Analysis of Neutral $B$-Meson Decays into Two Muons

R. Aaij et al. (*LHCb Collaboration*)

(Received 24 August 2021; accepted 8 December 2021; published 25 January 2022)

Branching fraction and effective lifetime measurements of the rare decay $B^0_s \rightarrow \mu^+\mu^-$ and searches for the decays $B^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-\gamma$ are reported using proton-proton collision data collected with the LHCb detector at center-of-mass energies of 7, 8, and 13 TeV, corresponding to a luminosity of $9 \, fb^{-1}$. The branching fraction $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-) = (3.69^{+0.46+0.11}_{-0.43-0.14}) \times 10^{-9}$ and the effective lifetime $\tau(B^0_s \rightarrow \mu^+\mu^-) = 2.07 \pm 0.29 \pm 0.03$ ps are measured, where the first uncertainty is statistical and the second systematic. No significant signal for $B^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-\gamma$ decays is found and upper limits $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 2.6 \times 10^{-10}$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-\gamma) < 2.0 \times 10^{-9}$ at the 95% C.L. are determined, where the latter is limited to the range $m_{\mu\mu} > 4.9$ GeV/$c^2$. The results are in agreement with the standard model expectations.

DOI: 10.1103/PhysRevLett.128.041801

The leptonic decays $B^0 \rightarrow \mu^+\mu^-$ and $B^0_s \rightarrow \mu^+\mu^-$ are very rare in the standard model (SM) of particle physics because they only proceed via quantum-loop transitions and are helicity and Cabibbo-Kobayashi-Maskawa (CKM) suppressed. The SM predictions of their time-integrated branching fractions, $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (3.69^{+0.46+0.11}_{-0.43-0.14}) \times 10^{-9}$ and $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-) = (1.03 \pm 0.05) \times 10^{-10}$ [1,2], have small uncertainties owing to the leptonic final state and to the progress in lattice QCD calculations [3–7]. Precise measurements of these observables may reveal discrepancies with the expected values due to the existence of new particles contributing to the decay amplitudes, such as heavy Z' gauge bosons, leptoquarks, or non-SM Higgs bosons (see, e.g., [8]). For these reasons, over the last decades the measurement of the $B^0 \rightarrow \mu^+\mu^-$ and $B^0_s \rightarrow \mu^+\mu^-$ rates has attracted considerable interest in both theoretical and experimental communities, culminating with the observation of the $B^0_s \rightarrow \mu^+\mu^-$ decay using the joint LHCb and CMS Run 1 data sets [9] followed by the first single-experiment observation by LHCb [10]. Recently, the LHCb measurement has been combined with the ATLAS and CMS measurements [11,12] resulting in $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (2.69^{+0.37}_{-0.33}) \times 10^{-9}$ and $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-) < 1.9 \times 10^{-10}$ at 95% confidence level (C.L.) [13], consistent with SM predictions within two standard deviations.

Full author list given at the end of the article.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP3.

0031-9007/22/128(4)/041801(13) 041801-1 © 2022 CERN, for the LHCb Collaboration
any value in the range \([-1, 1]\) may be possible in new physics scenarios. As a consequence, the effective lifetime of \(B_s^0 \rightarrow \mu^+\mu^-\) decays can probe new physics in a way complementary to the branching fraction [25].

This Letter reports improved measurements of the \(B_s^0 \rightarrow \mu^+\mu^-\) and \(B^0 \rightarrow \mu^+\mu^\ast\) time-integrated branching fractions and of the \(B_s^0 \rightarrow \mu^-\) effective lifetime, which supersede the previous LHCb results [10], and a first search for \(B_s^0 \rightarrow \mu^+\mu^-\gamma\) decays. A more comprehensive description of these measurements is reported in a companion article [26].

Inclusion of charge-conjugated processes is implied throughout the Letter. Results are based on data collected with the LHCb detector in the years 2011–2012 and 2015–2018, corresponding to an integrated luminosity of 1 fb\(^{-1}\) of proton-proton (\(pp\)) collisions at a center-of-mass energy \(\sqrt{s} = 7\) TeV, 2 fb\(^{-1}\) at \(\sqrt{s} = 8\) TeV and 6 fb\(^{-1}\) recorded at \(\sqrt{s} = 13\) TeV. The first two data sets are referred to as Run 1 and the latter as Run 2.

The LHCb detector is a single-arm forward spectrometer covering the pseudorapidity range \(2 < \eta < 5\), described in detail in Refs. [27,28]. The simulated events used in this analysis are produced with the software described in Refs. [29–33] taking into account the variations of the accelerator and detector conditions over time. In particular, FSR is simulated using PHOTOS [34]. ISR \(B_s^0 \rightarrow \mu^+\mu^-\gamma\) decays are simulated according to the study in Ref. [14].

The analysis strategy is similar to that employed in Ref. [10], optimized to enhance the sensitivity to both \(B_s^0\) and \(B^0\) decays to \(\mu^+\mu^-\). After loose trigger and selection requirements, \(B_s^0 \rightarrow \mu^+\mu^-\) candidates are classified based on the dimuon mass and the output variable BDT of a boosted decision tree classifier [35,36] designed to distinguish signal from combinatorial background. To avoid the experimenter’s bias, the candidates in the region \([5200, 5445]\) MeV/\(c^2\), where the \(B_s^0 \rightarrow \mu^+\mu^-\) and \(B^0 \rightarrow \mu^+\mu^-\) signal processes peak, were not examined until the full procedure had been finalized. The signal yields are determined from a maximum-likelihood fit to the dimuon mass distribution of the candidates in regions of BDT, and are converted into branching fractions using the decays \(B^0 \rightarrow K^+\pi^-\) and \(B^+ \rightarrow J/\psi K^+\), with \(J/\psi \rightarrow \mu^+\mu^-\), as normalization modes. These decays have been chosen for their relatively large and well-measured branching fractions and because they share the same topology or a dimuon pair in the final state with the signal. The \(B_s^0 \rightarrow \mu^+\mu^-\) effective lifetime is measured from the background-subtracted decay-time distribution of signal candidates.

Events are selected by a hardware trigger followed by a software trigger [37]. The \(B^0 \rightarrow \mu^+\mu^-\) candidates are predominantly selected by single-muon and dimuon triggers. The \(B^+ \rightarrow J/\psi K^+\) candidates are selected in the same way except for a different dimuon mass requirement in the software trigger. Candidate \(B_s^0 \rightarrow h^+h^-\) decays, with \(h^0 = \pi\) or \(K\), are used as control and normalization channels and are triggered independently of the \(B_s^0\) decay products to avoid selection biases.

The \(B_s^0 \rightarrow \mu^+\mu^-\) candidates are reconstructed by combining two oppositely charged tracks with transverse momentum with respect to the proton beam direction \(p_T\) in the range \(0.25 < p_T < 40\) GeV/\(c\), momentum \(p < 500\) GeV/\(c\), and high-quality muon identification [38]. The muon candidates are required to be inconsistent with originating from any primary \(pp\) interaction vertex (PV) and to form a good quality secondary vertex well displaced from any PV. In the selection, \(B_s^0\) candidates must have a decay time less than 13.25 ps, \(p_T > 0.5\) GeV/\(c\) and they must be consistent with originating from at least one PV. A \(B_s^0\) candidate is rejected if either of the two candidate muons combined with any other oppositely charged muon candidate in the event has a mass consistent with the \(J/\psi\) mass [39]. The \(B_s^0 \rightarrow h^+h^-\) selection is the same as that of \(B_s^0 \rightarrow \mu^+\mu^-\), except that the muon identification criteria are replaced with hadron identification requirements and the \(J/\psi\) veto is not applied. The \(B^+ \rightarrow J/\psi K^+\) decay is reconstructed by combining a muon pair, consistent with a \(J/\psi\) from a detached vertex, and a kaon candidate inconsistent with originating from any PV in the event. The selection criteria for signal and normalization candidates include a loose requirement on the response of a different multivariate classifier, described in Refs. [26,40].

The selected events are dominated by combinatorial background, mainly composed of muons originating from two semileptonic \(b\)-hadron decays. The separation between signal and combinatorial background is achieved by means of the BDT classifier, which is optimized using simulated samples of \(B_s^0 \rightarrow \mu^+\mu^-\) events for signal and of \(b\bar{b} \rightarrow \mu^+\mu^-X\) events for background. The classifier combines information from the following input variables: \(\sqrt{\Delta\phi^2 + \Delta\eta^2}\), where \(\Delta\phi\) and \(\Delta\eta\) are the azimuthal angle and pseudorapidity differences between the two muon candidates; the minimum \(\chi^2_{ib}\) of the two muons candidates with respect to the PV\(_B\), where PV\(_B\) is the PV most compatible with the \(B_s^0\) candidate trajectory and \(\chi^2_{ib}\) is defined as the difference between the vertex-fit \(\chi^2\) of the PV formed with and without the particle in question; the angle between the direction of the \(B_s^0\) candidate momentum and the vector joining the \(B_s^0\) decay vertex and PV\(_B\); the \(B_s^0\) candidate vertex-fit \(\chi^2\) and impact parameter significance with respect to the PV\(_B\); and two isolation variables that quantify how much the other tracks of the event are likely to originate from the same hadron decay as the signal tracks. The BDT variable is constructed to be approximately uniform in the range \([0, 1]\) for signal, and to peak strongly at zero for background. Its linear correlation with the dimuon mass is below 5%. The Run 1 and Run 2 data sets are each divided into six subsets based on BDT regions.
with boundaries 0.0, 0.25, 0.4, 0.5, 0.6, 0.7, and 1.0; candidates having BDT < 0.25 are not included in the fit to the dimuon mass distribution. The mass distribution of the $B^0_{(s)} \to \mu^+\mu^-$ candidates with BDT > 0.5 is shown in Fig. 1.

The BDT distributions of $B^0_{(s)} \to \mu^+\mu^-$ decays are calibrated using simulated samples which have been reweighted to improve the agreement with the data. The $p_T$, $\eta$, and $x_{IP}$ quantities of simulated $B^0$ and $B^0_s$ samples are corrected [41] using data samples of $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi \phi$ decays, respectively. The event occupancy is also corrected, separately for each BDT region, by comparing the fraction of $B^+ \to J/\psi K^+$ candidates in four intervals of the number of tracks in simulated events and in data. To align the reconstruction with that of the $B^0_{(s)} \to \mu^+\mu^-$ signal, the BDT response for the $B^+ \to J/\psi K^+$ candidates is evaluated using the information from the final state muons and the $B^+$ candidate, with two exceptions: the $B$ vertex-fit $\chi^2$ is replaced with that of the $J/\psi$, and the muon isolation variables are computed without considering the final-state kaon. The effect of the trigger selection on the BDT distribution is estimated using control channels in data. The resulting $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ BDT variable distributions are found to be compatible with that of $B^0 \to K^+\pi^-$ selected in data when corrected for the different trigger and particle identification selection and, in the case of $B^0_s \to \mu^+\mu^-$, the different lifetime.

The mass distributions of the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ signals are described by two-sided Crystal Ball functions [42] with core Gaussian parameters calibrated from the mass distributions of $B^0 \to K^+K^-$ and $B^0_s \to K^+\pi^-$ data samples, respectively. A mass resolution of about 22 MeV/$c^2$ is determined by interpolating the measured resolutions of charmonium and bottomonium resonances decaying into two muons. The radiative tails are obtained from simulation [43]. Small differences in the resolution and tail parameters of the mass shape for the different BDT regions are taken into account. The mass distribution of the $B^0 \to \mu^+\mu^-$ decays is described with a threshold function modeled on simulated events that were generated using the theoretical predictions of Refs. [14,15], convoluted with the experimental resolution.

The signal branching fractions are determined using the relation

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) = \frac{\mathcal{B}_\text{norm}}{\epsilon}\frac{N_{\text{norm}}}{d(s)} N_{\text{norm}} B^{\text{fit}}(B^0_{(s)} \to \mu^+\mu^-),$$

where $N_{\text{norm}} B^{\text{fit}}(B^0_{(s)} \to \mu^+\mu^-)$ is the signal yield determined in the mass fit, $N_{\text{norm}}$ is the number of selected normalization decays ($B^+ \to J/\psi K^+$ or $B^0 \to K^+\pi^-$), $B_{\text{norm}}$ the corresponding branching fraction [44], and $\epsilon$ ($\epsilon_{\text{norm}}$) is the total efficiency for the signal (normalization) channel. For each signal mode, the two single event sensitivities, $\alpha_{B_s^{0}}^{\mu^+\mu^-}$, are then averaged in a combined $\alpha_{B_s^{0}}^{\mu^+\mu^-}$ taking the correlations into account. The fraction $f_{d(s)}$ indicates the probability for a $b$ quark to fragment into a $B^0_{(s)}$ meson. The value of $f_{s}/f_{d}$ has been measured by LHCb to be $0.254 \pm 0.008$ in $pp$ collision data at $\sqrt{s} = 13$ TeV, while the average value in Run 1 is lower by a factor of 1.064 ± 0.007 [45]. The fragmentation probabilities for the $B^0$ and $B^+$ are assumed to be equal, hence $f_{\text{norm}} = f_{d}$ for both normalization modes.

The acceptance, reconstruction, and selection efficiencies are computed with samples of simulated events generated with the decay-time distribution predicted by the SM. The tracking and particle identification efficiencies are determined using control channels in data [46,47]. The trigger efficiencies are evaluated with control channels in data [48].

The yields of selected $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ decays are $(4733 \pm 3) \times 10^3$ and $(94 \pm 1) \times 10^3$, respectively. The normalization factors measured with the two channels are consistent and their weighted averages, taking correlations into account, are $\alpha_{B^0_{(s)} \to \mu^+\mu^-} = (3.51 \pm 0.13) \times 10^{-11}$, $\alpha_{B^0_{(s)} \to \mu^+\mu^-} = (9.20 \pm 0.17) \times 10^{-12}$, and $\alpha_{B^0_{(s)} \to \mu^+\mu^-} = (4.57 \pm 0.17) \times 10^{-11}$. Assuming SM predictions for the branching fractions, the analyzed data sample is expected to contain an average of $104 \pm 6$ $B^0 \to \mu^+\mu^-$. 

FIG. 1. Mass distribution of the selected $B^0_{(s)} \to \mu^+\mu^-$ candidates (black dots) with BDT > 0.5. The result of the fit is overlaid and the different components are detailed: $B^0 \to \mu^+\mu^-$ (red solid line), $B^0 \to \mu^+\mu^-$ (green solid line), $B^0_s \to \mu^+\mu^-$ (magenta dashed line), $B^0 \to \pi^+\pi^-\mu^-\nu_\mu$, $B^0_s \to K^+\mu^+\nu_\mu$, $B^0 \to J/\psi\mu^-\bar{\nu_\mu}$, and $N^0_s \to p\mu^-\bar{\nu_\mu}$ (orange dashed line), and $B^{0(+)} \to \pi^0\mu^+\mu^-$ (cyan dashed line). The solid bands around the signal shapes represent the variation of the branching fractions by their total uncertainty.
$11 \pm 1 \, B_0^0 \to \mu^+\mu^-$, and about 2 $B_0^0 \to \mu^+\mu^-\gamma$ decays in the BDT $> 0.25$ range and in the mass range $[4900, 6000]$ MeV/c$^2$.

The combinatorial background is distributed exponentially over the whole mass range. In addition, the $B^0$ and $B_0^0$ signal regions and the low-mass sideband are populated by background from specific $b$-hadron decays divided into two categories: those with the misidentification of at least one hadron as a muon and those where two real muons are present and the decay is partially reconstructed. The first category includes $B_{(s)}^0 \to h^+h^-$, $B^0 \to \pi^-\mu^+\nu_\mu$, $B_{(s)}^0 \to K^-\mu^+\nu_\mu$, and $B_0^0 \to p\mu^+\bar{\nu}_\mu$ decays, of which branching fractions are taken from Refs. [44,49,50]. The mass and BDT distributions of these decays are determined from simulated samples after calibrating the $K \to \mu$, $\pi \to \mu$, and $p \to \mu$ momentum-dependent misidentification probabilities using control channels in data. An independent estimate of the $B_{(s)}^0 \to h^+h^-$ background yield is obtained by extracting the yields of misidentified $B^0_{(s)} \to h^+h^-$ decays from the mass spectrum of $\pi^+\mu^-$ or $K^+\mu^-$ combinations in data, and rescaling the observed yields according to the misidentification probabilities. The difference with respect to the result from the first method is assigned as a systematic uncertainty. The second category of background in the low-mass sideband includes the decays $B^+_c \to J/\psi\mu^+\nu_\mu$, with $J/\psi \to \mu^+\mu^-$, and $B^{0(+)} \to \pi^{0(+)}\mu^+\nu_\mu$, which have at least two muons in the final state. The rate of $B^-_c \to J/\psi\mu^+\nu_\mu$ decays is evaluated from Refs. [51,52] and those of $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ decays from Refs. [53,54]. The expected yields of the background contributions originating from specific processes are estimated by normalizing to the $B^+ \to J/\psi K^+$ decay, except for the $B^0_{(s)} \to h^+h^-$ decays, which are normalized to the $B^0 \to K^+\pi^-$ channel. Their expected yields with BDT $> 0.25$ in the full mass range are $37 \pm 2$ $B^0_{(s)} \to h^+h^-$, $161 \pm 6$ $B^0 \to \pi^-\mu^+\nu_\mu$, $31 \pm 3$ $B^0 \to K^-\mu^+\nu_\mu$, $53 \pm 4$ $B^{0(+)} \to \pi^{0(+)}\mu^+\nu_\mu$, $7 \pm 3$ $B_0^0 \to p\mu^+\bar{\nu}_\mu$, and $28 \pm 1$ $B^+_c \to J/\psi\mu^+\nu_\mu$ decays.

The $B^+_c \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, and $B^0 \to \mu^+\mu^-\gamma$ branching fractions are determined with a simultaneous unbinned maximum-likelihood fit [55] to the dimuon mass distribution in the BDT regions of the Run 1 and Run 2 data sets, with BDT $> 0.25$. The fractions of $B_{(s)}^0 \to \mu^+\mu^-$ yield in each BDT region and the parameters of the Crystal Ball functions [42] describing the shapes of the mass distribution are Gaussian constrained according to their expected values and uncertainties. The combinatorial background in each BDT region is described by an exponential function with the yield and slope allowed to vary freely, but the slope parameter is common to all regions within a given data set. Each other background is included as a separate component in the fit. Their yields as well as the fractions in each BDT region are Gaussian-constrained according to their expected values, while their mass shapes are determined from simulation and fixed in the fit, separately in each BDT region. Figure 1 shows the fit results projected on the dimuon mass distribution for BDT $> 0.5$.

The branching fractions of the $B_0^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, and $B^0 \to \mu^+\mu^-\gamma$ decays obtained from the fit are

$$\begin{align*}
B(B_0^0 \to \mu^+\mu^-) &= (3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9}, \\
B(B^0 \to \mu^+\mu^-) &= (1.2^{+0.8}_{-0.7} \pm 0.1) \times 10^{-10}, \\
B(B_0^0 \to \mu^+\mu^-\gamma) &= (2.5 \pm 1.4 \pm 0.8) \times 10^{-9}
\end{align*}$$

with $m_{\mu\mu} > 4.9$ GeV/c$^2$.

The statistical uncertainty is obtained by rerunning the fit with all nuisance parameters fixed to the values found in the default fit. The systematic uncertainties of $B(B_0^0 \to \mu^+\mu^-)$ and $B(B^0 \to \mu^+\mu^-)$ are dominated by the uncertainty on $f_1/f_2$ (3%) and the knowledge of the background from specific processes (9%), respectively. The correlation between the $B^0 \to \mu^+\mu^-$ and $B^{0(+)} \to \mu^+\mu^-$ branching fractions is $-11\%$ while that between the $B_0^0 \to \mu^+\mu^-\gamma$ and $B^0 \to \mu^+\mu^-$ ($B_0^0 \to \mu^+\mu^-\gamma$) branching fractions is $-25\%$ (9%).

Two-dimensional profile likelihoods are evaluated by taking the ratio of the likelihood value of a fit where the parameters of interest are fixed and the likelihood value of the standard fit. They are shown in Fig. 2 for the possible combinations of two branching fractions.

An excess of $B_0^0 \to \mu^+\mu^-$ decays with respect to the expectation from background is observed with a significance of about ten standard deviations ($\sigma$), while the significance of the $B^0 \to \mu^+\mu^-$ signal is $1.7\sigma$, as determined using Wilks’ theorem [56] from the difference in likelihood between fits with and without the specific signal component. The negative fluctuation of the $B_0^0 \to \mu^+\mu^-\gamma$ signal has a $1.6\sigma$ significance.

Since the $B^0 \to \mu^+\mu^-$ and $B_0^0 \to \mu^+\mu^-\gamma$ signals are not significant, an upper limit on each branching fraction is set using the C.L.s method [57] with a profile likelihood ratio as a one-sided test statistic [58]. The likelihoods are computed with the nuisance parameters Gaussian-constrained to their fit values. The test statistic is then evaluated on an ensemble of pseudoexperiments where the nuisance parameters are floated according to their uncertainties. The resulting upper limit on $B(B_0^0 \to \mu^+\mu^-)$ is $2.6 \times 10^{-10}$ at 95% C.L., obtained without constraining the $B_0^0 \to \mu^+\mu^-\gamma$ yield. Similarly, the upper limit on $B(B_0^0 \to \mu^+\mu^-\gamma)$ with $m_{\mu\mu} > 4.9$ GeV/c$^2$ is evaluated to be $2.0 \times 10^{-9}$ at 95% C.L.

Fixing the $B_0^0 \to \mu^+\mu^-\gamma$ signal to zero, the $B_0^0 \to \mu^+\mu^-$ branching fraction increases by about 2% and the upper limit on $B(B_0^0 \to \mu^+\mu^-)$ decreases by about 10%.

The selection efficiency of $B_0^0 \to \mu^+\mu^-$ decays depends on the lifetime, introducing a model dependence in the
The dependence is approximately linear in the physically allowed $A_{\Delta f}$ range. A similar dependence is present for the $B_{s}^{0} \to \mu^{+}\mu^{-}\gamma$ decay with a negligible impact on the branching fraction limit.

The criteria used to select data for the $B_{s}^{0} \to \mu^{+}\mu^{-}$ lifetime measurement differ slightly from those used in the branching fraction measurement. As shown in Fig. 1, the contribution from the misidentified background is negligible under the peak, and therefore a narrower dimuon mass range of [5320, 6000] MeV/c² is selected, while particle-identification requirements are relaxed slightly due to the lower expected contamination from the misidentified background in the $B_{s}^{0} \to \mu^{+}\mu^{-}$ signal region, with a corresponding increase in signal efficiency. Finally, candidate $B_{s}^{0} \to \mu^{+}\mu^{-}$ decays are required to fall into two trigger categories: the trigger requirements must be satisfied entirely either by the $B_{s}^{0} \to \mu^{+}\mu^{-}$ candidates themselves, or by objects from the $p p$ collision that do not form part of the $B_{s}^{0} \to \mu^{+}\mu^{-}$ candidate. These more restrictive trigger requirements are imposed in order to improve the modeling of the decay-time dependence of the trigger efficiency in simulation.

In order to determine the $B_{s}^{0} \to \mu^{+}\mu^{-}$ effective lifetime the data are divided into two BDT regions [0.35, 0.55] and [0.55, 1.00], with boundaries optimized to achieve the best precision. Fits are performed to the dimuon mass distribution in each BDT region in order to extract background-subtracted decay time distributions using the sPlot technique [59]. The mass fits used in the background subtraction include $B_{s}^{0} \to \mu^{+}\mu^{-}$ and combinatorial background components, where the signal is modeled with the same function as in the branching fraction analysis and the background with exponential functions, with freely floating slope parameters in each BDT region. The correlation between the reconstructed mass and the reconstructed decay time of the selected candidates is consistent with zero in both data and simulation, as required by the sPlot technique.

A simultaneous fit is then performed to the two background-subtracted decay-time distributions, where each distribution is modeled by a single exponential multiplied by an acceptance function that models the decay time dependence of the reconstruction and selection efficiency. The acceptance functions are determined in each BDT region by fitting parametric functions to the efficiency distribution of simulated $B_{s}^{0} \to \mu^{+}\mu^{-}$ decays that have been weighted in order to improve the agreement with the data. The correction for the acceptance is validated by measured time-integrated branching fraction. In the fit the SM value for $\tau_{\mu}$, 1.620 ± 0.007 ps [44], is assumed, corresponding to $A_{\Delta f}^{\mu}=1$. The model dependence is evaluated by repeating the fit under the assumptions $A_{\Delta f}^{\mu}=0$ and $-1$, finding an increase of the branching fraction with respect to the SM hypothesis of 4.7% and 10.9%, respectively.

The acceptance functions are determined in each BDT region by fitting parametric functions to the efficiency distribution of simulated $B_{s}^{0} \to \mu^{+}\mu^{-}$ decays that have been weighted in order to improve the agreement with the data. The correction for the acceptance is validated by

FIG. 2. Two-dimensional profile likelihood of the branching fractions for the decays (top) $B^{0} \to \mu^{+}\mu^{-}$ and $B^{0} \to \mu^{+}\mu^{-}$, (center) $B^{0} \to \mu^{+}\mu^{-}$ and $B^{0} \to \mu^{+}\mu^{-}\gamma$ and (bottom) $B^{0} \to \mu^{+}\mu^{-}$ and $B^{0} \to \mu^{+}\mu^{-}\gamma$. The $B^{0} \to \mu^{+}\mu^{-}\gamma$ branching fraction is limited to the range $m_{\mu\mu}$ > 4.9 GeV/c². The measured central values of the branching fractions are indicated with a blue dot. The profile likelihood contours for 68% C.L., 95% C.L., and 99% C.L. regions of the result are shown as blue contours, while in the top plot the brown contours indicate the previous measurement [10] and the red cross shows the SM prediction.
measuring the lifetimes of $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ decays in data. The resulting values are $1.510 \pm 0.015$ ps and $1.435 \pm 0.026$ ps, respectively, where uncertainties are statistical only. These are consistent with the world averages [44]. The statistical uncertainty on the measured $B_s^0 \rightarrow K^+K^-$ lifetime is taken as the systematic uncertainty associated with the use of simulated events to determine the $B_s^0 \rightarrow \mu^+\mu^-$ acceptance function.

A number of sources of systematic bias are evaluated using a large number of simulated pseudoexperiments. The fit procedure is found to produce an unbiased estimate of the lifetime with uncertainties that provide the correct coverage. The effect of the contamination from $B^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h^-$, and semileptonic $b$-hadron decays in the mass fit is found to introduce a small bias of up to 0.012 ps. The effect of the acceptance on the relative admixture of light and heavy mass eigenstates in the decay-time distribution is found to be negligible. Likewise, the uncertainty in the decay-time distribution of the combinatorial background, the production asymmetry between $B^0_s$ and $\bar{B}^0_s$ mesons, and the mismodeling of the acceptance function in simulation is found to have a small effect on the final result. Together, these sources result in a systematic uncertainty of 0.031 ps, which is dominated by the uncertainty on the measured $B^0_s \rightarrow K^+K^-$ lifetime.

The mass distributions of the selected $B^0_s \rightarrow \mu^+\mu^-$ candidates are shown in Fig. 3 (top) for the two BDT regions. Figure 3 (bottom) shows the corresponding background-subtracted $B_s^0 \rightarrow \mu^+\mu^-$ decay-time distribution with the fit function superimposed [55]. The effective lifetime is found to be $2.07 \pm 0.29 \pm 0.03$ ps, where the first uncertainty is statistical and the second systematic. This value lies outside the range between the lifetimes of the light ($A_{\Delta\Gamma} = -1$) and heavy ($A_{\Delta\Gamma} = +1$) mass eigenstates, which are $\tau_L = 1.423 \pm 0.005$ ps and $\tau_H = 1.620 \pm 0.007$ ps [44], but is consistent with these values at 2.2 and 1.5 standard deviations, respectively.

In summary, a new measurement of the rare decay $B_s^0 \rightarrow \mu^+\mu^-$ and a search for $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-\gamma$ decays has been performed using the full dataset collected by the LHCb experiment during Run 1 and Run 2, corresponding to a total integrated luminosity of 9 fb$^{-1}$. The time-integrated branching fraction of $B_s^0 \rightarrow \mu^+\mu^-$ is measured to be $(3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9}$. The $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime is $2.07 \pm 0.29 \pm 0.03$ ps. No evidence for $B^0 \rightarrow \mu^+\mu^-$ or $B_s^0 \rightarrow \mu^+\mu^-\gamma$ signals is found, and the upper limits $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 2.6 \times 10^{-10}$ and $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-\gamma) < 2.0 \times 10^{-9}$ at 95% C.L. are set, where the latter is limited to the range $m_{\mu^+\mu^-} > 4.9$ GeV/$c^2$. The results are in agreement with the SM predictions and can be used to further constrain possible new physics contributions to these observables.

We express our gratitude to our colleagues in the CERN accelerator departments for the excellent performance of the LHC. We thank the technical and administrative staff at the LHCb institutes. We acknowledge support from CERN and from the national agencies: CAPES, CNPq, FAPERJ,
...and FINEP (Brazil); MOST and NSFC (China); CNRS/IN2P3 (France); BMBF, DFG, and MPG (Germany); INFN (Italy); NWO (Netherlands); MNiSW and NCN (Poland); MEN/IFA (Romania); MSHE (Russia); MICINN (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (U.K.); DOE NP and NSF (USA). We acknowledge the computing resources that are provided by CERN, IN2P3 (France), KIT and DESY (Germany), INFN (Italy), SURF (Netherlands), PIC (Spain), GridPP (U.K.), RRCKI and Yandex LLC (Russia), CSCS (Switzerland), IFIN-HH (Romania), CBPF (Brazil), PL-GRID (Poland), and NERSC (U.S.). We are indebted to the communities behind the multiple open-source software packages on which we depend. Individual groups or members have received support from ARC and ARDC (Australia); AvH Foundation (Germany); EPLANET, Marie Skłodowska-Curie Actions and ERC (E.U.); A*MIDEX, ANR, IPhU and Labex P2IO, and Région Auvergne-Rhône-Alpes (France); Key Research Program of Frontier Sciences of CAS, CAS PIFI, CAS CCEPP, Fundamental Research Funds for the Central Universities, and Sci. & Tech. Program of Guangzhou (China); RFBR, RSF and Yandex LLC (Russia); GVA, XuntaGal and GENCAT (Spain); the Leverhulme Trust, the Royal Society and UKRI (U.K.).


[5] A. Bussone et al. (ETM Collaboration), Mass of the $b$ quark and $B$ meson decay constants from $N_f = 2 + 1 + 1$ twisted-mass lattice QCD, Phys. Rev. D 93, 114505 (2016).


[9] V. Khachatryan et al. (CMS and LHCb Collaborations), Observation of the rare $B_{s}^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data, Nature (London) 522, 68 (2015).

[10] R. Aaij et al. (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, 191801 (2017).


[12] A. M. Sirunyan et al. (CMS Collaboration), Measurement of properties of $B^0 \to \mu^+\mu^-$ decays and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment, J. High Energy Phys. 04 (2020) 188.


[14] D. Melikhov and N. Nikitin, Rare radiative leptonic decays $B_{ds} \to \ell^+\ell^-\gamma$, Phys. Rev. D 70, 114028 (2004).


[16] F. Dettori, D. Guadagnoli, and M. Rebourg, $B_{d}^0 \to \mu^+\gamma$ from $B_{s}^0 \to \mu^+\mu^-$, Phys. Lett. B 768, 163 (2017).


[26] R. Aaij et al. (LHCb Collaboration), companion paper, Measurement of the $B_{s}^0 \to \mu^+\mu^-$ decay properties and search for the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ decays, Phys. Rev. D 105, 012010 (2022).


I. Belyaev et al.


R. Aaij et al. (LHCb Collaboration), Strong Constraints on the Rare Decays B^0 \rightarrow μ^-μ^+ and B^0 \rightarrow μ^-μ^+, Phys. Rev. Lett. 108, 231801 (2012).


W. Verkerke and D. Kirkby, The RooFit toolkit for data modeling, eConf C 0303241, MOLT007 (2003), https://ui.adsabs.harvard.edu/abs/2003physics...6116V.


National University of Science and Technology “MISIS”, Moscow, Russia
[associated with Institute of Theoretical and Experimental Physics NRC Kurchatov Institute (ITEP NRC KI), Moscow, Russia]

National Research Tomsk Polytechnic University, Tomsk, Russia
[associated with Institute of Theoretical and Experimental Physics NRC Kurchatov Institute (ITEP NRC KI), Moscow, Russia]

DS4DS, La Salle, Universitat Ramon Llull, Barcelona, Spain (associated with ICCUB, Universitat de Barcelona, Barcelona, Spain)

University of Michigan, Ann Arbor, Michigan, USA (associated with Syracuse University, Syracuse, New York, USA)

Also at Università di Genova, Genova, Italy.

Also at Università di Modena e Reggio Emilia, Modena, Italy.

Also at Università di Ferrara, Ferrara, Italy.

Also at Università di Milano Bicocca, Milano, Italy.

Also at Università di Bologna, Bologna, Italy.

Also at Università di Bari, Bari, Italy.

Also at Università di Cagliari, Cagliari, Italy.

Also at Novosibirsk State University, Novosibirsk, Russia.

Also at Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden.

Also at Università di Roma Tor Vergata, Roma, Italy.

Also at Universidade Federal do Triângulo Mineiro (UFTM), Uberaba-MG, Brazil.

Also at Hangzhou Institute for Advanced Study, UCAS, Hangzhou, China.

Also at AGH—University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Kraków, Poland.

Also at Università di Siena, Siena, Italy.

Also at Università di Padova, Padova, Italy.

Also at Scuola Normale Superiore, Pisa, Italy.

Also at Università degli Studi di Milano, Milano, Italy.

Also at MSU—Iligan Institute of Technology (MSU-IIT), Iligan, Philippines.

Also at Università di Firenze, Firenze, Italy.

Also at Hanoi University of Science, Hanoi, Vietnam.

Also at P.N. Lebedev Physical Institute, Russian Academy of Science (LPI RAS), Moscow, Russia.

Also at Università di Pisa, Pisa, Italy.

Also at Università della Basilicata, Potenza, Italy.

Also at Università di Urbino, Urbino, Italy.