

New results on the search for $B_s^0 \rightarrow \mu^+ \mu^-$ from LHCb

Johannes Albrecht^{1,2,a} on behalf of the LHCb collaboration

¹CERN, Geneva, Switzerland

²TU Dortmund, Dortmund, Germany

Abstract. A search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ is performed with the LHCb experiment using 1.1 fb^{-1} of data collected at $\sqrt{s} = 8 \text{ TeV}$ and 1.0 fb^{-1} of data collected at $\sqrt{s} = 7 \text{ TeV}$. An excess of $B_s^0 \rightarrow \mu^+ \mu^-$ candidates with respect to the background expectations is observed with a statistical significance of 3.5 standard deviations. A branching fraction of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.2_{-1.2}^{+1.5}) \times 10^{-9}$ is measured with an unbinned maximum likelihood fit. The measured branching fraction is in agreement with the expectation from the Standard Model. The observed number of $B^0 \rightarrow \mu^+ \mu^-$ candidates is consistent with the background expectation and an upper limit on the branching fraction of $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 9.4 \times 10^{-10}$ is obtained.

1 Introduction

One of the most important goals of the LHCb experiment at the LHC is to search for phenomena that cannot be explained by the Standard Model (SM) of particle physics. Precise measurements of the branching fractions of the two Flavour Changing Neutral Current (FCNC) decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ belong to the most promising of these searches. Both decays are strongly suppressed by loop and helicity factors, making the SM branching fraction small [1]

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.23 \pm 0.27) \times 10^{-9} \text{ and } (1)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (0.11 \pm 0.01) \times 10^{-9}. \quad (2)$$

These theoretical predictions are the CP -averaged branching fractions. As pointed out in Ref. [2], the finite width difference of the B_s^0 system needs to be considered. The time integrated branching fraction is evaluated to be

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{SM,t} = 3.4 \times 10^{-9}, \quad (3)$$

for which the SM prediction (Eq. 1) and the LHCb measurement of the width difference $\Delta\Gamma_s$ [3] are used. This is the expected value which is to be compared with an experimental measurement.

Enhancements of the branching fractions of these decays are predicted in a variety of different extensions of the Standard Model, an overview is given in Ref. [4]. In one popular example, the Minimal Supersymmetric Standard Model (MSSM), the enhancement is proportional to $\tan^6 \beta$, where $\tan \beta$ is the ratio of the vacuum expectation values of the two Higgs fields. For large values of $\tan \beta$, this search belongs to the most sensitive probes for physics beyond the SM which can be performed at collider experiments. A review of the experimental status of the searches for $B_{s,d}^0 \rightarrow \mu^+ \mu^-$ can be found in [5].

^ae-mail: albrecht@cern.ch

2 Dataset and analysis strategy

The measurements [6] presented here use data recorded by the LHCb experiment: 1 fb^{-1} , recorded in 2011 at a center of mass energy of $\sqrt{s} = 7 \text{ TeV}$, combined with 1.1 fb^{-1} of data recorded in 2012 at $\sqrt{s} = 8 \text{ TeV}$. The first part of the dataset has already been analyzed [7] and was used to produce the lowest published limit on the decay rate of both decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$.

Two main improvements have been implemented over the previous analysis: the use of particle identification to select $B_{(s)}^0 \rightarrow h^+ h'^-$ decays which are used to calibrate the geometrical and kinematic variables, and a refined estimate of the exclusive backgrounds. The updated estimate of the exclusive backgrounds is also applied to the 2011 data and the results reevaluated. The results obtained with the combined 2011 and 2012 data sets supersede those of Ref. [7].

Candidate $B_{s,d}^0 \rightarrow \mu^+ \mu^-$ events are required to be selected by a hardware and a subsequent software trigger [8], dominantly by single and dimuon lines. The first step of the final analysis is a simple selection, which removes the dominant part of the background and keeps about 60% of the reconstructed signal events. A second selection step, based on a Boosted Decision Tree (BDT) reduces 80% of the remaining background while retaining 92% of the signal. More details on the selections are given in Ref. [6]

2.1 Signal discrimination

Each event is then given a probability to be signal or background in a two-dimensional probability space defined by the dimuon invariant mass and a multivariate discriminant operator. This likelihood combines kinematic and topological variables of the $B_{(s)}^0$ decay using a BDT. The BDT is defined and trained on simulated events for both signal

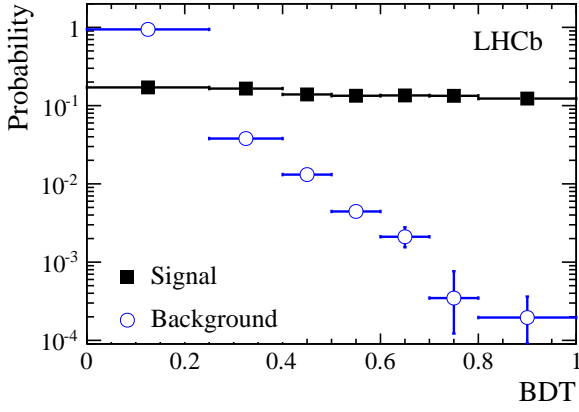


Figure 1. BDT distribution for the 2021 dataset, for the signal (black squares) and combinatorial background (blue open points). Values normalized to the bin size.

and background. The signal BDT shape is then calibrated using decays of the type $B_{(s)}^0 \rightarrow h^+h'^-$, where h^\pm represents a K^\pm or π^\pm . These decays have an identical topology to the signal. The calibrated BDT signal and background shape is shown in Fig. 1. It is designed to be flat in the signal, whereas the shape in the background falls over four orders of magnitude.

The invariant mass line shape of the signal events is described by a Crystal Ball function [9]. The mass resolution is calibrated with a combination of two methods: an interpolation of J/ψ , $\psi(2S)$ and $\Upsilon(1S)$, $\Upsilon(2S)$ and $\Upsilon(3S)$ decays to two muons and from exclusive $B_{(s)}^0 \rightarrow h^+h'^-$ samples. The results are $\sigma_{B_s^0} = 25.0 \pm 0.4 \text{ MeV}/c^2$ and $\sigma_{B^0} = 24.6 \pm 0.4 \text{ MeV}/c^2$. The transition point of the radiative tail is obtained from simulated $B_s^0 \rightarrow \mu^+\mu^-$ events smeared to reproduce the mass resolution measured in the data.

The background shapes are calibrated simultaneously in the mass and the BDT using the invariant mass sidebands. This procedure ensures that even though the BDT is defined using simulated events, the result will not be biased by discrepancies between data and simulation.

2.2 Binning

The binning of the BDT and invariant mass distributions is optimized using simulation, to maximize the separation between the median of the test statistic distribution expected for background and SM $B_s^0 \rightarrow \mu^+\mu^-$ signal, and that expected for background only. The chosen number and size of the bins are a compromise between maximizing the number of bins and the necessity to have enough $B_{(s)}^0 \rightarrow h^+h'^-$ events to calibrate the $B_s^0 \rightarrow \mu^+\mu^-$ BDT and enough background in the mass sidebands.

3 Normalization

The number of expected signal events is evaluated by normalizing with channels of known branching fraction.

Two independent channels are used: $B^- \rightarrow J/\psi K^-$ and $B^0 \rightarrow K^+\pi^-$. The first decay has similar trigger and muon identification efficiency to the signal but a different number of particles in the final state, while the second channel has the same two-body topology but is selected with a hadronic trigger. The event selection for these channels is specifically designed to be as close as possible to the signal selection. The normalization for $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ is then given as

$$\begin{aligned} \mathcal{B}(B_{s,d}^0 \rightarrow \mu^+\mu^-) &= \mathcal{B}_{\text{norm}} \times \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} \times \frac{f_{\text{norm}}}{f_{d(s)}} \times \frac{N_{B_{s,d}^0 \rightarrow \mu^+\mu^-}}{N_{\text{norm}}} \\ &= \alpha_{B_{s,d}^0}^{\text{norm}} \times N_{B_{s,d}^0 \rightarrow \mu^+\mu^-}, \end{aligned} \quad (4)$$

where $f_{d(s)}$ and f_{norm} are the probabilities that a b quark fragments into a $B_{(s)}^0$ and into the hadron involved in the given normalization mode respectively. The recently updated value $f_s/f_d = 0.256 \pm 0.020$ [10] is used. $\mathcal{B}_{\text{norm}}$ indicates the branching fraction and N_{norm} the number of signal events in the normalization channel obtained from a fit to the invariant mass distribution. The efficiency $\epsilon_{\text{sig(norm)}}$ for the signal (normalization channel) is the product of the reconstruction efficiency of all the final state particles of

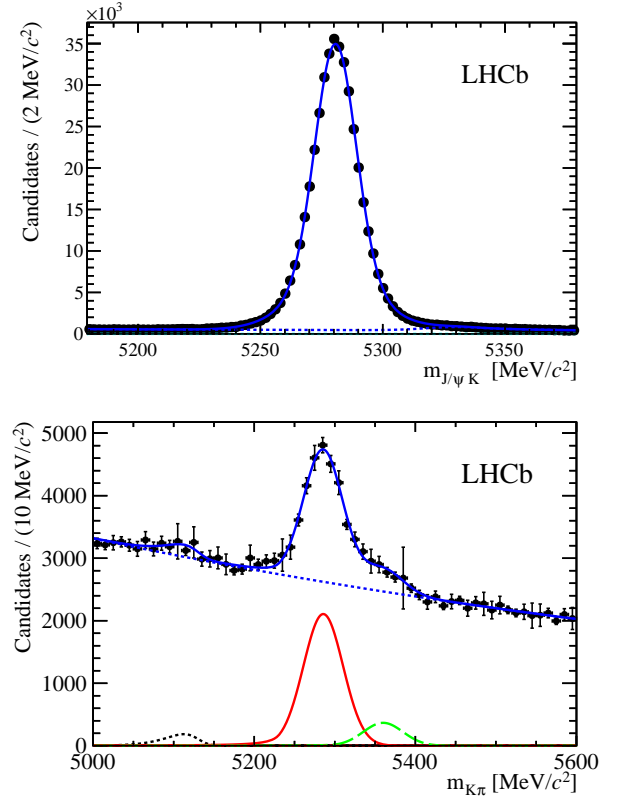


Figure 2. Invariant mass distribution of $B^- \rightarrow J/\psi K^-$ (top) and $B^0 \rightarrow K^+\pi^-$ (bottom) candidates. The fit to data is superimposed in blue while the blue dotted line is the combinatorial background. In the fit to the $B^0 \rightarrow K^+\pi^-$ data, the $B^0 \rightarrow K^+\pi^-$ component (red line), the $B_s^0 \rightarrow K^+\pi^-$ (green dashed line), and the partially reconstructed background (black dotted line) are also shown.

the decay including the geometric acceptance of the detector, the selection efficiency for reconstructed events, and the trigger efficiency for reconstructed and selected events. $N_{B_{s,d}^0 \rightarrow \mu^+ \mu^-}$ is the number of observed signal events. The ratios of reconstruction and selection efficiencies are estimated from the simulation, while the ratios of trigger efficiencies on selected events are determined from data.

The fit to the two normalization channels is shown in Fig. 2. The observed numbers of $B^- \rightarrow J/\psi K^-$ and $B^0 \rightarrow K^+ \pi^-$ candidates are $424\,222 \pm 1\,452$ and $14\,579 \pm 1\,110$. The two normalization factors are in agreement within the uncertainties and their weighted average, taking correlations into account, is

$$\begin{aligned} \alpha_{B_s^0}^{\text{norm}} &= (2.80 \pm 0.25) \times 10^{-10} \text{ and} \\ \alpha_{B^0}^{\text{norm}} &= (7.16 \pm 0.34) \times 10^{-11}, \end{aligned} \quad (5)$$

for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ candidates inside a signal window of $\pm 60 \text{ MeV}/c^2$ around the mass central value. These normalization factors are used for the limit computation. The normalization factors in the full mass range, used in the fit for the branching fraction, are 10% lower.

4 Background characterization

Partially reconstructed decays of beauty mesons or baryons can pollute the low mass sidebands. The dominant modes are:

- $B^0 \rightarrow \pi^- \mu^+ \bar{\nu}$,
- $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ and $B^+ \rightarrow \pi^+ \mu^+ \mu^-$
- $B_{(s)}^0 \rightarrow h^+ h'^-$ (with $h^{(\prime)} = K, \pi$) and
- $\Lambda_b \rightarrow p \mu^- \bar{\nu}$.

In some of these modes kaons, pions and protons are misidentified as muons. The contributions of these decays to the $B_{s,d}^0 \rightarrow \mu^+ \mu^-$ analysis is estimated from Monte Carlo simulated samples by folding the $K \rightarrow \mu$, $\pi \rightarrow \mu$ and $p \rightarrow \mu$ fake rates extracted from $D^0 \rightarrow K^+ \pi^-$ and $\Lambda \rightarrow p \pi^-$ data samples into the spectrum of simulated events. The fractional yields in BDT bins and the parameters that describe the mass lineshape are used as nuisance parameters in the unbinned maximum likelihood fit used to determine the branching fraction.

The exclusive background contribution in the most sensitive region of the analysis ($\text{BDT} > 0.8$) is dominated by the $B^0 \rightarrow \pi^- \mu^+ \bar{\nu}$, $B_{(s)}^0 \rightarrow h^+ h'^-$ and $B^{0,+} \rightarrow \pi^{0,+} \mu^+ \mu^-$ decays.

5 Results

The observed pattern of events in the 15 BDT bins (8 in the 2011 data and 7 in the 2012 data) is shown in Fig. 3 for $B_s^0 \rightarrow \mu^+ \mu^-$ (top) and $B^0 \rightarrow \mu^+ \mu^-$ (bottom) together with the fit for the branching fraction, which includes components for the exclusive background components discussed in Sec. 4.

The number of expected combinatorial background events in the B^0 and B_s^0 search windows is determined

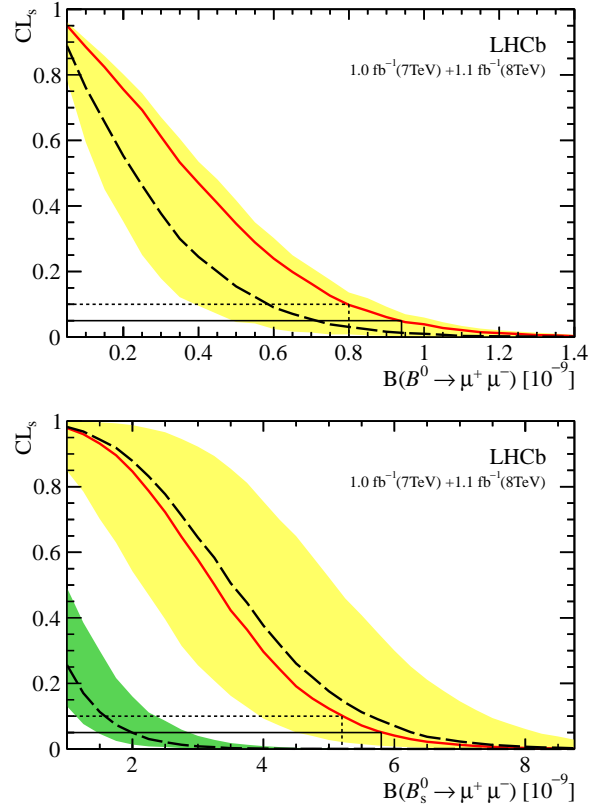


Figure 4. CLs as a function of the assumed $\mathcal{B}(f)$ or $B^0 \rightarrow \mu^+ \mu^-$ (top) and $B_s^0 \rightarrow \mu^+ \mu^-$ (bottom) decays for the combined 2011+2012 dataset. The long dashed gray curves are the medians of the expected CLs distributions if background and SM signal were observed. The yellow area covers the 1σ area around the median. The solid red curves are the observed CLs. For the $B_s^0 \rightarrow \mu^+ \mu^-$ (bottom), the long dashed gray curve in the green area is the expected CLs distribution if background only was observed with its 1σ interval.

from a simultaneous unbinned likelihood fit to the mass projections in the BDT bins. The same fit is then performed on the full mass range to extract the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ branching fractions.

In this fit the parameters that describe the mass distributions of the exclusive backgrounds, their fractional yield in each BDT bin and their overall yields are constrained to vary within $\pm 1\sigma$ with respect to the expected values. The combinatorial background is parameterized with an exponential function with a slope and a normalization which are free parameters of the fit.

The $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ signal yields are free parameters of the fit. Their fractional yields in BDT bins are constrained to the BDT fractions calibrated with the $B_{(s)}^0 \rightarrow h^+ h'^-$ sample and the parameters of the Crystal Ball functions that describe the mass lineshape are constrained to vary within $\pm 1\sigma$ with respect to the expected values.

The systematic uncertainties in the exclusive background and signal predictions in each bin are computed by fluctuating the mass parameters, the BDT fractional yields and the normalization factors along the Gaussian

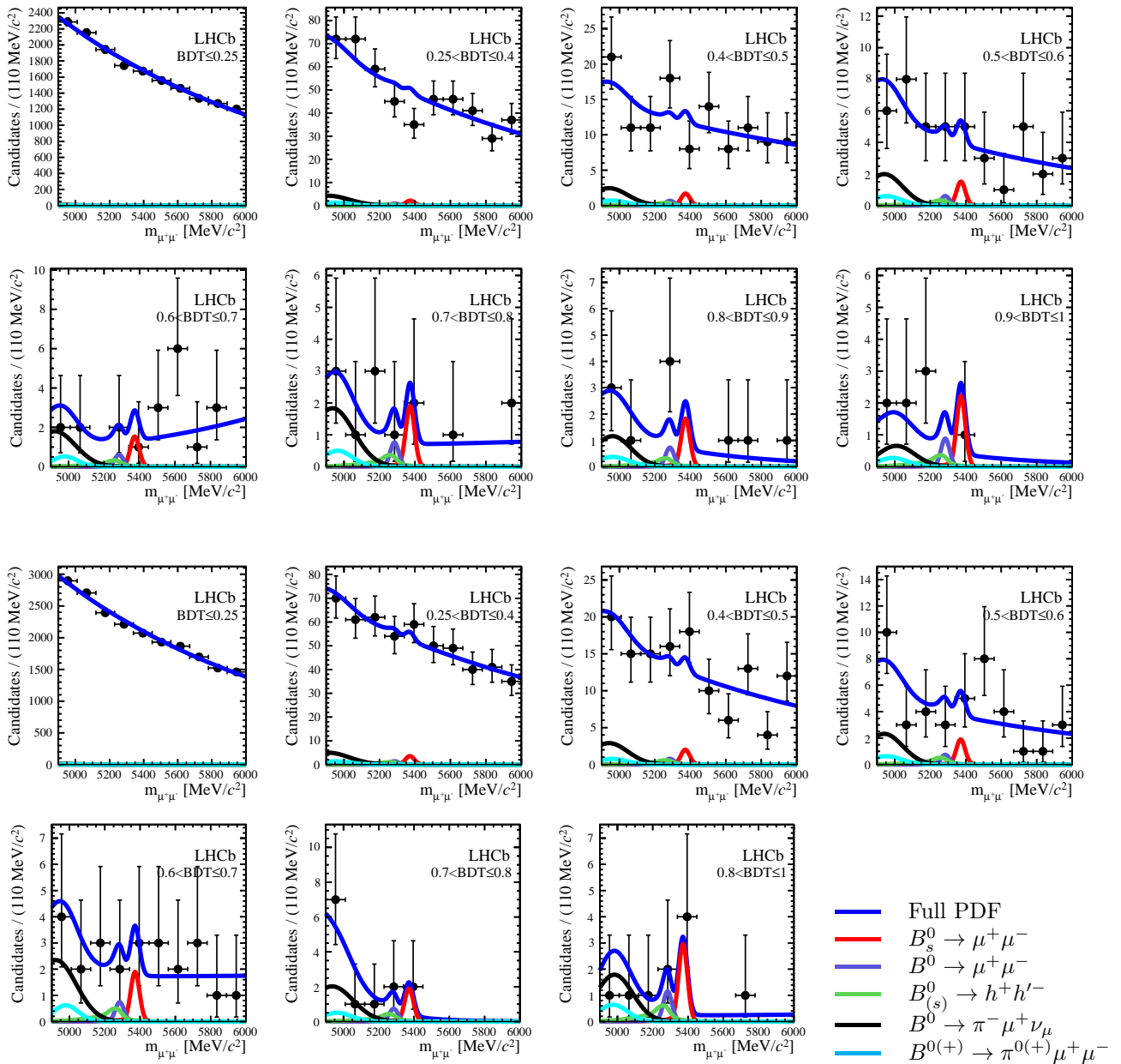


Figure 3. Simultaneous fit of the invariant mass distribution in the 8 BDT bins of 2011 (top) and 7 BDT bins of 2012 data (bottom). The fit result is superimposed in blue, the individual components are given in the legend.

distributions defined by their associated uncertainties. The systematic uncertainty on the estimated number of combinatorial background events in the search windows is computed by fluctuating with a Poissonian distribution the number of events measured in the sidebands, and by varying the value of the exponent accordingly to the its uncertainty.

The compatibility of the observed distribution of events with a given branching fraction hypothesis is computed using the CLs method [11, 12]. The pattern observed for $B^0 \rightarrow \mu^+\mu^-$ decays is compatible with the background only hypothesis. The CLs curve is shown in Fig. 4 (top). The observed CLb value at $CL_{s+b} = 0.5$ is 89%.

An excess of $B_s^0 \rightarrow \mu^+\mu^-$ candidates is observed, the CL_s curve to evaluate its significance is shown in Fig. 4 (bottom). The probability that background processes can produce the observed number of $B_s^0 \rightarrow \mu^+\mu^-$ candidates or more is 5×10^{-4} and corresponds to a statistical significance of about 3.5 standard deviations. The values of the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction extracted from the fit is $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.2_{-1.2}^{+1.5}) \times 10^{-9}$, in good agreement with the SM prediction.

6 Conclusions

A search for the rare decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ has been performed with 1.1 fb^{-1} of data collected at $\sqrt{s} = 8 \text{ TeV}$ and 1.0 fb^{-1} of data collected at $\sqrt{s} = 7 \text{ TeV}$. The data in the B^0 search window are consistent with the background expectations and an upper limit of $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 9.4 \times 10^{-10}$ is obtained at 95% CL. This is the most stringent published limit on this decay rate. The data in the B_s^0 search window show an excess of events with respect to the background expectation with a statistical significance of 3.5σ . A branching fraction of $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.2^{+1.5}_{-1.2}) \times 10^{-9}$ is measured. This is the first evidence of the $B_s^0 \rightarrow \mu^+\mu^-$ decay.

The next step is a precision measurement of the decay rate of $B_s^0 \rightarrow \mu^+\mu^-$ and then to limit and then measure the ratio of the decay rates of $B_s^0 \rightarrow \mu^+\mu^-/B^0 \rightarrow \mu^+\mu^-$. This ratio allows a stringent test of the hypothesis of minimal flavor violation and a good discrimination between various extensions of the Standard Model.

It should be stated that the precise measurement of $\mathcal{B}(B_{s,d}^0 \rightarrow \mu^+\mu^-)$ provides complementary information to the searches performed at high p_T experiments.

References

- [1] A.J. Buras, J. Girrbach, D. Guadagnoli, G. Isidori (2012), 1208.0934
- [2] K. de Bruyn, R. Fleischer, R. Kneijens, P. Koppenburg, M. Merk et al. (2012), 1204.1737
- [3] LHCb collaboration (2012), LHCb-CONF-2012-002
- [4] I. Bediaga et al. (2012), 1208.3355
- [5] J. Albrecht, Mod. Phys. Lett. A **27** (2012), 1207.4287
- [6] R. Aaij et al. (LHCb Collaboration), Phys. Rev. Lett. **110**, 021801 (2013), 1211.2674
- [7] R. Aaij et al. (LHCb collaboration), Phys. Rev. Lett. **108**, 231801 (2012), 1203.4493
- [8] R. Aaij, J. Albrecht, F. Alessio, S. Amato, E. Aslanides et al. (2012), 1211.3055
- [9] T. Skwarnicki, Institute of Nuclear Physics Krakow; Report No. DESY-F31-86-02. (1986)
- [10] R. Aaij et al. (LHCb collaboration) (2013), 1301.5286
- [11] T. Junk, Nucl.Instrum.Meth. **A434**, 435 (1999), hep-ex/9902006
- [12] A.L. Read, Journal of Physics G: Nuclear and Particle Physics **28**, 2693 (2002)