	2016b	2017	2018
Efficiency modeling		0.01		
Lifetime dependence		0.01		
Decay time distribution mismodeling	0.10	0.06	0.02	0.02
Lifetime fit bias	0.04	0.04	0.05	0.04
Total	0.11	0.07	0.05	0.04

9 Results

Using the result of the $B^+ \to J/\psi K^+$ normalization fit with Equations 1 and 3, we find the branching fractions to be:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \left[3.83^{+0.38}_{-0.36}\ (\text{stat})\ ^{+0.19}_{-0.16}\,(\text{syst})\ ^{+0.14}_{-0.13}\,(f_s/f_u)\right] \times 10^{-9},$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = \left[0.37^{+0.75}_{-0.67}\ (\text{stat})\ ^{+0.08}_{-0.09}\,(\text{syst})\right] \times 10^{-10}.$$

These results are based on the following external inputs

- $\mathcal{B}(B^+ \to J/\psi K^+) = (1.020 \pm 0.019) \times 10^{-3}$,
- $\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033) \times 10^{-2}$, and
- $f_s/f_u = 0.231 \pm 0.008$.

The branching fractions are taken from the PDG [41]. The f_s/f_u ratio is derived from the p_T-dependent measurement of the f_s/f_u ratio by LHCb [42]. We are using the effective p_T distribution observed in this measurement shown in Figure 2 to compute an effective f_s/f_u ratio for the phasespace used in this measurement.

The mass projections with all four data-taking periods merged together are shown in Figure 3. The profile likelihood scans as functions of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions for 1D and 2D cases are shown in Figure 4. The event yields for each component of the fit are summarized in Table 5.

We also estimate the branching fractions using the $B_s^0 \to J/\psi\phi$ decays for the normalization. This result is free from the systematic uncertainty on the f_s/f_u ratio, but it depends on the $B_s^0 \to J/\psi\phi$ branching fraction. At the moment the branching fraction measurement is correlated with the f_s/f_u ratio measurement, but in future this may become a more precise test of the SM when

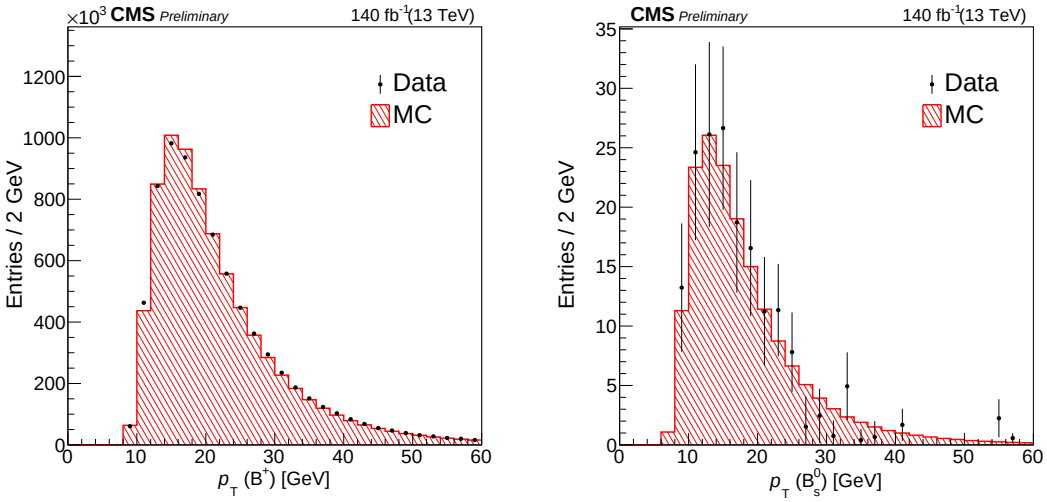

Figure 2: The distribution of the B meson p_{T} after the *sPlot* background subtraction in data (points with the error bars) and simulation (hatched histogram) for $\mathrm{B}^{+} \to \mathrm{J}/\psi\mathrm{K}^{+}$ (left) and $\mathrm{B}^{0}_{\mathrm{s}} \to \mu^{+}\mu^{-}$ (right) events. The MC distributions are normalized to the data integral.

new independent measurements of the $\mathrm{B}^{0}_{\mathrm{s}} \to \mathrm{J}/\psi\phi$ branching fraction become available such as the measurement performed at the Y(5S) resonance by the Belle collaboration at the KEKB $\mathrm{e}^{+}\mathrm{e}^{-}$ collider [43]. Experimentally, the measurement has a bit larger systematic uncertainties due to presence of the second kaon in the final state.

Assuming $\mathcal{B}(\mathrm{B}^{0}_{\mathrm{s}} \to \mathrm{J}/\psi\phi) = (1.018 \pm 0.050) \times 10^{-3}$ [42] and using the result of the $\mathrm{B}^{0}_{\mathrm{s}} \to \mathrm{J}/\psi\phi$ normalization fit with Equation 2, we get:

$$\mathcal{B}(\mathrm{B}^{0}_{\mathrm{s}} \to \mu^{+}\mu^{-}) = \left[3.95^{+0.39}_{-0.37}\,(\mathrm{stat})\,^{+0.27}_{-0.22}\,(\mathrm{syst})\,^{+0.21}_{-0.19}\,(\mathrm{BF})\right] \times 10^{-9}.$$

The observed (expected) signal significance is found to be 12.5 (13.1) and 0.5 (1.7) standard deviations for the $\mathrm{B}^{0}_{\mathrm{s}} \to \mu^{+}\mu^{-}$ and $\mathrm{B}^{0} \to \mu^{+}\mu^{-}$ decays, respectively. The 90 and 95% confidence level (CL) upper limits on $\mathcal{B}$ $(\mathrm{B}^{0} \to \mu^{+}\mu^{-})$ are evaluated using the $\mathrm{CL_s}$ approach [44, 45] and found to be

$$\mathcal{B}(\mathrm{B}^{0} \to \mu^{+}\mu^{-}) < 1.5 \times 10^{-10} \text{ at } 90\%\ \mathrm{CL},$$
$$\mathcal{B}(\mathrm{B}^{0} \to \mu^{+}\mu^{-}) < 1.9 \times 10^{-10} \text{ at } 95\%\ \mathrm{CL},$$

as shown in Figure 5.

The effective lifetime for the $\mathrm{B}^{0}_{\mathrm{s}} \to \mu^{+}\mu^{-}$ decay is found to be

$$\tau = 1.83\,^{+0.23}_{-0.20}\,(\mathrm{stat})\,^{+0.04}_{-0.04}\,(\mathrm{syst})\ \mathrm{ps}.$$

The UML fit projection on the decay time axis for the signal region $5.28 < m_{\mu\mu} < 5.48\,\mathrm{GeV}$ is shown in Figure 6. The observed lifetime is consistent with the SM value within 1 standard deviation, and therefore we do not correct the corresponding selection efficiency when performing the branching fraction measurement.

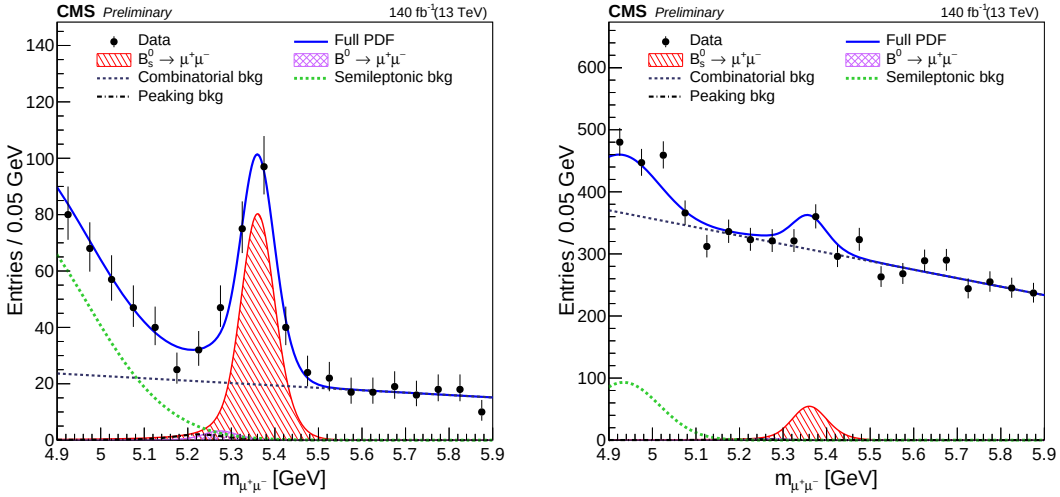

Figure 3: The projections on the dimuon mass axis for the branching fraction fit for the $\mathrm{MVA_B} > 0.99$ category (left) and $0.99 > \mathrm{MVA_B} > 0.90$ category (right). The blue curves represent the corresponding projections of the final fit model.

Table 5: Expected and observed event yields for each category (postfit). Channel 0 and 1 refer to the $|\eta_\mathrm{F}|$ ranges of [0.0,0.7] and [0.7,1.4], respectively.

Data set	Channel	$N(\mathrm{B_s^0})$	$N(\mathrm{B^0})$	N(comb)	N(peak)	N(semi)	N(total)	Data
$\mathrm{MVA_B} > 0.99$								
2016a	0	5.3	0.2	2.8	0.2	6.0	14.5	16
2016a	1	9.4	0.4	16.2	0.4	9.9	36.3	35
2016b	0	6.3	0.3	1.7	0.2	7.9	16.4	12
2016b	1	9.9	0.4	8.6	0.4	13.3	32.6	32
2017	0	23.5	1.0	51.4	0.8	29.6	106.3	114
2017	1	33.9	1.3	89.6	1.4	44.0	170.2	165
2018	0	34.5	1.4	64.8	1.3	38.4	140.4	143
2018	1	50.0	2.0	151.0	2.5	50.9	256.4	252
$0.99 > \mathrm{MVA_B} > 0.9$								
2016a	0	4.8	0.2	118.0	0.2	8.4	131.6	132
2016a	1	8.9	0.4	324.8	0.4	16.5	351.0	352
2016b	0	5.6	0.2	107.6	0.2	10.9	124.5	126
2016b	1	9.2	0.4	257.1	0.4	18.2	285.3	287
2017	0	15.2	0.6	637.7	0.7	26.4	680.6	683
2017	1	21.7	0.9	1430.5	1.1	44.3	1498.5	1498
2018	0	23.3	1.0	936.2	1.2	52.5	1014.2	1017
2018	1	34.2	1.4	2222.5	1.8	79.7	2339.6	2340

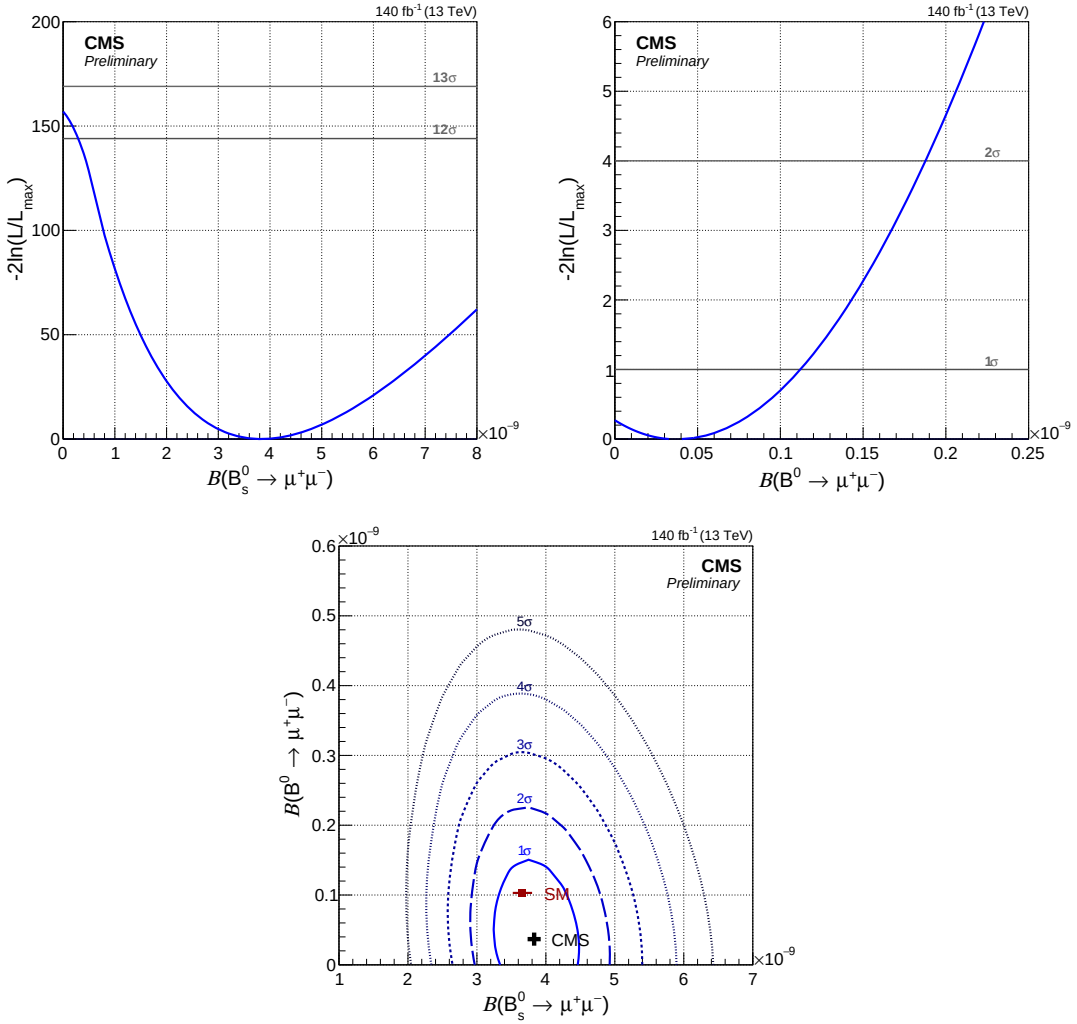

Figure 4: The profile likelihood scan as a function of $B_s^0 \to \mu^+\mu^-$ (left) and $B^0 \to \mu^+\mu^-$ (right) decay branching fractions in 1D (upper plots) and in 2D (lower plot). The contours in 2D enclose the regions with 1-5 standard deviation coverage, where 1σ, 2σ and 3σ regions correspond to 68.3%, 95.4%, and 99.7% confidence level, respectively.

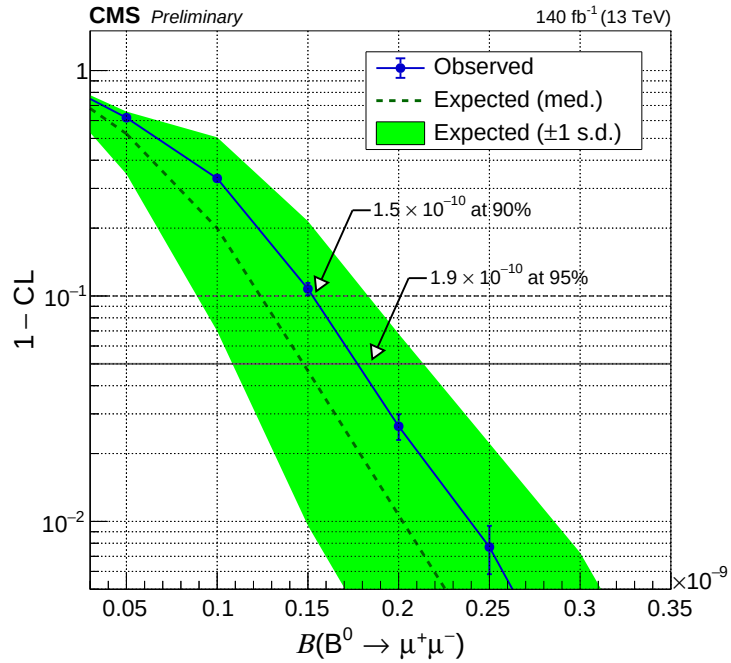

Figure 5: The upper limits on the $B^0 \to \mu^+\mu^-$ decay branching fraction using the CL_s method.

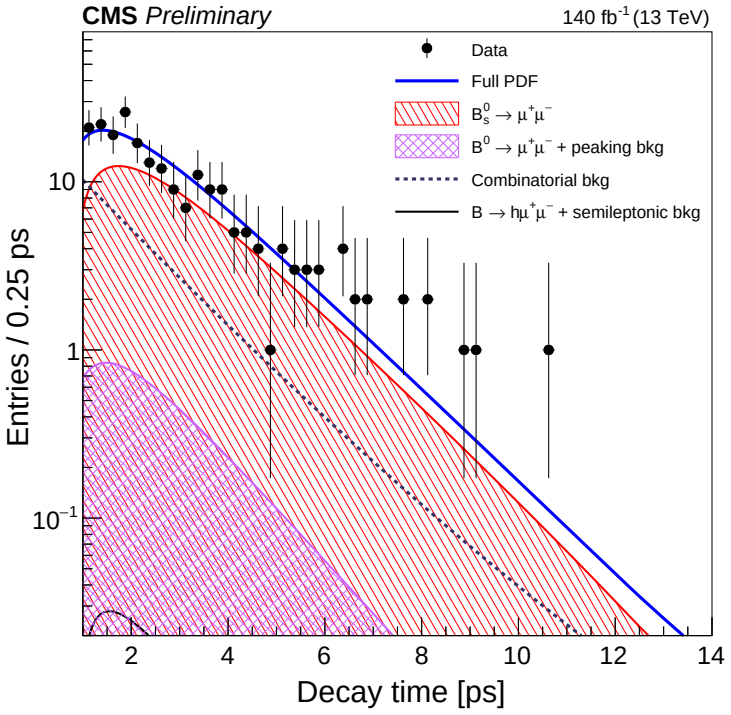

Figure 6: The UML fit projection on the decay time axis for the signal region $5.28 < m_{\mu\mu} < 5.48\,\text{GeV}$.

10 Summary

A measurement of the branching fraction of the $B_s^0 \to \mu^+\mu^-$ decay based on the data set of proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ corresponding to an integrated luminosity of $140\,\text{fb}^{-1}$ has been presented and is found consistent with the standard model (SM) predictions within one standard deviation. The relative combined uncertainty is reduced from 23 to 11% compared with the previous CMS measurement [4] based on the partial 13 TeV data set, while the central value is found to be somewhat higher. The new analysis applied to the data used in Ref. [4] yields a result similar to the original measurement, indicating that the shift in the central value is driven mostly by the new data.

The search for the $B^0 \to \mu^+\mu^-$ decay has not revealed a significant event excess with respect to the dominant combinatorial background prediction. More data will be required to establish its existence and compare the result with the SM predictions.

Compared with the latest LHCb measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.09^{+0.46\,+0.15}_{-0.43\,-0.11}) \times 10^{-9}$ [6], our result is about 1.2 standard deviations higher, which is going to shift the world average from its current value $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (2.69^{+0.37}_{-0.35}) \times 10^{-9}$ [7] to a larger value, more consistent with the SM prediction, thus reducing the overall tension. This implies that the potential beyond-the-SM physics contribution preferred by the global fits to the rare $b \to s\ell^+\ell^-$ decay data [20] is most likely constrained to a single semileptonic operator O_9.

The uncertainties in the branching fraction and effective lifetime measurements are dominated by the statistical component, which allows us to expect significant improvements in the precision of the future measurements with the LHC Run 3 data.

The effective lifetime measurement of the $B_s^0 \to \mu^+\mu^-$ decay is also consistent with the SM predictions and has achieved the level of uncertainty comparable with the lifetime difference between the heavy and light B_s meson mass eigenstates, thus offering sensitivity to potential beyond-the-SM physics effects in the effective lifetime.

References

[1] M. Beneke, C. Bobeth, and R. Szafron, "Power-enhanced leading-logarithmic QED corrections to $B_q \to \mu^+\mu^-$", *JHEP* **10** (2019) 232, doi:10.1007/JHEP10(2019)232, arXiv:1908.07011.

[2] CMS, 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.

[3] 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.

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

[5] 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), no. 19, 191801, doi:10.1103/PhysRevLett.118.191801, arXiv:1703.05747.

[6] 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.

[7] CMS Collaboration, "Combination of the ATLAS, CMS and LHCb results on the $B_{(s)}^0 \to \mu^+\mu^-$ decays", 2020. CMS-PAS-BPH-20-003, LHCb-CONF-2020-002, ATLAS-CONF-2020-049.

[8] LHCb Collaboration, "Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays", *JHEP* **06** (2014) 133, doi:10.1007/JHEP06(2014)133, arXiv:1403.8044.

[9] LHCb Collaboration, "Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$", *JHEP* **09** (2015) 179, doi:10.1007/JHEP09(2015)179, arXiv:1506.08777.

[10] LHCb Collaboration, "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, doi:10.1007/JHEP11(2016)047, arXiv:1606.04731. [Erratum: doi:10.1007/JHEP04(2017)142.

[11] 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.

[12] 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.

[13] 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.

[14] LHCb Collaboration, "Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays", *JHEP* **08** (2017) 055, doi:10.1007/JHEP08(2017)055, arXiv:1705.05802.

[15] LHCb Collaboration, "Test of lepton universality in beauty-quark decays", *Nature Phys.* **18** (2022) 277, doi:10.1038/s41567-021-01478-8, arXiv:2103.11769.

[16] Belle Collaboration, "Test of Lepton-Flavor Universality in $B \to K^*\ell^+\ell^-$ Decays at Belle", *Phys. Rev. Lett.* **126** (2021) 161801, doi:10.1103/PhysRevLett.126.161801, arXiv:1904.02440.

[17] BELLE Collaboration, "Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays", *JHEP* **03** (2021) 105, doi:10.1007/JHEP03(2021)105, arXiv:1908.01848.

[18] CMS Collaboration, "Angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Lett. B* **753** (2016) 424–448, doi:10.1016/j.physletb.2015.12.020, arXiv:1507.08126.

[19] CMS Collaboration, "Measurement of angular parameters from the decay $\mathrm{B}^0 \to \mathrm{K}^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Phys. Lett. B* **781** (2018) 517–541, doi:10.1016/j.physletb.2018.04.030, arXiv:1710.02846.

[20] W. Altmannshofer and P. Stangl, "New physics in rare B decays after Moriond 2021", *Eur. Phys. J. C* **81** (2021), no. 10, 952, doi:10.1140/epjc/s10052-021-09725-1, arXiv:2103.13370.

[21] HFLAV Collaboration, "Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016", *Eur. Phys. J.* **C77** (2017) 895, doi:10.1140/epjc/s10052-017-5058-4, arXiv:1612.07233.

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

[23] 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.

[24] CMS Tracker Group Collaboration, "The CMS phase-1 pixel detector upgrade", *JINST* **16** (2021) P02027, doi:10.1088/1748-0221/16/02/P02027, arXiv:2012.14304.

[25] CMS Collaboration, "Track impact parameter resolution for the full pseudo rapidity coverage in the 2017 dataset with the CMS phase-1 pixel detector", CMS Detector Performance Summary CMS-DP-2020-049, 2020.

[26] 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.

[27] 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.

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

[29] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS", *Eur. Phys. J. C* **81** (2021) 800, doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927.

[30] CMS Collaboration, "CMS luminosity measurements for the 2017 data-taking period at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

[31] CMS Collaboration, "CMS luminosity measurements for the 2018 data-taking period at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-LUM-18-001, 2018.

[32] 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.

[33] 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.

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

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

[36] 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.

[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, pp. 785–794. ACM, New York, NY, USA, 2016. doi:10.1145/2939672.2939785.

[39] M. Pivk and F. R. Le Diberder, "SPlot: A Statistical tool to unfold data distributions", *Nucl. Instrum. Meth.* **A555** (2005) 356, doi:10.1016/j.nima.2005.08.106, arXiv:physics/0402083.

[40] CMS Collaboration, "Tracking performances for charged pions with run2 legacy data", CMS Detector Performance Summary CMS-DP-2022-012, 2022.

[41] Particle Data Group Collaboration, "Review of Particle Physics", *PTEP* **2020** (2020), no. 8, 083C01, doi:10.1093/ptep/ptaa104.

[42] LHCb Collaboration, "Precise measurement of the f_s/f_d ratio of fragmentation fractions and of B^0_s decay branching fractions", *Phys. Rev. D* **104** (2021) 032005, doi:10.1103/PhysRevD.104.032005, arXiv:2103.06810.

[43] Belle Collaboration, "Measurement of the decays $B^0_s \to J/\psi\phi(1020)$, $B^0_s \to J/\psi f'_2(1525)$ and $B^0_s \to J/\psi K^+K^-$ at Belle", *Phys. Rev. D* **88** (2013), no. 11, 114006, doi:10.1103/PhysRevD.88.114006, arXiv:1309.0704.

[44] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[45] A. L. Read, "Presentation of search results: The CL_s technique", *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.