Searches for rare $B^0_s$ and $B^0$ decays into four muons

The LHCb collaboration

E-mail: cliff@hep.phy.cam.ac.uk

ABSTRACT: Searches for rare $B^0_s$ and $B^0$ decays into four muons are performed using proton-proton collision data recorded by the LHCb experiment, corresponding to an integrated luminosity of $9\text{ fb}^{-1}$. Direct decays and decays via light scalar and $J/\psi$ resonances are considered. No evidence for the six decays searched for is found and upper limits at the 95% confidence level on their branching fractions ranging between $1.8 \times 10^{-10}$ and $2.6 \times 10^{-9}$ are set.

KEYWORDS: B Physics, Branching fraction, Flavour Changing Neutral Currents, Hadron-Hadron Scattering, Rare Decay

ArXiv ePrint: 2111.11339

This paper is dedicated to the memory of our friend and colleague Alexei Vorobyev.
1 Introduction

Decays of neutral $B^0_s$ and $B^0$ mesons into four muons that are not mediated by intermediate resonances proceed through $b \to s$ and $b \to d$ quark flavour-changing neutral current (FCNC) transitions. These decays proceed by electroweak loop amplitudes in the Standard Model (SM), as illustrated in figure 1, due to the absence of tree-level FCNCs and hence are highly suppressed, with predicted branching fractions of $\mathcal{B}(B^0_s \to \mu^+\mu^-\mu^+\mu^-) = (0.9 - 1.0) \times 10^{-10}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-\mu^+\mu^-) = (0.4 - 4.0) \times 10^{-12}$ [1].

However, new particles beyond the Standard Model (BSM) may significantly enhance these branching fractions. For example, decays via scalar and pseudoscalar sgoldstino particles into a pair of dimuons in the Minimal Supersymmetric Standard Model (MSSM) may lead to significant enhancements of the branching fractions [2]. Furthermore, rare $B^0_s$ and $B^0$ decays into a pair of dimuons mediated by BSM light narrow scalar resonances (a) of the form $B^0_{(s)} \to a (\mu^+\mu^-)$ naturally occur in extensions of the SM involving a new strongly interacting sector. In particular, such models [3, 4] can account for the long-standing tension between the SM prediction [5] and the observed [6, 7] value of the anomalous magnetic dipole moment of the muon, as well as the widely discussed anomalies in $b \to s\ell^+\ell^-$ transitions [8–14]. This motivates a search for $B$ decays into two light scalars with masses around 1 GeV/$c^2$ [15].

This article presents searches for $B^0_{(s)} \to \mu^+\mu^-\mu^+\mu^-$ decays using proton-proton ($pp$) collision data recorded by the LHCb experiment in 2011, 2012 (Run 1) and 2015-2018
Figure 1. Feynman diagrams for resonant $B_{s,0} \rightarrow \mu^+ \mu^- \mu^+ \mu^-$ decays.

(Run 2) at centre-of-mass energies of 7, 8 and 13 TeV, respectively, corresponding to a total integrated luminosity of 9 fb$^{-1}$. Previously, the most sensitive searches for $B_{s,0} \rightarrow \mu^+ \mu^- \mu^+ \mu^-$ decays were performed by the LHCb collaboration using data recorded during Run 1 corresponding to an integrated luminosity of 3 fb$^{-1}$. No signals were observed and the limits $\mathcal{B}(B_{s,0} \rightarrow \mu^+ \mu^- \mu^+ \mu^-) < 2.5 \times 10^{-9}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^-) < 6.9 \times 10^{-10}$ were set [16]. These searches were insensitive to $B_{s,0} \rightarrow a(\mu^+ \mu^-) a(\mu^+ \mu^-)$ decays with $m_a \approx 1$ GeV/$c^2$ due to requirements imposed to remove decays via $\phi$ mesons. Now, a specific selection is used that avoids vetoing the dimuon mass region around 1 GeV/$c^2$.

In addition, searches are performed for $B_{s,0} \rightarrow J/\psi (\mu^+ \mu^-) \mu^+ \mu^-$ decays, which are tree-level $b \rightarrow c$ transitions that proceed via $W$ boson exchange, as demonstrated in figure 2. No SM predictions for these branching fractions are currently available. However, an estimate for the $B_{s,0}^0$ decay may be obtained using a prediction for the branching fraction of the $B_{s,0}^0 \rightarrow J/\psi e^+ e^-$ decay [17], the measured $J/\psi \rightarrow \mu^+ \mu^-$ branching fraction [18], and assuming that the $B_{s,0}^0 \rightarrow J/\psi \mu^+ \mu^-$ and $B^0_{s,0} \rightarrow J/\psi e^+ e^-$ branching fractions are equal, yielding $\mathcal{B}(B_{s,0}^0 \rightarrow J/\psi (\mu^+ \mu^-) \mu^+ \mu^-) \sim 10^{-11}$. The equivalent $B^0$ branching fraction may be estimated by multiplying this value by $|V_{cb}V_{cs}| / |V_{cb}V_{cd}| \approx 0.05$, leading to $\mathcal{B}(B^0 \rightarrow J/\psi (\mu^+ \mu^-) \mu^+ \mu^-) \sim 10^{-13}$. Since these branching fractions are significantly below the sensitivity of this analysis, the presence of a signal would indicate BSM physics.

The branching fractions are measured using $B^0_{s,0} \rightarrow J/\psi (\mu^+ \mu^-) \phi (\mu^+ \mu^-)$ decays as a normalisation channel, which has a measured branching fraction of
(1.74 ± 0.14) × 10^{-8} [18, 19]. Since this decay has the same final state as the signal modes, many sources of systematic uncertainty cancel. First, the yield of the $B_{s}^{0} \to J/\psi (\mu^{+}\mu^{-}) \phi (\mu^{+}\mu^{-})$ control mode is estimated using an unbinned maximum-likelihood fit to its invariant mass spectrum. Then, searches for $B_{s}^{0} \to \mu^{+}\mu^{-}\mu^{+}\mu^{-}$ and $B_{s}^{0} \to J/\psi (\mu^{+}\mu^{-}) \mu^{+}\mu^{-}$ decays are made using unbinned maximum-likelihood fits to the invariant mass spectra of each decay. These are performed simultaneously in multiple intervals of the response of a multivariate classifier constructed to separate signal and combinatorial background, where combinatorial background arises from muons that do not all originate from the same $b$-hadron decay. In contrast, $B_{s}^{0} \to a (\mu^{+}\mu^{-}) a (\mu^{+}\mu^{-})$ decays are searched for by imposing a minimum requirement on the multivariate classifier response due to the very low levels of background, followed by a single fit to the mass spectrum of each signal mode. For all six signal modes, the branching fractions are normalised using the yield of the control mode, its measured branching fraction, and the ratio of efficiencies between the signal and control modes calculated using simulated events. In the case of $B_{s}^{0}$ decays, the ratio of fragmentation fractions $f_{s}/f_{d}$ is also taken into account.

2 Detector and simulation

The LHCb detector is a single-arm forward spectrometer covering the pseudorapidity range $2 < \eta < 5$, described in detail in refs. [20, 21]. It includes a high-precision tracking system consisting of a silicon-strip vertex locator (VELO), 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. Particle identification is provided by two ring-imaging Cherenkov (RICH) detectors, an electromagnetic and a hadronic calorimeter, and a muon system composed of alternating layers of iron and multiwire proportional chambers.

In the simulation, $pp$ collisions are generated using PYTHIA [22, 23] with a specific LHCb configuration [24]. Decays of unstable particles are described by EvtGen [25], in which final-state radiation is generated using PHOTOS [26]. The interaction of the generated particles with the detector, and its response, are implemented using the GEANT4 toolkit [27, 28] as described in ref. [29]. Simulated $B_{s}^{0} \to \mu^{+}\mu^{-}\mu^{+}\mu^{-}$ decays are generated using a phase-space model for the kinematics of the muons, due to the lack of a SM
prediction for the decay dynamics. The \(B_s^0 \rightarrow a (\mu^+\mu^-) a (\mu^+\mu^-)\) decays are simulated with an a state with \(m_a = 1\text{ GeV}/c^2\), a lifetime of 1 fs and a natural width of zero. With these settings it is assumed that the a natural width is significantly below the experimental resolution of approximately 20 MeV/c^2. Simulated \(B_s^0 \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-\) decays are generated using the BTOSLLBALL model [30].

3 Event selection

The online selection is composed of a hardware trigger, followed by a two-stage software trigger. Candidate \(B^0_{(s)} \rightarrow \mu^+\mu^-\mu^+\mu^-\) decays are required to pass triggers designed to select decays involving muons, which are identified from tracks that penetrate the calorimeters and the iron layers between the muon stations. In the hardware stage, candidates must pass at least one of two selections: one requiring the presence of at least one high transverse momentum (\(p_T\)) muon, the other requiring a pair of muons with a product of their respective transverse momenta above a threshold. In the first software trigger stage, candidates must include a high-\(p_T\) muon that is inconsistent with having originated at a primary \(pp\) interaction vertex (PV). In the second stage, requirements are imposed on pairs of muons in order to select candidates consistent with \(B\) meson decay vertices that are significantly displaced from the PV.

Candidate \(B^0_{(s)} \rightarrow \mu^+\mu^-\mu^+\mu^-\) decays are formed by combining four tracks identified as muons that originate from a common decay vertex. The muons forming the candidate are all required to be inconsistent with originating from a PV, have a \(p_T > 250\text{ MeV}/c\), and a good track fit quality. The maximum distance of closest approach between the four tracks is required to be below 0.3 mm. The resulting \(B\) candidates are required to have an invariant mass within 1 GeV/c^2 of the known \(B^0_s\) mass [18] in the Run 1 dataset or larger than 4 GeV/c^2 in the Run 2 dataset, a good quality vertex fit, to be consistent with originating from a PV, have a significant flight distance, and the cosine of the angle between its momentum vector and the vector pointing from the PV to its decay vertex greater than zero. Hadronic background is suppressed using particle identification (PID) information provided by the muon system, calorimeters and RICH detectors, which is used to select well identified muons [31].

The different signal modes and the control channel are separated using requirements on the four invariant masses of pairs of oppositely charged muons \(q_{ij}\), where \(i\) and \(j\) index the positively and negatively charged muons, respectively, such that \(ij\) can take the values \{11, 12, 21, 22\}. Regions around the \(\phi, J/\psi\) and \(\psi(2S)\) mesons are defined as \(950\text{ MeV}/c^2 < q_{ij} < 1990\text{ MeV}/c^2\), \(|q_{ij} - m(J/\psi)| < 100\text{ MeV}/c^2\) and \(|q_{ij} - m(\psi(2S))| < 100\text{ MeV}/c^2\), respectively, where \(m(J/\psi)\) and \(m(\psi(2S))\) are the known \(J/\psi\) and \(\psi(2S)\) masses [18]. The \(B^0_{(s)} \rightarrow \mu^+\mu^-\mu^+\mu^-\) signal selection requires that none of the four pairs of opposite-sign muons has an invariant mass within any of the \(\phi, J/\psi\) or \(\psi(2S)\) regions. This requirement is approximately 64% efficient on simulated \(B^0_s \rightarrow \mu^+\mu^-\mu^+\mu^-\) candidates passing the trigger and offline selection, while rejecting 99.94% of the \(B^0_s \rightarrow J/\psi (\mu^+\mu^-) \phi (\mu^+\mu^-)\) decays. The control mode selection requires that at least one of the mutually exclusive pairs of dimuons (i.e. \(\mu^+_1\mu^-_1\) and \(\mu^+_2\mu^-_2\) or \(\mu^+_1\mu^-_2\)
and \( \mu_1^+ \mu_1^- \) has one dimuon mass in the \( \phi \) region while the other falls in the \( J/\psi \) region. This selection is around 95\% efficient on simulated \( B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (\mu^+ \mu^-) \) decays.

Candidate \( B_s^0 (s) \rightarrow a (\mu^+ \mu^-) a (\mu^+ \mu^-) \) decays are selected by requiring two mutually exclusive pairs of opposite-sign muons with similar invariant masses, satisfying

\[
|q_{ij}^2 - q_{kl}^2| < 2\sqrt{\sigma^2(q_{ij}^2) + \sigma^2(q_{kl}^2)}, \tag{3.1}
\]

where \( \sigma(q_{ij}^2) \) is the dimuon invariant mass squared resolution evaluated at \( q_{ij}^2 \). The resolution increases roughly linearly as a function of \( q^2 \), from around 0.003 GeV\(^2\) to around 0.15 GeV\(^2\), and is estimated as a function of \( q^2 \) separately in each data-taking year using simulated \( B_s^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^- \) decays. This requirement does not completely eliminate contamination from \( B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (\mu^+ \mu^-) \) decays, which may satisfy the requirement if muons from the intermediate \( J/\psi \) and \( \phi \) mesons are paired incorrectly, giving two intermediate combinations with similar masses. This remaining background is vetoed by requiring that none of the four pairs of oppositely charged muons has an invariant mass in the \( J/\psi \) region.

The \( B_s^0 (s) \rightarrow J/\psi (\mu^+ \mu^-) \mu^+ \mu^- \) candidates are selected by requiring that at least one of the pairs of opposite-sign muons has a mass in the \( J/\psi \) region. Meanwhile the corresponding opposite-sign pair of muons is required to have a mass that falls outside the \( \phi \) region under both a dimuon and dikaon mass hypothesis, in order to remove background from \( B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (\mu^+ \mu^-) \) and \( B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (K^+ K^-) \) decays, respectively. To suppress additional hadronic background such as \( B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \pi^+ \pi^- \) and \( B^0 \rightarrow J/\psi (\mu^+ \mu^-) K^{*0} (K^+ \pi^-) \) decays, the two muons not forming the \( J/\psi \) candidate are required to satisfy more stringent PID criteria.

Finally, a multivariate classifier based on a boosted decision tree (BDT) algorithm from the TMVA package \[32\] is used to separate signal from combinatorial background. The classifier is trained using a mixture of simulated \( B_s^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^- \) and \( B^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^- \) decays as the signal proxy, while the background proxy comprises data with an invariant \( B \) mass, \( m(\mu^+ \mu^- \mu^+ \mu^-) \), in the regions \( m(\mu^+ \mu^- \mu^+ \mu^-) > 5426 \text{ MeV}/c^2 \) and \( m(\mu^+ \mu^- \mu^+ \mu^-) < 5020 \text{ MeV}/c^2 \). The simulated \( B_s^0 (s) \rightarrow \mu^+ \mu^- \mu^+ \mu^- \) decays used in the training are corrected to improve agreement with data as described in section 5. Two separate BDT classifiers are trained for the Run 1 and Run 2 datasets due to the different data-taking conditions. The size of the available training sample is maximised in each case using the k-folding technique \[33\], which ensures that the BDT used to classify a given event in the data background sample is trained independent of the event itself. The following properties of the \( B \) candidate are taken as inputs to the classifiers: the logarithm of its impact parameter \( \chi^2 \) with respect to the associated PV, the logarithm of its flight distance \( \chi^2 \), its flight distance, pseudorapidity, \( p_T \), decay time, decay vertex \( \chi^2 \) per degree of freedom, and the minimum impact parameter \( \chi^2 \) with respect to the associated PV of the four muons. Here, the impact parameter \( \chi^2 \) is defined as the difference in the vertex-fit \( \chi^2 \) of a given PV reconstructed with and without the track or candidate being considered. In Run 2, two variables reflecting the isolation of the \( B \) candidate from other tracks in the event are also included in the training: the changes in the decay vertex \( \chi^2 \) when either one or two of the closest additional tracks in the underlying event are added to the decay vertex.
fit. Finally, the classifier response is transformed to give a uniform distribution between 0 and 1 for simulated $B^0_s \rightarrow \mu^+\mu^-\mu^+\mu^-$ candidates.

In the searches for $B^0_{(s)} \rightarrow \mu^+\mu^-\mu^+\mu^-$ and $B_0^0 \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ decays the data are split into four intervals in the BDT classifier response with boundaries 0.2, 0.3, 0.5, 0.6 and 1.0, and a simultaneous fit to all four BDT intervals is used to search for the decays. Due to the low background levels in the $B^0_{(s)} \rightarrow a (\mu^+\mu^-) a (\mu^+\mu^-)$ search region, the BDT response in this case is only required to be greater than 0.1. Similarly, the selection for $B^0_s \rightarrow J/\psi (\mu^+\mu^-) \phi (\mu^+\mu^-)$ decays requires the BDT classifier response to exceed 0.05.

4 Background

A range of background sources have the potential to contaminate the signal search regions. The four-muon decays $B^0_s \rightarrow J/\psi (\mu^+\mu^-) \phi (\mu^+\mu^-)$ and $B^0_s \rightarrow \phi (\mu^+\mu^-) \mu^+\mu^-$, which have branching fractions of $1.74 \times 10^{-8}$ [19] and $2.3 \times 10^{-10}$ [18], respectively, are reduced to negligible levels by the $\phi$ and $J/\psi$ vetoes. Potential background to the $B^0_{(s)} \rightarrow a (\mu^+\mu^-) a (\mu^+\mu^-)$ search comes from $B^0_{(s)} \rightarrow \phi (\mu^+\mu^-) \phi (\mu^+\mu^-)$ decays, however, the corresponding branching fraction of $1.5 \times 10^{-12}$ is well below the sensitivity of this search.

A second category of background arises from decays of $b$-hadrons of the form $H_b \rightarrow h^+h^-\mu^+\mu^-$ where two hadrons ($h^+$ and $h^-$) are misidentified as muons. Such background includes $B^0_s \rightarrow J/\psi (\mu^+\mu^-) \phi (K^+K^-)$, $B^0_s \rightarrow J/\psi (\mu^+\mu^-) K^+K^-$, $B^0_s \rightarrow J/\psi (\mu^+\mu^-) \pi^+\pi^-$, $B^0 \rightarrow J/\psi (\mu^+\mu^-) K^{*0} (K^+\pi^-)$, $B^0 \rightarrow K^{*0} (K^+\pi^-) \mu^+\mu^-$, $B^0 \rightarrow \rho (\pi^+\pi^-) \mu^+\mu^-$, $A^0 \rightarrow A (p\pi^-) \mu^+\mu^-$ and $A^0 \rightarrow A (pK^-) \mu^+\mu^-$ decays. The typical probability for a hadron to be misidentified as a muon by the LHCb detector is around 1% [31], and this rate is further reduced by the PID requirements described in section 3. Studies performed using simulation demonstrate that all these decays are reduced to negligible levels by the signal and the control sample selections. As such, this analysis is essentially free of background from other $b$-hadron decays, which are therefore not modelled in the fits for the signal and control decays.

5 Fit procedure

The signal yields in the branching fraction measurements are normalised with respect to the $B^0_{(s)} \rightarrow J/\psi (\mu^+\mu^-) \phi (\mu^+\mu^-)$ control mode such that the branching fraction of a given signal decay is calculated as

$$B_{\text{sig}} = \alpha_{s/d}^n \times N_{\text{sig}}^n,$$

where the normalisation factors for $B^0_s$ and $B^0$ modes, $\alpha_s^n$ and $\alpha_d^n$, are defined as

$$\alpha_s^n = \frac{B_{\text{norm}}}{N_{\text{norm}}} \cdot \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}},$$

$$\alpha_d^n = \frac{B_{\text{norm}}}{N_{\text{norm}}} \cdot \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} \cdot \frac{f_s}{f_d}.$$

The symbols sig and norm refer to the signal and normalisation decays, respectively, $B$ is the branching fraction, $N$ is the yield, $\epsilon$ is the total efficiency to reconstruct and select a given decay mode and $n$ indexes the BDT response intervals used in the simul-
taneous fits for $B_{(s)}^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow J/\psi (\mu^+ \mu^-) \mu^+ \mu^-$ decays (in the case of $B_{(s)}^0 \rightarrow a (\mu^+ \mu^-) a (\mu^+ \mu^-)$ there is only a single BDT response interval).

The branching fraction of the $B_{s}^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (\mu^+ \mu^-)$ normalisation mode is calculated from the product of the branching fractions of the $B_{s}^0 \rightarrow J/\psi \phi$ [19], $J/\psi \rightarrow \mu^+ \mu^-$ and $\phi \rightarrow \mu^+ \mu^-$ [18] decays, yielding $B_{\text{norm}} = (1.74 \pm 0.14) \times 10^{-5}$. The ratio of fragmentation fractions is calculated as a weighted average of values measured at centre-of-mass energies of 7, 8 and 13 TeV [19], giving $f_s/f_d = 0.250 \pm 0.008$. The correlation in the measurements of $B (B_{s}^0 \rightarrow J/\psi \phi)$ and $f_s/f_d$ is taken into account when calculating the uncertainty in the normalisation term $\alpha_{d}^s$.

The efficiencies are calculated using simulated decays, to which weights are applied in order to improve concordance with data. These weights are calculated by comparing $B_{(s)}^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (K^+ K^-)$ decays in data and simulation, for which the branching fraction is three orders of magnitude larger than for $B_{s}^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (\mu^+ \mu^-)$ and therefore allows for a much more precise determination of differences between data and simulation. The trigger and offline selections for these decays are similar to those used for $B_{s}^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (\mu^+ \mu^-)$ decays, but no PID requirement is applied on the kaons. The distributions of the variables of interest for $B_{s}^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (K^+ K^-)$ decays are separated from background in data using the sPlot method [34], using the $B$ candidate invariant mass as the discriminating variable.

A first set of weights, referred to as the generator weights, corrects the $p_T$ and pseudorapidity distributions of $B$ mesons as generated by PYTHIA 8 [22], along with the multiplicity of the underlying events. A second set of weights, referred to as the reconstruction weights, is used to correct the vertex $\chi^2$ per degree of freedom and the impact parameter $\chi^2$ of the $B$ mesons. The efficiencies for each mode in each data-taking year are then calculated as

$$\epsilon = \epsilon_{\text{presel}} \cdot \epsilon_{\text{BDT}} = \frac{\sum_{\text{presel}} \omega_{\text{gen}}}{\sum_{\text{gen}} \omega_{\text{gen}}} \cdot \frac{\sum_{\text{BDT}} \omega_{\text{gen}} \omega_{\text{rec}}}{\sum_{\text{presel}} \omega_{\text{gen}} \omega_{\text{rec}}},$$

(5.3)

where $\sum_{\text{presel}}$, $\sum_{\text{gen}}$ and $\sum_{\text{BDT}}$ refer to sums over the generated simulation sample before any reconstruction or selection, the offline reconstructed simulation sample passing the full selection excluding the BDT requirements, and the offline reconstructed simulation sample passing the full selection including the BDT requirements, respectively. Meanwhile, $\omega_{\text{gen}}$ and $\omega_{\text{rec}}$ refer to the generator and reconstruction weights. Note that the reconstruction weights are only used in the calculation of the efficiencies of the BDT requirements, since requirements on the $B$ candidate’s vertex $\chi^2$ per degree of freedom and the impact parameter $\chi^2$ imposed in prior stages of the selection are highly efficient for the signal modes. The individual efficiencies for each year are then combined in a weighted average according to the integrated luminosity recorded and the $b\bar{b}$ production cross-section in that year [35, 36].

The invariant mass distribution of $B_{s}^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (\mu^+ \mu^-)$ candidates in the range $5100 < m(\mu^+ \mu^- \mu^+ \mu^-) < 6000 \text{ MeV}/c^2$ is shown in figure 3. The yield of the normalisation mode is determined using an extended unbinned maximum-likelihood fit to this invariant mass spectrum. The $B_{s}^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (\mu^+ \mu^-)$ signal is modelled using the sum of two Crystal Ball functions [37] with common mean, referred to as a double Crystal Ball.
Figure 3. Distribution of the $\mu^+\mu^-\mu^+\mu^-$ invariant mass of candidates passing the $B_s^0 \rightarrow J/\psi (\mu^+\mu^-) \phi (\mu^+\mu^-)$ selection, with the fit model used to determine the control sample yield overlaid.

The parameters of this function are determined from a fit to simulated $B_s^0 \rightarrow J/\psi (\mu^+\mu^-) \phi (\mu^+\mu^-)$ decays, in which the values are fixed in the fit to data, with the exception of the mean and widths of the DCB function, which are allowed to vary freely in order to account for differences in resolution and mass scale between data and simulation. The resulting ratio of the mass resolution in data over simulation is found to be $1.11 \pm 0.09$. The combinatorial background is modelled with an exponential function. Background from the six signal decays in the control mode mass distribution is negligible due to the low branching fractions of the signal decays and the requirements on the $J/\psi$ and $\phi$ masses imposed in the control mode selection. The yield of $B_s^0 \rightarrow J/\psi (\mu^+\mu^-) \phi (\mu^+\mu^-)$ decays is found to be $218 \pm 16$, where the uncertainty is statistical only. The invariant mass spectrum and the fit projection are shown in figure 3.

As described in section 3, the data used to search for $B^0_{(s)} \rightarrow \mu^+\mu^-\mu^+\mu^-$ and $B^0_{(s)} \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ decays are split into four intervals in the BDT classifier response, while for the $B^0_{(s)} \rightarrow a (\mu^+\mu^-) a (\mu^+\mu^-)$ data sample a single BDT interval is defined. The normalisation factors $\alpha^b_i$ and $\alpha^d_i$ are calculated using eq. 5.1 and the inputs described above, with their values Gaussian constrained to their central values according to their statistical uncertainties. Additional Gaussian constraints are used to include the effects of the systematic uncertainties in the inputs described in section 6.

For each of the $B^0_{(s)} \rightarrow \mu^+\mu^-\mu^+\mu^-$ and $B^0_{(s)} \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ signal modes, an un-binned extended maximum-likelihood fit is performed to the $B$ candidate invariant mass distribution in the range $4900 < m(\mu^+\mu^-\mu^+\mu^-) < 6000 \text{ MeV}/c^2$, simultaneously in all four BDT intervals, in order to search for the decays and measure their branching fractions. In the case of the $B^0_{(s)} \rightarrow a (\mu^+\mu^-) a (\mu^+\mu^-)$ modes, fits are performed to the invariant mass distribution in a single region of the BDT response. All signal mass shapes are modelled using DCB functions, in which the parameters are fixed to values obtained from simula-
tion, with the exception of the mean and widths, which are offset and scaled respectively according to the results of the fit for the control mode. Due to their overlap, the fits to the $B_0^s$ and $B^0$ decays are performed separately in each case, i.e. the $B_0^s$ component is set to zero when performing the fit to the $B^0$ component and vice versa. The yields of each signal mode are not fitted directly, rather they are expressed in terms of the normalisation factors and the decay branching fractions (see eq. 5.1), which are allowed to vary freely. The branching fraction is constrained to be greater than or equal to zero in each fit. The combinatorial background is modelled using exponential functions, in which the slope parameters and yields are allowed to vary freely independently in each BDT interval.

6 Systematic uncertainties

Due to the similarity between the signal and control decay modes, many sources of systematic uncertainty cancel in the ratio of efficiencies. Nonetheless, a number of sources of systematic uncertainty have the potential to affect the measured branching fractions. Two sources of uncertainty arise from the models used to generate the simulated signal decays. In particular, since no theoretical prediction for the decay dynamics of $B_{(s)}^0 \to \mu^+\mu^-\mu^+\mu^-$ decays is currently available, decays are generated using a phase-space model. This may lead to a bias in the efficiency calculation if the true kinematic distributions of these decays differs from the one used in the simulation. To estimate the representative scale of the bias arising from such an effect, the dimuon invariant mass squared ($q_{ij}^2$) distributions for $B_{(s)}^0 \to \mu^+\mu^-\mu^+\mu^-$ decays are weighted to be uniform in the $(q_{11}^2,q_{22}^2)$ plane, and the efficiencies are evaluated. A similar effect is evaluated for $B_{(s)}^0 \to J/\psi (\mu^+\mu^-) \mu^+\mu^-$ decays, where the invariant mass squared of the two muons not originating from the $J/\psi$ meson is weighted to give a uniform distribution. This leads to shifts relative to the nominal efficiencies of up to about 20% depending on the decay mode and BDT interval, which are taken as systematic uncertainties. This represents the largest source of systematic uncertainty. Similarly, the effective lifetimes of the $B_0^s$ meson decays, $B_0^s \to \mu^+\mu^-\mu^+\mu^-$, $B_0^s \to a (\mu^+\mu^-) a (\mu^+\mu^-)$ and $B_0^s \to J/\psi (\mu^+\mu^-) \mu^+\mu^-$, depend on the CP-even and CP-odd contributions in the decay amplitudes, which are a priori unknown. These decays are generated assuming the mean $B_0^s$ lifetime. If their true effective lifetimes differ significantly from this value, then the efficiency estimates will be biased. The maximum size of such an effect is estimated by weighting simulated decay modes to have their effective lifetimes equal to the lifetimes of the heavy and light $B_0^s$ mass eigenstates. The largest shift with respect to the nominal efficiency is again taken as a source of systematic uncertainty, with relative shifts in the efficiencies at around the 5% level.

Further sources of uncertainty arise due to data-simulation differences. The effect of mismodelling of the PID response in simulation is evaluated using calibration samples of muons from data to compare data-driven efficiencies with those obtained directly from simulation, with the differences in the ratios of efficiencies of 1–2% taken as systematic uncertainties. The effect of the difference in mass resolution between simulation and data on the selection efficiency is estimated to be below 1% using the results of the fits to the control mode mass distribution. Finally, the efficiencies are evaluated without the
application of the generator and reconstruction weights described in section 5 in order to get an upper bound on the likely effect of mismodelling in simulation. Since the effect on the efficiency ratio is found to be small, up to around 10% depending on the decay and BDT interval, this shift is conservatively taken as a source of systematic uncertainty in the efficiency. All these systematic uncertainties are included in the branching fraction fit by applying Gaussian constraints to the efficiencies or efficiency ratios.

7 Results

The invariant mass distributions of \( B_0^0 \) and \( B^0 \) candidates passing the \( B_0^0 \to \mu^+\mu^-\mu^+\mu^- \), \( B_0^0 \to a (\mu^+\mu^-) a (\mu^+\mu^-) \) and \( B_0^0 \to J/\psi (\mu^+\mu^-) \mu^+\mu^- \) selections in the most sensitive BDT intervals are shown in figures 5, 4 and 6, respectively. The projections of the fits used to search for the signal decays and measure their branching fractions are overlaid. The distributions and fit projections in the remaining BDT intervals are shown in appendix A. No

---

Figure 4. Distribution of the \( \mu^+\mu^-\mu^+\mu^- \) invariant mass of candidates passing the \( B_0^0 \to a (\mu^+\mu^-) a (\mu^+\mu^-) \) selection, with the fit models used to determine the branching fractions of (left) \( B_0^0 \to a (\mu^+\mu^-) a (\mu^+\mu^-) \) and (right) \( B^0 \to a (\mu^+\mu^-) a (\mu^+\mu^-) \) decays overlaid.

Figure 5. Distribution of the \( \mu^+\mu^-\mu^+\mu^- \) invariant mass of candidates passing the \( B_0^0 \to \mu^+\mu^-\mu^+\mu^- \) selection in the most sensitive BDT interval, with the fit models used to determine the branching fractions of (left) \( B_0^0 \to \mu^+\mu^-\mu^+\mu^- \) and (right) \( B^0 \to \mu^+\mu^-\mu^+\mu^- \) decays overlaid.
Figure 6. Distribution of the $\mu^+\mu^-\mu^+\mu^-$ invariant mass of candidates passing the $B^0_s \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ selection in the most sensitive BDT interval, with the fit models used to determine the branching fractions of (left) $B^0_s \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ and (right) $B^0 \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ decays overlaid.

evidence for any of the six signal decay modes is found, with the most significant excesses found in the $B^0_s \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ and $B^0 \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ searches, amounting to two standard deviations in both cases, calculated using Wilks’ theorem [38]. Limits are set on the branching fractions using the CL$_s$ method [39] as implemented in the GAMMACOMBO package [40, 41] using a one-sided test statistic, with 80 scan points and 2000 pseudoexperiments per scan point. The limits at 95% confidence level are

\[
\begin{align*}
\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-\mu^+\mu^-) &< 8.6 \times 10^{-10}, \\
\mathcal{B}(B^0 \rightarrow \mu^+\mu^-\mu^+\mu^-) &< 1.8 \times 10^{-10}, \\
\mathcal{B}(B^0_s \rightarrow a(\mu^+\mu^-)a(\mu^+\mu^-)) &< 5.8 \times 10^{-10}, \\
\mathcal{B}(B^0 \rightarrow a(\mu^+\mu^-)a(\mu^+\mu^-)) &< 2.3 \times 10^{-10}, \\
\mathcal{B}(B^0_s \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-) &< 2.6 \times 10^{-9}, \\
\mathcal{B}(B^0 \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-) &< 1.0 \times 10^{-9}.
\end{align*}
\]

The corresponding CL$_s$ scans for all six decays are shown in figure 7. Note that in the case of $B^0_s \rightarrow a(\mu^+\mu^-)a(\mu^+\mu^-)$ decays, these limits are evaluated assuming a promptly decaying intermediate scalar with a mass of 1 GeV/$c^2$. The limits quoted for $B^0_s \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ decays include the $J/\psi \rightarrow \mu^+\mu^-$ branching fraction. These results constitute the most stringent limits on each of the six $B^0_s \rightarrow \mu^+\mu^-\mu^+\mu^-$ decays to date and supersede previous results by the LHCb experiment [16].
Figure 7. Confidence levels (CLs) obtained from the CLs method to set limits on the branching fractions of (top left) $B_0^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^-$, (top right) $B^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^-$, (middle left) $B_s^0 \rightarrow (\mu^+ \mu^-) a (\mu^+ \mu^-)$, (middle right) $B^0 \rightarrow a (\mu^+ \mu^-) a (\mu^+ \mu^-)$, (bottom left) $B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \mu^+ \mu^-$ and (bottom right) $B^0 \rightarrow J/\psi (\mu^+ \mu^-) \mu^+ \mu^-$ decays.
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 (U.S.A.). 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 (U.S.A.). 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).
A Additional figures

This appendix contains the distributions of $\mu^+\mu^-\mu^+\mu^-$ invariant mass for candidates passing the $B^0_{(s)} \rightarrow \mu^+\mu^-\mu^+\mu^-$ and $B^0_{(s)} \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ selections, in the lower three intervals of the BDT classifier response, with the results of the maximum-likelihood fits overlaid.

Figure 8. Distribution of the $\mu^+\mu^-\mu^+\mu^-$ invariant mass of candidates passing the $B^0_{(s)} \rightarrow \mu^+\mu^-\mu^+\mu^-$ selection in (top left) the lowest BDT interval, (top right) the second lowest BDT interval and (bottom) the second highest BDT interval, with the fit models used to determine the branching fraction of $B^0_{(s)} \rightarrow \mu^+\mu^-\mu^+\mu^-$ overlaid.
Figure 9. Distribution of the $\mu^+\mu^+\mu^-\mu^-$ invariant mass of candidates passing the $B_s^0 \rightarrow \mu^+\mu^-\mu^+\mu^-$ selection in (top left) the lowest BDT interval, (top right) the second lowest BDT interval and (bottom) the second highest BDT interval, with the fit models used to determine the branching fraction of $B_s^0 \rightarrow \mu^+\mu^-\mu^+\mu^-$ overlaid.
Figure 10. Distribution of the $\mu^+\mu^-\mu^+\mu^-$ invariant mass of candidates passing the $B^0_s \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ selection in (top left) the lowest BDT interval, (top right) the second lowest BDT interval and (bottom) the second highest BDT interval, with the fit models used to determine the branching fraction of $B^0_s \rightarrow J/\psi (\mu^+\mu^-) \mu^+\mu^-$ overlaid.
Figure 11. Distribution of the $\mu^+\mu^-\mu^+\mu^-$ invariant mass of candidates passing the $B^{0}_{(s)} \to J/\psi (\mu^+\mu^-) \mu^+\mu^-$ selection in (top left) the lowest BDT interval, (top right) the second lowest BDT interval and (bottom) the second highest BDT interval, with the fit models used to determine the branching fraction of $B^0 \to J/\psi (\mu^+\mu^-) \mu^+\mu^-$ overlaid.
Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] A.V. Danilina and N.V. Nikitin, Four-leptonic decays of charged and neutral B mesons within the standard model, Phys. Atom. Nucl. 81 (2018) 347 [Yad. Fiz. 81 (2018) 331] [inSPIRE].

[2] S.V. Demidov and D.S. Gorbunov, Flavor violating processes with sgoldstino pair production, Phys. Rev. D 85 (2012) 077701 [arXiv:1112.5230] [inSPIRE].

[3] M. Bauer, M. Neubert and A. Thamm, LHC as an axion factory: probing an axion explanation for $(g-2)_\mu$ with exotic Higgs decays, Phys. Rev. Lett. 119 (2017) 031802 [arXiv:1704.08207] [inSPIRE].

[4] J. Liu, C.E.M. Wagner and X.-P. Wang, A light complex scalar for the electron and muon anomalous magnetic moments, JHEP 03 (2019) 008 [arXiv:1810.11028] [inSPIRE].

[5] T. Aoyama et al., The anomalous magnetic moment of the muon in the Standard Model, Phys. Rept. 887 (2020) 1 [arXiv:2006.04822] [inSPIRE].

[6] Muon $g-2$ collaboration, Measurement of the positive muon anomalous magnetic moment to 0.46 ppm, Phys. Rev. Lett. 126 (2021) 141801 [arXiv:2107.13428] [inSPIRE].

[7] Muon $g-2$ collaboration, Final report of the muon ES21 anomalous magnetic moment measurement at BNL, Phys. Rev. D 73 (2006) 072003 [hep-ex/0602035] [inSPIRE].

[8] LHCb collaboration, Tests of lepton universality using $B^0 \rightarrow K^0_\ell^+\ell^-\text{ and } B^+ \rightarrow K^{*+}\ell^+\ell^-$ decays, arXiv:2110.09501 [inSPIRE].

[9] LHCb collaboration, Test of lepton universality in beauty-quark decays, arXiv:2103.11769 [inSPIRE].

[10] LHCb collaboration, Test of lepton universality with $\Lambda^0_b \rightarrow pK^-\ell^+\ell^-\text{ decays}$, JHEP 05 (2020) 040 [arXiv:1912.08139] [inSPIRE].

[11] LHCb collaboration, Test of lepton universality with $B^0 \rightarrow K^{*0}\ell^+\ell^-\text{ decays}$, JHEP 08 (2017) 055 [arXiv:1705.05802] [inSPIRE].

[12] LHCb collaboration, Angular analysis of the rare decay $B_s^0 \rightarrow \phi\mu^+\mu^-$, JHEP 11 (2021) 043 [arXiv:2107.13428] [inSPIRE].

[13] LHCb collaboration, Angular analysis of the $B^+ \rightarrow K^{*+}\mu^+\mu^-$ decay, Phys. Rev. Lett. 126 (2021) 161802 [arXiv:2012.13241] [inSPIRE].

[14] LHCb collaboration, Measurement of $C\bar{P}$-averaged observables in the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay, Phys. Rev. Lett. 125 (2020) 011802 [arXiv:2003.04831] [inSPIRE].

[15] M. Chala, U. Egede and M. Spannowsky, Searching new physics in rare $B$-meson decays into multiple muons, Eur. Phys. J. C 79 (2019) 431 [arXiv:1902.10156] [inSPIRE].

[16] LHCb collaboration, Search for decays of neutral beauty mesons into four muons, JHEP 03 (2017) 001 [arXiv:1611.07704] [inSPIRE].

[17] D.H. Evans, B. Grinstein and D.R. Nolte, Short distance analysis of $\bar{B} \rightarrow J/\psi e^+e^-$, $\bar{B} \rightarrow \eta_c e^+e^-$, $\bar{B} \rightarrow D^{0*}e^+e^-$ and $\bar{B} \rightarrow D^0 e^+e^-$, Nucl. Phys. B 577 (2000) 240 [hep-ph/9906528] [inSPIRE].
[18] Particle Data Group collaboration, *Review of particle physics*, *PTEP* 2020 (2020) 083C01 [arXiv:2005.02597] [SPIRE].

[19] LHCb collaboration, Precise measurement of the $f_s/f_d$ ratio of fragmentation fractions and of $B^0_s$ decay branching fractions, *Phys. Rev. D* 104 (2021) 032005 [arXiv:2103.06810] [SPIRE].

[20] LHCb collaboration, *The LHCb detector at the LHC*, 2008 *JINST* 3 S08005 [SPIRE].

[21] LHCb collaboration, LHCb detector performance, *Int. J. Mod. Phys. A* 30 (2015) 1530022 [arXiv:1412.6352] [SPIRE].

[22] T. Sjöstrand, S. Mrenna and P.Z. Skands, A brief introduction to PYTHIA 8.1, *Comput. Phys. Commun.* 178 (2008) 852 [arXiv:0710.3820] [SPIRE].

[23] T. Sjöstrand, S. Mrenna and P.Z. Skands, PYTHIA 6.4 physics and manual, *JHEP* 05 (2006) 026 [hep-ph/0603175] [SPIRE].

[24] LHCb collaboration, Handling of the generation of primary events in Gauss, the LHCb simulation framework, *J. Phys. Conf. Ser.* 331 (2011) 032047 [SPIRE].

[25] D.J. Lange, The EvtGen particle decay simulation package, *Nucl. Instrum. Meth. A* 462 (2001) 152 [SPIRE].

[26] N. Davidson, T. Przedzinski and Z. Was, PHOTOS interface in C++: Technical and Physics Documentation, *Comput. Phys. Commun.* 199 (2016) 86 [arXiv:1011.0937] [SPIRE].

[27] GEANT4 collaboration, GEANT4 developments and applications, *IEEE Trans. Nucl. Sci.* 53 (2006) 270.

[28] GEANT4 collaboration, GEANT4 — a simulation toolkit, *Nucl. Instrum. Meth. A* 506 (2003) 250 [SPIRE].

[29] LHCb collaboration, The LHCb simulation application, Gauss: Design, evolution and experience, *J. Phys. Conf. Ser.* 331 (2011) 032023 [SPIRE].

[30] A. Ali, P. Ball, L.T. Handoko and G. Hiller, A comparative study of the decays $B \to (K, K^*) e^+e^-$ in standard model and supersymmetric theories, *Phys. Rev. D* 61 (2000) 074024 [hep-ph/9910221] [SPIRE].

[31] F. Archilli et al., Performance of the muon identification at LHCb, 2013 *JINST* 8 P10020 [arXiv:1306.0249] [SPIRE].

[32] H. Voss, A. Hocker, J. Stelzer and F. Tegenfeldt, TMVA, the Toolkit for Multivariate Data Analysis with ROOT, *PoS* ACAT (2007) 040 [SPIRE].

[33] A. Blum, A. Kalai, and J. Langford, Beating the hold-out: bounds for k-fold and progressive cross-validation, in proceedings of the 12th annual conference on computational learning theory (COLT’99), New York, U.S.A. (1999).

[34] M. Pivk and F.R. Le Diberder, SPlot: a statistical tool to unfold data distributions, *Nucl. Instrum. Meth. A* 555 (2005) 356 [physics/0402083] [SPIRE].

[35] LHCb collaboration, Measurement of the $b$-quark production cross-section in 7 and 13 TeV $pp$ collisions, *Phys. Rev. Lett.* 118 (2017) 052002 [Erratum ibid. 119 (2017) 169901] [arXiv:1612.05140] [SPIRE].

[36] LHCb collaboration, Measurement of differential $b\bar{b}$ and $c\bar{c}$-dijet cross-sections in the forward region of $pp$ collisions at $\sqrt{s} = 13$ TeV, *JHEP* 02 (2021) 023 [arXiv:2010.09437] [SPIRE].
[37] T. Skwarnicki, *A study of the radiative cascade transitions between the \( \Upsilon' \) and \( \Upsilon \) resonances*, Ph.D. thesis, Institute of Nuclear Physics, Krakow, Poland (1986) [DESY-F31-86-02].

[38] S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *Annals Math. Statist.* 9 (1938) 60 [inSPIRE].

[39] A.L. Read, *Presentation of search results: the CLs technique*, *J. Phys. G* 28 (2002) 2693 [inSPIRE].

[40] M. Kenzie et al., *GammaCombo: a statistical analysis framework for combining measurements, fitting datasets and producing confidence intervals*, Zenodo.

[41] LHCb collaboration, *Measurement of the CKM angle \( \gamma \) from a combination of LHCb results*, *JHEP* 12 (2016) 087 [arXiv:1611.03076] [inSPIRE].
