Search for the rare decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$

LHCb Collaboration

**Abstract**

A search for the decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ is performed with 0.37 fb$^{-1}$ of $pp$ collisions at $\sqrt{s} = 7$ TeV collected by the LHCb experiment in 2011. The upper limits on the branching fractions are $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) < 1.6 \times 10^{-8}$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 3.6 \times 10^{-9}$ at 95% confidence level. A combination of these results with the LHCb limits obtained with the 2010 dataset leads to $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) < 1.4 \times 10^{-8}$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 3.2 \times 10^{-9}$ at 95% confidence level.

© 2012 CERN. Published by Elsevier B.V. All rights reserved.

1. Introduction

Measurements of low-energy processes can provide indirect constraints on particles that are too heavy to be produced directly. This is particularly true for Flavour Changing Neutral Current (FCNC) processes which are highly suppressed in the Standard Model (SM) and can only occur through higher-order diagrams. The SM predictions for the branching fractions of the FCNC decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ are $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.2 \pm 0.2) \times 10^{-9}$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (0.10 \pm 0.01) \times 10^{-9}$ [1]. However, contributions from new processes or new heavy particles can significantly enhance these values. For example, within Minimal Supersymmetric extensions of the SM (MSSM), in the large $\tan\beta$ regime, $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ is found to be approximately proportional to $\tan^6\beta$ [2], where $\tan\beta$ is the ratio of the vacuum expectation values of the two neutral CP-even Higgs fields. The branching fractions could therefore be enhanced by orders of magnitude for large values of $\tan\beta$.

The best published limits from the Tevatron are $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) < 5.1 \times 10^{-8}$ at 95% confidence level (CL) by the D$0$ Collaboration using 6.1 fb$^{-1}$ of data [3], and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 6.0 \times 10^{-9}$ at 95% CL by the CDF Collaboration using 6.9 fb$^{-1}$ of data [4]. In the same dataset the CDF Collaboration observes an excess of $B_s^0 \rightarrow \mu^+\mu^-$ candidates compatible with $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (1.8^{+1.1}_{-0.9}) \times 10^{-8}$ and with an upper limit of $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) < 4.0 \times 10^{-8}$ at 95% CL. The CMS Collaboration has recently published $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) < 1.9 \times 10^{-8}$ at 95% CL and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 4.6 \times 10^{-9}$ at 95% CL using 1.14 fb$^{-1}$ of data [5]. The LHCb Collaboration has published the limits [6] $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) < 5.4 \times 10^{-8}$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 1.5 \times 10^{-8}$ at 95% CL based on about 37 pb$^{-1}$ of integrated luminosity collected in the 2010 run.

This Letter presents an analysis of the data recorded by LHCb in the first half of 2011 which correspond to an integrated luminosity of $\sim 0.37$ fb$^{-1}$. The results of this analysis are then combined with those published from the 2010 dataset.

2. The LHCb detector

The LHCb detector [7] is a single-arm forward spectrometer designed to study production and decays of hadrons containing $b$ or $c$ quarks. The detector consists of a vertex locator (VELO) providing precise locations of primary $pp$ interaction vertices and detached vertices of long lived hadrons.

The momenta of charged particles are determined using information from the VELO together with the rest of the tracking system, composed of a large area silicon tracker located before a warm dipole magnet with a bending power of $\sim 4$ Tm, and a combination of silicon strip detectors and straw drift chambers located after the magnet. Two Ring Imaging Cherenkov (RICH) detectors are used for charged hadron identification in the momentum range 2–100 GeV/c. Photon, electron and hadron candidates are identified by electromagnetic and hadronic calorimeters. A muon system of alternating layers of iron and drift chambers provides muon identification. The two calorimeters and the muon system provide the energy and momentum information to implement a first level (L0) hardware trigger. An additional trigger level (HLT) is software...
based, and its algorithms are tuned to the experimental operating condition.

Events with a muon final states are triggered using two L0 trigger
decisions: the single-muon decision, which requires one muon
candidate with a transverse momentum $p_T$ larger than 1.5 GeV/c,
and the di-muon decision, which requires two muon candidates
with transverse momenta $p_{T1}$ and $p_{T2}$ satisfying the relation
$\sqrt{p_{T1}^2 - p_{T2}^2} > 1.3$ GeV/c.

The single muon trigger decision in the second trigger level
(HLT) includes a cut on the impact parameter (IP) with respect to
the primary vertex, which allows for a lower $p_T$ requirement ($p_T >$
1.0 GeV/c, IP > 0.1 mm). The di-muon trigger decision requires
muon pairs of opposite charge with $p_T > 500$ MeV/c, forming a
common vertex and with an invariant mass $m_{\mu\mu} > 4.7$ GeV/c$^2$.
A second trigger decision, primarily to select $J/\psi$ events, requires
2.97 $< m_{\mu\mu} < 3.21$ GeV/c$^2$. A remaining region of the di-muon
invariant mass range is also covered by trigger decisions that in ad-
dition require the di-muon secondary vertex to be well separated
from the primary vertex.

Events with purely hadronic final states are triggered by the
L0 trigger if there is a calorimeter cluster with transverse energy
$E_T > 3.6$ GeV. Other HLT trigger decisions select generic displaced
vertices, providing high efficiency for purely hadronic decays.

3. Analysis strategy

Assuming the branching fractions predicted by the SM, and using
the $b\bar{b}$ cross-section measured by LHCb in the pseudorapidity
interval $2 < \eta < 6$ and integrated over all transverse momenta of
$\sigma_{bg} = 75 \pm 14$ mb [8], approximately 3.9 $B^0 \rightarrow \mu^+\mu^-$ and 0.4
$B^0 \rightarrow \tau^+\tau^-$ events are expected to be triggered, reconstructed
and selected in the analyzed sample embedded in a large back-
ground.

The general structure of the analysis is based upon the one
described in Ref. [6]. First a very efficient selection removes
the biggest amount of background while keeping most of the sig-
nal within the LHCb acceptance. The number of observed events
is compared to the number of expected signal and background
events in bins of two independent variables, the invariant mass
and the output of a multi-variate discriminant. The discriminant is
a Boosted Decision Tree (BDT) constructed using the TMVA pack-
age [9]. It supersedes the Geometrical Likelihood (GL) used in the
previous analysis [6] as it has been found more performant in dis-
criminating between signal and background events in simulated
samples. No data were used in the choice of the multivariate dis-
criminant in order not to bias the result.

The combination of variables entering the BDT discriminant is
optimized using simulated events. The probability for a signal or
background event to have a given value of the BDT output is ob-
tained from data using $B^{(*)} ightarrow J/\psi K^+$ candidates (where $h^+$ can
be a pion or a kaon) as signal and sidebands $B^{(*)} ightarrow \mu^+\mu^-$ candidates
as background.

The invariant mass line shape of the signal is described by a
Crystal Ball function [10] with parameters extracted from data
control samples. The central values of the masses are obtained
from $B^0 \rightarrow K^+\pi^-$ and $B^{(*)} \rightarrow K^+\pi^-$ samples. The $B^0$ and $B^\pm$ mass
resolutions are estimated by interpolating those obtained with
di-muon resonances ($J/\psi$, $\psi(2S)$ and $\Upsilon(1S, 2S, 3S)$) and cross-
checked with a fit to the invariant mass distributions of both in-
clusive $B^{(*)} \rightarrow J/\psi K^+$ and exclusive $B^0 \rightarrow K^+\pi^-$ decays.
The central values of the masses and the mass resolution are used
to define the signal regions.

The number of expected signal events, for a given branch-
ing fraction hypothesis, is obtained by normalizing to channels
of known branching fractions: $B^+ \rightarrow J/\psi K^+$, $B^{0} \rightarrow J/\psi \phi$ and
$B^0 \rightarrow K^+\pi^-$. These channels are selected in a way similar as
possible to the signals in order to minimize the systematic uncer-
tainty related to the different phase space accessible to each final
state.

The BDT output and invariant mass distributions for combina-
torial background events in the signal regions are obtained using fits
of the mass distribution of events in the mass sidebands in bins of the
BDT output.

The two-dimensional space formed by the invariant mass and
the BDT output is binned. For each bin we count the number of
candidates observed in the data, and compute the expected number of
signal events and the expected number of background events. The binning is unchanged with respect to the 2010 analy-
sis [6]. The compatibility of the observed distribution of events in
all bins with the distribution expected for a given branching frac-
tion hypothesis is computed using the CLs method [11], which al-
lows a given hypothesis to be excluded at a given confidence level.

4. Selection

The $B^{(*)} \rightarrow \mu^+\mu^-$ selections require two muon candidates of
opposite charge. Tracks are required to be of good quality and to be
displaced with respect to any primary vertex. The secondary vertex
is required to be well fitted ($\chi^2$/NDOF < 9) and must be separated
from the primary vertex in the forward direction by a distance
of flight significance ($L/sigma(L)$) greater than 15. When more than
one primary vertex is reconstructed, the one that gives the mini-

The reconstructed candidate has to point to this primary vertex
(IP/\sigma(IP) < 5).

Improvements have been made to the selection developed for
2010 data [6]. The RICH is used to identify kaons in the $B^{0} \rightarrow
J/\psi \phi$ normalization channel and the Kullback-Leibler (KL) dis-
tance [12] is used to suppress duplicated tracks created by the
reconstruction. This procedure compares the parameters and cor-
relation matrices of the reconstructed tracks and where two are
found to be similar, in this case with a symmetrized KL divergence
less than 5000, only the one with the higher track fit quality is
considered.

The inclusive $B^{(*)} \rightarrow h^+h^-$ sample is the main control sam-
ple for the determination from data of the probability distribution
function (PDF) of the BDT output. This sample is selected in exactly
the same way as the $B^{(*)} \rightarrow \mu^+\mu^-$ signals apart from the muon
identification requirement. The same selection is also applied to
the $B^0 \rightarrow K^+\pi^-$ normalization channel.

The muon identification efficiency is uniform within ~1% in
the considered phase space therefore no correction is added to the
BDT PDF extracted from the $B^{(*)} \rightarrow h^+h^-$ sample. The remaining
phase space dependence of the muon identification efficiency is
instead taken into account in the computation of the normalization
factor when the $B^{(*)} \rightarrow K^+\pi^-$ channel is considered.

The $J/\psi \rightarrow \mu^+\mu^-$ decay in the $B^{(*)} \rightarrow J/\psi K^+$ and $B^{(*)} \rightarrow J/\psi \phi$
normalization channels is selected in a very similar way to the
$B^{(*)} \rightarrow \mu^+\mu^-$ channels, apart from the pointing requirement.
$K^{\pm}$ candidates are required to be identified by the RICH detector
and to pass track quality and impact parameter cuts.

To avoid pathological events, all tracks from selected candi-
dates are required to have a momentum less than 1 TeV/c. Only
B candidates with decay times less than 5 $\tau_{B^{(*)}}$, where $\tau_{B^{(*)}}$ is the
B lifetime [13], are accepted for further analysis. Di-muon candi-
dates coming from elastic di-photon production are removed by
requiring a minimum transverse momentum of the B candidate of
500 MeV/c.
three new variables are: the two muons with respect to any other track in the event. The distance of closest approach between the two muons and the isolation of signal candidates while is peaked at zero for $IP_{\text{b}}$, the impact parameter, transverse momentum, the minimum impact parameter significance ($IP/\sigma(\text{IP})$) of the muons, the distance of closest approach between the two muons and the isolation of the two muons with respect to any other track in the event. The three new variables are:

1. the minimum $p_T$ of the two muons;
2. the cosine of the angle between the muon momentum in the $B$ rest frame and the vector perpendicular to the $B$ momentum and the beam axis:

$$\cos P = \frac{p_{y,\mu_1} p_{x,\mu_2} - p_{x,\mu_1} p_{y,\mu_2}}{p_{r,\mu_2}/(m_{\mu_1}/2)}$$

where $\mu_1$ labels one of the muons and $m_{\mu_1}$ is the reconstructed $B$ candidate mass$^2$;
3. the $B$ isolation [14]

$$I_B = \frac{p_{r}(B)}{p_{r}(B) + \sum_i p_{r,i}}$$

where $p_{r}(B)$ is the $B$ transverse momentum with respect to the beam line and the sum is over all the tracks, excluding the muon candidates, that satisfy $\sqrt{\delta \eta^2 + \delta \phi^2} < 1.0$, where $\delta \eta$ and $\delta \phi$ denote respectively the difference in pseudorapidity and azimuthal angle between the track and the $B$ candidate.

The BDT output is found to be independent of the invariant mass for both signal and background and is defined such that the signal is uniformly distributed between zero and one and the background peaks at zero. The BDT range is then divided in four bins of equal width. The BDT is trained using simulated samples ($B^0_{(d)} \rightarrow \mu^+ \mu^-$ for signals and $b\bar{b} \rightarrow \mu^+ \mu^- X$ for background where $X$ is any other set of particles) and the PDF obtained from data as explained below.

5.1. Combinatorial background PDFs

The BDT and invariant mass shapes for the combinatorial background inside the signal regions are determined from data by interpolating the number of expected events using the invariant mass sidebands for each BDT bin. The boundaries of the signal regions are defined as $m_{B^0} \pm 60 \text{MeV}/c^2$ and $m_{B^0} \pm 600 \text{MeV}/c^2$ and the mass sidebands as $[m_{B^0} - 600 \text{MeV}/c^2, m_{B^0} - 60 \text{MeV}/c^2]$ and $[m_{B^0} + 60 \text{MeV}/c^2, m_{B^0} + 600 \text{MeV}/c^2]$.

Fig. 1 shows the invariant mass distribution for events that lie in each BDT output bin. In each case the fit model used to estimate the expected number of combinatorial background events in the signal regions is superimposed. Aside from combinatorial background, the low-mass sideband is potentially polluted by two other contributions: cascading $b \rightarrow c\mu\nu \rightarrow \mu\mu X$ decays below $4900 \text{MeV}/c^2$ and peaking background from $B_{(d)}^{0} \rightarrow h^+h^-$ candidates with the two hadrons misidentified as muons above $5000 \text{MeV}/c^2$. To avoid these contaminations, the number of expected combinatorial background events is obtained by fitting a single exponential function to the events in the reduced low-mass sideband $[4900, 5000] \text{MeV}/c^2$ and in the full

![Fig. 1. Distribution of the $\mu^+ \mu^-$ invariant mass for events in each BDT output bin. The curve shows the model used to fit the sidebands and extract the expected number of combinatorial background events in the $B^{0}_{(d)}$ and $B^0$ signal regions, delimited by the vertical dotted orange and dashed green lines respectively. Only events in the region in which the line is solid have been considered in the fit. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this Letter.)](image-url)
high-mass sideband. As a cross-check, two other models, a single exponential function and the sum of two exponential functions, have been used to fit the events in different ranges of sidebands providing consistent background estimates inside the signal regions.

5.2. Peaking background PDFs

The peaking backgrounds due to \( B_0 \to h^+h^- \) events in which both hadrons are misidentified as muons have been evaluated from data and simulated events to be \( N_{gh} = 5.0 \pm 0.9 \) events and \( N_{gh} = 1.0 \pm 0.4 \) events and \( N_B = 0.0 \pm 0.9 \) events within the two mass windows and in the whole BDT output range. The mass line shape of the peaking background is obtained from a simulated sample of doubly-misidentified \( B_0 \to h^+h^- \) events and normalized to the number of events expected in the two search windows from data, \( N_{gh} \) and \( N_{gh} \). The BDT PDF of the peaking background is assumed to be the same as for the signal.

5.3. Signal PDFs

The BDT PDF for signal events is determined using an inclusive \( B_0 \to h^+h^- \) sample. Only events which are triggered independently on the signal candidates have been considered (TIS events).

The number of \( B_0 \to h^+h^- \) signal events in each BDT output bin is determined by fitting the \( hh \) invariant mass distribution under the \( \mu\mu \) mass hypothesis [15]. Fig. 2 shows the fit to the mass distribution of the full sample and for the three highest BDT output bins for \( B_0 \to h^+h^- \) TIS events. The \( B_0 \to h^+h^- \) exclusive decays, the combinatorial background and the physical background components are drawn under the fit to the data; the physical background is due to the partial reconstruction of three-body \( B \) meson decays.

In order to cross-check this result, two other fits have been performed on the same dataset. The signal line shape is parametrized either by a single or a double Crystal Ball function [10], the combinatorial background by an exponential function and the physical background by an ARGUS function [16]. In addition, exclusive \( B_0 \to \pi^-K^+, \pi^+K^-, K^-K^+ \) channels, selected using the \( K^-\pi^+ \) separation capability of the RICH system, are used to cross-check the calibration of the BDT output both using the \( \pi^-K^+, \pi^-\pi^+ \), \( K^-K^+ \) inclusive yields without separating \( B \) and \( B_0 \) and using the \( B_0 \to K^+\pi^- \) exclusive channel alone. The maximum spread in the fractional yield obtained among the different models has been used as a systematic uncertainty in the signal BDT PDF. The BDT PDFs for signals and combinatorial background are shown in Fig. 3.

The invariant mass shape for the signal is parametrized as a Crystal Ball function. The mean value is determined using the \( B_0 \to K^+\pi^- \) and \( B_0 \to K^+\pi^- \) exclusive channels and the transition point of the radiative tail is obtained from simulated events [6]. The central values are

\[
\begin{align*}
    m_{B0} &= 5358.0 \pm 1.0 \text{ MeV}/c^2, \\
    m_{B0} &= 5272.0 \pm 1.0 \text{ MeV}/c^2.
\end{align*}
\]

The measured values of \( m_{B0} \) and \( m_{B0} \) are 7–8 MeV/c^2 below the PDG values [13] due to the fact that the momentum scale is uncalibrated in the dataset used in this analysis. The mass resolutions are extracted from data with a linear interpolation between the measured resolution of charmonium and bottomonium resonances decaying into two muons: \( J/\psi, \psi(2S), \Upsilon(15), \Upsilon(2S), \Upsilon(2S) \) and \( \Upsilon(3S) \). The mass line shapes for quarkonium resonances are

Fig. 2. Invariant mass distributions of \( B_0 \to h^+h^- \) candidates in the \( \mu^+\mu^- \) mass hypothesis for the whole sample (top left) and for the samples in the three highest bins of the BDT output (top right, bottom left, bottom right). The \( B_0 \to h^+h^- \) exclusive decays, the combinatorial background and the physical background components are drawn under the fit to the data (solid blue line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this Letter.)
shown in Fig. 4. Each resonance is fitted with two Crystal Ball functions with common mean value and common resolution but different parameterization of the tails. The background is fitted with an exponential function.

The results of the interpolation at the $m_{h\phi}$ and $m_{h\gamma}$ masses are

$$\sigma(m_{h\phi}) = 24.6 \pm 0.2_{\text{stat}} \pm 1.0_{\text{syst}} \text{ MeV/c}^2,$$

$$\sigma(m_{h\gamma}) = 24.3 \pm 0.2_{\text{stat}} \pm 1.0_{\text{syst}} \text{ MeV/c}^2.$$

This result has been checked using both the fits to the $B_{(s)}^0 \rightarrow h^+h^-$ inclusive decay line shape and the $B^0 \rightarrow K^+\pi^-$ exclusive decay. The results are in agreement within the uncertainties.

6. Normalization

To estimate the signal branching fraction, the number of observed signal events is normalized to the number of events of a channel with a well-known branching fraction. Three complementary normalization channels are used: $B^+ \rightarrow J/\psi(K^+\mu^-\mu^+)$, $B_s^0 \rightarrow J/\psi(\mu^-\mu^+\phi(K^+K^-))$ and $B^0 \rightarrow K^+\pi^-$. The first two channels have similar trigger and muon identification efficiencies to the signal but different number of particles in the final state. The third channel has a similar topology but is selected by different trigger lines.

The numbers of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ candidates are translated into a branching fractions ($\mathcal{B}$) using the equation

$$\mathcal{B} = \mathcal{B}_{\text{norm}} \times \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{norm}}} \times \frac{f_{\text{norm}}}{f_{d(s)}} \times \frac{N_{B_{(s)}^0 \rightarrow \mu^+\mu^-}}{N_{\text{norm}}}$$

where $f_{d(s)}$ and $f_{\text{norm}}$ are the probabilities that a $b$ quark fragments into a $B_{(s)}^0$ and into the $b$ hadron involved for the chosen normalization mode. LHCb has measured $f_s/f_d = 0.26_{-0.02}^{+0.01}$ [17]. $\mathcal{B}_{\text{norm}}$ is the branching fraction and $N_{\text{norm}}$ is the number of selected events of the normalization channel. The efficiency is the product of three factors: $\epsilon_{\text{REC}}$ is the reconstruction efficiency of all the final state particles of the decay including the geometric acceptance of the detector; $\epsilon_{\text{SEL}}$ is the selection efficiency for reconstructed events; $\epsilon_{\text{TRIG}}$ is the trigger efficiency for reconstructed and selected events. The subscript (sig,norm) indicates whether the efficiency refers to the signal or the normalization channel. Finally, $\alpha_{\text{norm}}^B_{B_{(s)}^0 \rightarrow \mu^+\mu^-}$ is the normalization factor (or single event sensitivity) and $N_{B_{(s)}^0 \rightarrow \mu^+\mu^-}$ the number of observed signal events.

For each normalization channel $N_{\text{norm}}$ is obtained from a fit to the invariant mass distribution. The invariant mass distributions for reconstructed $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ candidates are shown in Fig. 5, while the $B^0 \rightarrow K^+\pi^-$ yield is obtained from the full $B_{(s)}^0 \rightarrow h^+h^-$ fit as shown in the top left of Fig. 2.

The numbers used to calculate the normalization factors are summarized in Table 1. A weighted average of the three normalization channels, assuming the tracking and trigger uncertainties to be correlated between the two $J/\psi$ normalization channels and the uncertainty on $f_s/f_d$ to be correlated between the $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+\pi^-$, gives

$$\alpha_{\text{norm}}^B_{B_{(s)}^0 \rightarrow \mu^+\mu^-} = (8.38 \pm 0.74) \times 10^{-10},$$

$$\alpha_{\text{norm}}^B_{B^0 \rightarrow \mu^+\mu^-} = (2.20 \pm 0.11) \times 10^{-10}.$$

These normalization factors are used to determine the limits.

7. Results

The results for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ are summarized in Tables 2 and 3 respectively and in each of the bins the expected number of combinatorial background, peaking background, signal events, with the SM prediction assumed, is shown together with the observations on the data. The uncertainties in the signal and
Table 1
Summary of the quantities and their uncertainties required to calculate the normalization factors \(a_{\text{norm}}^{B^0 \rightarrow \mu^+\mu^-}\) for the three normalization channels considered. The branching fractions are taken from Refs. [13,18]. The trigger efficiency and the number of \(B^0 \rightarrow K^+\pi^-\) candidates correspond to TIS events.

| Channel | \(|B| \times 10^{-3}\) | Trigger eff. | \(N_{\text{norm}}\) | \(a_{\text{norm}}^{B^0 \rightarrow \mu^+\mu^-}\) |
|---------|----------------|---------------|----------------|--------------------------------|
| \(B^+ \rightarrow J/\psi K^+\) | \(6.01 \pm 0.21\) | \(0.48 \pm 0.014\) | \(0.95 \pm 0.01\) | \(124518 \pm 2025\) | \(2.23 \pm 0.11\) | \(0.83 \pm 0.08\) |
| \(B^0 \rightarrow J/\psi \phi\) | \(3.4 \pm 0.9\) | \(0.24 \pm 0.014\) | \(0.95 \pm 0.01\) | \(6940 \pm 93\) | \(2.96 \pm 0.84\) | \(1.11 \pm 0.30\) |
| \(B^0 \rightarrow K^+\pi^-\) | \(1.94 \pm 0.06\) | \(0.86 \pm 0.02\) | \(0.049 \pm 0.004\) | \(4146 \pm 608\) | \(1.98 \pm 0.34\) | \(0.74 \pm 0.14\) |

Table 2
Expected combinatorial background events, expected peaking (\(B^0 \rightarrow h^+h^-\)) background events, expected signal events assuming the SM branching fraction prediction, and observed events in the \(B^0 \rightarrow \mu^+\mu^-\) search window.

<table>
<thead>
<tr>
<th>Invariant mass [MeV/c^2]</th>
<th>BDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25 – 0.5</td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td></td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>575.5 (\pm 6.5)</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.126 (\pm 0.017)</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.059 (\pm 0.022)</td>
</tr>
<tr>
<td>Observed</td>
<td>531</td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>566.8 (\pm 6.3)</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.052 (\pm 0.018)</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.205 (\pm 0.074)</td>
</tr>
<tr>
<td>Observed</td>
<td>525</td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>558.2 (\pm 6.1)</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.024 (\pm 0.012)</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.38 (\pm 0.14)</td>
</tr>
<tr>
<td>Observed</td>
<td>561</td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>549.8 (\pm 6.0)</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.0145 (\pm 0.0014)</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.36 (\pm 0.14)</td>
</tr>
<tr>
<td>Observed</td>
<td>515</td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>541.5 (\pm 5.8)</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.0115 (\pm 0.0017)</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.204 (\pm 0.074)</td>
</tr>
<tr>
<td>Observed</td>
<td>547</td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>533.4 (\pm 5.7)</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.0089 (\pm 0.0013)</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.058 (\pm 0.021)</td>
</tr>
<tr>
<td>Observed</td>
<td>501</td>
</tr>
</tbody>
</table>
Table 3

Expected combinatorial background events, expected peaking ($B^0_s \rightarrow h^+h^-$) background events, expected $B^0 \rightarrow \mu^+\mu^-$ signal events assuming the SM branching fraction, expected cross-feed events from $B^0 \rightarrow \mu^+\mu^-$ assuming the SM branching fraction and observed events in the $B^0 \rightarrow \mu^+\mu^-$ search window.

<table>
<thead>
<tr>
<th>Invariant mass [MeV/c^2]</th>
<th>BDT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0–0.25</td>
</tr>
<tr>
<td>5212–5232</td>
<td></td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>614.2±7.5</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.203±0.034</td>
</tr>
<tr>
<td>Cross-feed</td>
<td>0.0056±0.0012</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.0070±0.0027</td>
</tr>
<tr>
<td>Observed</td>
<td>554</td>
</tr>
<tr>
<td>5232–5252</td>
<td></td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>605.0±7.2</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.281±0.049</td>
</tr>
<tr>
<td>Cross-feed</td>
<td>0.0071±0.0026</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.024±0.0086</td>
</tr>
<tr>
<td>Observed</td>
<td>556</td>
</tr>
<tr>
<td>5252–5272</td>
<td></td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>595.9±7.0</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.323±0.061</td>
</tr>
<tr>
<td>Cross-feed</td>
<td>0.0097±0.0036</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.045±0.016</td>
</tr>
<tr>
<td>Observed</td>
<td>588</td>
</tr>
<tr>
<td>5272–5292</td>
<td></td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>586.9±6.7</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.252±0.058</td>
</tr>
<tr>
<td>Cross-feed</td>
<td>0.015±0.005</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.045±0.016</td>
</tr>
<tr>
<td>Observed</td>
<td>616</td>
</tr>
<tr>
<td>5292–5312</td>
<td></td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>578.1±6.6</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.124±0.023</td>
</tr>
<tr>
<td>Cross-feed</td>
<td>0.038±0.014</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.024±0.0086</td>
</tr>
<tr>
<td>Observed</td>
<td>549</td>
</tr>
<tr>
<td>5312–5332</td>
<td></td>
</tr>
<tr>
<td>Expected comb. bkg</td>
<td>569.3±6.5</td>
</tr>
<tr>
<td>Expected peak. bkg</td>
<td>0.047±0.012</td>
</tr>
<tr>
<td>Cross-feed</td>
<td>0.149±0.054</td>
</tr>
<tr>
<td>Expected signal</td>
<td>0.006±0.0026</td>
</tr>
<tr>
<td>Observed</td>
<td>509</td>
</tr>
</tbody>
</table>

Fig. 6. Distribution of selected di-muon events in the invariant mass–BDT plane. The orange short-dashed (green long-dashed) lines indicate the ±60 MeV/c^2 search window around the mean $B^0_s$ ($B^0$) mass. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this Letter.)

The two-dimensional (mass, BDT) distribution of selected events can be seen in Fig. 6. The distribution of the invariant mass in the four BDT bins is shown in Fig. 7 for $B^0_s \rightarrow \mu^+\mu^-$ and in Fig. 8 for $B^0 \rightarrow \mu^+\mu^-$ selected candidates.

The compatibility of the distribution of events inside the search window in the invariant mass–BDT plane with a given branching fraction hypothesis is evaluated using the CL_s method [11]. This method provides three estimators: CL_{s+b}, a measure of the compatibility of the observed distribution with the signal and background hypotheses, CL_{b}, a measure of the compatibility with the background-only hypothesis and CL_{s}, a measure of the compatibility of the observed distribution with the signal and background hypotheses normalized to the background-only hypothesis.

The expected CL_s values are shown in Fig. 9 for $B^0_s \rightarrow \mu^+\mu^-$ and for $B^0 \rightarrow \mu^+\mu^-$ as dashed black lines under the hypothesis that background and SM events are observed. The shaded areas cover the region of ±1σ of compatible observations. The observed values of CL_s, as a function of the assumed branching ratio is shown as dotted blue lines on both plots.

The expected limits and the measured limits for $B^0_s \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ at 90% and 95% CL are shown in Tables 4 and 5, respectively. For the $B^0_s \rightarrow \mu^+\mu^-$ decay, the expected limits are computed allowing the presence of $B^0_s \rightarrow \mu^+\mu^-$ events according to the SM branching fraction. For the $B^0 \rightarrow \mu^+\mu^-$ decay the expected limit is computed in the background-only hypothesis and also allowing the presence of $B^0 \rightarrow \mu^+\mu^-$ events with the SM rate: the two results are identical. In the determination of the
Fig. 7. Distribution of selected di-muon events in the $B^0_s \rightarrow \mu^+\mu^-$ mass window for the four BDT output bins. The black dots are data, the light grey histogram shows the contribution of the combinatorial background, the black filled histogram shows the contribution of the $B^0_s \rightarrow h^+h^-$ background and the dark grey filled histogram the contribution of $B^0_s \rightarrow \mu^+\mu^-$ signal events according to the SM rate. The hatched area depicts the uncertainty on the sum of the expected contributions.

Fig. 8. Distribution of selected di-muon events in the $B^0 \rightarrow \mu^+\mu^-$ mass window for the four BDT output bins. The black dots are data, the light grey histogram shows the contribution of the combinatorial background, the black filled histogram shows the contribution of the $B^0_s \rightarrow h^+h^-$ background and the dark grey filled histogram shows the cross-feed of $B^0 \rightarrow \mu^+\mu^-$ events in the $B^0$ mass window assuming the SM rate. The hatched area depicts the uncertainty on the sum of the expected contributions.
are shown in Table 4 for the 2010 and observed limits for 90% and 95% CL for the combined results.

The observed CLs values are shown in the same tables. The expected background distribution results in a p-value (comparison of the observed distribution of events with the expected background distribution) of 5% for the $B^0 \rightarrow \mu^+ \mu^-$ decay, the probability that the observed events are compatible with the sum of expected background events and signal events according to the SM rate is measured by $1 - \text{CL}_{s+b}$ and it is 33%.

The result obtained in 2011 with 0.37 fb$^{-1}$ has been combined with the published result based on $\sim 37$ pb$^{-1}$ [6]. The expected and observed limits for 90% and 95% CL for the combined results are shown in Table 4 for the $B^0 \rightarrow \mu^+ \mu^-$ decay and in Table 5 for the $B^0_\text{s} \rightarrow \mu^+ \mu^-$ decay.

8. Conclusions

With 0.37 fb$^{-1}$ of integrated luminosity, a search for the rare decays $B^0_\text{s} \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ has been performed and sensitivities better than the existing limits have been obtained.

The observed events in the $B^0_\text{s}$ and in the $B^0$ mass windows are compatible with the background expectations at 5% and 32% confidence level, respectively. For the $B^0_\text{s} \rightarrow \mu^+ \mu^-$ decay, the probability that the observed events are compatible with the sum of expected background events and signal events according to the SM rate is 33%. The upper limits for the branching fractions are evaluated to be

$$B(B^0_\text{s} \rightarrow \mu^+ \mu^-) < 1.3 (1.6) \times 10^{-8} \text{ at } 90\% (95\%) \text{ CL},$$

$$B(B^0 \rightarrow \mu^+ \mu^-) < 3.0 (3.6) \times 10^{-9} \text{ at } 90\% (95\%) \text{ CL}.$$  

The $B(B^0_\text{s} \rightarrow \mu^+ \mu^-)$ and $B(B^0 \rightarrow \mu^+ \mu^-)$ upper limits have been combined with those published previously by LHCb [6] and the results are

$$B(B^0_\text{s} \rightarrow \mu^+ \mu^-)(2010 + 2011) < 1.2 (1.4) \times 10^{-8} \text{ at } 90\% (95\%) \text{ CL},$$

$$B(B^0 \rightarrow \mu^+ \mu^-)(2010 + 2011) < 2.6 (3.2) \times 10^{-9} \text{ at } 90\% (95\%) \text{ CL}.$$  

The above 90% (95%) CL upper limits are still about 3.8 (4.4) times the SM branching fractions for the $B^0_\text{s}$ and 26 (32) times for the $B^0$. These results represent the best upper limits to date.

Acknowledgements

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 CERN and at the LHCb institutes, and acknowledge support from the National Agencies: CAPES, CNPq, FAPERJ and FINEP (Brazil); CERN; NSFC (China); CNRS/IN2P3 (France); BMBF, DFG, HGF and MPG (Germany); SFI (Ireland); INFN (Italy); FOM and NWO (The Netherlands); SCRR (Poland); ANCS (Romania); MNES Russia and Rosatom (Russia); MICINN, XuntaGal and GENCAT (Spain); SNSF and SER (Switzerland); NAS Ukraine (Ukraine); STFC (United Kingdom); NSF (USA). We also acknowledge the support received from the ERC under FP7 and the Region Auvergne.

Open access

This article is published Open Access at sciencedirect.com. It is distributed under the terms of the Creative Commons Attribution License 3.0, which permits unrestricted use, distribution, and
reproduction in any medium, provided the original authors and source are credited.

References


The LHCb Collaboration
Hanoi University of Science, Hanoi, Viet Nam.

Associated member.

Associated to Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.