Editors' Suggestion | Featured in Physics

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

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

(Received 23 August 2021; accepted 8 December 2021; published 25 January 2022)

An improved measurement of the decay $B_s^0 \to \mu^+\mu^-$ and searches for the decays $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ are performed at the LHCb experiment using data collected in proton-proton collisions at $\sqrt{s} = 7$, 8 and 13 TeV, corresponding to integrated luminosities of 1, 2 and 6 fb^{-1}, respectively. The $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime are measured to be $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9}$ and $\tau(B_s^0 \to \mu^+\mu^-) = (2.07 \pm 0.29 \pm 0.03)$ ps, respectively, where the uncertainties include both statistical and systematic contributions. No significant signal for $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays is found and the upper limits $\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.6 \times 10^{-10}$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) < 2.0 \times 10^{-9}$ at 95% confidence level are determined, where the latter is limited to the range $m_{\mu\mu} > 4.9$ GeV/c^2. Additionally, the ratio between the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ branching fractions is measured to be $\mathcal{R}_{\mu^+\mu^-} < 0.095$ at 95% confidence level. The results are in agreement with the Standard Model predictions.

DOI: 10.1103/PhysRevD.105.012010

I. INTRODUCTION

Decays mediated by a quark flavor-changing neutral interaction are not allowed at tree level in the Standard Model (SM) of particle physics but can proceed through quantum loops, making them rare processes. The leptonic $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ decays (the inclusion of charge-conjugated processes is implied throughout this paper) are even rarer because they are additionally helicity-suppressed. As they are characterized by a purely leptonic final state, and thanks to the progress in lattice QCD calculations [1–5], their time-integrated branching fractions are predicted in the SM with small uncertainties to be $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.66 \pm 0.14) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.03 \pm 0.05) \times 10^{-10}$ [6,7]. This makes these processes powerful probes for detecting deviations from the SM due to new physics (NP) contributions mediated, for instance, by heavy Z' gauge bosons, leptoquarks or non-SM Higgs bosons (see e.g., Ref. [8]).

An effective field theory description of $b \to s\mu^+\mu^-$ transitions makes it possible to tightly constrain the currents contributing to their amplitudes in a model-independent way. In this framework, the branching fractions of $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ decays are sensitive to axial-vector, scalar and pseudoscalar operators and their chirality-flipped counterparts [9,10]. Of these, only the left-handed axial-vector current is present in the SM at a significant level. Examples of SM Feynman diagrams contributing to the $B_s^0 \to \mu^+\mu^-$ amplitude are shown in Figs. 1(a) and 1(b).

Given the low rate of the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ decays, their branching fractions are measured without distinguishing between $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ at production. Both B^0 and B_s^0 mesons oscillate into their antiparticles but, in contrast to the B^0 system, the light and heavy mass eigenstates of B_s^0 mesons are characterized by a sizeable difference between their decay widths, $\Delta\Gamma_s = 0.085 \pm 0.004$ ps^{-1} [11], and thus have different lifetimes. This gives rise to the relation [12]

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \left[\frac{1 + A^{\mu\mu}_{\Delta\Gamma_s} y_s}{1 - y_s^2}\right] \mathcal{B}(B_s^0 \to \mu^+\mu^-)_{t=0}, \qquad (1)$$

between the flavor-untagged and time-integrated branching fraction and the value at decay time $t = 0$. The theoretical prediction mentioned at the beginning of this section includes this correction. In Eq. (1), $y_s \equiv \Delta\Gamma_s/(2\Gamma_s) = 0.065 \pm 0.003$ [11] and the parameter $A^{\mu\mu}_{\Delta\Gamma_s}$ is defined as $A^{\mu\mu}_{\Delta\Gamma_s} \equiv -2\Re(\lambda)/(1 + |\lambda|^2)$, with $\lambda = (q/p)(A(\bar{B}_s^0 \to \mu^+\mu^-)/A(B_s^0 \to \mu^+\mu^-))$. The complex coefficients p and q relate the mass and CP eigenstates of the $B_s^0 - \bar{B}_s^0$ system with the flavor eigenstates B_s^0 and $\bar{B}_s^0$ (see, e.g., Ref. [11]), and A is the amplitude of the process. In the SM, only the CP-odd eigenstate (which, except for small modifications from CP violation, corresponds to the heavy mass eigenstate) decays

[*]Full author list given at the end of the article.

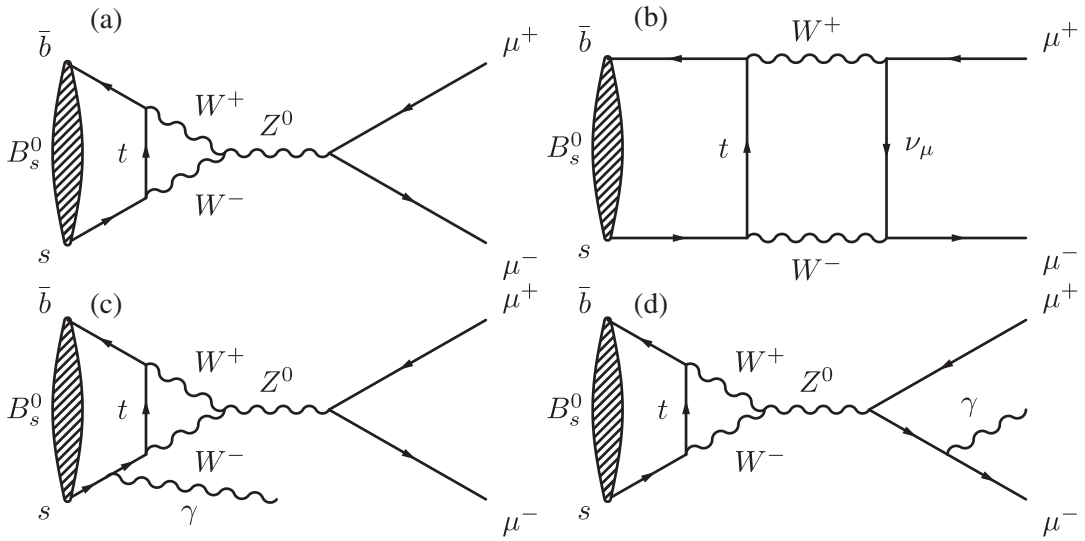

FIG. 1. SM Feynman diagrams mediating (top) the $B_s^0 \to \mu^+\mu^-$ and (bottom) the $B_s^0 \to \mu^+\mu^-\gamma$ processes. Subpanels show (a) the so-called "penguin" diagram and (b) the "box" diagram for $B_s^0 \to \mu^+\mu^-$, and (c) an ISR contribution and (d) an FSR contribution to $B_s^0 \to \mu^+\mu^-\gamma$.

to $\mu^+\mu^-$ and the quantity $A^{\mu\mu}_{\Delta\Gamma_s}$ is equal to unity. However, in the presence of NP contributions it can assume any value in the range $-1 \le A^{\mu\mu}_{\Delta\Gamma_s} \le 1$ [12]. Thus the $B_s^0 \to \mu^+\mu^-$ branching fraction might differ from the SM prediction in either of the two factors in the right-hand side of Eq. (1).

The $B_s^0 \to \mu^+\mu^-$ effective lifetime is defined as

$$\tau_{\mu^+\mu^-} \equiv \frac{\int_0^\infty t\Gamma(B_s(t) \to \mu^+\mu^-)\mathrm{d}t}{\int_0^\infty \Gamma(B_s(t) \to \mu^+\mu^-)\mathrm{d}t} = \frac{\tau_{B_s^0}}{1-y_s^2}\left[\frac{1+2A^{\mu\mu}_{\Delta\Gamma_s}y_s+y_s^2}{1+A^{\mu\mu}_{\Delta\Gamma_s}y_s}\right], \quad (2)$$

where t is the decay time of the B_s^0 or $\bar{B}_s^0$ meson, the decay-time distribution $\Gamma(B_s(t) \to \mu^+\mu^-)$ for $B_s^0 \to \mu^+\mu^-$ and $\bar{B}_s^0 \to \mu^+\mu^-$ decays with and without oscillations is defined as $\Gamma(B_s(t) \to \mu^+\mu^-) \equiv \Gamma(B_s^0(t) \to \mu^+\mu^-) + \Gamma(\bar{B}_s^0(t) \to \mu^+\mu^-)$, and $\tau_{B_s^0} = 1.515 \pm 0.004$ ps [11] is the mean B_s^0 lifetime. By measuring the $B_s^0 \to \mu^+\mu^-$ effective lifetime, the contribution of each mass eigenstate, and thus the CP structure of the interaction involved in the decay, can be inferred, and a direct evaluation of $A^{\mu\mu}_{\Delta\Gamma_s}$ can be performed. The lifetime thus makes it possible to discriminate between contributions from scalar or pseudoscalar interactions in a complementary way to the branching ratio. Similar effects are not significant for $B^0 \to \mu^+\mu^-$ decays due to the negligible decay width difference of the B^0 mass eigenstates.

The ratio of the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ branching fractions also provides powerful discrimination between NP theories [13]. This quantity is theoretically more precise than the two individual branching fractions due to the cancellation of common theoretical uncertainties. It can be obtained as

$$\mathcal{R}_{\mu^+\mu^-} \equiv \frac{\mathcal{B}(B^0 \to \mu^+\mu^-)}{\mathcal{B}(B_s^0 \to \mu^+\mu^-)} = \frac{\tau_{B^0}}{1/\Gamma_H^s}\left(\frac{f_{B^0}}{f_{B_s^0}}\right)^2\left|\frac{V_{td}}{V_{ts}}\right|^2\frac{\sqrt{M_{B^0}^2-4m_\mu^2}}{\sqrt{M_{B_s^0}^2-4m_\mu^2}}, \quad (3)$$

where τ_{B^0} is the lifetime of the B^0, Γ_H^s is the width of the heavy-mass eigenstate of the B_s^0 meson, M_{B^0} and $M_{B_s^0}$ are the masses of the $B^0_{(s)}$ mesons, f_{B^0} and $f_{B_s^0}$ are the $B^0_{(s)}$ meson decay constants, V_{td} and V_{ts} are Cabibbo-Kobayashi-Maskawa (CKM) matrix elements and m_μ is the mass of the muon. In the SM, $\mathcal{R}_{\mu^+\mu^-}$ is predicted to be 0.0281 ± 0.0016 [7] and it assumes the same value in all NP models with the same flavor structure as the SM [14].

The $B_s^0 \to \mu^+\mu^-\gamma$ decay is also rare in the SM. Compared to the $B_s^0 \to \mu^+\mu^-$ amplitude, the additional suppression arising from the photon is compensated by the fact that the amplitude is no longer helicity-suppressed, increasing the total predicted branching fraction to $\mathcal{O}(10^{-8})$ [15–21]. Two groups of amplitudes contribute to this decay: those where the photon is emitted from the initial state (initial-state radiation or ISR), an example of which is shown in Fig. 1(c), and those in which it is emitted from the final state (final-state radiation, FSR), as in Fig. 1(d). Their interference is evaluated to be negligible due to their combined helicity and kinematic suppression [18,19,22]. The FSR contribution to the $B_s^0 \to \mu^+\mu^-\gamma$ process is experimentally included in the $B_s^0 \to \mu^+\mu^-$ decay through the description of the radiative tail in its mass distribution. The ISR component is sensitive to a wider range of interactions and is treated as a separate contribution to the mass fit. In the mass region of interest its

contribution decreases as the mass increases, becoming null for values larger than the B_s^0 mass. Similar to other multibody $b \to s\ell\ell$ decays, the sensitivity to different interactions depends on the dimuon mass squared, q^2, of the decay. At low q^2, the decay is mostly sensitive to tensor and pseudotensor interactions, while at high q^2 vector and axial-vector contributions dominate [23,24]. This makes the ISR $B_s^0 \to \mu^+\mu^-\gamma$ decay at high q^2 an ideal place to probe the same interactions that drive the anomalies seen in some $b \to s\ell\ell$ decays [25–28]. In the rest of this paper, $B_s^0 \to \mu^+\mu^-\gamma$ refers only to the ISR process.

Measurements of $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ processes have attracted considerable experimental interest since the first search for these decays at the CLEO experiment [29] almost forty years ago. The first evidence for the $B_s^0 \to \mu^+\mu^-$ decay was obtained at LHCb [30] with data corresponding to 2 fb^{-1} of proton-proton (pp) collisions, and the decay was then observed through a combined analysis of data taken by the LHCb and CMS experiments [31]. Subsequent measurements were performed by the LHCb collaboration [32] with 4.4 fb^{-1}, by the ATLAS collaboration [33] with 51.3 fb^{-1}, and by the CMS collaboration [34] with 63 fb^{-1}. These last three measurements are combined in Ref. [35], yielding $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (2.69^{+0.37}_{-0.35}) \times 10^{-9}$ and an upper limit on the $B^0 \to \mu^+\mu^-$ decay of $\mathcal{B}(B^0 \to \mu^+\mu^-) < 1.9 \times 10^{-10}$ at 95% confidence level (CL). In the two-dimensional plane of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$, the consistency of the profile likelihood minimum with the SM prediction is measured to be 2.1 standard deviations (σ). To date, no experimental search has been performed for the $B_s^0 \to \mu^+\mu^-\gamma$ decay, while the corresponding B^0 decay has been probed by the *BABAR* experiment, yielding $\mathcal{B}(B^0 \to \mu^+\mu^-\gamma) < 1.5 \times 10^{-7}$ at 90% CL in the whole q^2 region [36], which is well above the SM prediction.

This paper presents improved measurements of the $B_s^0 \to \mu^+\mu^-$ time-integrated branching fraction and effective lifetime, as well as a search for the $B^0 \to \mu^+\mu^-$ decay, superseding the results in Ref. [32]. Moreover, a first search for the $B_s^0 \to \mu^+\mu^-\gamma$ decay at high dimuon mass is also presented. Due to the large uncertainty on the form factors, no attempt is made to extrapolate the $B_s^0 \to \mu^+\mu^-\gamma$ result to the full dimuon range. These results, also reported in Ref. [37], are based on data collected with the LHCb detector, corresponding to an integrated luminosity of 1 fb^{-1} of pp collisions at a center-of-mass energy $\sqrt{s} = 7$ TeV, 2 fb^{-1} at $\sqrt{s} = 8$ TeV and 6 fb^{-1} at $\sqrt{s} = 13$ TeV. The first two datasets are referred to as Run 1 and the latter as Run 2. Throughout this paper, $B_{(s)}^0 \to \mu^+\mu^-$ candidates include $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ or $B_s^0 \to \mu^+\mu^-\gamma$ decays with the dimuon pair selected in the mass range $[4900, 6000]$ MeV$/c^2$ and the photon not reconstructed. The contribution from $B^0 \to \mu^+\mu^-\gamma$ decays is considered negligible compared to that from $B_s^0 \to \mu^+\mu^-\gamma$ because of the additional CKM suppression and the mass shift to lower values.

II. ANALYSIS STRATEGY

The signature of $B_{(s)}^0 \to \mu^+\mu^-$ decays in the LHCb detector consists of two oppositely charged muons with a dimuon mass in the B_s^0 or B^0 mass region, and a decay vertex displaced with respect to any pp interaction vertex as a result of the significant average flight distance of the B mesons. The $B_s^0 \to \mu^+\mu^-\gamma$ channel is searched for with the same signature, without reconstructing the photon, as was proposed in Ref. [22].

The main background can be divided into two categories: combinatorial background arising from random combinations of muons from two distinct b-hadron decays in the same event, and physical background comprising b-hadron decays where one or more final state particles has either been misidentified as a muon or not reconstructed. Combinatorial background candidates are distributed across the entire search region from low to high mass, while physical background contributions tend to populate the region below the B_s^0 mass. The dominant physical background sources are: $B_{(s)}^0 \to h^+h'^-$ decays where the hadrons $h, h' = K, \pi$ are misidentified as muons, which mainly contribute to the B^0 mass region; and partially reconstructed b-hadron decays, which populate the same lower dimuon mass region as the $B_s^0 \to \mu^+\mu^-\gamma$ signal. The most important partially reconstructed background sources are semileptonic $H_b \to h^+\mu^-\bar{\nu}_\mu$ (where H_b is a b hadron), $B_c^+ \to J/\psi\mu^+\nu_\mu$ (with $J/\psi \to \mu^+\mu^-$) and $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ decays.

Combinatorial background is separated from the $B_{(s)}^0 \to \mu^+\mu^-$ signal by exploiting differences between their topologies and the relative isolation from other tracks in the event of the muons forming the B candidate. This information is combined in a multivariate classifier based on a boosted decision tree [38], the output response of which, BDT, is used to classify the events as described in Sec. IV. A stringent particle identification (PID) requirement is used to suppress physical background from $B_{(s)}^0 \to h^+h'^-$ and semileptonic decays, described in Sec. VII.

As presented in Sec. VIII, the signal yields are estimated using an extended unbinned maximum-likelihood fit to the dimuon mass distribution, which is performed simultaneously in intervals of the BDT response to increase the sensitivity of the measurement. The BDT and mass distributions of the signals are calibrated and validated using data, as detailed in Sec. V.

To measure the branching fractions, the yields of $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays are

normalized relative to those of $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ decays, with $J/\psi \to \mu^+\mu^-$, reported in Sec. VI.

The measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime uses a similar selection strategy, which is optimized to achieve the highest sensitivity. After the selection, a maximum-likelihood fit is performed to the dimuon mass distribution in two BDT regions to subtract the background. The decay-time acceptance in each BDT region is calibrated on corrected simulation samples and validated by applying the analysis procedure to $B^0_{(s)} \to h^+h'^-$ candidates from data. Finally, the effective lifetime is extracted using a maximum-likelihood fit to the background-subtracted decay-time distribution, performed simultaneously across both BDT regions, as presented in Sec. IX.

III. DETECTOR AND SIMULATION

The LHCb detector [39,40] is a single-arm forward spectrometer covering the pseudorapidity range $2 < \eta < 5$, designed for the study of particles containing b or c quarks. The detector includes a high-precision tracking system consisting of a silicon-strip vertex detector surrounding the pp interaction region, a large-area silicon-strip detector located upstream of a dipole magnet with a bending power of about 4 Tm, and three stations of silicon-strip detectors and straw drift tubes placed downstream of the magnet. The tracking system provides a measurement of the momentum, p, of charged particles with a relative uncertainty that varies from 0.5% at low momentum to 1.0% at 200 GeV/c. The minimum distance of a track to a primary pp collision vertex (PV), the impact parameter (IP), is measured with a resolution of $(15 + 29/p_{\mathrm{T}})$ μm, where p_{T} is the component of the momentum transverse to the beam, in GeV/c. Different types of charged hadrons are distinguished using information from two ring-imaging Cherenkov detectors. Photons, electrons and hadrons are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers. The online event selection is performed by a trigger, which consists of a hardware stage, based on information from the calorimeter and muon systems, followed by two software stages. The first software stage performs a preliminary event reconstruction using only part of the available event information, while the second stage performs a full event reconstruction.

Simulation is used to estimate the acceptance, reconstruction and selection efficiencies and to optimise the analysis strategy. The pp collisions are generated using PYTHIA [41] with a specific LHCb configuration [42]. Decays of particles are described by EVTGEN [43]. Decays of B_c^+ mesons are generated using the dedicated BCVEGPY generator [44,45]. Final-state radiation in the decay of particles is simulated using PHOTOS [46], which is observed to agree with a full quantum electrodynamics calculation at the level of 1% [47]. ISR $B_s^0 \to \mu^+\mu^-\gamma$ decays are simulated according to the study in Ref. [18]. The interaction of the generated particles with the detector, and its response, are implemented using the GEANT4 toolkit [48], as described in Ref. [49].

IV. SIGNAL SELECTION

In the online event selection, signal candidates are first required to pass the hardware trigger, which selects events with at least one muon with high transverse momentum, followed by a two-level software stage, which applies a full event reconstruction. The software stage imposes minimum requirements on the muon transverse momentum and impact parameter with respect to all PV. However, to maximize the signal selection efficiency, events triggered by particles not related to the signal candidates are also retained for further analysis. Events used in the branching fraction measurement may also be triggered by a combination of particles forming the signal candidate and particles from the rest of the event. However, events that are only triggered from such a combination of signal and background particles are excluded from the effective lifetime measurement in order to ensure accurate modeling of the trigger efficiency as a function of decay time in simulation.

Candidate $B^0_{(s)} \to \mu^+\mu^-$ decays are selected offline by combining two well-reconstructed oppositely charged particles identified as muons [50], with transverse momentum in the range $0.25 < p_{\mathrm{T}} < 40$ GeV/c, and momentum $p < 500$ GeV/c. The muon candidates are required to form a secondary vertex (SV) with a vertex-fit χ^2 per degree of freedom smaller than 9 and separated from any PV with a significance greater than 15. Only muon candidate tracks with $\chi^2_{\mathrm{IP}} > 25$ for any PV are selected, where χ^2_{IP} is defined as the difference between the vertex-fit χ^2 of the PV formed with and without the particle in question.

The resulting $B^0_{(s)}$ candidates must have a decay time lower than 13.25 ps, $\chi^2_{\mathrm{IP}} < 25$ with respect to the PV for which the χ^2_{IP} is minimal (henceforth referred to as the PV associated with the $B^0_{(s)}$ candidate) and $p_{\mathrm{T}} > 0.5$ GeV/c. To suppress the $B_c^+ \to J/\psi\mu^+\nu_\mu$ background, a $B^0_{(s)}$ candidate is rejected if either of the two candidate muons combined with any other oppositely charged muon candidate in the event has a mass within 30 MeV/c^2 of the J/ψ mass [51] (J/ψ veto). Further requirements on the particle identification (PID) information of the two muons are imposed in order to reject misidentified hadronic background. PID identification uses multivariate techniques to combine information from different subsystems taking correlations into account [40].

Candidate $B^0_{(s)} \to \mu^+\mu^-$ decays used in the branching fraction measurements are selected in the dimuon mass range $4900 \leq m(\mu^+\mu^-) \leq 6000$ MeV/c^2, while those used

in the lifetime measurement are selected in a narrower range, $5320 \leq m(\mu^+\mu^-) \leq 6000$ MeV/c^2. The reduced mass range used in the lifetime measurement excludes most of the $B^0 \to \mu^+\mu^-$ and physical background decays that populate the lower dimuon mass region, greatly simplifying the fit and making it possible to impose less stringent PID requirements used to reject misidentified background, thus increasing the signal selection efficiency. To avoid potential biases, the candidates in the mass region $5200 \leq m(\mu^+\mu^-) \leq 5445$ MeV/c^2, where $B^0_{(s)} \to \mu^+\mu^-$ candidates peak, were not examined until the analysis procedure was finalized.

In addition to the signal channels, $B^0_{(s)} \to h^+h'^-$ and $B^+ \to J/\psi K^+$ decays are selected as normalization and control channels, and $B_s^0 \to J/\psi\phi$ as control channel. Candidate $B^0_{(s)} \to h^+h'^-$ decays are selected using the same requirements as the signal channels, except that the muon identification criteria are replaced with hadron identification, the events are triggered independently of the decay final state, and the J/ψ veto is not applied. Candidate $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ decays are formed by combining a muon pair, with mass close to the J/ψ mass [51], with one track ($B^+ \to J/\psi K^+$) or two oppositely charged tracks consistent with originating from a ϕ decay ($B_s^0 \to J/\psi\phi$), with the kaon mass hypothesis assigned. All tracks forming $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ candidates are selected with the same requirements as those applied to select the $B^0_{(s)} \to \mu^+\mu^-$ candidates, except for the dimuon mass range and the particle identification criteria. The muons are only required to pass the muon system identification criteria [50] and no kaon identification criteria are required, given the already excellent signal purity achieved. The same trigger strategy as for the signal decays is used for these two channels.

Background events are further rejected using a loose requirement on the response of a boosted decision tree [38,52,53], which was first described in Ref. [54] and has remained unchanged. The classifier takes as input: the angle between the direction of the momentum of the $B^0_{(s)}$ candidate and the direction defined by the vector joining the primary and the secondary vertices; the $B^0_{(s)}$ candidate IP and its vertex χ^2; the minimum IP of the muons with respect to any PV; the minimum distance between the two muon tracks; the χ^2 of the SV. This classifier is also applied to the control and normalization channels, where for the $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ modes the χ^2 of the SV is replaced with that of the J/ψ vertex. The selected sample of $B^0_{(s)} \to \mu^+\mu^-$ candidates is dominated by random combinations of two muons (combinatorial background), mainly originating from semileptonic decays of two different b hadrons. The reconstruction and selection efficiencies for the signal and normalization modes are reported in Sec. VI C.

Isolation variables, which quantify the possibility that other tracks in the event originated from the same hadron decay as the signal muon candidates, are constructed in order to further reject background. Most combinatorial background candidates arise from semileptonic b-hadron decays, where other charged particles produced in the decay may be reconstructed close to the signal muon candidate. Two isolation variables are designed to recognize these particles, each considering a different type of track: one uses additional tracks that have been reconstructed both before and after the magnet (long tracks), while the other considers tracks reconstructed only in the vertex detector (VELO tracks). These isolation variables are defined based on the proximity of the two muons forming the $B^0_{(s)}$ candidate to other tracks in the event.

The closeness of each muon candidate to either a long track or a VELO track is measured using two dedicated multivariate classifiers that take the following quantities as inputs: the minimum $\chi^2_{\rm IP}$ of the track with respect to any PV; the signed distance between the muon-track vertex and the PV associated to the $B^0_{(s)}$ candidate; the signed distance between the muon-track vertex and the $B^0_{(s)}$ decay vertex; the distance of closest approach between the track and the muon; the angular separation between the track and the muon; a quantity that measures the compatibility of the muon-track system with having originated from the PV associated to the $B^0_{(s)}$ candidate. The long-track isolation classifier takes three additional variables as input: the absolute difference between the azimuthal angles of the track and the muon; the absolute difference between the pseudorapidities of the track and the muon; and the track $p_{\rm T}$. The classifiers are trained on collections of track-muon pairs from simulated $B_s^0 \to \mu^+\mu^-$ decays and from simulated $b\bar{b} \to \mu^+\mu^- X$ events. The latter sample includes decays with two muons originating from two different b-hadrons or from the same b-hadron. In both cases the muons can either originate directly from the hadron containing the b-quark or from intermediate resonances. Only tracks originating from the same b-hadron decay as the muon candidate are considered to train the classifier with the simulated $b\bar{b} \to \mu^+\mu^- X$ events. The output value of the classifiers for a given track-muon pair is defined to be higher when the track is "closer" to the muon. Defining $I(\mu^\pm)$ as the maximum value of a given classifier over all the track-muon pairs in the event, the long and VELO-track isolation variables are each defined as $I(\mu^+) + I(\mu^-)$.

Signal and background events are separated using a final boosted decision tree classifier that combines kinematic, topological and isolation information, defined as in the previous measurement [32]. The BDT response is used to divide the data into samples of varying signal purity, which are then fitted simultaneously as described in Secs. VIII and IX. The classifier is trained using simulated samples of

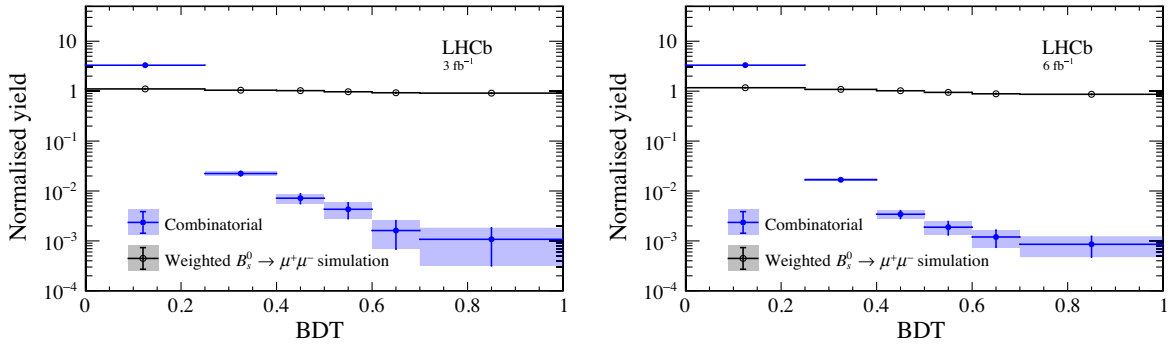

FIG. 2. BDT distribution calibrated using corrected simulated $B_s^0 \to \mu^+\mu^-$ decays (black circles) and combinatorial background from high dimuon-mass data sidebands (blue filled circles) in (left) Run 1 and (right) Run 2 data. Blue error bands represent the statistical uncertainty.

$B_s^0 \to \mu^+\mu^-$ decays as signal and of inclusive $b\bar{b} \to \mu^+\mu^- X$ events as proxy for the combinatorial background. It 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 $\chi^2_{\rm IP}$ of the two muons with respect to the $B^0_{(s)}$ associated PV; the angle between the $B^0_{(s)}$ candidate momentum and the vector joining the $B^0_{(s)}$ decay vertex and $B^0_{(s)}$ associated PV; the $B^0_{(s)}$ candidate vertex-fit χ^2; the $B^0_{(s)}$ impact parameter significance with respect to the $B^0_{(s)}$ associated PV; the long- and VELO-track isolation variables. The BDT classifier response is defined to have an approximately uniform distribution in the range $0 \leq$ BDT ≤ 1 for signal, and to peak at zero for background. Its correlation with the dimuon mass is below 5%. The branching fraction measurement is performed by dividing the Run 1 and Run 2 data samples into six subsets each, based on regions in the BDT response with boundaries 0, 0.25, 0.4, 0.5, 0.6, 0.7 and 1. Figure 2 shows the expected BDT distribution for $B_s^0 \to \mu^+\mu^-$ decays, as determined in Sec. V B, and combinatorial background. The sample with $0 \leq$ BDT < 0.25 is discarded as it is dominated by background. The data used in the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime are split into two regions in the BDT response with the ranges $0.35 \leq$ BDT < 0.55 and $0.55 \leq$ BDT ≤ 1, which are chosen to minimize the expected statistical uncertainty on the effective lifetime based on the results of pseudoexperiments.

V. SIGNAL CALIBRATION

The dimuon mass and the BDT classifier response are used to separate signal from background in the determination of the branching fractions and the $B_s^0 \to \mu^+\mu^-$ effective lifetime. It is therefore essential that these variables are accurately calibrated in order to account for possible discrepancies between data and simulation. The calibration procedures for these two variables are described in the following sections.

A. Mass shape calibration

The mass shape of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ signals is described with a double-sided Crystal Ball (DSCB) function [55]

$$f(m|\mu,\sigma,\alpha_l,n_l,\alpha_r,n_r) = N \begin{cases} \left(\frac{n_l}{\alpha_l}\right)^{n_l} \exp\left[-\frac{\alpha_l^2}{2}\right] \left(-\frac{m-\mu}{\sigma} + \frac{n_l}{\alpha_l} - \alpha_l\right)^{-n_l}, & \text{if } \frac{m-\mu}{\sigma} < -\alpha_l \\ \left(\frac{n_r}{\alpha_r}\right)^{n_r} \exp\left[-\frac{\alpha_r^2}{2}\right] \left(\frac{m-\mu}{\sigma} + \frac{n_r}{\alpha_r} - \alpha_r\right)^{-n_r}, & \text{if } \frac{m-\mu}{\sigma} > \alpha_r \\ \exp\left[-\frac{(m-\mu)^2}{2\sigma^2}\right], & \text{otherwise,} \end{cases} \tag{4}$$

where m is the dimuon mass and all the parameters are positive. The function has a Gaussian core with mean μ and resolution σ and power-law tails on both sides defined by two starting points in units of σ, α_l and α_r, and two slopes, n_l and n_r for the left and right side, respectively.

The mean of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ signal peaks are calibrated with data samples containing $B_s^0 \to K^+K^-$ and $B^0 \to K^+\pi^-$ decays, respectively. Besides the contamination from combinatorial background, the $B^0 \to K^+\pi^-$ sample contains contributions from $B_s^0 \to K^-\pi^+$ decays and partially reconstructed background decays, while the $B_s^0 \to K^+K^-$ sample contains contribution from misidentified $\Lambda_b^0 \to ph^-$ decays. The $m_{K^+\pi^-}$ and $m_{K^+K^-}$ mass distributions of the $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ decays, shown in Fig. 3, are modeled with a DSCB function. The difference between the B^0 and B_s^0 mass values, taken from Refs. [51,56–58], is used to constrain the $B_s^0 \to K^-\pi^+$ mean with respect to the $B^0 \to K^+\pi^-$ mode. The mass resolutions for the two modes are constrained using the mass resolution calibration described

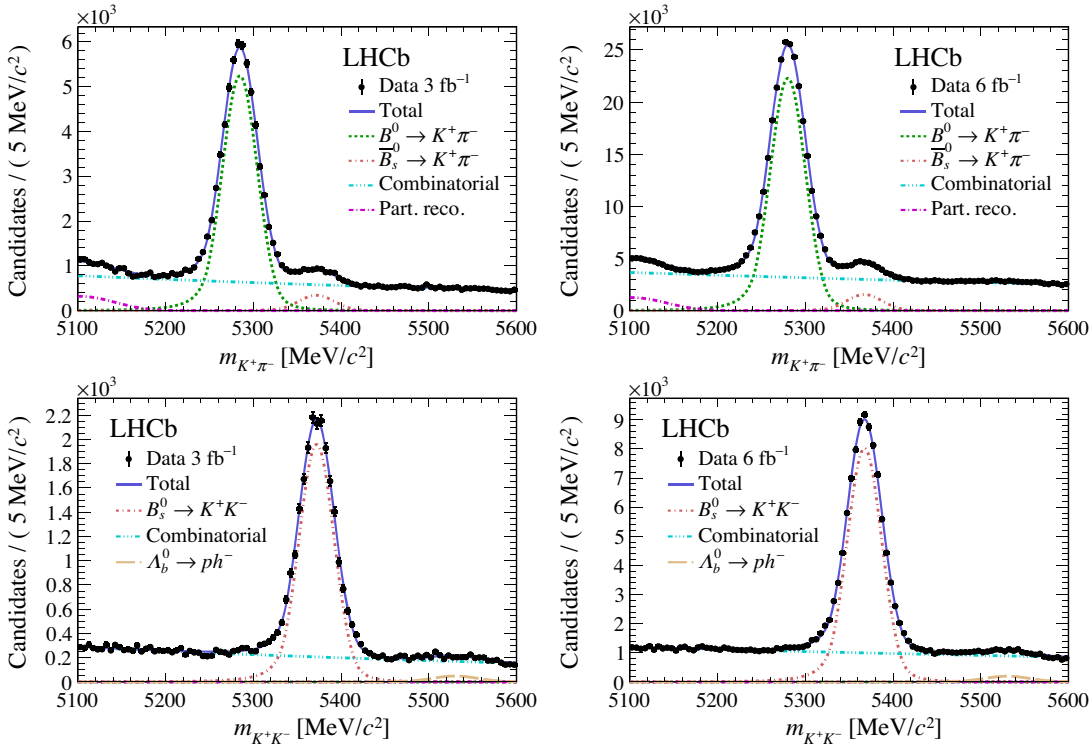

FIG. 3. Mass distributions of selected (top) $B^0 \to K^+\pi^-$ and (bottom) $B_s^0 \to K^+K^-$ candidates in (left) Run 1 and (right) Run 2 data. The results of the fits used to determine the means of the $B^0_{(s)} \to \mu^+\mu^-$ mass distributions are overlaid and the different components are detailed in the legends.

below. The combinatorial background is modeled with an exponential shape with its slope parameter left to vary freely. The partially reconstructed background component is described by an ARGUS shape [59], while the component for $\Lambda_b^0 \to ph^-$ decays is modeled as the sum of two Crystal Ball (CB) functions [55], with all parameters, except the total yield, fixed from simulation. The results of the fits are shown in Fig. 3. To check the correlation between the PID selection and the mean of the $B^0_{(s)} \to h^+h'^-$ signal peak, and to study the effect of possible contamination from misidentified background, the fit is repeated after tightening the PID requirements. The variation of the mean value is assigned as a systematic uncertainty. The mean is found to be uncorrelated with the BDT response.

The mass resolution is calibrated with data samples containing charmonium (J/ψ, $\psi(2S)$) and bottomonium [$\Upsilon(1S)$, $\Upsilon(2S)$ and $\Upsilon(3S)$] resonances decaying into two muons, selected similarly to the signal. The natural widths of all these resonances are negligible compared to the mass resolution of the LHCb experiment. The resolution of each resonance is obtained from a mass fit to the data. The distributions of the dimuon mass, $m_{\mu^+\mu^-}$, of quarkonium decays, shown in Fig. 4, are modeled with a DSCB function. The combinatorial background is modeled with an exponential shape. In the bottomonium fits, the tails are constrained from simulation. A second-order Chebychev polynomial is used as alternative shape for the combinatorial background. The difference between the mass resolutions measured with the two background descriptions is taken as systematic uncertainty. The power-law function $\sigma_{\mu^+\mu^-}(m_{\mu^+\mu^-}) = a_0 + a_1 \cdot (m_{\mu^+\mu^-})^{a_2}$ is found to describe the mass resolution of simulated Drell-Yan events accurately. This function is fitted to the measured resolutions of the quarkonia, including their systematic uncertainties, and used to determine the mass resolution at the B_s^0 and B^0 mass, as shown in Fig 5. The mass resolution in Run 2 is found to be slightly better than in Run 1, which is explained by improvements in the track reconstruction. The $\Upsilon(3S)$ resolution is larger than expected from the power-law function and thus an interpolation with a third-order polynomial was also performed. The differences of the interpolated B^0 and B_s^0 mass resolutions with respect to their default widths are assigned as a systematic uncertainty.

The four tail parameters of the $B^0_{(s)} \to \mu^+\mu^-$ signal shape are in common between the B^0 and B_s^0 decays. They are determined from the mass distributions in simulation, after convoluting them with a Gaussian function to match their core resolutions with the values found in data as described in the previous paragraph.

Each step of the mass calibration is performed separately for each data-taking year. Subsequently, average results for the mass shape parameters for Run 1 and Run 2 are calculated by weighting the year-by-year values by the

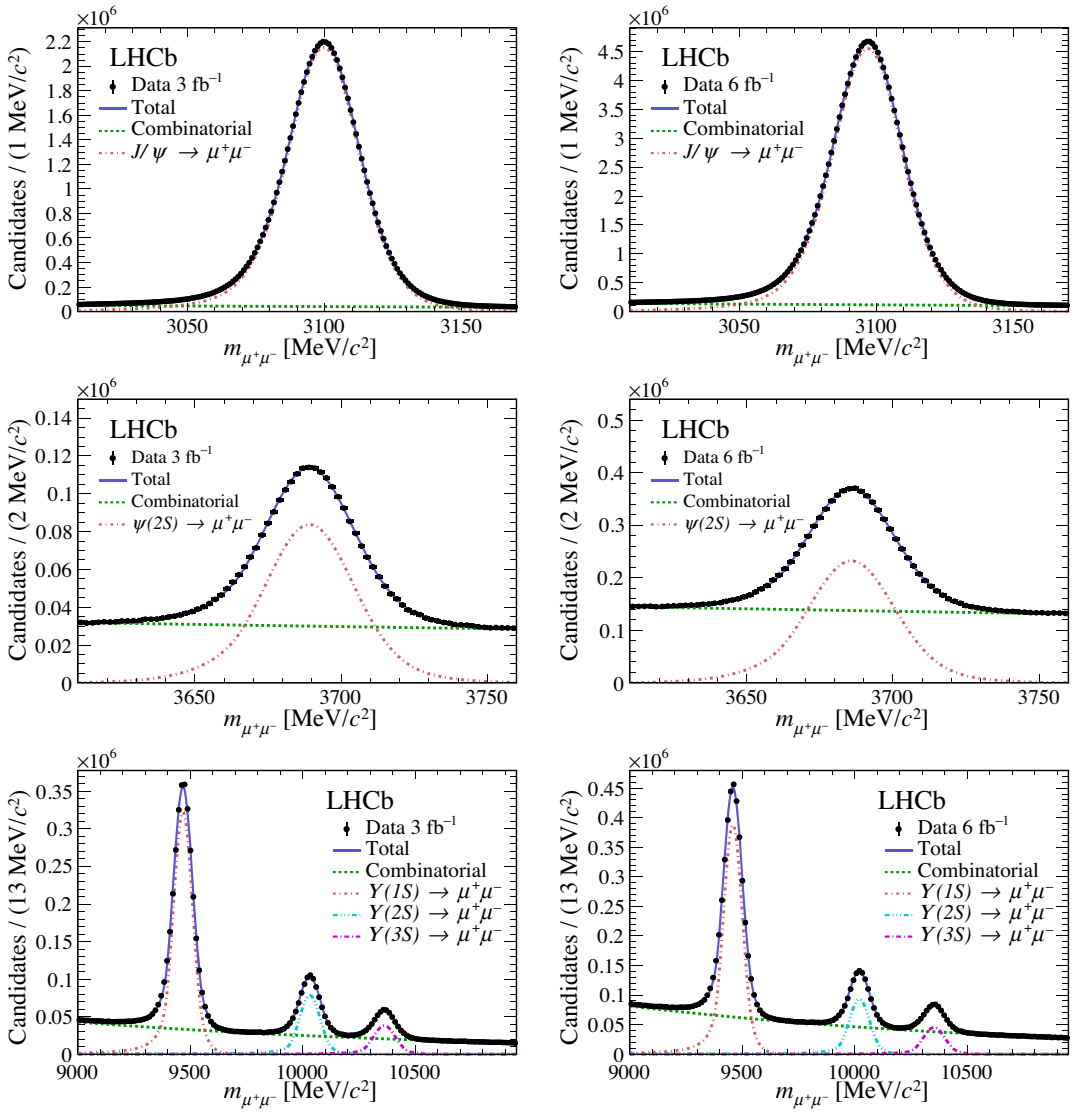

FIG. 4. Mass distributions of (top) $J/\psi \to \mu^+\mu^-$, (center) $\psi(2S) \to \mu^+\mu^-$, (bottom) $\Upsilon(1S, 2S, 3S) \to \mu^+\mu^-$ candidates in (left) Run 1 and (right) Run 2 data. The result from the fit to determine the mass resolutions to each sample is overlaid, and the components are detailed in the legend.

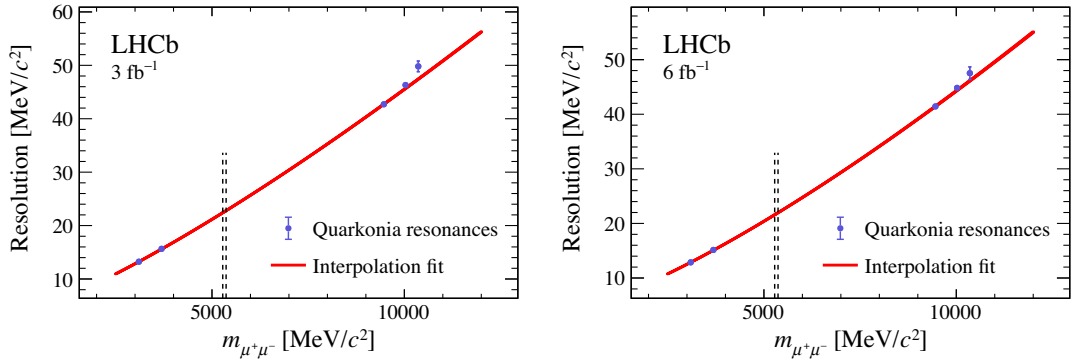

FIG. 5. Fit to the measured mass resolutions of quarkonia resonances (blue dots) with a power-law function to obtain the mass resolution at the B^0 and B_s^0 masses as indicated by the two dashed lines in (left) Run 1 and (right) Run 2 data.

TABLE I. Luminosity-weighted signal mass shape parameter combinations per dataset, including propagated uncertainties. Where appropriate, statistical and systematic uncertainties are added in quadrature. As the tail parameters determined in Run 1 and Run 2 are consistent, they are combined into common estimates.

	Run 1	Run 2	Common
B^0 mean (MeV/c^2)	5284.61 ± 0.18	5280.13 ± 0.16	…
B_s^0 mean (MeV/c^2)	5372.27 ± 0.36	5367.54 ± 0.26	…
B^0 width (MeV/c^2)	22.4 ± 0.7	21.6 ± 0.6	…
B_s^0 width (MeV/c^2)	22.8 ± 0.7	22.0 ± 0.6	…
n_l tail	1.55 ± 0.06	1.49 ± 0.02	1.50 ± 0.04
n_r tail	5.86 ± 0.31	5.80 ± 0.24	5.81 ± 0.26
α_l tail	1.79 ± 0.03	1.79 ± 0.01	1.79 ± 0.02
α_r tail	2.12 ± 0.04	2.14 ± 0.03	2.14 ± 0.04

integrated luminosity in each year. These averages are then used in the final mass fit, with their values shown in Table I. A small but significant dependence of the mass resolution with the BDT response is found, for which correction factors ranging from 0.97 to 1.03 are applied. For the left-hand tail parameters, BDT-dependent correction factors ranging from 0.9 to 1.2 are obtained.

B. BDT calibration

The BDT response is determined using simulated decays, to which corrections are applied to account for possible discrepancies between simulation and data. This calibration procedure is performed in three steps. First, the B -meson kinematics and the detector occupancy of the simulation are corrected using control channels in data. Second, the effect of the PID and trigger selections are evaluated on data and used to correct the BDT response. Finally, since the BDT response for real $B_s^0 \to \mu^+\mu^-$ decays is strongly correlated with their decay-time distribution and hence with the $B_s^0 \to \mu^+\mu^-$ effective lifetime, which is *a priori* unknown, an additional correction is applied under different effective lifetime hypotheses. Taking these three correction factors into account, the fraction of signal decays in each BDT region, $f_{\mathrm{BDT},i}$, can be expressed as

$$f_{\mathrm{BDT},i} = f^{\mu\mu}_{\mathrm{sim},i} \cdot f^{\mu\mu}_{\mathrm{PID},i} \cdot f^{\mu\mu}_{\mathrm{trig},i}(\cdot k_i), \tag{5}$$

where $f^{\mu\mu}_{\mathrm{sim},i}$ is the fraction of events per BDT region in the corrected simulation, $f^{\mu\mu}_{\mathrm{PID},i}$ and $f^{\mu\mu}_{\mathrm{trig},i}$ are the weights used to correct for the PID and trigger selections, respectively, and the effective lifetime correction k_i is used for $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays. No such correction is needed for $B^0 \to \mu^+\mu^-$ decays due to the small width difference between the B^0 mass eigenstates.

The B meson kinematics of the simulated signals are corrected using a gradient boosting reweighter. This technique consists of training a boosted decision tree classifier to align two samples, in this case data and simulation, as described in Ref [60]. The transverse momentum p_{T}, the pseudorapidity η, and the χ^2_{IP} of the B candidate are used as input variables for the gradient boosting reweighter, as these are the variables required to correct the simulation. The weights obtained from this procedure are applied to all simulation samples used for calibration and normalization. The kinematic distributions for B^0 and B_s^0 mesons differ, as determined in hadronization fraction measurements [61], thus they are corrected with a sample of $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ decays, respectively.

An additional correction to the BDT distribution shape stems from the event occupancy, measured as the number of tracks in the event. As this affects the muon track isolation variables, which are important inputs to the BDT classifier, the correction is determined in four intervals of the total number of reconstructed tracks. The correction weights are determined by comparing the relative number of $B^+ \to J/\psi K^+$ decays in background-subtracted data and simulated samples in these intervals. It is ensured that the input variable distributions of $B^+ \to J/\psi K^+$ candidates match those of $B^0_{(s)} \to \mu^+\mu^-$ candidates as closely as possible by evaluating the BDT response based on the final state muons and the B^+ candidate, with two exceptions: the final state kaon is excluded from the calculation of the isolation variables, and the decay vertex χ^2 is determined on the J/ψ candidate, to match the number of degrees of freedom of the signal.

The PID efficiency correction per BDT region for $B^0_{(s)} \to \mu^+\mu^-$ decays, $f^{\mu\mu}_{\mathrm{PID},i}$, is determined on dedicated calibration samples and convolved with the muon kinematics of simulated signal per region, as described in Sec. VI C. As the total PID efficiency is part of the normalization (Sec. VI), the PID efficiency correction of the BDT response is determined as the relative PID efficiency per BDT region; no uncertainty is assigned on this correction, as it is already included in the total efficiency.

A similar procedure is adopted for the trigger efficiency per BDT regions, $f^{\mu\mu}_{\mathrm{trig},i}$. The trigger efficiencies are determined on data samples containing $B^+ \to J/\psi K^+$ decays with the trigger calibration method (see Sec. VI C) [62]. They are calculated in ranges of the maximum p_{T} of the

two muons and the product of the p_{T} of the two muons, which are the variables used for the muon and dimuon hardware trigger. The trigger efficiency per BDT region is determined by the convolution of the obtained efficiencies with the kinematics of simulated signal. The details of this method, also employed for the full efficiency determination, are given in Sec. VI C.

In simulation, $B_s^0 \to \mu^+\mu^-(\gamma)$ decays are generated using the mean B_s^0 lifetime, while the effective lifetime can have any value between the lifetime of the light and the heavy mass eigenstates. As the BDT classifier is correlated with the B_s^0 candidate decay time, an additional correction is included for the $B_s^0 \to \mu^+\mu^-(\gamma)$ BDT response distributions. The correction is evaluated for $A_{\Delta\Gamma_s}^{\mu\mu} = -1$, 0 and 1, corresponding to $\tau_{\mu^+\mu^-} = 1.423$, 1.527 and 1.620 ps, and covering the full physically allowed range. Simulated candidates selected with the procedure described in Sec. IV are weighted according to

$$\omega_j = \frac{\tau_{\mathrm{gen}}}{\tau_{\mu^+\mu^-}} e^{-t_j(1/\tau_{\mu^+\mu^-} - 1/\tau_{\mathrm{gen}})}, \tag{6}$$

where t_j is the reconstructed decay time of the candidate j, τ_{gen} is the lifetime used for generation, and $\tau_{\mu^+\mu^-}$ is the effective lifetime calculated from y_s, $\tau_{B_s^0}$ and $A_{\Delta\Gamma_s}^{\mu\mu}$. A correction factor, k_i, is then calculated for each BDT region according to

$$k_i = \sum_{j=1}^{N_i} \frac{\omega_j}{N_i} \tag{7}$$

where N_i is the number of signal decays in each BDT region.

The calibrated BDT response for $B_s^0 \to \mu^+\mu^-$ decays under the SM hypothesis $A_{\Delta\Gamma_s}^{\mu\mu} = 1$ and for $B^0 \to \mu^+\mu^-$ decays are shown in Fig. 6. The main systematic uncertainties on the calibrated BDT distribution arise from the limited samples used for the trigger efficiency and the event occupancy corrections; they are summed in quadrature with each other and with the statistical uncertainties to determine the total uncertainty on the BDT distribution. The systematic uncertainty on the trigger efficiency correction has been evaluated using $B^+ \to J/\psi K^+$ as a control channel and is related to possible discrepancies between the signal decay and the control channel and to mismodeling in the simulation used in the TISTOS method described in Sec. VI C. The systematic uncertainty due to the event occupancy is obtained by comparing the correction obtained with the default and an alternative interval scheme used to determine the occupancy weights. The systematic uncertainties from the PID selection correction and kinematic reweighting are found to be negligible.

To validate the BDT calibration procedure, an alternative calibration is performed using a data sample containing $B_{(s)}^0 \to h^+h'^-$ decays. While this procedure directly measures the BDT distribution on data and the BDT response is expected to be very similar for any two-body B decay, $B_{(s)}^0 \to h^+h'^-$ decays require significant corrections to be compared to $B_{(s)}^0 \to \mu^+\mu^-$ decays, making it less precise than the default strategy.

The two most frequent B-meson decays into two hadrons, $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$, are considered for the BDT calibration cross-check of $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$, respectively. The same selection is required for $B_{(s)}^0 \to h^+h'^-$ candidates as for the signal, except for the trigger and the particle identification requirements. A trigger selection independent of the $B_{(s)}^0$ decay products is applied to the candidates to select $B_{(s)}^0 \to h^+h'^-$ decays in order to avoid selection biases. Then, a PID selection is applied to separate $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ decays from other $B_{(s)}^0 \to h^+h'^-$ decays.

The fraction of $B_{(s)}^0 \to h^+h'^-$ decays in each BDT region, $f_{\mathrm{data},i}^{hh'}$, is determined by fitting the mass distribution of the

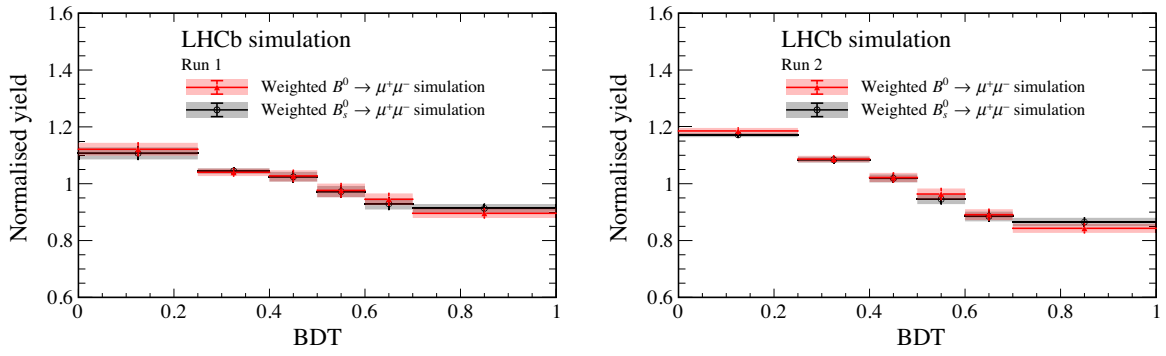

FIG. 6. The calibrated BDT distribution for $B_s^0 \to \mu^+\mu^-$ decays with $A_{\Delta\Gamma_s}^{\mu\mu} = 1$ (black) and $B^0 \to \mu^+\mu^-$ decays (red) for (left) Run 1 and (right) Run 2, including the total uncertainty on the fraction per BDT region. The $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ distributions are determined on corrected simulated samples, as described in the text.

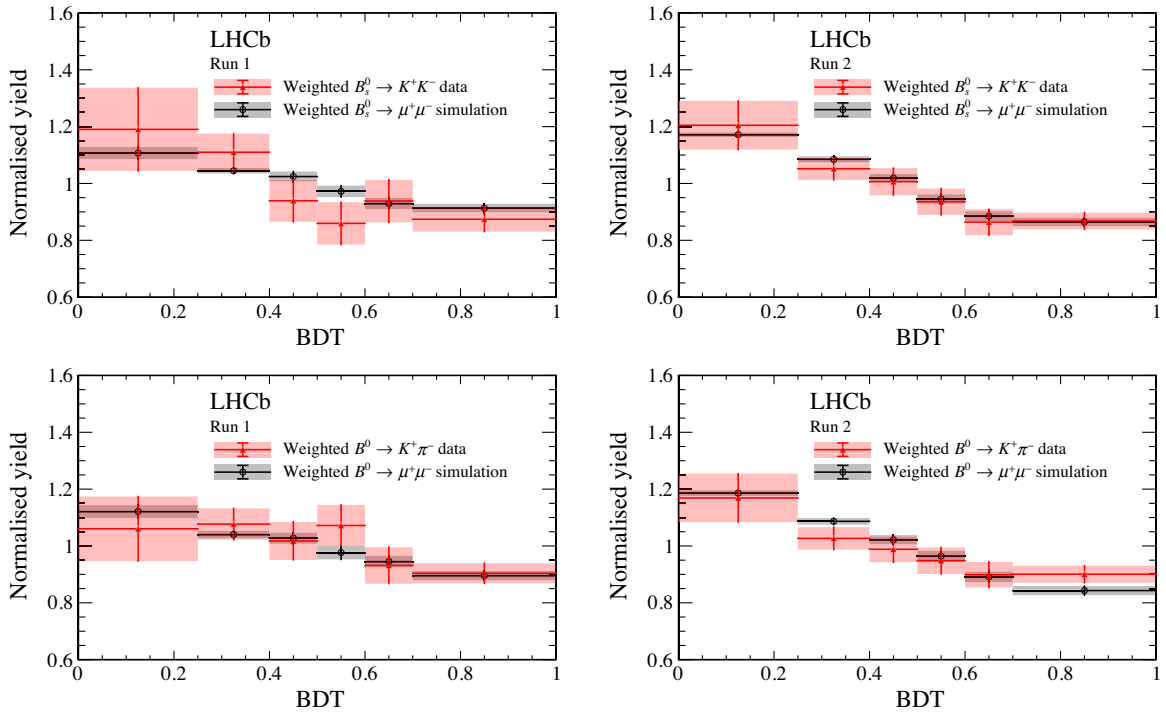

FIG. 7. The BDT distributions of (top) $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to K^+K^-$ decays and (bottom) $B^0 \to \mu^+\mu^-$ and $B^0 \to K^+\pi^-$ decays in (left) Run 1 and (right) Run 2 data, including the total uncertainty on the fraction per BDT region. For $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays, the distributions are determined on corrected simulated samples, as described in the text, and are shown in black. The $B_s^0 \to K^+K^-$ and $B^0 \to K^+\pi^-$ distributions are determined using fits to data as described in the text and are shown in red.

two hadrons with the corresponding mass hypothesis. A detailed description of the mass fit for $B^0_{(s)} \to h^+h'^-$ decays is given in Sec. V A.

To avoid correlated uncertainties, the BDT distribution from $B^0_{(s)} \to h^+h'^-$ data corrected for PID and trigger efficiencies is compared with the distribution of $B^0_{(s)} \to \mu^+\mu^-$ decays from corrected simulation samples. The $B^0_{(s)} \to h^+h'^-$ PID efficiencies are determined with a dedicated procedure [63], while the trigger efficiency is evaluated on $B^+ \to J/\psi K^+$ decays with the trigger calibration method reported in Sec. VI C for the hardware trigger and first software trigger requirements, and determined with $B^0_{(s)} \to h^+h'^-$ simulated samples for the second software trigger selection.

Therefore, the fraction of $B^0_{(s)} \to h^+h'^-$ decays per BDT region, $f'_{\mathrm{BDT},i}$, can be described as

$$f'_{\mathrm{BDT},i} = f^{hh'}_{\mathrm{data},i} \cdot f^{hh'}_{\mathrm{PID},i} \cdot f^{hh'}_{\mathrm{trig},i}, \tag{8}$$

where $f^{hh'}_{\mathrm{PID},i}$ and $f^{hh'}_{\mathrm{trig},i}$ are the relative PID and trigger efficiencies for $B^0_{(s)} \to h^+h'^-$ decays versus $B^0_{(s)} \to \mu^+\mu^-$ decays. The corrected $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ distributions are compared with those determined on corrected simulated $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ samples, respectively, as shown in Fig. 7. The BDT distributions of $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to K^+K^-$ decays are compared for the same effective lifetime, namely for $A^{\mu\mu}_{\Delta\Gamma_s} = 1$. Because of the good agreement between the two different methods used to calibrate the BDT response, no additional systematic uncertainty is assigned to the BDT distribution.

VI. NORMALIZATION

The branching fractions of the signal channels are estimated by comparing their yields in data with those of two normalization channels with well-known branching fractions, $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$, according to

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-(\gamma)) = \frac{f_{\mathrm{norm}}}{f_{\mathrm{sig}}} \frac{\varepsilon_{\mathrm{norm}}}{\varepsilon_{\mathrm{sig}}} \frac{N_{\mathrm{sig}}}{N_{\mathrm{norm}}} \mathcal{B}_{\mathrm{norm}} = \alpha_{\mathrm{sig}} N_{\mathrm{sig}}, \tag{9}$$

where $\mathcal{B}$, ε and N are the branching fraction, efficiency and yield of the corresponding channel and $f_{\mathrm{sig(norm)}}$ indicates the fragmentation fraction of the relevant B meson. Signal candidates having BDT < 0.25 are not included in the fit to the dimuon mass distribution. The parameter α_{sig} is the single-event sensitivity. In the following, the different elements entering Eq. (9) are described. The final single-event sensitivity is obtained for each signal channel as the weighted average of those obtained with the two normalization channels, taking the correlations between the inputs

into account. The branching fractions of the two normalization channels are taken as $\mathcal{B}(B^+ \to J/\psi K^+) = (6.02 \pm 0.17) \times 10^{-5}$ [64–66], including the $J/\psi \to \mu^+\mu^-$ branching fraction, and $\mathcal{B}(B^0 \to K^+\pi^-) = (1.96 \pm 0.05) \times 10^{-5}$ [51,67–75]. The efficiency for the $B_s^0 \to \mu^+\mu^-\gamma$ channel is calculated only for those signal decays in the region $m_{\mu\mu} > 4.9\ \mathrm{GeV}/c^2$ where the branching fraction is measured.

A. Normalization channel yields

The yields of the normalization channels are obtained through unbinned extended maximum-likelihood fits to the mass distributions of the candidates for each data-taking year separately, after the corresponding selection described in Sec. IV. To improve the mass resolution of $B^+ \to J/\psi K^+$ candidates, the J/ψ mass is constrained to its known value [51,76–79]. The mass distributions for selected $B^+ \to J/\psi K^+$ candidates are shown in Fig. 8 for the different data taking periods. The mass distribution of signal $B^+ \to J/\psi K^+$ decays is described by a Hypatia function [80] with parameters Gaussian-constrained to the values derived from simulation within their uncertainties, except for the mean and width, which are free to vary in the fit. In addition to signal, the selected candidates contain a contribution of combinatorial background and a small contamination from the $B^+ \to J/\psi\pi^+$ decay, where a pion is misidentified as a kaon. The mass distribution of the combinatorial background is described by an exponential function with the slope left free to vary in the fit. The $B^+ \to J/\psi\pi^+$ decay, which is expected to peak at higher mass

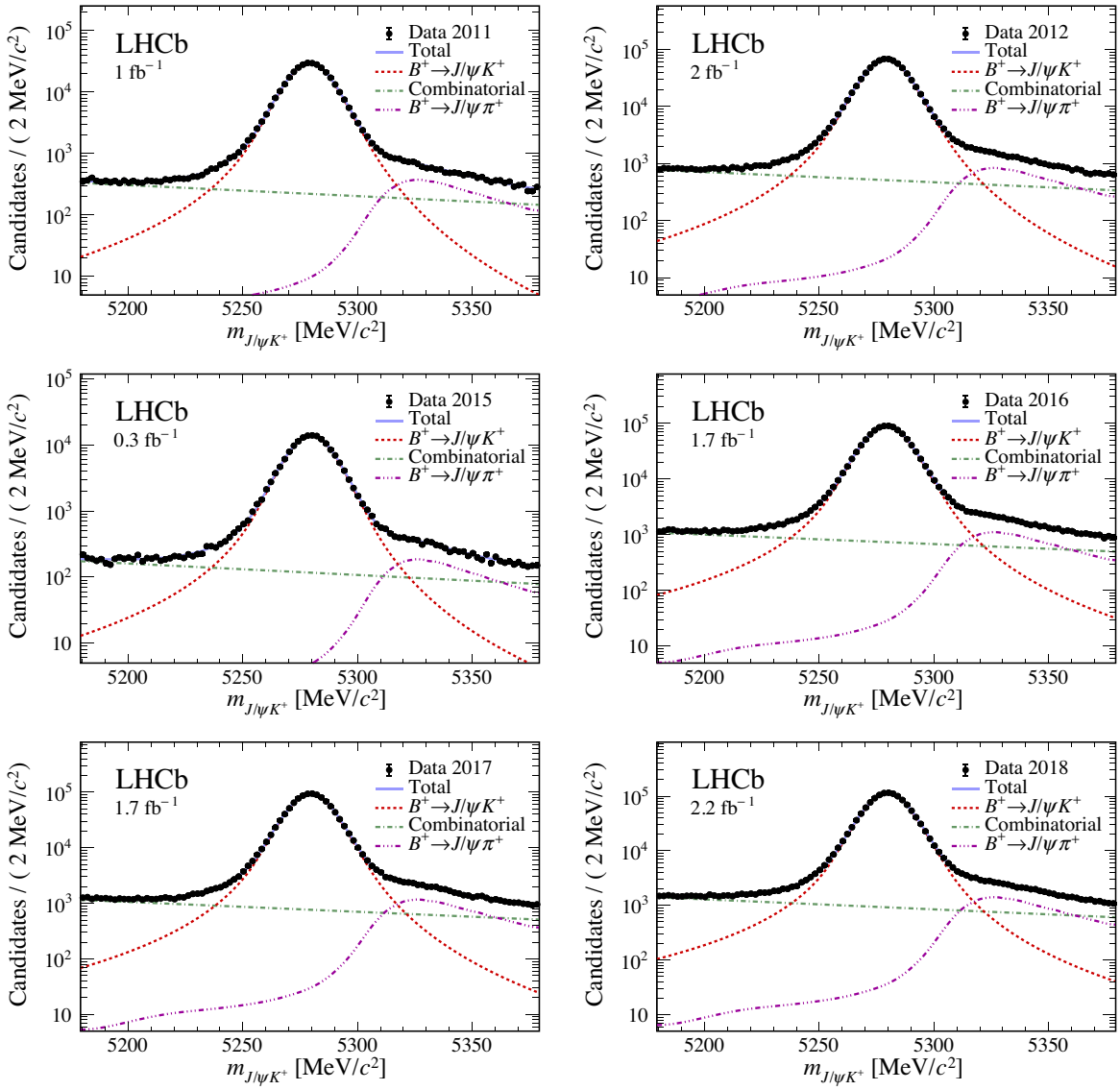

FIG. 8. Mass distribution of $B^+ \to J/\psi K^+$ candidates in data for different data-taking years. Superimposed is a fit to the distribution: the blue line shows the total fit, the red dashed line is the $B^+ \to J/\psi K^+$ component, the green dash-dotted line is the combinatorial background, the purple dash-three-dotted line is the $B^+ \to J/\psi\pi^+$ misidentified background.

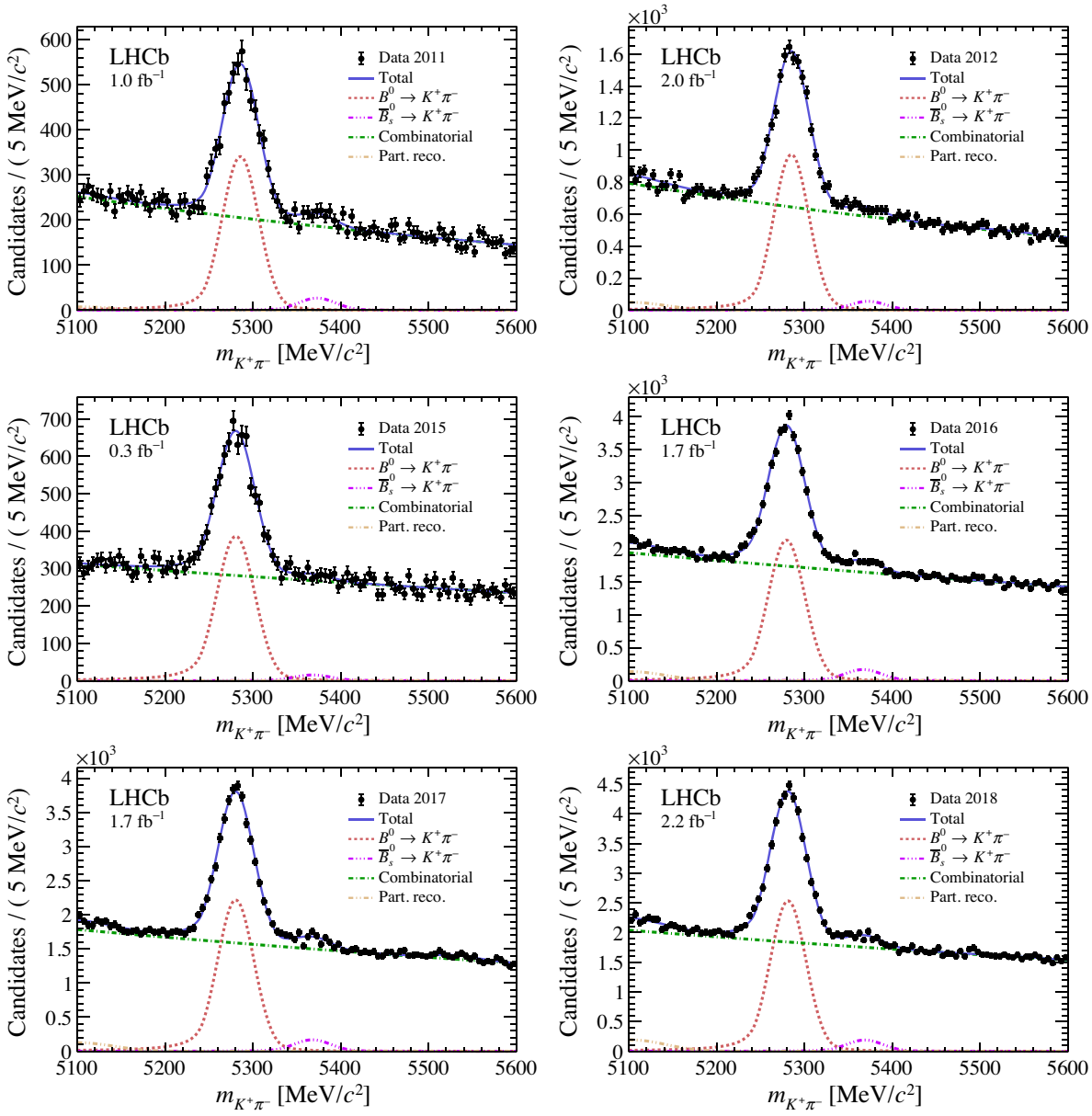

FIG. 9. Mass distribution of $B^0 \to K^+\pi^-$ candidates in data for different data-taking years, triggered independently of the signal. Superimposed is a fit to the distribution: the blue line shows the total fit, the red dashed line is the $B^0 \to K^+\pi^-$ component, the magenta dashed line is the $\bar{B}_s^0 \to K^+\pi^-$ component, the green dashed line is the combinatorial background, and the brown dashed line is the partially reconstructed background component.

values than the B^+ mass due to assigning the kaon mass to a pion track, is described by an analytical function developed in Ref. [81]. This function is obtained by transforming a Gaussian mass distribution under the pion hypothesis to one under the incorrectly-assigned kaon mass hypothesis, using an analytical description of the candidate kinematics. As an alternative model a non-parametric function tuned on a simulated $B^+ \to J/\psi\pi^+$ sample, where the events were reconstructed under the $B^+ \to J/\psi K^+$ hypothesis as in Ref. [82], has been used to cross check the results. The fits with the two different descriptions for $B^+ \to J/\psi\pi^+$ decays give compatible results for the $B^+ \to J/\psi K^+$ yields.

Additional possible background from $B_c^+ \to J/\psi K^+K^-\pi^+$ and $\Lambda_b^0 \to J/\psi pK^-$ decays has been investigated and found to be small and evenly distributed in the considered mass range, and hence is considered negligible.

The yield of $B^0 \to K^+\pi^-$ candidates is determined with a binned maximum-likelihood fit to the data, using the fit model for $B^0 \to K^+\pi^-$ candidates described in Sec. V A. In contrast to the mass calibration, events are required to be triggered independently of the $B^0 \to K^+\pi^-$ signal, such

TABLE II. Yields of the two normalization channels with their combined statistical and systematic errors.

Period	$B^+ \to J/\psi K^+$	$B^0 \to K^+\pi^-$
2011	$(3.479 \pm 0.008) \times 10^5$	$(3.73 \pm 0.13) \times 10^3$
2012	$(7.780 \pm 0.012) \times 10^5$	$(10.32 \pm 0.23) \times 10^3$
2015	$(1.676 \pm 0.005) \times 10^5$	$(4.43 \pm 0.14) \times 10^3$
2016	$(10.369 \pm 0.015) \times 10^5$	$(2.37 \pm 0.06) \times 10^4$
2017	$(10.820 \pm 0.014) \times 10^5$	$(2.43 \pm 0.06) \times 10^4$
2018	$(13.208 \pm 0.015) \times 10^5$	$(2.75 \pm 0.06) \times 10^4$
Run 1	$(11.259 \pm 0.015) \times 10^5$	$(14.05 \pm 0.26) \times 10^3$
Run 2	$(36.072 \pm 0.026) \times 10^5$	$(7.99 \pm 0.10) \times 10^4$

that the trigger efficiency for this hadronic channel can be determined on $B^+ \to J/\psi K^+$ decays with the same TISTOS method that is used to determine the signal efficiency. The mass distribution of $B^0 \to K^+\pi^-$ candidates is shown in Fig. 9 together with the result of the fit, for the different data-taking years. The yields of the two normalization channels are reported in Table II, for the different data-taking years and for the two data-taking periods combined.

B. Fragmentation fractions

The fragmentation fractions, denoted as f_u, f_d, f_s, and f_{baryon}, are the probabilities for a b quark to hadronize into a B^+, B^0, B_s^0 meson or a b baryon, respectively. These fractions include contributions from intermediate states decaying to the aforementioned hadrons via the strong or electromagnetic interactions. The ratio of fragmentation fractions, f_s/f_d, used in this analysis has been measured at LHCb using several B decay modes: semileptonic $B \to D\mu X$ decays at 7 TeV [83] and at 13 TeV [84]; hadronic $B \to Dh$ decays, where $h = \pi$, K, at 7, 8 and 13 TeV [85,86]; $B \to J/\psi X$ decays at 7, 8 and 13 TeV [82]. These measurements have been combined in Ref. [61]. The value of f_s/f_d is found to be dependent on B transverse momentum and pp collision centre-of-mass energy, while it is found not to be dependent on pseudorapidity. Here only the integrated values at different energies are used since the average p_{T} of b-hadrons in this analysis is found to be compatible with those used in the determination of the fragmentation fractions [61].

Since the reported values at 7, 8 and 13 TeV are strongly correlated because their uncertainties are dominated by external measurements, only the 13 TeV value of

$$f_s/f_d(13\ \text{TeV}) = 0.254 \pm 0.008$$

is used, while the Run 1 value is normalized with respect to f_s/f_d (13 TeV) by the ratio [61]

$$\frac{f_s/f_d(13\ \text{TeV})}{f_s/f_d(\text{Run 1})} = 1.064 \pm 0.007.$$

TABLE III. Efficiencies of reconstruction within the LHCb detector and selection for the signal and normalization channels, averaged for the two running periods. The uncertainties include the statistical uncertainty from the simulated samples and the uncertainty of the tracking efficiency corrections.

	$\varepsilon_{\text{RecSel}}$	
	Run 1	Run 2
$B_s^0 \to \mu^+\mu^-$	0.0602 ± 0.0003	0.0640 ± 0.0004
$B^0 \to \mu^+\mu^-$	0.0594 ± 0.0003	0.0635 ± 0.0004
$B_s^0 \to \mu^+\mu^-\gamma$	0.0508 ± 0.0003	0.0546 ± 0.0004
$B^0 \to K^+\pi^-$	0.0462 ± 0.0007	0.0500 ± 0.0006
$B^+ \to J/\psi K^+$	0.0290 ± 0.0003	0.0305 ± 0.0003

Following the approach adopted in Ref. [61], isospin symmetry is assumed to hold in b quark hadronization at the LHC such that $f_u = f_d$, and hence the same f_s/f_d values are used in relation to the $B^+ \to J/\psi K^+$ normalization channel.

C. Efficiencies

The efficiencies to detect the signal and normalization channels can be factorized as

$$\varepsilon_{\text{norm(sig)}} = \varepsilon_{\text{RecSel}} \cdot \varepsilon_{\text{PID}} \cdot \varepsilon_{\text{Trig}}(\cdot\varepsilon_{\text{BDT}}) \tag{10}$$

into reconstruction within the LHCb detector and selection (RecSel), PID, trigger (Trig) efficiencies and exclusion of the first BDT region (BDT > 0.25) on signal candidates. These are evaluated separately on top of each preceding stage.

The acceptance, reconstruction and selection efficiencies are evaluated using simulation with corrections applied to improve the agreement with data. The efficiency to detect and reconstruct tracks is evaluated on a sample of $J/\psi \to \mu^+\mu^-$ decays in data [87], using a tag-and-probe method. These samples are used to determine efficiency correction factors as a function of the particle kinematics, which are convolved with the simulated samples to calculate the total efficiency correction. The corrections are at the level of 1% for all channels and data-taking years. When considering the ratio of signal and normalization channels in the normalization formulae [Eq. (9)], uncertainties on these corrections are treated as 100% correlated. The total efficiencies for the reconstruction within the LHCb detector and selection are listed for the relevant channels in Table III, where the efficiency with which the muon system detects muons is included. Correction factors for the imperfect modeling of the muon system efficiency simulation are estimated using a sample of $B^+ \to J/\psi K^+$ decays in data, which are selected without particle identification criteria [63,88]. These corrections are applied to the signal, $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ channels, modifying the efficiencies by 1%–3%.

TABLE IV. Particle identification efficiencies for the signal and normalization channels, averaged for the two running periods, where the first uncertainty is statistical and the second systematic. A data-simulation correction of the muon-system identification is included for channels with muons. For the $B^+ \to J/\psi K^+$ channel only the data-simulation correction part of the muon identification is reported, as no multivariate PID requirement is applied to this channel.

	$\varepsilon_{\rm PID}$	
	Run 1	Run 2
$B_s^0 \to \mu^+\mu^-$	$0.8580 \pm 0.0006 \pm 0.0053$	$0.8822 \pm 0.0003 \pm 0.0039$
$B^0 \to \mu^+\mu^-$	$0.8518 \pm 0.0007 \pm 0.0063$	$0.8759 \pm 0.0004 \pm 0.0046$
$B_s^0 \to \mu^+\mu^-\gamma$	$0.8487 \pm 0.0006 \pm 0.0088$	$0.8785 \pm 0.0003 \pm 0.0064$
$B^0 \to K^+\pi^-$	$0.4741 \pm 0.0049 \pm 0.0010$	$0.5004 \pm 0.0027 \pm 0.0012$
$B^+ \to J/\psi K^+$	1.0096 ± 0.0005	1.00260 ± 0.00018

The efficiency of the PID requirements described in Sec. IV is measured using high-purity control samples of each particle species obtained from data [63,88]. These control samples are obtained by means of kinematic requirements only, with muons obtained from $J/\psi \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decays, pions and kaons from $D^0 \to K^-\pi^+$ decays selected via $D^{*+} \to D^0\pi^+$, and protons from $\Lambda \to p\pi^-$ and $\Lambda_c^+ \to pK^-\pi^+$ decays. The muon PID efficiencies are evaluated as a function of the muon momentum and pseudorapidity, as well as the track multiplicity of the event using a dedicated procedure [63]. The resulting efficiency maps are then applied to simulated samples to determine the integrated efficiency for a specific channel. The efficiency measurements for the different hadronic species are described in Sec. VII A. The results for the signal and normalization channels are shown in Table IV and include for the channels with muons the data-simulation correction of the muon system identification. For $B^+ \to J/\psi K^+$ candidates, only the correction to the muon system identification efficiency is computed, as no further requirements on the multivariate PID classifier are applied when selecting these decays. The systematic uncertainties arise from modeling the dependencies of the PID efficiency maps.

The trigger efficiencies are determined from data with the TISTOS method [62]. Trigger information is associated to the reconstructed candidates during the offline processing. The event of a reconstructed signal candidate can be classified into three categories: events triggered on signal (TOS), triggered on part of the underlying event that is independent of the tracks forming the signal candidate (TIS), or triggered on both elements of the signal candidate and the underlying event.

The trigger efficiency can be estimated by exploiting the overlap between the TIS and TOS categories (TIS&TOS) and assuming signal decays uncorrelated with the rest of the event. The trigger efficiency, $\varepsilon_{\rm trig}$, of a given decay channel, with respect to a total of $N_{\rm Tot}$ events, can be computed as

$$\varepsilon_{\rm trig} = \frac{N_{\rm trig}}{N_{\rm Tot}} = \frac{N_{\rm trig}}{N_{\rm TIS}} \cdot \varepsilon_{\rm TIS} = \frac{N_{\rm trig}}{N_{\rm TIS}} \frac{N_{\rm TIS\&TOS}}{N_{\rm TOS}}, \quad (11)$$

where N_X is the number of background-subtracted candidates triggered within the category X and the efficiency $\varepsilon_{\rm TIS} = N_{\rm TIS\&TOS}/N_{\rm TOS}$ is estimated under the already mentioned independence assumption, which is verified in Ref. [62].

The trigger efficiency estimation is done in two steps: the hardware and first software-level trigger efficiencies are estimated from data as described in the following; the second software-level efficiency, being aligned to the offline selection, is estimated from simulation and included in the full trigger efficiency presented in this paragraph. The trigger efficiencies for signal and normalization channels with muons are calibrated using the $B^+ \to J/\psi K^+$ channel. In order to reduce residual kinematic correlations between the decay in question and the rest of the event, the calibration is performed in intervals of kinematic quantities. Yields of $B^+ \to J/\psi K^+$ decays for each trigger category and different kinematic ranges are obtained by performing a mass fit as described in Sec. VI A. Efficiency tables are obtained as a function of the maximum $p_{\rm T}$ of the two muons and of the product of the $p_{\rm T}$ of the two muons, as these are the variables used in the muon hardware trigger. These efficiency distributions are then convolved with the simulated samples of the relevant channels. The trigger efficiencies for the $B^0 \to K^+\pi^-$ channel are also obtained in data by measuring the TIS trigger efficiency in Eq. (11) through the more abundant $B^+ \to J/\psi K^+$ channel, since the TIS efficiencies do not depend on the control channel used to evaluate it.

The trigger efficiencies for the signal and normalization channels in each data-taking period are presented in Table V. The systematic uncertainty on the trigger efficiency is comprised of a number of sources. A systematic effect is associated with the choice of the mass model used for the $B^+ \to J/\psi K^+$ channel. This is estimated by fitting the $B^+ \to J/\psi K^+$ data using a double Crystal Ball function

TABLE V. Trigger efficiencies for the signal and normalization channels, averaged for the two data taking periods. The first uncertainty is statistical and the second systematic.

	$\varepsilon_{\text{Trig}}$	
	Run 1	Run 2
$B_s^0 \to \mu^+\mu^-$	$0.9579 \pm 0.0033 \pm 0.0164$	$0.9712 \pm 0.0014 \pm 0.0093$
$B^0 \to \mu^+\mu^-$	$0.9570 \pm 0.0032 \pm 0.0176$	$0.9708 \pm 0.0014 \pm 0.0097$
$B_s^0 \to \mu^+\mu^-\gamma$	$0.9538 \pm 0.0032 \pm 0.0195$	$0.9694 \pm 0.0013 \pm 0.0111$
$B^0 \to K^+\pi^-$	$0.0433 \pm 0.0002 \pm 0.0016$	$0.0727 \pm 0.0002 \pm 0.0020$
$B^+ \to J/\psi K^+$	$0.8810 \pm 0.0040 \pm 0.0080$	$0.9033 \pm 0.0016 \pm 0.0089$

TABLE VI. Efficiency on the signal channels of excluding the BDT region BDT < 0.25, averaged for the two data taking periods. The uncertainties combine statistical and systematic contributions.

	ε_{BDT}	
	Run 1	Run 2
$B_s^0 \to \mu^+\mu^-$	0.723 ± 0.006	0.7071 ± 0.0026
$B^0 \to \mu^+\mu^-$	0.720 ± 0.006	0.7036 ± 0.0027
$B_s^0 \to \mu^+\mu^-\gamma$	0.656 ± 0.007	0.6531 ± 0.0035

to model the signal and taking the difference with the default fit as systematic uncertainty. A second systematic uncertainty stems from the precision of the TISTOS method and is obtained by comparing the efficiency determined by applying the TISTOS method to simulation. A third source of systematic effect is due to the difference in the phase space between the $B^+ \to J/\psi K^+$ decay and the decay channel for which the trigger efficiency is evaluated. The corresponding systematic uncertainty is estimated by comparing the results of the method applied to simulated events using the $B^+ \to J/\psi K^+$ channel or the considered channel. The last source of systematic effect arises from the choice of the kinematic ranges for which the efficiencies are evaluated, and its uncertainty is determined from the change of the efficiency when these ranges are varied. The resulting shifts in trigger efficiency from each source are added in quadrature, and assigned as the total systematic uncertainty in Table V.

The efficiencies of the exclusion of the first BDT region on the signal decays are evaluated using the calibrated BDT response described in Sec. V B and are listed in Table VI, combining statistical and systematic uncertainties.

D. Single-event sensitivities

The single-event sensitivities [defined in Eq. (9)] for the three signal channels in Run 1, Run 2 and the full data sample, are reported in Table VII. Single-event sensitivities

TABLE VII. Single-event sensitivities, $\alpha(B^+)$, $\alpha(B^0)$ and $\alpha(\text{Comb})$ for the three signal channels obtained for BDT > 0.25 with the two normalization channels, $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$, and combined, for Run 1, Run 2 and the full dataset. The first uncertainty is statistical and the second systematic. The expected yields assuming SM branching fractions, N_{exp}, are also reported. For $B_s^0 \to \mu^+\mu^-\gamma$, an approximate branching fraction of 1×10^{-10} for $m_{\mu\mu} > 4.9$ GeV/c^2 has been assumed.

	$B^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-\gamma$
		Run 1	
$\alpha(B^+)$	$(3.96 \pm 0.13 \pm 0.09) \times 10^{-11}$	$(1.57 \pm 0.07 \pm 0.03) \times 10^{-10}$	$(2.11 \pm 0.10 \pm 0.05) \times 10^{-10}$
$\alpha(B^0)$	$(3.79 \pm 0.14 \pm 0.16) \times 10^{-11}$	$(1.50 \pm 0.07 \pm 0.06) \times 10^{-10}$	$(2.01 \pm 0.10 \pm 0.09) \times 10^{-10}$
$\alpha(\text{Comb})$	$(3.93 \pm 0.10 \pm 0.08) \times 10^{-11}$	$(1.56 \pm 0.06 \pm 0.03) \times 10^{-10}$	$(2.09 \pm 0.09 \pm 0.05) \times 10^{-10}$
N_{exp}	$2.62 \pm 0.14 \pm 0.05$	$23.5 \pm 1.3 \pm 0.4$	~ 0.5
		Run 2	
$\alpha(B^+)$	$(1.214 \pm 0.037 \pm 0.018) \times 10^{-11}$	$(4.54 \pm 0.20 \pm 0.07) \times 10^{-11}$	$(5.86 \pm 0.26 \pm 0.10) \times 10^{-11}$
$\alpha(B^0)$	$(1.176 \pm 0.035 \pm 0.037) \times 10^{-11}$	$(4.40 \pm 0.19 \pm 0.14) \times 10^{-11}$	$(5.67 \pm 0.25 \pm 0.18) \times 10^{-11}$
$\alpha(\text{Comb})$	$(1.204 \pm 0.023 \pm 0.014) \times 10^{-11}$	$(4.50 \pm 0.16 \pm 0.05) \times 10^{-11}$	$(5.81 \pm 0.21 \pm 0.08) \times 10^{-11}$
N_{exp}	$8.55 \pm 0.45 \pm 0.10$	$81.3 \pm 4.3 \pm 0.9$	~ 1.7
		All	
$\alpha(B^+)$	$(9.27 \pm 0.28 \pm 0.12) \times 10^{-12}$	$(3.53 \pm 0.15 \pm 0.04) \times 10^{-11}$	$(4.61 \pm 0.20 \pm 0.07) \times 10^{-11}$
$\alpha(B^0)$	$(8.95 \pm 0.26 \pm 0.23) \times 10^{-12}$	$(3.41 \pm 0.15 \pm 0.09) \times 10^{-11}$	$(4.45 \pm 0.19 \pm 0.12) \times 10^{-11}$
$\alpha(\text{Comb})$	$(9.20 \pm 0.14 \pm 0.09) \times 10^{-12}$	$(3.51 \pm 0.12 \pm 0.03) \times 10^{-11}$	$(4.57 \pm 0.16 \pm 0.05) \times 10^{-11}$
N_{exp}	$11.20 \pm 0.57 \pm 0.11$	$104.4 \pm 5.4 \pm 1.0$	~ 2.2

for BDT > 0.25 are obtained using the two normalization channels $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ separately, which are combined into a normalization using a weighted average, taking into account the relevant correlations. In the same table, the expected number of signal candidates for BDT > 0.25 is reported, assuming the SM branching fraction.

In order to cross-check the normalization of the signal channels, the ratio of the efficiency-corrected yields of the two normalization channels

$$\frac{\mathcal{B}(B^0 \to K^+\pi^-)}{\mathcal{B}(B^+ \to J/\psi K^+)} = \frac{N_{B^0\to K^+\pi^-}}{N_{B^+\to J/\psi K^+}} \frac{\varepsilon_{B^+\to J/\psi K^+}}{\varepsilon_{B^0\to K^+\pi^-}} \frac{f_u}{f_d}$$

is measured using data, where $f_u = f_d$ is assumed. This ratio is found to be $0.340 \pm 0.016(\text{stat})$ and $0.336 \pm 0.012(\text{stat})$ in Run 1 and Run 2, respectively, in agreement with the ratio of the world averages of these branching fractions, 0.326 ± 0.012 [51].

To cross-check the ratio of the B_s^0 and B^+ fragmentation fractions and its stability over the data taking, the ratio of $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ efficiency-corrected yields,

$$\mathcal{R} = \frac{N_{B_s^0\to J/\psi\phi}}{N_{B^+\to J/\psi K^+}} \frac{\varepsilon_{B^+\to J/\psi K^+}}{\varepsilon_{B_s^0\to J/\psi\phi}} = \frac{f_s}{f_u} \frac{\mathcal{B}(B_s^0 \to J/\psi\phi)}{\mathcal{B}(B^+ \to J/\psi K^+)}, \quad (12)$$

is also measured, following a similar approach to Ref. [82]. The ratios are found to be similar to those measured in Ref. [82], although the two methods explore different kinematic regions. A dependence on the center-of-mass energy is seen and found to be consistent with Ref. [82] and the combined analysis of Ref. [61], justifying the use of different f_s/f_d values for the Run 1 and Run 2 data samples.

VII. BACKGROUND

Three classes of background events are considered in the analysis: combinatorial background; $B^0_{(s)} \to h^+h'^-$ decays; semileptonic b-hadron decays. The combinatorial background, mainly composed of real muons originating from two different B decays, is modeled using an exponential function with the slope left free to float in the mass fit, as described in Sec. VIII. The other background sources are included as separate components in the fit, with mass shapes evaluated on simulated events and with yields that are Gaussian-constrained to their estimated values, as explained in the following sections.

A. Hadron misidentification rates

To estimate the yield of physical background where one or two final-state particles are misidentified as a muon, it is crucial to perform unbiased measurements of the probability for protons, pions and kaons to pass the muon identification requirements. These measurements are carried out as a function of the track momentum and transverse momentum, using the data control samples listed in Sec. VI C. The dedicated procedure from Ref. [63] is used for protons, while a different method is developed to determine the pion and kaon misidentification rates, using $D^0 \to K^-\pi^+$ from $D^{*+} \to D^0\pi^+$ decays. For these particles, especially at low momenta, a sizeable contribution to the misidentification rate originates from hadrons decaying to muons. When the hadron decays in flight, the momentum resolution of the reconstructed track degrades by an amount that depends on the distance the hadron has travelled before decaying and on the fraction of energy inherited by the daughter muon. As a consequence, the mass distribution of the D^0 candidates broadens significantly and the efficiency to select $D^0 \to K^-\pi^+$ decays in the D^0 mass selection window, $1825 \le m_{K\pi} \le 1910\ \text{MeV}/c^2$, decreases. If this effect is not taken into account, a significant underestimation of the misidentified hadron yield would occur. The misidentification efficiencies for pions and kaons are determined by measuring the D^0 yield from a two-dimensional fit to the $m_{K\pi\pi} - m_{K\pi}$, $m_{K\pi}$ distribution with or without the muon requirement applied to the particle in question. The shape of the signal D^0 mass distribution when the PID selection is applied includes the tail arising from hadron decays-in-flight, estimated from simulated events. The resulting misidentification probability is then corrected for the fraction of $D^0 \to K^-\pi^+$ decays with the $K^-\pi^+$ mass falling outside the D^0 selection window, estimated from simulated events.

B. $B^0_{(s)} \to h^+h'^-$ decays

The $B^0_{(s)} \to h^+h'^-$ decays can appear as background when both final state hadrons are misidentified as muons. These candidates have a broad mass distribution centered close to the B^0 mass, as determined from simulation where each of the four modes is weighted according to its expected yield.

The expected yield of doubly misidentified $B^0_{(s)} \to h^+h'^-$ events is estimated by normalizing to the $B^0 \to K^+\pi^-$ channel as

$$N_{B^0_{(s)}\to h^+h'^-\to\mu^+\mu^-} = \varepsilon^{\text{trig}}_{B_s^0\to\mu^+\mu^-} \cdot \frac{N^{\text{TIS}}_{hh}}{\varepsilon^{\text{TIS}}} \cdot \varepsilon_{hh\to\mu\mu}, \quad (13)$$

where $\varepsilon^{\text{trig}}_{B_s^0\to\mu^+\mu^-}$ is the signal trigger efficiency, N^{TIS}_{hh} is the number of $B^0_{(s)} \to h^+h'^-$ TIS events evaluated by correcting the $B^0 \to K^+\pi^-$ TIS yield by the expected fraction of this mode, ε^{TIS} is the TIS efficiency (Sec. V), and $\varepsilon_{hh\to\mu\mu}$ represents the double misidentification rate, which is estimated using data control samples (Sec. VII A) and found to be in the range $10^{-6} - 10^{-5}$, depending on the dataset and BDT region. An independent estimate of the $B^0_{(s)} \to h^+h'^-$ background is performed on $\pi\mu$ and $K\mu$ combinations, selected from data samples of B candidates

TABLE VIII. Expected background yields per BDT region and for BDT > 0.25 with their total estimated uncertainties for Run 1 data.

BDT region	$B^0_{(s)} \to h^+h'^-$	$B^0 \to \pi^-\mu^+\nu_\mu$	$B^0_s \to K^-\mu^+\nu_\mu$	$B^+ \to \pi^+\mu^+\mu^-$	$B^0 \to \pi^0\mu^+\mu^-$	$\Lambda^0_b \to p\mu^-\bar{\nu}_\mu$	$B^+_c \to J/\psi\mu^+\nu_\mu$
[0.25–0.4]	3.8 ± 0.9	8.8 ± 0.7	0.94 ± 0.20	2.72 ± 0.35	0.9 ± 0.4	0.29 ± 0.32	2.84 ± 0.21
[0.4–0.5]	1.57 ± 0.20	6.0 ± 0.4	0.76 ± 0.16	1.47 ± 0.19	0.65 ± 0.29	0.18 ± 0.20	1.38 ± 0.13
[0.5–0.6]	1.61 ± 0.21	5.8 ± 0.4	0.90 ± 0.19	1.22 ± 0.16	0.65 ± 0.28	0.20 ± 0.22	0.80 ± 0.09
[0.6–0.7]	1.65 ± 0.21	5.8 ± 0.4	1.03 ± 0.22	0.99 ± 0.13	0.65 ± 0.29	0.19 ± 0.21	0.58 ± 0.08
[0.7–1.0]	5.3 ± 0.7	11.6 ± 0.8	2.5 ± 0.5	1.46 ± 0.19	1.4 ± 0.6	0.32 ± 0.35	0.41 ± 0.06
[0.25-1.0]	13.9 ± 1.2	30.8 ± 1.8	9.0 ± 1.1	7.6 ± 0.5	4.2 ± 0.9	1.2 ± 0.6	6.01 ± 0.28

TABLE IX. Expected background yields per BDT region and for BDT > 0.25 with their total estimated uncertainties for Run 2 data.

BDT region	$B^0_{(s)} \to h^+h'^-$	$B^0 \to \pi^-\mu^+\nu_\mu$	$B^0_s \to K^-\mu^+\nu_\mu$	$B^+ \to \pi^+\mu^+\mu^-$	$B^0 \to \pi^0\mu^+\mu^-$	$\Lambda^0_b \to p\mu^-\bar{\nu}_\mu$	$B^+_c \to J/\psi\mu^+\nu_\mu$
[0.25–0.4]	6.2 ± 1.0	40.1 ± 2.6	3.3 ± 0.7	9.9 ± 1.3	3.5 ± 1.5	1.1 ± 1.2	11.7 ± 0.6
[0.4–0.5]	2.90 ± 0.26	24.6 ± 1.6	2.5 ± 0.5	5.2 ± 0.7	2.3 ± 1.0	0.8 ± 0.8	4.59 ± 0.27
[0.5–0.6]	2.82 ± 0.26	24.4 ± 1.6	2.9 ± 0.6	4.2 ± 0.5	2.3 ± 1.0	0.9 ± 0.9	2.85 ± 0.19
[0.6–0.7]	2.68 ± 0.24	23.5 ± 1.5	3.3 ± 0.7	3.3 ± 0.4	2.3 ± 1.0	0.9 ± 1.0	1.56 ± 0.12
[0.7–1.0]	8.1 ± 0.7	44.6 ± 2.9	7.7 ± 1.6	4.8 ± 0.6	4.7 ± 2.1	1.6 ± 1.8	0.85 ± 0.08
[0.25–1.0]	22.7 ± 1.3	130 ± 5	21.5 ± 2.5	26.5 ± 1.7	14.7 ± 3.0	5.3 ± 2.7	21.6 ± 0.7

with two tracks in the final states applying strong muon and hadron identification requirements on the tracks. Their mass spectra are fitted and the resulting yields are scaled by the $\pi \to \mu$ and $K \to \mu$ misidentification rates. The ratio between this result and the default estimate is assigned as a correction factor to the misidentification efficiency. The correction factor is consistent with 1 with an uncertainty of 0.03 for BDT > 0.4. The estimated $B^0_{(s)} \to h^+h'^-$ background yields in each BDT region with BDT > 0.25 are summarized in Table VIII for Run 1 and Table IX for Run 2 data. The background yield per fb^{-1} is smaller in Run 2 compared to Run 1 due to the improvement of the muon identification algorithm, which has a better charged-kaon rejection power with a similar efficiency to select muons.

C. Semileptonic decays

Several semileptonic b-hadron decays, with branching fractions ranging from 10^{-8} to 10^{-4}, are considered in the fit: $B^+_c \to J/\psi\mu^+\nu_\mu$, with $J/\psi \to \mu^+\mu^-$, and $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ decays have two real muons in the final state, while $B^0 \to \pi^-\mu^+\nu_\mu$, $B^0_s \to K^-\mu^+\nu_\mu$ and $\Lambda^0_b \to p\mu^-\bar{\nu}_\mu$ decays represent non-negligible background when the final-state hadron is misidentified as a muon. When reconstructed as dimuon candidates, these decays are partially reconstructed and therefore populate the lower B^0/B^0_s sideband, but can have tails reaching into the signal region.

For each of the above channels, the number of expected candidates is estimated by normalizing to the yield of $B^+ \to J/\psi K^+$ decays, according to

$$N_x = \frac{f_x}{f_u} \cdot \frac{N_{B^+\to J/\psi K^+}}{\mathcal{B}_{B^+\to J/\psi K^+} \cdot \varepsilon^{\rm tot}_{B^+\to J/\psi K^+}} \cdot \mathcal{B}_x \cdot \varepsilon^{\rm tot}_x, \quad (14)$$

where f_x is the hadronization fraction of the initial-state hadron for the decay mode x, $\mathcal{B}_x$ is the branching fraction and $\varepsilon^{\rm tot}_x$ its total selection and trigger efficiency. The efficiencies are estimated from simulation, except for the PID, which is estimated from data control samples as described in Secs. VI C and VII A. The branching fraction, including the hadronization fraction of the $B^+_c \to J/\psi\mu^+\nu_\mu$ channel, is taken from Ref. [89], while those of $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ and $B^0 \to \pi^-\mu^+\nu_\mu$ channels are obtained from Refs. [51,90], assuming $f_u = f_d$. LHCb measurements for the $\Lambda^0_b \to p\mu^-\bar{\nu}_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ branching fractions [91,92] and hadronization fractions [61,84] are used. The estimated yields in each BDT region with BDT > 0.25 are shown in Table VIII for Run 1 and Table IX for Run 2 data.

VIII. MEASUREMENT OF SIGNAL BRANCHING FRACTIONS

The data sample containing $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates is divided into the two data-taking periods, Run 1 and Run 2, which are further divided into six subsets based on the BDT response, using the intervals defined in Sec. IV. The branching fractions of the signal decays are determined using an unbinned extended maximum-likelihood fit to the dimuon mass distributions, performed simultaneously on the subsets with BDT > 0.25. Due to substantial contamination from combinatorial background, the lowest BDT region, $0 \leq$ BDT < 0.25, is excluded from the dataset but

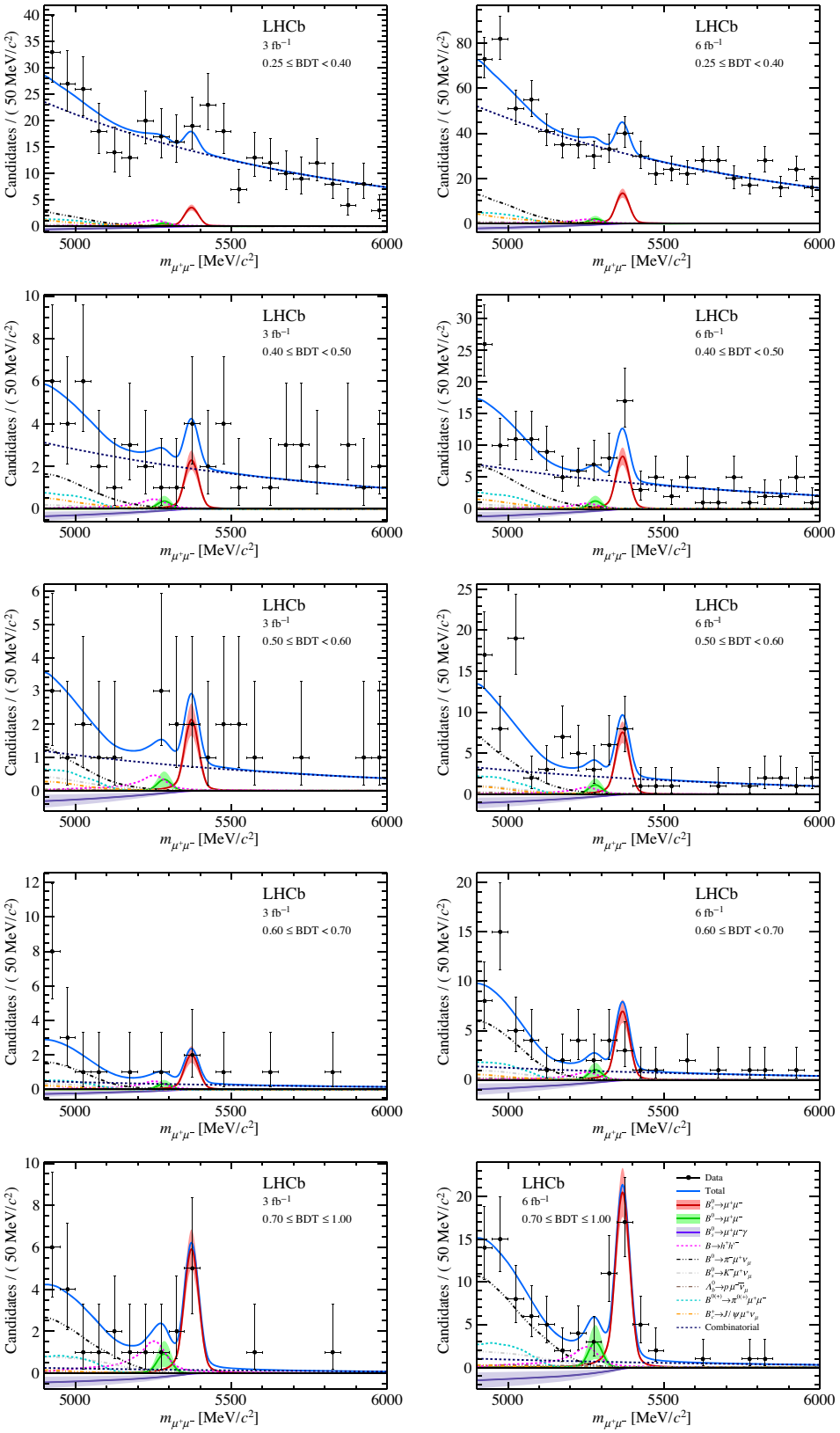

FIG. 10. Mass distribution of signal candidates (black dots) for (left) Run 1 and (right) Run 2 samples in regions of BDT. The result of the fit is overlaid (blue line) and the different components detailed in the legend. The solid bands represent the variation of the signal branching fractions within their total uncertainty.

its fraction is taken into account in the total normalization of the BDT shape. The fit is performed in a mass window of $4900 \leq m_{\mu\mu} \leq 6000$ MeV/c^2. The dimuon mass distribution is shown in Fig. 10 for the Run 1 and Run 2 samples in all BDT intervals. The low mass region is populated by the partially reconstructed background and $B_s^0 \to \mu^+\mu^-\gamma$ decays, while the higher mass region is dominated by combinatorial background.

The probability density functions (PDFs) of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays are described by DSCB functions, defined in Eq. (4), with their parameters Gaussian-constrained to the values measured in Sec. V. The mass distribution of the $B_s^0 \to \mu^+\mu^-\gamma$ decay is described using an empirical threshold function

$$f(m_{\mu\mu}) \propto \left(1 - \frac{m_{\mu\mu}}{M_{B_s}}\right)^b - a\sqrt{1 - e^{\frac{m_{\mu\mu} - M_{B_s}}{s}}}, \qquad (15)$$

where the parameters a, b and s, are determined from simulation, which is based on the theoretical predictions and form factors of Ref. [18]. The parameter b is found to be close to 0.5, while the other parameters vary across the BDT regions. This threshold function is convolved with a Gaussian resolution function which models the effect of the detector resolution. The parameter values are estimated from kinematically weighted simulated events, and fixed in the fit. While the branching fraction prediction for the $B_s^0 \to \mu^+\mu^-\gamma$ decay is dependent on the exact form-factor parametrization used [19], the distribution of the dimuon mass at high q^2 is found to not depend significantly on the choice of the form-factor, and so the same threshold function can be used for a range of scenarios. Moreover, varying the detector resolution parameter within the known uncertainties has a negligible effect on the yield of $B_s^0 \to \mu^+\mu^-\gamma$ decays. The combinatorial background is modeled with a single exponential function with an independent yield in each BDT region but with common slope parameters for each data-taking period. Both the yields and the parameters are free to float in the fit. The common slope assumption across the BDT intervals is introduced to stabilize the fit and it is estimated to be correct at the level of 20%. The systematic uncertainty on the signal branching fractions associated to this assumption is negligible. The mass distribution is found to be well described by an exponential function in all BDT regions both in data with $m_{\mu\mu} > 5450$ MeV/c^2, where the combinatorial background dominates, and in simulated $b\bar{b} \to \mu^+\mu^- X$ events. The $B_{(s)}^0 \to h^+h'^-$ and semileptonic b-hadron contributions are described using kernel estimation techniques [93] applied to simulated events in each BDT region. Their expected yields in each BDT region are Gaussian-constrained according to the values reported in Sec. VII. Moreover, common parameters, such as the yields of the normalization channels, efficiencies and branching fractions are shared across all BDT regions and their values are Gaussian-constrained to their estimated values and uncertainties.

The result of the fit in each subset is shown in Fig. 10. The resulting branching fractions of the $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays are

$$\begin{aligned}
\mathcal{B}(B_s^0 \to \mu^+\mu^-) &= (3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9}, \\
\mathcal{B}(B^0 \to \mu^+\mu^-) &= (1.20^{+0.83}_{-0.74} \pm 0.14) \times 10^{-10}, \\
\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) &= (-2.5 \pm 1.4 \pm 0.8) \times 10^{-9} \\
&\text{with} \quad m_{\mu\mu} > 4.9 \text{ GeV}/c^2.
\end{aligned}$$

The statistical uncertainties are evaluated by repeating the fit with all nuisance parameters fixed to the value obtained in the standard fit, where all nuisance parameters are free to float within their constraints. The systematic uncertainties are then computed by subtracting in quadrature the statistical uncertainties from the total ones. The main contribution to the systematic uncertainty of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ originates from the knowledge of f_s/f_d (3%), while that of $\mathcal{B}(B^0 \to \mu^+\mu^-)$ is dominated by the knowledge of the $B_{(s)}^0 \to h^+h'^-$ and semileptonic b-hadron background (9%). The correlation between the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ branching fractions is found to be -11%. The correlation between the $B_s^0 \to \mu^+\mu^-\gamma$ and $B^0 \to \mu^+\mu^-$ branching fractions is -25%, while that with $B_s^0 \to \mu^+\mu^-$ is below 10%. Using the α factors in Table VII, the measured branching fractions translate approximately to 88 ± 13 $B_s^0 \to \mu^+\mu^-$, 13 ± 8 $B^0 \to \mu^+\mu^-$ and -55 ± 35 $B_s^0 \to \mu^+\mu^-\gamma$ signal events.

Two-dimensional profile likelihoods are evaluated in the plane of the possible combinations of two branching fractions. These are obtained 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. The results are shown in Fig. 11.

The difference between the logarithm of the likelihood values under the presence or the absence of a specific signal component is used to evaluate the significance with Wilks' theorem [94]. The $B_s^0 \to \mu^+\mu^-$ signal exceeds the background-only hypothesis more than 10 standard deviations (σ), while the statistical significance of the $B^0 \to \mu^+\mu^-$ signal is 1.7σ and the $B_s^0 \to \mu^+\mu^-\gamma$ signal is compatible with the background-only hypothesis within 1.5σ. Upper limits on the branching fractions of $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ are evaluated using the $\mathrm{CL_s}$ method [95] with a one-sided test statistic [96] as implemented in Refs. [97,98]. The one-sided test statistic for a given branching fraction value is defined as twice the negative logarithm of the profile likelihood ratio if it is larger than the measured branching fraction and zero otherwise. Its distribution is determined from pseudoexperiments, where nuisance parameters are set to their best fit values for toy generation

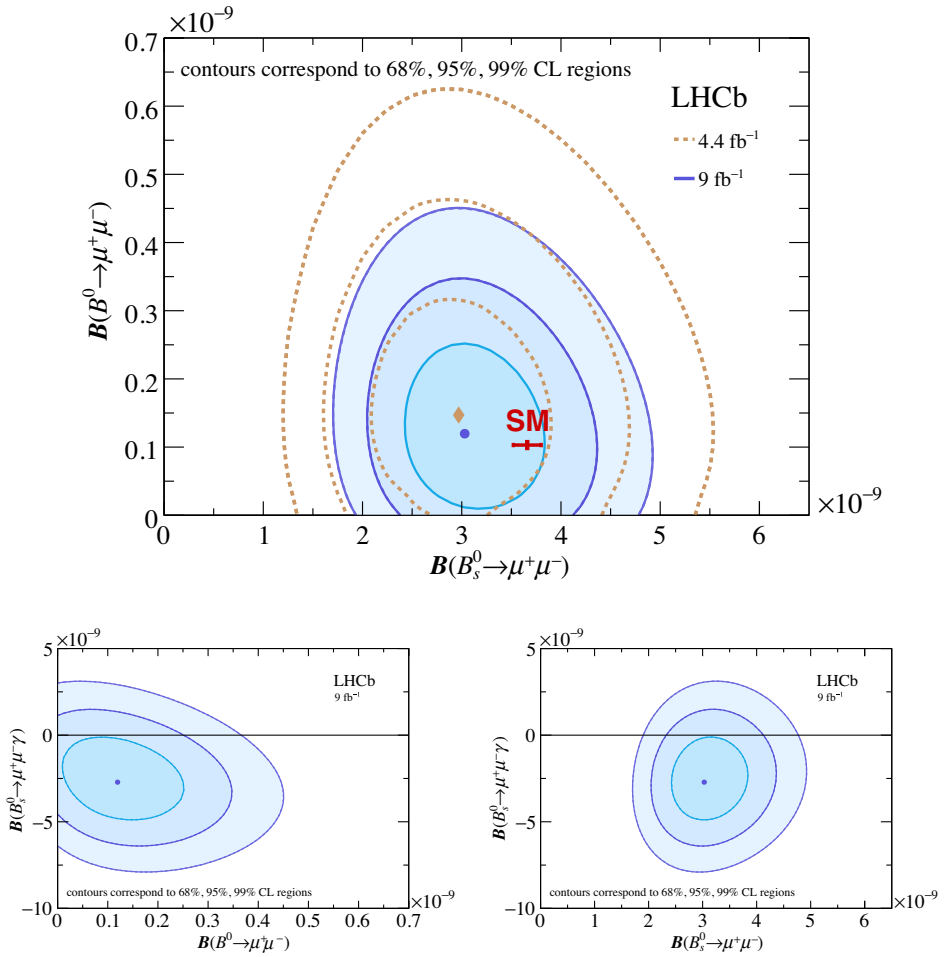

FIG. 11. Two-dimensional representations of the branching fraction measurements for the decays (top) $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, (bottom left) $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ and (bottom right) $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$. The $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction is limited to the range $m_{\mu\mu} > 4.9$ GeV$/c^2$. The measured central values of the branching fractions are indicated with a blue dot. The profile likelihood contours for 68%, 95% and 99% CL regions of the result presented in this paper are shown as blue contours, while in the top plot the brown contours indicate the previous measurement [32] and the red cross shows the SM prediction. Figure reproduced from Ref. [37].

while central values of the Gaussian-constraints are independently fluctuated within their uncertainty for each pseudoexperiment as described in Ref. [99]. The CL_s curves are shown in Fig. 12 from which the limit on the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ branching fractions are found to be

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.3(2.6) \times 10^{-10},$$
$$\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) < 1.5(2.0) \times 10^{-9}$$
$$\text{with} \quad m_{\mu\mu} > 4.9 \text{ GeV}/c^2,$$

at 90% (95%) CL. The measured upper limits are shown in Fig. 12, together with the expected ones.

To quantify the impact of the $B_s^0 \to \mu^+\mu^-\gamma$ component on the other signal modes, the fit is repeated by fixing its branching fraction to zero. Using this configuration, $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ increases by 2% while the limit on $\mathcal{B}(B^0 \to \mu^+\mu^-)$ decreases by 12%.

As described in Sec. V B, the BDT calibration of $B_s^0 \to \mu^+\mu^-$ decays depends on the effective lifetime which introduces a model dependence in the measured time-integrated branching fraction. In the fit, the SM value of $A_{\Delta\Gamma_s}^{\mu\mu} = 1$ is assumed for $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays. The model dependence is evaluated by repeating the fit under the assumptions $A_{\Delta\Gamma_s}^{\mu\mu} = 0$ and -1, finding an increase of the $B_s^0 \to \mu^+\mu^-$ branching fraction with respect to the SM hypothesis of 4.7% and 10.9%, respectively. On the contrary, the $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction decreases with respect to the $A_{\Delta\Gamma_s}^{\mu\mu} = 1$ hypothesis by 2% and 5% with a negligible impact on its limit. The dependence for $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ is approximately linear in the physically allowed $A_{\Delta\Gamma_s}^{\mu\mu}$ range. To evaluate the ratio of

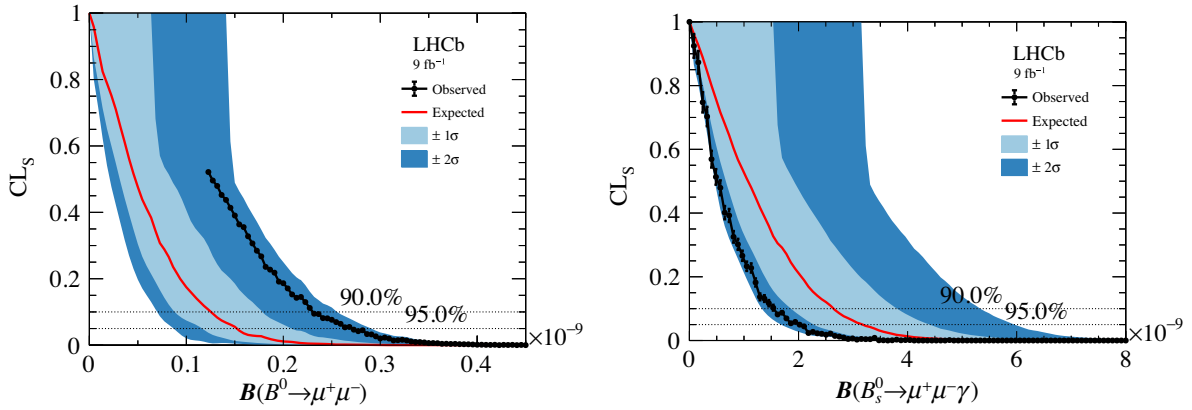

FIG. 12. Results from the CL_s scan used to obtain the limit on (left) $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and (right) $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$. The background-only expectation is shown by the red line and the 1σ and 2σ bands are shown as light blue and blue bands respectively. The observation is shown as the solid black line. The two dashed lines intersecting with the observation indicate the limits at 90% and 95% CL for the upper and lower line, respectively.

the branching fractions $\mathcal{R}_{\mu^+\mu^-}$ defined in Eq. (3), the fit is modified such that $\mathcal{R}_{\mu^+\mu^-}$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ are floating observables, which allows for the cancellation of common uncertainties, while $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$ is kept as a floating observable. The ratio is found to be

$$\mathcal{R}_{\mu^+\mu^-} = 0.039^{+0.030+0.006}_{-0.024-0.004},$$

where the first uncertainty is statistical and the second systematic. Using the CL_s method described above, the upper limit is evaluated to be

$$\mathcal{R}_{\mu^+\mu^-} < 0.081(0.095)$$

at 90%(95%) CL.

IX. MEASUREMENT OF $B_s^0 \to \mu^+\mu^-$ EFFECTIVE LIFETIME

The effective lifetime of the $B_s^0 \to \mu^+\mu^-$ decay is measured using the same data sample as for the branching fraction measurement but with a slightly different selection, described in Sec. IV. The data are divided into two regions of the BDT classifier response and unbinned extended maximum-likelihood fits are performed to the dimuon mass distribution in each region in order to calculate weights using the *sPlot* method [100]. These weights are then used to extract the $B_s^0 \to \mu^+\mu^-$ signal decay-time distributions. Finally, the effective lifetime is measured using an unbinned maximum-likelihood fit to the weighted decay-time distributions, performed simultaneously to both BDT regions.

The fits to the dimuon mass used to extract the weights are performed in the range $5320 \leq m_{\mu\mu} \leq 6000$ MeV$/c^2$. The lower limit of 5320 MeV$/c^2$ removes the low mass region containing most of the physical background, including $B^0 \to \mu^+\mu^-$ and $B^0_{(s)} \to h^+h'^-$ decays, so only $B_s^0 \to \mu^+\mu^-$ decay and combinatorial background components are included in the fit. Residual contamination from physical background in the fit region is low and is treated as a source of systematic uncertainty. The $B_s^0 \to \mu^+\mu^-$ signal is modeled using a DSCB PDF and the background with an exponential PDF. The parameters of the signal PDF are determined using the same method as for the branching fraction (Sec. VIII) and are fixed in the fit, while the signal and background yields and the decay constant of the combinatorial background exponential are allowed to float freely. The distributions of the dimuon mass in the two BDT regions are shown in Fig. 13.

To make an unbiased measurement of the effective lifetime, the decay-time dependence of the combined trigger, reconstruction and selection efficiency must be accurately estimated. This decay-time acceptance is calculated using simulated $B_s^0 \to \mu^+\mu^-$ candidates, which have been weighted in order to improve agreement with data. Since the LHCb PID response is known to be imperfectly modeled in simulation, the efficiencies of the PID requirements are calculated using calibration samples of muons in data and applied as weights to simulated $B_s^0 \to \mu^+\mu^-$ candidates. Weights are also calculated to correct the number of tracks in the pp collision, the kinematics of the B meson and inputs to the BDT classifier using candidates from the $B^0 \to K^+\pi^-$ control channel in data.

The parameters of the functions used to model the decay-time efficiency are extracted using unbinned maximum-likelihood fits to the decay-time distributions of simulated $B_s^0 \to \mu^+\mu^-$ candidates, with the effective lifetime fixed to its true value. Since the decay-time efficiency has a different form in the two BDT regions, two different

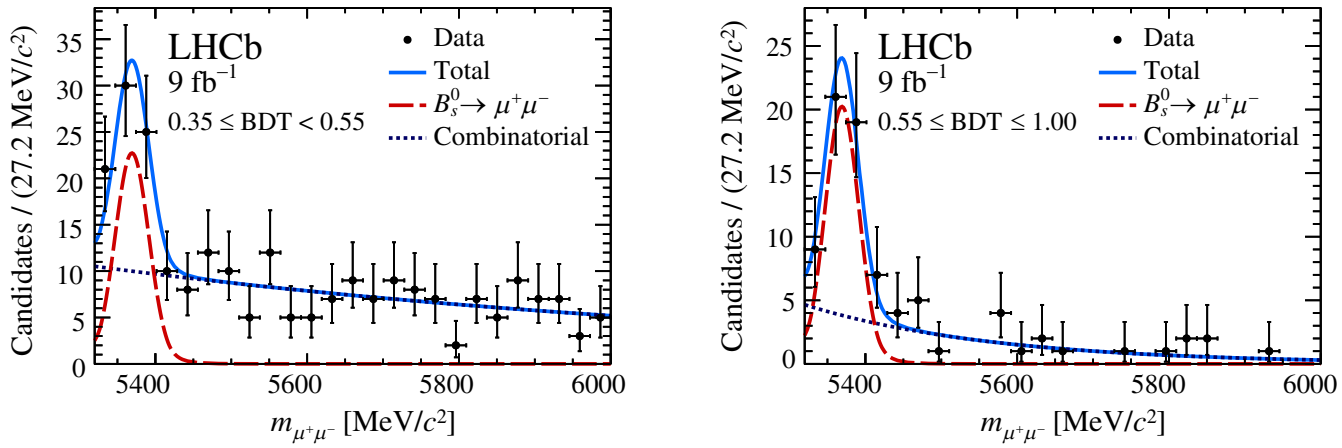

FIG. 13. Dimuon mass distributions of $B_s^0 \to \mu^+\mu^-$ candidates with the fit model used to perform the background subtraction for the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime superimposed in the (left) low and (right) high BDT regions.

empirical functions are used. In the low BDT region the efficiency is modeled as

$$\varepsilon(t) = a \times \mathrm{Erf}\left(t\sqrt{b \times \tanh(ct^3)}\right) + \exp\left(-dt^e\right) - 1, \quad (16)$$

where Erf is the error function, t is the reconstructed decay time, a, b, c, d and e are free parameters and $\varepsilon(t) = 0$ when $t < 0.26$ ps. The acceptance in the high region is modeled using

$$\varepsilon(t) = \exp\left(-\frac{1}{2}\left(\frac{\ln(t - t_0) - f}{g}\right)^2\right), \quad (17)$$

where f, g and t_0 are free parameters and $\varepsilon(t) = 0$ when $t \leq t_0$. The forms of these functions with respect to the B_s^0 meson decay time are shown in Fig. 14. The different behavior in the two intervals reflect the positive correlation of the B_s^0 -meson decay time with the BDT response, so that the low (high) BDT region contains more signal decays with small (large) decay times.

Finally, the $B_s^0 \to \mu^+\mu^-$ effective lifetime is determined using a simultaneous fit to the background-subtracted decay-time distributions in the two BDT regions in data, where the decay-time distributions are modeled by the acceptance functions above multiplied by an exponential function. Only the effective lifetime is allowed to float freely in the fit, while the parameters of the acceptance function are Gaussian constrained to the results of the fits to simulation.

Pseudoexperiments are used to evaluate several systematic effects that have the potential to bias the measurement. The fit procedure is found to return an unbiased estimate of the lifetime to a precision of 0.009 ps and good coverage. The effects of residual contamination from physical background, predominantly $B_{(s)}^0 \to h^+h'^-$ and $\Lambda_b^0 \to p\mu^-\bar{\nu}_\mu$ decays, is found to introduce a bias of around 0.012 ps. The effect of the decay-time acceptance on the mixture of the light and heavy mass eigenstates is evaluated and found to be negligible [32]. A further source of uncertainty is related to the decay-time distribution of the combinatorial background, which is unknown *a priori* and can bias the

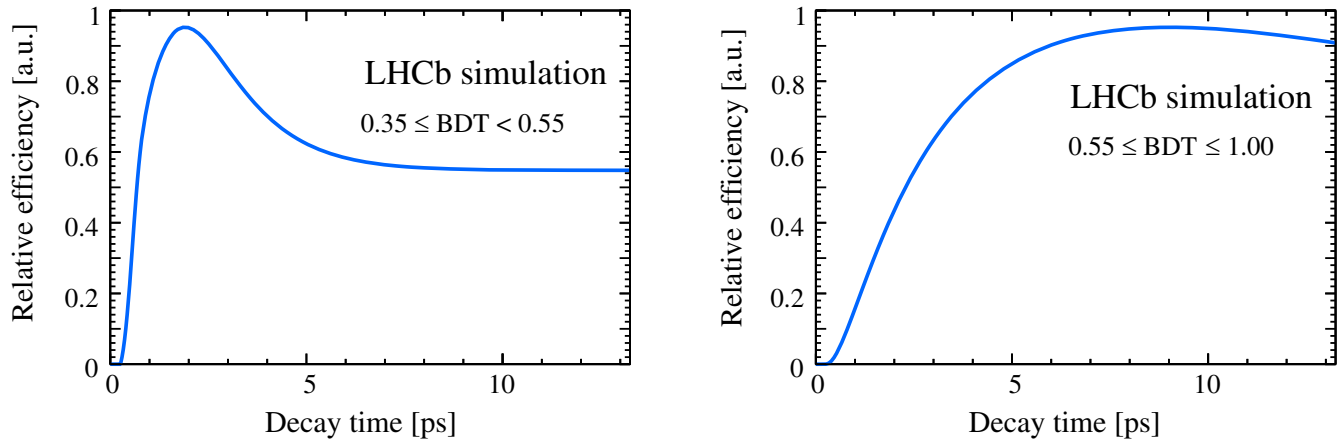

FIG. 14. The functions used to model the decay-time efficiency in the (left) low and (right) high BDT regions in the fit for the $B_s^0 \to \mu^+\mu^-$ effective lifetime.

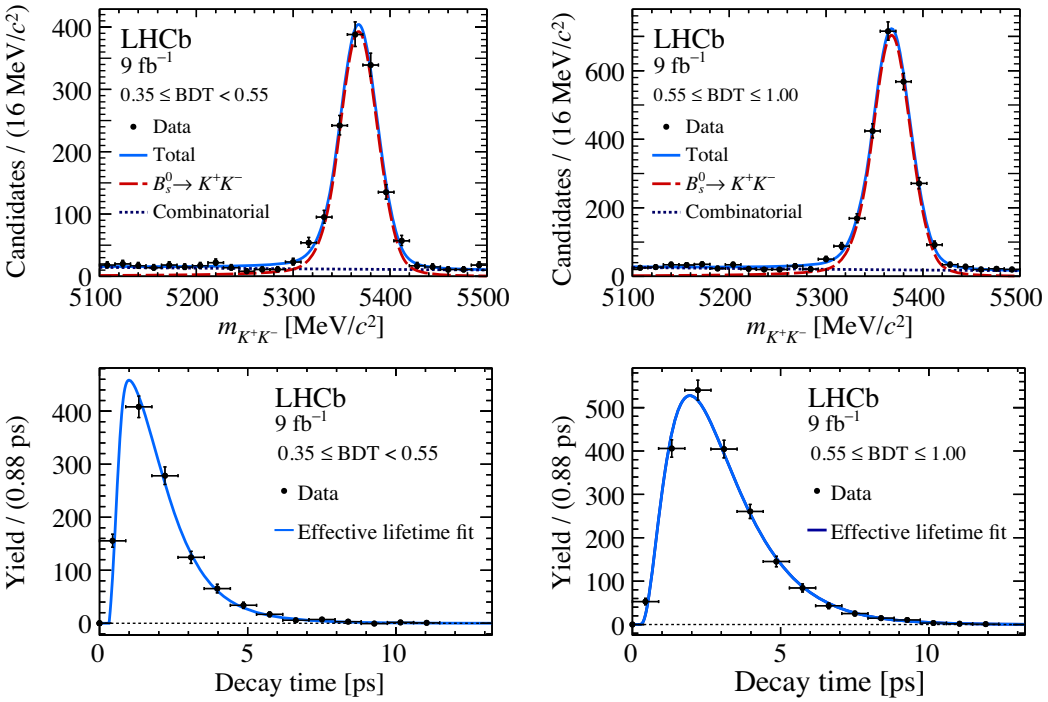

FIG. 15. Top: distribution of K^+K^- mass with the fit models used to perform the background subtraction superimposed and bottom: the background-subtracted decay-time distributions with the fit model used to determine the $B_s^0 \to K^+K^-$ effective lifetime superimposed (bottom row). The distributions in the low and high BDT regions are shown in the left and right columns, respectively.

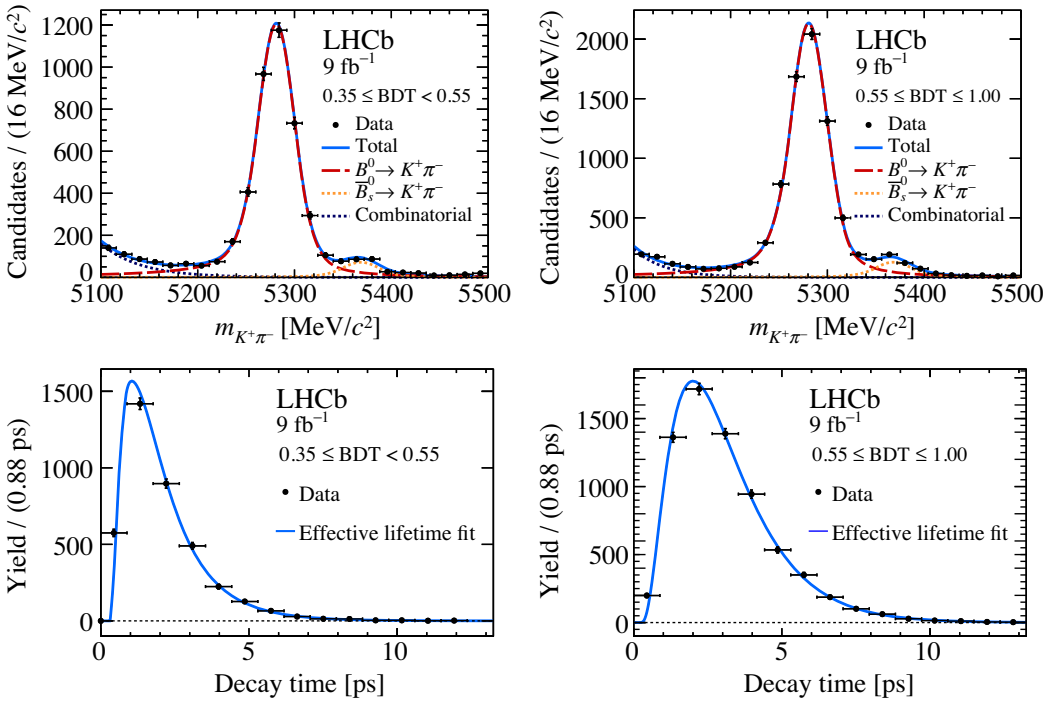

FIG. 16. Top: distribution of $K^+\pi^-$ mass with the fit models used to perform the background subtraction superimposed and bottom: the background-subtracted decay-time distributions with the fit model used to determine the $B^0 \to K^+\pi^-$ lifetime superimposed. The distributions in the low and high BDT regions are shown in the left and right columns, respectively.

TABLE X. Summary of the systematic uncertainties for the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime.

Source	Systematic uncertainty on $\tau_{\mu^+\mu^-}$ (ps)
Fit accuracy	0.009
Background contamination	0.012
Background decay-time model	0.003
Production asymmetry	0.002
Decay-time acceptance accuracy	0.026
Total	0.030

lifetime measurement if the background lifetime is significantly longer than that of the signal. The decay-time distribution of combinatorial background in the dimuon sample cannot be determined directly from data due to the very small number of candidates and so instead the decay-time model used in the pseudoexperiments is taken from a fit to the decay-time distributions of candidates in the high mass region of the higher yield dihadron sample, which includes both a short and long lived component. A systematic uncertainty is evaluated by fluctuating the mean lifetimes of both components upwards by 1σ, which results in an overall bias on the $B_s^0 \to \mu^+\mu^-$ effective lifetime of around 0.003 ps. The effect of a production asymmetry between B_s^0 and $\bar{B}_s^0$ mesons [101] is found to be small, at 0.002 ps. The effects of ignoring the detector decay-time resolution and also the choice of signal mass PDF are evaluated and are both found to be negligible. Any correlation between mass and decay-time is found to be negligible, as required by the *sPlot* method.

The entire procedure used to measure the lifetime is cross-checked by performing measurements of the lifetimes of the $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ decays, which have much larger branching fractions and have already been precisely measured. Candidates are selected using similar requirements to those used to select $B_s^0 \to \mu^+\mu^-$ decays, with a few differences. While the efficiency of the $B_s^0 \to \mu^+\mu^-$ trigger selection is independent of decay time, this is not true for $B_{(s)}^0 \to h^+h'^-$ decays where flight distance requirements are imposed on candidates. In order to match the $B_s^0 \to \mu^+\mu^-$ trigger requirements as closely as possible, $B_{(s)}^0 \to h^+h'^-$ events are selected with TIS requirements, eliminating any dependence of the trigger efficiency on decay time. Furthermore, different requirements on particle identification information are imposed in order to separate the $K^+\pi^-$ and K^+K^- final states. The decay-time acceptance is evaluated using weighted simulation in the same way as for $B_s^0 \to \mu^+\mu^-$ and the fit procedure is identical apart from some differences in the fit ranges and the inclusion of a $B_s^0 \to K^-\pi^+$ component in the $B^0 \to K^+\pi^-$ fit. The fits to the $B_s^0 \to K^+K^-$ and $B^0 \to K^+\pi^-$ mass and weighted decay-time distributions are shown in Figs. 15 and 16. The measured values of the $B_s^0 \to K^+K^-$ and $B^0 \to K^+\pi^-$ lifetimes are found to be

$$\tau_{B_s^0 \to K^+K^-} = 1.435 \pm 0.026 \text{ ps},$$
$$\tau_{B^0 \to K^+\pi^-} = 1.510 \pm 0.015 \text{ ps},$$

where the uncertainties are statistical only. Systematic uncertainties on these cross-check measurements are not evaluated. The results are in agreement with the values measured previously by the LHCb collaboration of $\tau_{B_s^0 \to K^+K^-} = 1.407 \pm 0.016$ ps and $\tau_{B^0 \to K^+\pi^-} = 1.524 \pm 0.011$ ps [102]. The measurements presented here are performed on data samples with very little overlap with those used to make the measurements published in Ref. [102] due to different data-taking periods and trigger requirements. The statistical uncertainty on the measured $B_s^0 \to K^+K^-$ lifetime is taken as the systematic uncertainty associated with the use of simulated events to determine the $B_s^0 \to \mu^+\mu^-$ acceptance function.

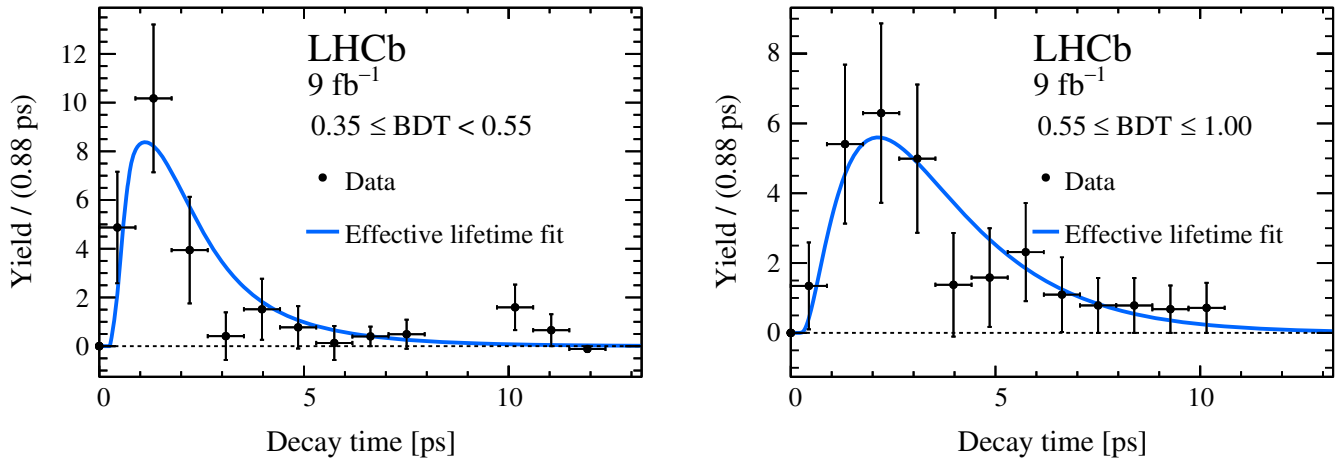

FIG. 17. Background-subtracted decay-time distributions with the fit model used to extract the $B_s^0 \to \mu^+\mu^-$ effective lifetime superimposed in the (left) low and (right) high BDT regions.

A summary of the systematic uncertainties is reported in Table X. Finally, the effective $B_s^0 \to \mu^+\mu^-$ lifetime is measured with the fits to the decay-time distributions shown in Fig. 17, as

$$\tau_{\mu^+\mu^-} = 2.07 \pm 0.29 \pm 0.03 \text{ ps},$$

where the first uncertainty is statistical and the second systematic. While this value lies above the physical range defined by the lifetimes of the light ($A_{\Delta\Gamma} = -1$) and heavy ($A_{\Delta\Gamma} = 1$, predicted by the SM) mass eigenstates, which are $\tau_L = 1.423 \pm 0.005$ ps and $\tau_H = 1.620 \pm 0.007$ ps [51], it is consistent with these lifetimes at 2.2σ and 1.5σ, respectively.

X. CONCLUSIONS

In summary, the full Run 1 and Run 2 data sample of the LHCb experiment was analysed to measure the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and to search for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays.

The branching fractions of the $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays are measured to be

$$\begin{aligned}
\mathcal{B}(B_s^0 \to \mu^+\mu^-) &= (3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9}, \\
\mathcal{B}(B^0 \to \mu^+\mu^-) &= (1.20^{+0.83}_{-0.74} \pm 0.14) \times 10^{-10}, \\
\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) &= (-2.5 \pm 1.4 \pm 0.8) \times 10^{-9} \\
&\text{with} \quad m_{\mu\mu} > 4.9 \text{ GeV}/c^2,
\end{aligned}$$

where the first uncertainty is statistical and the second systematic. The systematic uncertainty on the $B_s^0 \to \mu^+\mu^-$ decay is significantly reduced compared to previous measurements thanks to a new precise value of the hadronization fraction ratio and a more precise calibration of the BDT response and of the particle misidentification rate. The $B_s^0 \to \mu^+\mu^-$ branching fraction is the most precise single-experiment measurement to date.

The $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ signals are not statistically significant, and consistent with the background-only hypothesis at 1.7 and 1.5σ level, respectively. Upper limits on the branching fractions are set to

$$\begin{aligned}
\mathcal{B}(B^0 \to \mu^+\mu^-) &< 2.6 \times 10^{-10} \\
\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) &< 2.0 \times 10^{-9}
\end{aligned}$$

at 95% CL, the latter with $m_{\mu\mu} > 4.9$ GeV$/c^2$. The limit on the $B^0 \to \mu^+\mu^-$ decay is the most stringent to date from a single experiment. An upper limit on the $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction is determined for the first time. This limit only constrains the high-q^2 region of this decay and no attempt is made here to extrapolate the result to the full branching fraction.

Using the same data sample, with a slightly different selection, the effective lifetime of the $B_s^0 \to \mu^+\mu^-$ decay is found to be

$$\tau_{\mu^+\mu^-} = 2.07 \pm 0.29 \pm 0.03 \text{ ps},$$

where the first uncertainty is statistical and the second systematic.

All the results are compatible with the predictions of the SM and with previous measurements of these quantities. In particular, in the two-dimensional space of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions the compatibility is at one standard deviation level. These results significantly constrain possible contributions to these decays from new interactions that would cause effective scalar, pseudoscalar or axial-vector currents, and thus limit the parameter space of new physics models.

ACKNOWLEDGMENTS

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 (United Kingdom); DOE NP and NSF (USA). We acknowledge the computing resources that are provided by CERN, IN2P3 (France), KIT and DESY (Germany), INFN (Italy), SURF (Netherlands), PIC (Spain), GridPP (United Kingdom), RRCKI and Yandex LLC (Russia), CSCS (Switzerland), IFIN-HH (Romania), CBPF (Brazil), PL-GRID (Poland) and NERSC (USA). We are indebted to the communities behind the multiple open-source software packages on which we depend. Individual groups or members have received support from ARC and ARDC (Australia); AvH Foundation (Germany); EPLANET, Marie Skłodowska-Curie Actions and ERC (European Union); 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 (United Kingdom).

[1] S. Aoki *et al.* (Flavour Lattice Averaging Group), FLAG review 2019, Eur. Phys. J. C **80**, 113 (2020).

[2] A. Bazavov *et al.* (Fermilab Lattice and MILC Collaborations), B- and D-meson leptonic decay constants from four-flavor lattice QCD, Phys. Rev. D **98**, 074512 (2018).

[3] 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).

[4] R. J. Dowdall, C. T. H. Davies, R. R. Horgan, C. J. Monahan, and J. Shigemitsu (HPQCD Collaboration), B-Meson Decay Constants from Improved Lattice Nonrelativistic QCD with Physical u, d, s, and c Quarks, Phys. Rev. Lett. **110**, 222003 (2013).

[5] C. Hughes, C. T. H. Davies, and C. J. Monahan, New methods for B meson decay constants and form factors from lattice NRQCD, Phys. Rev. D **97**, 054509 (2018).

[6] C. Bobeth, M. Gorbahn, T. Hermann, M. Misiak, E. Stamou, and M. Steinhauser, $B_{s,d} \to l^+l^-$ in the Standard Model with Reduced Theoretical Uncertainty, Phys. Rev. Lett. **112**, 101801 (2014).

[7] M. Beneke, C. Bobeth, and R. Szafron, Power-enhanced leading-logarithmic QED corrections to $B_q \to \mu^+\mu^-$, J. High Energy Phys. 10 (2019) 232.

[8] W. Altmannshofer, C. Niehoff, and D. M. Straub, $B_s \to \mu^+\mu^-$ as current and future probe of new physics, J. High Energy Phys. 05 (2017) 076.

[9] H. E. Logan and U. Nierste, $B_{s,d} \to \ell^+\ell^-$ in a two-higgs-doublet model, Nucl. Phys. **B586**, 39 (2000).

[10] R. Alonso, B. Grinstein, and J. Martin Camalich, $SU(2) \times U(1)$ Gauge Invariance and the Shape of New Physics in Rare B Decays, Phys. Rev. Lett. **113**, 241802 (2014).

[11] Y. Amhis *et al.* (Heavy Flavor Averaging Group), Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018, Eur. Phys. J. C **81**, 226 (2021).

[12] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, A. Pellegrino, and N. Tuning, Probing New Physics via the $B_s^0 \to \mu^+\mu^-$ Effective Lifetime, Phys. Rev. Lett. **109**, 041801 (2012).

[13] A. J. Buras, Relations between $\Delta M_{s,d}$ and $B_{s,d} \to \mu\bar{\mu}$ in models with minimal flavour violation, Phys. Lett. B **566**, 115 (2003).

[14] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia, Minimal flavor violation: An effective field theory approach, Nucl. Phys. **B645**, 155 (2002).

[15] G. Eilam, C.-D. Lu, and D.-X. Zhang, Radiative dileptonic decays of B mesons, Phys. Lett. B **391**, 461 (1997).

[16] T. M. Aliev, A. Ozpineci, and M. Savci, $B_q \to l^+l^-\gamma$ decays in light cone QCD, Phys. Rev. D **55**, 7059 (1997).

[17] C. Q. Geng, C. C. Lih, and W.-M. Zhang, Study of $B_{s,d} \to l^+l^-\gamma$ decays, Phys. Rev. D **62**, 074017 (2000).

[18] D. Melikhov and N. Nikitin, Rare radiative leptonic decays $B_{d,s} \to l^+l^-\gamma$, Phys. Rev. D **70**, 114028 (2004).

[19] A. Kozachuk, D. Melikhov, and N. Nikitin, Rare FCNC radiative leptonic $B_{s,d} \to \gamma l^+l^-$ decays in the standard model, Phys. Rev. D **97**, 053007 (2018).

[20] S. Dubnička, A. Z. Dubničková, M. A. Ivanov, A. Liptaj, P. Santorelli, and C. T. Tran, Study of $B_s \to \ell^+\ell^-\gamma$ decays in covariant quark model, Phys. Rev. D **99**, 014042 (2019).

[21] M. Beneke, C. Bobeth, and Y.-M. Wang, $B_{d,s} \to \gamma\ell\bar{\ell}$ decay with an energetic photon, J. High Energy Phys. 12 (2020) 148.

[22] F. Dettori, D. Guadagnoli, and M. Reboud, $B_s^0 \to \mu^+\mu^-\gamma$ from $B_s^0 \to \mu^+\mu^-$, Phys. Lett. B **768**, 163 (2017).

[23] T. M. Aliev, A. Ozpineci, and M. Savci, Rare $B \to l^+l^-\gamma$ decay and new physics effects, Phys. Lett. B **520**, 69 (2001).

[24] D. Guadagnoli, M. Reboud, and R. Zwicky, $B_s^0 \to \ell^+\ell^-\gamma$ as a test of lepton flavor universality, J. High Energy Phys. 11 (2017) 184.

[25] R. Aaij *et al.* (LHCb Collaboration), Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays, J. High Energy Phys. 08 (2017) 055.

[26] R. Aaij *et al.* (LHCb Collaboration), Search for Lepton-Universality Violation in $B^+ \to K^+\ell^+\ell^-$ Decays, Phys. Rev. Lett. **122**, 191801 (2019).

[27] R. Aaij *et al.* (LHCb Collaboration), Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay, Phys. Rev. Lett. **125**, 011802 (2020).

[28] R. Aaij *et al.* (LHCb Collaboration), Angular Analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ Decay, Phys. Rev. Lett. **126**, 161802 (2021).

[29] R. Giles *et al.*, Two-body decays of B mesons, Phys. Rev. D **30**, 2279 (1984).

[30] R. Aaij *et al.* (LHCb Collaboration), First Evidence for the Decay $B_s^0 \to \mu^+\mu^-$, Phys. Rev. Lett. **110**, 021801 (2013).

[31] 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).

[32] 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).

[33] M. Aaboud *et al.* (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, J. High Energy Phys. 04 (2019) 098.

[34] A. M. Sirunyan *et al.* (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, J. High Energy Phys. 04 (2020) 188.

[35] ATLAS, CMS, and LHCb Collaborations, Combination of the ATLAS, CMS and LHCb results on the $B_{(s)}^0 \to \mu^+\mu^-$ decays, Report Nos. LHCb-CONF-2020-002, ATLAS-CONF-2020-049, and CMS-PAS-BPH-20-003, 2020, https://cds.cern.ch/record/2727207.

[36] B. Aubert *et al.* (*BABAR* Collaboration), Search for the decays $B^0 \to e^+e^-\gamma$ and $B^0 \to \mu^+\mu^-\gamma$, Phys. Rev. D **77**, 011104 (2008).

[37] R. Aaij *et al.* (LHCb Collaboration), companion Letter, Analysis of Neutral B-Meson Decays into Two Muons, Phys. Rev. Lett. **128**, 041801 (2022).

[38] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees* (Wadsworth International Group, Belmont, California, USA, 1984).

[39] A. A. Alves, Jr. *et al.* (LHCb Collaboration), The LHCb detector at the LHC, J. Instrum. **3**, S08005 (2008).

[40] R. Aaij *et al.* (LHCb Collaboration), LHCb detector performance, Int. J. Mod. Phys. A **30**, 1530022 (2015).

[41] T. Sjöstrand, S. Mrenna, and P. Skands, PYTHIA 6.4 physics and manual, J. High Energy Phys. 05 (2006) 026; T. Sjöstrand, S. Mrenna, and P. Skands, A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. **178**, 852 (2008).

[42] I. Belyaev *et al.*, Handling of the generation of primary events in Gauss, the LHCb simulation framework, J. Phys. Conf. Ser. **331**, 032047 (2011).

[43] D. J. Lange, The EvtGen particle decay simulation package, Nucl. Instrum. Methods Phys. Res., Sect. A **462**, 152 (2001).

[44] C.-H. Chang, C. Driouichi, P. Eerola, and X. G. Wu, BCVEGPY: An event generator for hadronic production of the B_c meson, Comput. Phys. Commun. **159**, 192 (2004).

[45] C.-H. Chang, X.-Y. Wang, and X.-G. Wu, BCVEGPY2.2: A newly upgraded version for hadronic production of the meson B_c and its excited states, Comput. Phys. Commun. **197**, 335 (2015).

[46] N. Davidson, T. Przedzinski, and Z. Was, PHOTOS interface in C++: Technical and physics documentation, Comput. Phys. Commun. **199**, 86 (2016).

[47] M. Bordone, G. Isidori, and A. Pattori, On the Standard Model predictions for R_K and R_{K^*}, Eur. Phys. J. C **76**, 440 (2016).

[48] J. Allison *et al.* (Geant4 Collaboration), Geant4 developments and applications, IEEE Trans. Nucl. Sci. **53**, 270 (2006); S. Agostinelli *et al.* (Geant4 Collaboration), Geant4: A simulation toolkit, Nucl. Instrum. Methods Phys. Res., Sect. A **506**, 250 (2003).

[49] M. Clemencic, G. Corti, S. Easo, C. R. Jones, S. Miglioranzi, M. Pappagallo, and P. Robbe, The LHCb simulation application, Gauss: Design, evolution and experience, J. Phys. Conf. Ser. **331**, 032023 (2011).

[50] F. Archilli *et al.*, Performance of the muon identification at LHCb, J. Instrum. **8**, P10020 (2013).

[51] P. A. Zyla *et al.* (Particle Data Group), Review of particle physics, Prog. Theor. Exp. Phys. **2020**, 083C01 (2020).

[52] Y. Freund and R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. **55**, 119 (1997).

[53] A. Hoecker *et al.*, TMVA 4 - toolkit for multivariate data analysis with ROOT. Users Guide, arXiv:physics/0703039.

[54] R. Aaij *et al.* (LHCb Collaboration), Strong Constraints on the Rare Decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, Phys. Rev. Lett. **108**, 231801 (2012).

[55] T. Skwarnicki, A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances, Ph.D. thesis, Institute of Nuclear Physics, Krakow [Report No. DESY-F31-86-02, 1986].

[56] R. Aaij *et al.* (LHCb Collaboration), Observation of the decay $\bar{B}_s^0 \to \psi(2S)K^+\pi^-$, Phys. Lett. B **747**, 484 (2015).

[57] R. Aaij *et al.* (LHCb Collaboration), Measurement of b-hadron masses, Phys. Lett. B **708**, 241 (2012).

[58] D. Acosta *et al.* (CDF Collaboration), Measurement of b Hadron Masses in Exclusive J/ψ Decays with the CDF Detector, Phys. Rev. Lett. **96**, 202001 (2006).

[59] H. Albrecht *et al.* (ARGUS Collaboration), Search for hadronic $b \to u$ decays, Phys. Lett. B **241**, 278 (1990).

[60] A. Rogozhnikov, Reweighting with boosted decision trees, J. Phys. Conf. Ser. **762**, 012036 (2016).

[61] R. Aaij *et al.* (LHCb Collaboration), Precise measurement of the f_s/f_d ratio of fragmentation fractions and of B_s^0 decay branching fractions, Phys. Rev. D **104**, 032005 (2021).

[62] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, Data driven trigger efficiency determination at LHCb, Report No. LHCb-PUB-2014-039, 2014.

[63] L. Anderlini *et al.*, The PIDCalib package, Report No. LHCb-PUB-2016-021, 2016.

[64] B. Aubert *et al.* (*BABAR* Collaboration), Improved measurements of the branching fractions for $B^0 \to \pi^+\pi^-$ and $B^0 \to K^+\pi^-$, and a search for $B^0 \to K^+K^-$, Phys. Rev. D **75**, 012008 (2007).

[65] Y.-T. Duh *et al.* (Belle Collaboration), Measurements of branching fractions and direct CP asymmetries for $B \to K\pi$, $B \to \pi\pi$ and $B \to KK$ decays, Phys. Rev. D **87**, 031103 (2013).

[66] A. Bornheim *et al.* (CLEO Collaboration), Measurements of charmless hadronic two body B meson decays and the ratio $\mathcal{B}(B \to DK)/\mathcal{B}(B \to D\pi)$, Phys. Rev. D **68**, 052002 (2003); **75**, 119907(E) (2007).

[67] K. Chilikin *et al.* (Belle Collaboration), Evidence for $B^+ \to h_c K^+$ and observation of $\eta_c(2S) \to p\bar{p}\pi^+\pi^-$, Phys. Rev. D **100**, 012001 (2019).

[68] Y. Kato *et al.* (Belle Collaboration), Measurements of the absolute branching fractions of $B^+ \to X_{c\bar{c}}K^+$ and $B^+ \to \bar{D}^{(*)0}\pi^+$ at Belle, Phys. Rev. D **97**, 012005 (2018).

[69] B. Aubert *et al.* (*BABAR* Collaboration), Measurements of the Absolute Branching Fractions of $B^\pm \to K^\pm$ X($c\bar{c}$), Phys. Rev. Lett. **96**, 052002 (2006).

[70] B. Aubert *et al.* (*BABAR* Collaboration), Measurement of the $B^+ \to p\bar{p}K^+$ branching fraction and study of the decay dynamics, Phys. Rev. D **72**, 051101 (2005).

[71] B. Aubert *et al.* (*BABAR* Collaboration), Measurement of Branching Fractions and Charge Asymmetries for Exclusive B Decays to Charmonium, Phys. Rev. Lett. **94**, 141801 (2005).

[72] K. Abe *et al.* (Belle Collaboration), Measurement of branching fractions and charge asymmetries for two-body B meson decays with charmonium, Phys. Rev. D **67**, 032003 (2003).

[73] C. P. Jessop *et al.* (CLEO Collaboration), Measurement of the Decay Amplitudes and Branching Fractions of $B \to J/\psi K^*$ and $B \to J/\psi K$ Decays, Phys. Rev. Lett. **79**, 4533 (1997).

[74] D. Bortoletto *et al.* (CLEO Collaboration), Inclusive and exclusive decays of B mesons to final states including charm and charmonium mesons, Phys. Rev. D **45**, 21 (1992).

[75] H. Albrecht *et al.* (ARGUS Collaboration), Exclusive hadronic decays of B mesons, Z. Phys. C **48**, 543 (1990).

[76] V. V. Anashin *et al.*, Final analysis of KEDR data on J/ψ and $\psi(2S)$ masses, Phys. Lett. B **749**, 50 (2015).

[77] T. A. Armstrong *et al.* (E760 Collaboration), Measurement of the J/ψ and ψ' resonance parameters in $\bar{p}p$ annihilation, Phys. Rev. D **47**, 772 (1993).

[78] A. S. Artamonov *et al.* (OLYA and MD-1 Collaborations), High precision mass measurements in ψ and Υ families revisited, Phys. Lett. B **474**, 427 (2000).

[79] C. Baglin *et al.*, J/ψ resonance formation and mass measurement in $\bar{p}$-p annihilations, Nucl. Phys. **B286**, 592 (1987).

[80] D. Martinez Santos and F. Dupertuis, Mass distributions marginalized over per-event errors, Nucl. Instrum. Methods Phys. Res., Sect. A **764**, 150 (2014).

[81] M. van Veghel, Pursuing forbidden beauty: Search for the lepton-flavour violating decays $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ and study of electron-reconstruction performance at LHCb, Ph.D. thesis, Groningen University [Report No. CERN-THESIS-2020-060, 2020].

[82] R. Aaij *et al.* (LHCb Collaboration), Measurement of f_s/f_u Variation with Proton-Proton Collision Energy and B-Meson Kinematics, Phys. Rev. Lett. **124**, 122002 (2020).

[83] R. Aaij *et al.* (LHCb Collaboration), Measurement of b hadron production fractions in 7 TeV pp collisions, Phys. Rev. D **85**, 032008 (2012).

[84] R. Aaij *et al.* (LHCb Collaboration), Measurement of b-hadron fractions in 13 TeV pp collisions, Phys. Rev. D **100**, 031102(R) (2019).

[85] R. Aaij *et al.* (LHCb Collaboration), Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics, J. High Energy Phys. 04 (2013) 001.

[86] R. Aaij *et al.* (LHCb Collaboration), Measurement of the branching fraction of the $B^0 \to D_s^+\pi^-$ decay, Eur. Phys. J. C **81**, 314 (2021).

[87] R. Aaij *et al.* (LHCb Collaboration), Measurement of the track reconstruction efficiency at LHCb, J. Instrum. **10**, P02007 (2015).

[88] R. Aaij *et al.*, Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2, Eur. Phys. J. Tech. Instrum. **6**, 1 (2019).

[89] R. Aaij *et al.* (LHCb Collaboration), Measurement of the B_c^- production fraction and asymmetry in 7 and 13 TeV pp collisions, Phys. Rev. D **100**, 112006 (2019).

[90] W.-F. Wang and Z.-J. Xiao, The semileptonic decays $B/B_s \to (\pi, K)(\ell^+\ell^-, \ell\nu, \nu\bar{\nu})$ in the perturbative QCD approach beyond the leading-order, Phys. Rev. D **86**, 114025 (2012).

[91] R. Aaij *et al.* (LHCb Collaboration), Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays, Nat. Phys. **11**, 743 (2015).

[92] R. Aaij *et al.* (LHCb Collaboration), First Observation of the Decay $B_s^0 \to K^-\mu^+\nu_\mu$ and Measurement of $|V_{ub}|/|V_{cb}|$, Phys. Rev. Lett. **126**, 081804 (2021).

[93] K. S. Cranmer, Kernel estimation in high-energy physics, Comput. Phys. Commun. **136**, 198 (2001).

[94] S. S. Wilks, The large-sample distribution of the likelihood ratio for testing composite hypotheses, Ann. Math. Stat. **9**, 60 (1938).

[95] A. L. Read, Presentation of search results: The CL_S technique, J. Phys. G **28**, 2693 (2002).

[96] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, Eur. Phys. J. C **71**, 1554 (2011); **73**, 2501(E) (2013).

[97] M. Kenzie *et al.*, GammaCombo: A statistical analysis framework for combining measurements, fitting datasets and producing confidence intervals, 10.5281/zenodo.3371421.

[98] R. Aaij *et al.* (LHCb Collaboration), Measurement of the CKM angle γ from a combination of LHCb results, J. High Energy Phys. 12 (2016) 087.

[99] S. Bodhisattva, M. Walker, and M. Woodroofe, On the unified method with nuisance parameters, Statistica Sinica **19**, 301 (2009), https://www.jstor.org/stable/24308721.

[100] M. Pivk and F. R. Le Diberder, sPlot: A statistical tool to unfold data distributions, Nucl. Instrum. Methods Phys. Res., Sect. A **555**, 356 (2005).

[101] R. Aaij *et al.* (LHCb Collaboration), Measurement of B^0, B_s^0, B^+ and Λ_b^0 production asymmetries in 7 and 8 TeV proton-proton collisions, Phys. Lett. B **774**, 139 (2017).

[102] R. Aaij *et al.* (LHCb Collaboration), Effective lifetime measurements in the $B_s^0 \to K^+K^-$, $B^0 \to K^+\pi^-$ and $B_s^0 \to \pi^+K^-$ decays, Phys. Lett. B **736**, 446 (2014).

R. Aaij,[32] C. Abellán Beteta,[50] T. Ackernley,[60] B. Adeva,[46] M. Adinolfi,[54] H. Afsharnia,[9] C. A. Aidala,[86] S. Aiola,[25] Z. Ajaltouni,[9] S. Akar,[65] J. Albrecht,[15] F. Alessio,[48] M. Alexander,[59] A. Alfonso Albero,[45] Z. Aliouche,[62] G. Alkhazov,[38] P. Alvarez Cartelle,[55] S. Amato,[2] Y. Amhis,[11] L. An,[48] L. Anderlini,[22] A. Andreianov,[38] M. Andreotti,[21] F. Archilli,[17] A. Artamonov,[44] M. Artuso,[68] K. Arzymatov,[42] E. Aslanides,[10] M. Atzeni,[50] B. Audurier,[12] S. Bachmann,[17] M. Bachmayer,[49] J. J. Back,[56] P. Baladron Rodriguez,[46] V. Balagura,[12] W. Baldini,[21] J. Baptista Leite,[1] R. J. Barlow,[62] S. Barsuk,[11] W. Barter,[61] M. Bartolini,[24,a] F. Baryshnikov,[83] J. M. Basels,[14] G. Bassi,[29] B. Batsukh,[68] A. Battig,[15] A. Bay,[49] M. Becker,[15] F. Bedeschi,[29] I. Bediaga,[1] A. Beiter,[68] V. Belavin,[42] S. Belin,[27] V. Bellee,[49] K. Belous,[44] I. Belov,[40] I. Belyaev,[41] G. Bencivenni,[23] E. Ben-Haim,[13] A. Berezhnoy,[40] R. Bernet,[50] D. Berninghoff,[17] H. C. Bernstein,[68] C. Bertella,[48] A. Bertolin,[28] C. Betancourt,[50] F. Betti,[48] Ia. Bezshyiko,[50] S. Bhasin,[54] J. Bhom,[35] L. Bian,[73] M. S. Bieker,[15] S. Bifani,[53] P. Billoir,[13] M. Birch,[61] F. C. R. Bishop,[55] A. Bitadze,[62] A. Bizzeti,[22,b] M. Bjørn,[63] M. P. Blago,[48] T. Blake,[56] F. Blanc,[49] S. Blusk,[68] D. Bobulska,[59] J. A. Boelhauve,[15] O. Boente Garcia,[46] T. Boettcher,[65] A. Boldyrev,[82] A. Bondar,[43] N. Bondar,[38,48] S. Borghi,[62] M. Borisyak,[42] M. Borsato,[17] J. T. Borsuk,[35] S. A. Bouchiba,[49] T. J. V. Bowcock,[60] A. Boyer,[48]

C. Bozzi,[21] M. J. Bradley,[61] S. Braun,[66] A. Brea Rodriguez,[46] M. Brodski,[48] J. Brodzicka,[35] A. Brossa Gonzalo,[56] D. Brundu,[27,48] A. Buonaura,[50] C. Burr,[48] A. Bursche,[72] A. Butkevich,[39] J. S. Butter,[32] J. Buytaert,[48] W. Byczynski,[48] S. Cadeddu,[27] H. Cai,[73] R. Calabrese,[21,c] L. Calefice,[15,13] L. Calero Diaz,[23] S. Cali,[23] R. Calladine,[53] M. Calvi,[26,d] M. Calvo Gomez,[85] P. Camargo Magalhaes,[54] A. Camboni,[45,85] P. Campana,[23] A. F. Campoverde Quezada,[6] S. Capelli,[26,d] L. Capriotti,[20,e] A. Carbone,[20,e] G. Carboni,[31] R. Cardinale,[24,a] A. Cardini,[27] I. Carli,[4] P. Carniti,[26,d] L. Carus,[14] K. Carvalho Akiba,[32] A. Casais Vidal,[46] G. Casse,[60] M. Cattaneo,[48] G. Cavallero,[48] S. Celani,[49] J. Cerasoli,[10] A. J. Chadwick,[60] M. G. Chapman,[54] M. Charles,[13] Ph. Charpentier,[48] G. Chatzikonstantinidis,[53] C. A. Chavez Barajas,[60] M. Chefdeville,[8] C. Chen,[3] S. Chen,[4] A. Chernov,[35] V. Chobanova,[46] S. Cholak,[49] M. Chrzaszcz,[35] A. Chubykin,[38] V. Chulikov,[38] P. Ciambrone,[23] M. F. Cicala,[56] X. Cid Vidal,[46] G. Ciezarek,[48] P. E. L. Clarke,[58] M. Clemencic,[48] H. V. Cliff,[55] J. Closier,[48] J. L. Cobbledick,[62] V. Coco,[48] J. A. B. Coelho,[11] J. Cogan,[10] E. Cogneras,[9] L. Cojocariu,[37] P. Collins,[48] T. Colombo,[48] L. Congedo,[19,f] A. Contu,[27] N. Cooke,[53] G. Coombs,[59] G. Corti,[48] C. M. Costa Sobral,[56] B. Couturier,[48] D. C. Craik,[64] J. Crkovská,[67] M. Cruz Torres,[1] R. Currie,[58] C. L. Da Silva,[67] S. Dadabaev,[83] E. Dall'Occo,[15] J. Dalseno,[46] C. D'Ambrosio,[48] A. Danilina,[41] P. d'Argent,[48] A. Davis,[62] O. De Aguiar Francisco,[62] K. De Bruyn,[79] S. De Capua,[62] M. De Cian,[49] J. M. De Miranda,[1] L. De Paula,[2] M. De Serio,[19,f] D. De Simone,[50] P. De Simone,[23] F. De Vellis,[15] J. A. de Vries,[80] C. T. Dean,[67] D. Decamp,[8] L. Del Buono,[13] B. Delaney,[55] H.-P. Dembinski,[15] A. Dendek,[34] V. Denysenko,[50] D. Derkach,[82] O. Deschamps,[9] F. Desse,[11] F. Dettori,[27,g] B. Dey,[77] A. Di Cicco,[23] P. Di Nezza,[23] S. Didenko,[83] L. Dieste Maronas,[46] H. Dijkstra,[48] V. Dobishuk,[52] A. M. Donohoe,[18] F. Dordei,[27] A. C. dos Reis,[1] L. Douglas,[59] A. Dovbnya,[51] A. G. Downes,[8] K. Dreimanis,[60] M. W. Dudek,[35] L. Dufour,[48] V. Duk,[78] P. Durante,[48] J. M. Durham,[67] D. Dutta,[62] A. Dziurda,[35] A. Dzyuba,[38] S. Easo,[57] U. Egede,[69] V. Egorychev,[41] S. Eidelman,[43,h] S. Eisenhardt,[58] S. Ek-In,[49] L. Eklund,[59,i] S. Ely,[68] A. Ene,[37] E. Epple,[67] S. Escher,[14] J. Eschle,[50] S. Esen,[13] T. Evans,[48] A. Falabella,[20] J. Fan,[3] Y. Fan,[6] B. Fang,[73] S. Farry,[60] D. Fazzini,[26,d] M. Féo,[48] A. Fernandez Prieto,[46] A. D. Fernez,[66] F. Ferrari,[20,e] L. Ferreira Lopes,[49] F. Ferreira Rodrigues,[2] S. Ferreres Sole,[32] M. Ferrillo,[50] M. Ferro-Luzzi,[48] S. Filippov,[39] R. A. Fini,[19] M. Fiorini,[21,c] M. Firlej,[34] K. M. Fischer,[63] D. S. Fitzgerald,[86] C. Fitzpatrick,[62] T. Fiutowski,[34] F. Fleuret,[12] M. Fontana,[13] F. Fontanelli,[24,a] R. Forty,[48] V. Franco Lima,[60] M. Franco Sevilla,[66] M. Frank,[48] E. Franzoso,[21] G. Frau,[17] C. Frei,[48] D. A. Friday,[59] J. Fu,[25] Q. Fuehring,[15] W. Funk,[48] E. Gabriel,[32] T. Gaintseva,[42] A. Gallas Torreira,[46] D. Galli,[20,e] S. Gambetta,[58,48] Y. Gan,[3] M. Gandelman,[2] P. Gandini,[25] Y. Gao,[5] M. Garau,[27] L. M. Garcia Martin,[56] P. Garcia Moreno,[45] J. García Pardiñas,[26,d] B. Garcia Plana,[46] F. A. Garcia Rosales,[12] L. Garrido,[45] C. Gaspar,[48] R. E. Geertsema,[32] D. Gerick,[17] L. L. Gerken,[15] E. Gersabeck,[62] M. Gersabeck,[62] T. Gershon,[56] D. Gerstel,[10] Ph. Ghez,[8] V. Gibson,[55] H. K. Giemza,[36] M. Giovannetti,[23,j] A. Gioventù,[46] P. Gironella Gironell,[45] L. Giubega,[37] C. Giugliano,[21,48,c] K. Gizdov,[58] E. L. Gkougkousis,[48] V. V. Gligorov,[13] C. Göbel,[70] E. Golobardes,[85] D. Golubkov,[41] A. Golutvin,[61,83] A. Gomes,[1,k] S. Gomez Fernandez,[45] F. Goncalves Abrantes,[63] M. Goncerz,[35] G. Gong,[3] P. Gorbounov,[41] I. V. Gorelov,[40] C. Gotti,[26] E. Govorkova,[48] J. P. Grabowski,[17] T. Grammatico,[13] L. A. Granado Cardoso,[48] E. Graugés,[45] E. Graverini,[49] G. Graziani,[22] A. Grecu,[37] L. M. Greeven,[32] P. Griffith,[21,c] L. Grillo,[62] S. Gromov,[83] B. R. Gruberg Cazon,[63] C. Gu,[3] M. Guarise,[21] P. A. Günther,[17] E. Gushchin,[39] A. Guth,[14] Y. Guz,[44] T. Gys,[48] T. Hadavizadeh,[69] G. Haefeli,[49] C. Haen,[48] J. Haimberger,[48] T. Halewood-leagas,[60] P. M. Hamilton,[66] J. P. Hammerich,[60] Q. Han,[7] X. Han,[17] T. H. Hancock,[63] S. Hansmann-Menzemer,[17] N. Harnew,[63] T. Harrison,[60] C. Hasse,[48] M. Hatch,[48] J. He,[6,l] M. Hecker,[61] K. Heijhoff,[32] K. Heinicke,[15] A. M. Hennequin,[48] K. Hennessy,[60] L. Henry,[48] J. Heuel,[14] A. Hicheur,[2] D. Hill,[49] M. Hilton,[62] S. E. Hollitt,[15] J. Hu,[17] J. Hu,[72] W. Hu,[7] X. Hu,[3] W. Huang,[6] X. Huang,[73] W. Hulsbergen,[32] R. J. Hunter,[56] M. Hushchyn,[82] D. Hutchcroft,[60] D. Hynds,[32] P. Ibis,[15] M. Idzik,[34] D. Ilin,[38] P. Ilten,[65] A. Inglessi,[38] A. Ishteev,[83] K. Ivshin,[38] R. Jacobsson,[48] S. Jakobsen,[48] E. Jans,[32] B. K. Jashal,[47] A. Jawahery,[66] V. Jevtic,[15] F. Jiang,[3] M. John,[63] D. Johnson,[48] C. R. Jones,[55] T. P. Jones,[56] B. Jost,[48] N. Jurik,[48] S. Kandybei,[51] Y. Kang,[3] M. Karacson,[48] M. Karpov,[82] F. Keizer,[48] M. Kenzie,[56] T. Ketel,[33] B. Khanji,[15] A. Kharisova,[84] S. Kholodenko,[44] T. Kirn,[14] V. S. Kirsebom,[49] O. Kitouni,[64] S. Klaver,[32] K. Klimaszewski,[36] S. Koliiev,[52] A. Kondybayeva,[83] A. Konoplyannikov,[41] P. Kopciewicz,[34] R. Kopecna,[17] P. Koppenburg,[32] M. Korolev,[40] I. Kostiuk,[32,52] O. Kot,[52] S. Kotriakhova,[21,38] P. Kravchenko,[38] L. Kravchuk,[39] R. D. Krawczyk,[48] M. Kreps,[56] F. Kress,[61] S. Kretzschmar,[14] P. Krokovny,[43,h] W. Krupa,[34] W. Krzemien,[36] W. Kucewicz,[35,m] M. Kucharczyk,[35] V. Kudryavtsev,[43,h] H. S. Kuindersma,[32,33] G. J. Kunde,[67] T. Kvaratskheliya,[41] D. Lacarrere,[48] G. Lafferty,[62] A. Lai,[27] A. Lampis,[27] D. Lancierini,[50] J. J. Lane,[62] R. Lane,[54] G. Lanfranchi,[23,48] C. Langenbruch,[14] J. Langer,[15] O. Lantwin,[50] T. Latham,[56] F. Lazzari,[29,n] R. Le Gac,[10] S. H. Lee,[86] R. Lefèvre,[9] A. Leflat,[40] S. Legotin,[83] O. Leroy,[10] T. Lesiak,[35] B. Leverington,[17] H. Li,[72] L. Li,[63] P. Li,[17] S. Li,[7] Y. Li,[4] Y. Li,[4] Z. Li,[68] X. Liang,[68] T. Lin,[61] R. Lindner,[48] V. Lisovskyi,[15] R. Litvinov,[27]

G. Liu,[72] H. Liu,[6] S. Liu,[4] A. Loi,[27] J. Lomba Castro,[46] I. Longstaff,[59] J. H. Lopes,[2] G. H. Lovell,[55] Y. Lu,[4] D. Lucchesi,[28,o] S. Luchuk,[39] M. Lucio Martinez,[32] V. Lukashenko,[32,52] Y. Luo,[3] A. Lupato,[62] E. Luppi,[21,c] O. Lupton,[56] A. Lusiani,[29,p] X. Lyu,[6] L. Ma,[4] R. Ma,[6] S. Maccolini,[20,e] F. Machefert,[11] F. Maciuc,[37] V. Macko,[49] P. Mackowiak,[15] S. Maddrell-Mander,[54] O. Madejczyk,[34] L. R. Madhan Mohan,[54] O. Maev,[38] A. Maevskiy,[82] D. Maisuzenko,[38] M. W. Majewski,[34] J. J. Malczewski,[35] S. Malde,[63] B. Malecki,[48] A. Malinin,[81] T. Maltsev,[43,h] H. Malygina,[17] G. Manca,[27,g] G. Mancinelli,[10] D. Manuzzi,[20,e] D. Marangotto,[25,q] J. Maratas,[9,r] J. F. Marchand,[8] U. Marconi,[20] S. Mariani,[22,s] C. Marin Benito,[48] M. Marinangeli,[49] J. Marks,[17] A. M. Marshall,[54] P. J. Marshall,[60] G. Martellotti,[30] L. Martinazzoli,[48,d] M. Martinelli,[26,d] D. Martinez Santos,[46] F. Martinez Vidal,[47] A. Massafferri,[1] M. Materok,[14] R. Matev,[48] A. Mathad,[50] Z. Mathe,[48] V. Matiunin,[41] C. Matteuzzi,[26] K. R. Mattioli,[86] A. Mauri,[32] E. Maurice,[12] J. Mauricio,[45] M. Mazurek,[48] M. McCann,[61] L. Mcconnell,[18] T. H. Mcgrath,[62] A. McNab,[62] R. McNulty,[18] J. V. Mead,[60] B. Meadows,[65] G. Meier,[15] N. Meinert,[76] D. Melnychuk,[36] S. Meloni,[26,d] M. Merk,[32,80] A. Merli,[25] L. Meyer Garcia,[2] M. Mikhasenko,[48] D. A. Milanes,[74] E. Millard,[56] M. Milovanovic,[48] M.-N. Minard,[8] A. Minotti,[21] L. Minzoni,[21,c] S. E. Mitchell,[58] B. Mitreska,[62] D. S. Mitzel,[48] A. Mödden,[15] R. A. Mohammed,[63] R. D. Moise,[61] T. Mombächer,[46] I. A. Monroy,[74] S. Monteil,[9] M. Morandin,[28] G. Morello,[23] M. J. Morello,[29,p] J. Moron,[34] A. B. Morris,[75] A. G. Morris,[56] R. Mountain,[68] H. Mu,[3] F. Muheim,[58,48] M. Mulder,[48] D. Müller,[48] K. Müller,[50] C. H. Murphy,[63] D. Murray,[62] P. Muzzetto,[27,48] P. Naik,[54] T. Nakada,[49] R. Nandakumar,[57] T. Nanut,[49] I. Nasteva,[2] M. Needham,[58] I. Neri,[21] N. Neri,[25,q] S. Neubert,[75] N. Neufeld,[48] R. Newcombe,[61] T. D. Nguyen,[49] C. Nguyen-Mau,[49,t] E. M. Niel,[11] S. Nieswand,[14] N. Nikitin,[40] N. S. Nolte,[64] C. Normand,[8] C. Nunez,[86] A. Oblakowska-Mucha,[34] V. Obraztsov,[44] D. P. O'Hanlon,[54] R. Oldeman,[27,g] M. E. Olivares,[68] C. J. G. Onderwater,[79] A. Ossowska,[35] J. M. Otalora Goicochea,[2] T. Ovsiannikova,[41] P. Owen,[50] A. Oyanguren,[47] B. Pagare,[56] P. R. Pais,[48] T. Pajero,[63] A. Palano,[19] M. Palutan,[23] Y. Pan,[62] G. Panshin,[84] A. Papanestis,[57] M. Pappagallo,[19,f] L. L. Pappalardo,[21,c] C. Pappenheimer,[65] W. Parker,[66] C. Parkes,[62] C. J. Parkinson,[46] B. Passalacqua,[21] G. Passaleva,[22] A. Pastore,[19] M. Patel,[61] C. Patrignani,[20,e] C. J. Pawley,[80] A. Pearce,[48] A. Pellegrino,[32] M. Pepe Altarelli,[48] S. Perazzini,[20] D. Pereima,[41] P. Perret,[9] M. Petric,[59,48] K. Petridis,[54] A. Petrolini,[24,a] A. Petrov,[81] S. Petrucci,[58] M. Petruzzo,[25] T. T. H. Pham,[68] A. Philippov,[42] L. Pica,[29,p] M. Piccini,[78] B. Pietrzyk,[8] G. Pietrzyk,[49] M. Pili,[63] D. Pinci,[30] F. Pisani,[48] Resmi P. K,[10] V. Placinta,[37] J. Plews,[53] M. Plo Casasus,[46] F. Polci,[13] M. Poli Lener,[23] M. Poliakova,[68] A. Poluektov,[10] N. Polukhina,[83,u] I. Polyakov,[68] E. Polycarpo,[2] G. J. Pomery,[54] S. Ponce,[48] D. Popov,[6,48] S. Popov,[42] S. Poslavskii,[44] K. Prasanth,[35] L. Promberger,[48] C. Prouve,[46] V. Pugatch,[52] H. Pullen,[63] G. Punzi,[29,v] H. Qi,[3] W. Qian,[6] J. Qin,[6] N. Qin,[3] R. Quagliani,[13] B. Quintana,[8] N. V. Raab,[18] R. I. Rabadan Trejo,[10] B. Rachwal,[34] J. H. Rademacker,[54] M. Rama,[29] M. Ramos Pernas,[56] M. S. Rangel,[2] F. Ratnikov,[42,82] G. Raven,[33] M. Reboud,[8] F. Redi,[49] F. Reiss,[62] C. Remon Alepuz,[47] Z. Ren,[3] V. Renaudin,[63] R. Ribatti,[29] S. Ricciardi,[57] K. Rinnert,[60] P. Robbe,[11] G. Robertson,[58] A. B. Rodrigues,[49] E. Rodrigues,[60] J. A. Rodriguez Lopez,[74] A. Rollings,[63] P. Roloff,[48] V. Romanovskiy,[44] M. Romero Lamas,[46] A. Romero Vidal,[46] J. D. Roth,[86] M. Rotondo,[23] M. S. Rudolph,[68] T. Ruf,[48] J. Ruiz Vidal,[47] A. Ryzhikov,[82] J. Ryzka,[34] J. J. Saborido Silva,[46] N. Sagidova,[38] N. Sahoo,[56] B. Saitta,[27,g] M. Salomoni,[48] C. Sanchez Gras,[32] R. Santacesaria,[30] C. Santamarina Rios,[46] M. Santimaria,[23] E. Santovetti,[31,j] D. Saranin,[83] G. Sarpis,[14] M. Sarpis,[75] A. Sarti,[30] C. Satriano,[30,w] A. Satta,[31] M. Saur,[15] D. Savrina,[41,40] H. Sazak,[9] L. G. Scantlebury Smead,[63] A. Scarabotto,[13] S. Schael,[14] M. Schellenberg,[15] M. Schiller,[59] H. Schindler,[48] M. Schmelling,[16] B. Schmidt,[48] O. Schneider,[49] A. Schopper,[48] M. Schubiger,[32] S. Schulte,[49] M. H. Schune,[11] R. Schwemmer,[48] B. Sciascia,[23] S. Sellam,[46] A. Semennikov,[41] M. Senghi Soares,[33] A. Sergi,[24,a] N. Serra,[50] L. Sestini,[28] A. Seuthe,[15] P. Seyfert,[48] Y. Shang,[5] D. M. Shangase,[86] M. Shapkin,[44] I. Shchemerov,[83] L. Shchutska,[49] T. Shears,[60] L. Shekhtman,[43,h] Z. Shen,[5] V. Shevchenko,[81] E. B. Shields,[26,d] E. Shmanin,[83] J. D. Shupperd,[68] B. G. Siddi,[21] R. Silva Coutinho,[50] G. Simi,[28] S. Simone,[19,f] N. Skidmore,[62] T. Skwarnicki,[68] M. W. Slater,[53] I. Slazyk,[21,c] J. C. Smallwood,[63] J. G. Smeaton,[55] A. Smetkina,[41] E. Smith,[50] M. Smith,[61] A. Snoch,[32] M. Soares,[20] L. Soares Lavra,[9] M. D. Sokoloff,[65] F. J. P. Soler,[59] A. Solovev,[38] I. Solovyev,[38] F. L. Souza De Almeida,[2] B. Souza De Paula,[2] B. Spaan,[15] E. Spadaro Norella,[25,q] P. Spradlin,[59] F. Stagni,[48] M. Stahl,[65] S. Stahl,[48] P. Stefko,[49] O. Steinkamp,[50,83] O. Stenyakin,[44] H. Stevens,[15] S. Stone,[68] M. E. Stramaglia,[49] M. Straticiuc,[37] D. Strekalina,[83] F. Suljik,[63] J. Sun,[27] L. Sun,[73] Y. Sun,[66] P. Svihra,[62] P. N. Swallow,[53] K. Swientek,[34] A. Szabelski,[36] T. Szumlak,[34] M. Szymanski,[48] S. Taneja,[62] A. Terentev,[83] F. Teubert,[48] E. Thomas,[48] K. A. Thomson,[60] V. Tisserand,[9] S. T'Jampens,[8] M. Tobin,[4] L. Tomassetti,[21,c] D. Torres Machado,[1] D. Y. Tou,[13] M. T. Tran,[49] E. Trifonova,[83] C. Trippl,[49] G. Tuci,[29,v] A. Tully,[49] N. Tuning,[32,48] A. Ukleja,[36] D. J. Unverzagt,[17] E. Ursov,[83] A. Usachov,[32] A. Ustyuzhanin,[42,82] U. Uwer,[17] A. Vagner,[84] V. Vagnoni,[20] A. Valassi,[48] G. Valenti,[20] N. Valls Canudas,[85] M. van Beuzekom,[32] M. Van Dijk,[49] E. van Herwijnen,[83] C. B. Van Hulse,[18] M. van Veghel,[79] R. Vazquez Gomez,[45]

P. Vazquez Regueiro,[46] C. Vázquez Sierra,[48] S. Vecchi,[21] J. J. Velthuis,[54] M. Veltri,[22,x] A. Venkateswaran,[68] M. Veronesi,[32] M. Vesterinen,[56] D. Vieira,[65] M. Vieites Diaz,[49] H. Viemann,[76] X. Vilasis-Cardona,[85] E. Vilella Figueras,[60] A. Villa,[20] P. Vincent,[13] D. Vom Bruch,[10] A. Vorobyev,[38] V. Vorobyev,[43,h] N. Voropaev,[38] K. Vos,[80] R. Waldi,[17] J. Walsh,[29] C. Wang,[17] J. Wang,[5] J. Wang,[4] J. Wang,[3] J. Wang,[73] M. Wang,[3] R. Wang,[54] Y. Wang,[7] Z. Wang,[50] Z. Wang,[3] H. M. Wark,[60] N. K. Watson,[53] S. G. Weber,[13] D. Websdale,[61] C. Weisser,[64] B. D. C. Westhenry,[54] D. J. White,[62] M. Whitehead,[54] D. Wiedner,[15] G. Wilkinson,[63] M. Wilkinson,[68] I. Williams,[55] M. Williams,[64] M. R. J. Williams,[58] F. F. Wilson,[57] W. Wislicki,[36] M. Witek,[35] L. Witola,[17] G. Wormser,[11] S. A. Wotton,[55] H. Wu,[68] K. Wyllie,[48] Z. Xiang,[6] D. Xiao,[7] Y. Xie,[7] A. Xu,[5] J. Xu,[6] L. Xu,[3] M. Xu,[7] Q. Xu,[6] Z. Xu,[5] Z. Xu,[6] D. Yang,[3] S. Yang,[6] Y. Yang,[6] Z. Yang,[3] Z. Yang,[66] Y. Yao,[68] L. E. Yeomans,[60] H. Yin,[7] J. Yu,[71] X. Yuan,[68] O. Yushchenko,[44] E. Zaffaroni,[49] M. Zavertyaev,[16,u] M. Zdybal,[35] O. Zenaiev,[48] M. Zeng,[3] D. Zhang,[7] L. Zhang,[3] S. Zhang,[5] Y. Zhang,[5] Y. Zhang,[63] A. Zharkova,[83] A. Zhelezov,[17] Y. Zheng,[6] X. Zhou,[6] Y. Zhou,[6] X. Zhu,[3] Z. Zhu,[6] V. Zhukov,[14,40] J. B. Zonneveld,[58] Q. Zou,[4] S. Zucchelli,[20,e] D. Zuliani,[28] and G. Zunica[62]

(LHCb Collaboration)

[1]*Centro Brasileiro de Pesquisas Físicas (CBPF), Rio de Janeiro, Brazil*
[2]*Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil*
[3]*Center for High Energy Physics, Tsinghua University, Beijing, China*
[4]*Institute Of High Energy Physics (IHEP), Beijing, China*
[5]*School of Physics State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, China*
[6]*University of Chinese Academy of Sciences, Beijing, China*
[7]*Institute of Particle Physics, Central China Normal University, Wuhan, Hubei, China*
[8]*Univ. Savoie Mont Blanc, CNRS, IN2P3-LAPP, Annecy, France*
[9]*Université Clermont Auvergne, CNRS/IN2P3, LPC, Clermont-Ferrand, France*
[10]*Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France*
[11]*Université Paris-Saclay, CNRS/IN2P3, IJCLab, Orsay, France*
[12]*Laboratoire Leprince-Ringuet, CNRS/IN2P3, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, France*
[13]*LPNHE, Sorbonne Université, Paris Diderot Sorbonne Paris Cité, CNRS/IN2P3, Paris, France*
[14]*I. Physikalisches Institut, RWTH Aachen University, Aachen, Germany*
[15]*Fakultät Physik, Technische Universität Dortmund, Dortmund, Germany*
[16]*Max-Planck-Institut für Kernphysik (MPIK), Heidelberg, Germany*
[17]*Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany*
[18]*School of Physics, University College Dublin, Dublin, Ireland*
[19]*INFN Sezione di Bari, Bari, Italy*
[20]*INFN Sezione di Bologna, Bologna, Italy*
[21]*INFN Sezione di Ferrara, Ferrara, Italy*
[22]*INFN Sezione di Firenze, Firenze, Italy*
[23]*INFN Laboratori Nazionali di Frascati, Frascati, Italy*
[24]*INFN Sezione di Genova, Genova, Italy*
[25]*INFN Sezione di Milano, Milano, Italy*
[26]*INFN Sezione di Milano-Bicocca, Milano, Italy*
[27]*INFN Sezione di Cagliari, Monserrato, Italy*
[28]*Universita degli Studi di Padova, Universita e INFN, Padova, Padova, Italy*
[29]*INFN Sezione di Pisa, Pisa, Italy*
[30]*INFN Sezione di Roma La Sapienza, Roma, Italy*
[31]*INFN Sezione di Roma Tor Vergata, Roma, Italy*
[32]*Nikhef National Institute for Subatomic Physics, Amsterdam, Netherlands*
[33]*Nikhef National Institute for Subatomic Physics and VU University Amsterdam, Amsterdam, Netherlands*
[34]*AGH—University of Science and Technology, Faculty of Physics and Applied Computer Science, Kraków, Poland*
[35]*Henryk Niewodniczanski Institute of Nuclear Physics Polish Academy of Sciences, Kraków, Poland*
[36]*National Center for Nuclear Research (NCBJ), Warsaw, Poland*
[37]*Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest-Magurele, Romania*
[38]*Petersburg Nuclear Physics Institute NRC Kurchatov Institute (PNPI NRC KI), Gatchina, Russia*
[39]*Institute for Nuclear Research of the Russian Academy of Sciences (INR RAS), Moscow, Russia*

[40]Institute of Nuclear Physics, Moscow State University (SINP MSU), Moscow, Russia
[41]Institute of Theoretical and Experimental Physics NRC Kurchatov Institute (ITEP NRC KI), Moscow, Russia
[42]Yandex School of Data Analysis, Moscow, Russia
[43]Budker Institute of Nuclear Physics (SB RAS), Novosibirsk, Russia
[44]Institute for High Energy Physics NRC Kurchatov Institute (IHEP NRC KI), Protvino, Russia, Protvino, Russia
[45]ICCUB, Universitat de Barcelona, Barcelona, Spain
[46]Instituto Galego de Física de Altas Enerxías (IGFAE), Universidade de Santiago de Compostela, Santiago de Compostela, Spain
[47]Instituto de Fisica Corpuscular, Centro Mixto Universidad de Valencia—CSIC, Valencia, Spain
[48]European Organization for Nuclear Research (CERN), Geneva, Switzerland
[49]Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[50]Physik-Institut, Universität Zürich, Zürich, Switzerland
[51]NSC Kharkiv Institute of Physics and Technology (NSC KIPT), Kharkiv, Ukraine
[52]Institute for Nuclear Research of the National Academy of Sciences (KINR), Kyiv, Ukraine
[53]University of Birmingham, Birmingham, United Kingdom
[54]H.H. Wills Physics Laboratory, University of Bristol, Bristol, United Kingdom
[55]Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
[56]Department of Physics, University of Warwick, Coventry, United Kingdom
[57]STFC Rutherford Appleton Laboratory, Didcot, United Kingdom
[58]School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
[59]School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom
[60]Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom
[61]Imperial College London, London, United Kingdom
[62]Department of Physics and Astronomy, University of Manchester, Manchester, United Kingdom
[63]Department of Physics, University of Oxford, Oxford, United Kingdom
[64]Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[65]University of Cincinnati, Cincinnati, Ohio, USA
[66]University of Maryland, College Park, Maryland, USA
[67]Los Alamos National Laboratory (LANL), Los Alamos, USA
[68]Syracuse University, Syracuse, New York, USA
[69]School of Physics and Astronomy, Monash University, Melbourne, Australia (associated with Department of Physics, University of Warwick, Coventry, United Kingdom)
[70]Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil (associated with Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil)
[71]Physics and Micro Electronic College, Hunan University, Changsha City, China (associated with Institute of Particle Physics, Central China Normal University, Wuhan, Hubei, China)
[72]Guangdong Provincial Key Laboratory of Nuclear Science, Guangdong-Hong Kong Joint Laboratory of Quantum Matter, Institute of Quantum Matter, South China Normal University, Guangzhou, China (associated with Center for High Energy Physics, Tsinghua University, Beijing, China)
[73]School of Physics and Technology, Wuhan University, Wuhan, China (associated with Center for High Energy Physics, Tsinghua University, Beijing, China)
[74]Departamento de Fisica, Universidad Nacional de Colombia, Bogota, Colombia (associated with LPNHE, Sorbonne Université, Paris Diderot Sorbonne Paris Cité, CNRS/IN2P3, Paris, France)
[75]Universität Bonn—Helmholtz-Institut für Strahlen und Kernphysik, Bonn, Germany (associated with Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany)
[76]Institut für Physik, Universität Rostock, Rostock, Germany (associated with Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany)
[77]Eotvos Lorand University, Budapest, Hungary (associated with European Organization for Nuclear Research (CERN), Geneva, Switzerland)
[78]INFN Sezione di Perugia, Perugia, Italy (associated with INFN Sezione di Ferrara, Ferrara, Italy)
[79]Van Swinderen Institute, University of Groningen, Groningen, Netherlands (associated with Nikhef National Institute for Subatomic Physics, Amsterdam, Netherlands)
[80]Universiteit Maastricht, Maastricht, Netherlands (associated with Nikhef National Institute for Subatomic Physics, Amsterdam, Netherlands)

[81]*National Research Centre Kurchatov Institute, Moscow, Russia (associated with Institute of Theoretical and Experimental Physics NRC Kurchatov Institute (ITEP NRC KI), Moscow, Russia)*
[82]*National Research University Higher School of Economics, Moscow, Russia (associated with Yandex School of Data Analysis, Moscow, Russia)*
[83]*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)*
[84]*National Research Tomsk Polytechnic University, Tomsk, Russia (associated with Institute of Theoretical and Experimental Physics NRC Kurchatov Institute (ITEP NRC KI), Moscow, Russia)*
[85]*DS4DS, La Salle, Universitat Ramon Llull, Barcelona, Spain (associated with ICCUB, Universitat de Barcelona, Barcelona, Spain)*
[86]*University of Michigan, Ann Arbor, USA (associated with Syracuse University, Syracuse, New York, USA)*

[a]Also at Università di Genova, Genova, Italy.
[b]Also at Università di Modena e Reggio Emilia, Modena, Italy.
[c]Also at Università di Ferrara, Ferrara, Italy.
[d]Also at Università di Milano Bicocca, Milano, Italy.
[e]Also at Università di Bologna, Bologna, Italy.
[f]Also at Università di Bari, Bari, Italy.
[g]Also at Università di Cagliari, Cagliari, Italy.
[h]Also at Novosibirsk State University, Novosibirsk, Russia.
[i]Also at Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden.
[j]Also at Università di Roma Tor Vergata, Roma, Italy.
[k]Also at Universidade Federal do Triângulo Mineiro (UFTM), Uberaba-MG, Brazil.
[l]Also at Hangzhou Institute for Advanced Study, UCAS, Hangzhou, China.
[m]Also at AGH—University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Kraków, Poland.
[n]Also at Università di Siena, Siena, Italy.
[o]Also at Università di Padova, Padova, Italy.
[p]Also at Scuola Normale Superiore, Pisa, Italy.
[q]Also at Università degli Studi di Milano, Milano, Italy.
[r]Also at MSU—Iligan Institute of Technology (MSU-IIT), Iligan, Philippines.
[s]Also at Università di Firenze, Firenze, Italy.
[t]Also at Hanoi University of Science, Hanoi, Vietnam.
[u]Also at P.N. Lebedev Physical Institute, Russian Academy of Science (LPI RAS), Moscow, Russia.
[v]Also at Università di Pisa, Pisa, Italy.
[w]Also at Università della Basilicata, Potenza, Italy.
[x]Also at Università di Urbino, Urbino, Italy.