$B^0_{(s)} \to \mu^+\mu^-$ combination at the LHC

Maximilian Schlupp[1]

Department of Physics, Technische Universität Dortmund, Otto-Hahn-Str 4, 44227 Dortmund, Germany

E-mail: maximilian.schlupp@cern.ch

Abstract. The combined analysis of the B^0_s and B^0 meson decays into two muons based on the full CMS and LHCb datasets is presented. The simultaneous fit to the datasets yields the most precise measurement of the branching fractions to date with a total uncertainty of about 23% for the B^0_s meson and 38% for the B^0 decay into two muons. The measured branching fractions are compatible with the Standard Model predictions within 1.2σ and 2.2σ for the B^0_s and B^0 signals, respectively. The statistical significance for the B^0_s signal is 6.2σ and 3.2σ for the B^0 signal. This results in the first observation of B^0_s meson decays into two muons and the first evidence for the B^0 sister decay.

1. Introduction

Rare decays of beauty mesons provide a unique environment to search for deviations in the Standard Model of particle physics (SM). These highly suppressed processes are sensitive to expansions of the SM as small deviations from the SM predictions can have a significant effect. Two of the most promising examples able to test for "New Physics" models are the decays $B^0_{(s)} \to \mu^+\mu^-$, where either a B^0 or a B^0_s meson decays into two muons. The unique role of these decays arise from the suppression mechanisms. Firstly, the $B^0_{(s)} \to \mu^+\mu^-$ decays proceed through flavour changing neutral currents which are forbidden at lowest order pertubation theory. Hence, the dominant SM contributions come from weak penguin or box-diagrams. Secondly, due to the spin configuration and the special chiral nature of the weak interaction $B^0_{(s)} \to \mu^+\mu^-$ decays are helicity suppressed. The SM predictions [1] for the branching fractions for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ incorporating the latest combined value of the top quark mass from the LHC and Tevatron experiments [2] are

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.66 \pm 0.23) \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.06 \pm 0.09) \times 10^{-10}\,.$$

This picture changes if beyond SM theories are realised in nature. Especially additional (pseudo-)scalar particle that break the helicity suppression can enhance the branching fraction by a significant amount. Consequently, $B^0_{(s)} \to \mu^+\mu^-$ decays are highly sensitive to extension of the SM with an enlarged scalar sector, e.g. supersymmetric or two Higgs doublet models.
Additionally the ratio of the branching fractions of $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ provides

[1] on behalf of the LHCb and CMS collaborations

an exceptional observable of the underlaying flavour structure. This quantity can precisely be predicted within the SM as

$$\mathcal{R} \equiv \frac{\mathcal{B}(B^0 \to \mu^+\mu^-)}{\mathcal{B}(B^0_s \to \mu^+\mu^-)} = 0.0295^{+0.0028}_{-0.0025} \, .$$

In particular it discriminates the SM and the so-called minimal flavour violating (MFV) extensions from new theories with a different flavour sector.
After thirty years of searches for $B^0_{(s)} \to \mu^+\mu^-$ decays in 2012 the LHCb collaboration published the first evidence for the $B^0_s \to \mu^+\mu^-$ decay with 3.5σ significance [3] using $2\,\mathrm{fb}^{-1}$ of proton-proton collisions at center of mass energies of $\sqrt{s} = 7$ and $8\,\mathrm{TeV}$. In 2013 the LHCb and CMS experiments simultaneously reported their analyses on the complete LHC-RunI dataset of $3\,\mathrm{fb}^{-1}$ and $25\,\mathrm{fb}^{-1}$, respectively [4, 5]. Both collaborations reached approximately the same sensitivity for both decay channels, however neither could claim an observation of $B^0_s \to \mu^+\mu^-$ nor an evidence for $B^0 \to \mu^+\mu^-$. In order to calculate the combined significance of the $B^0_{(s)} \to \mu^+\mu^-$ signals at the LHC, the datasets of the LHCb and CMS experiment are combined and a joint analysis is performed. The results are presented in this proceeding and are published in Ref. [6]. More details on the individual publications can be found in Ref. [4, 5].

2. Joint CMS and LHCb analysis

This analysis presents the full combination of the individual publications realised by merging the two datasets and describing the data with a common probability density function. Hence, the two different fit models need to be harmonised. The $\Lambda^0_b \to p\mu^-\bar{\nu}$ background is added to the default background description including randomly selected muon pairs (combinatorial background), semileptonic B meson decays such as $B^0_{(s)} \to h^-\mu^+\nu$, partially reconstructed decays $B^{+,0} \to \pi^{+,0}\mu^+\mu^-$, and the peaking component of $B \to hh'$, where h could be a kaon or pion. Additionally the branching fraction prediction for $\Lambda^0_b \to p\mu^-\bar{\nu}$ is updated to the latest calculation [7, 8] and the kinematic properties of the simulated decays are changed in order to get more realistic momentum distributions for the final state muons. Finally, the lifetime bias correction [9, 10, 11] is applied to the $B^0_s \to \mu^+\mu^-$ signal descriptions.
Besides these harmonisations, the analysis strategy is unchanged. Both experiments select signal candidates with a soft preselection followed by a multivariate classifier, a boosted decision tree (BDT), to further separate signal from background. Then the invariant mass distribution of the muon pair is fit in categories of the multivariate classifier. Due to the approximately constant running conditions, the full LHCb dataset is only split into eight BDT bins. In case of CMS, the data is also categorised dependent on the year of data-taking and the detector region. The invariant mass distributions of the muon pairs are split in up to three bins of the BDT classifiers resulting in twelve total categories for the CMS dataset.
The twenty combined CMS and LHCb categories of invariant mass distributions are fit using the simultaneous extended maximum likelihood method. The shared parameters between all the categories are the signals branching fractions $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$, the branching fraction of the normalisation channel $B^+ \to J/\psi K^+$, and the ratio of hadronisation fractions f_s/f_d.
The combined fit yields

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9} \text{ and}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10},$$

where the errors combine the statistical and systematic uncertainties. The latter accounts for 35% and 18% of the total uncertainty for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, respectively. Using

Wilks' theorem [12], the statistical significances for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are 6.2σ and 3.2σ, respectively. The expected significances assuming the SM values are 7.4σ for the B^0_s and 0.8σ for the B^0 signal. Consequently, the decay of $B^0_s \to \mu^+\mu^-$ is discovered for the first time and the first evidence for $B^0 \to \mu^+\mu^-$ decays is found.
The total dataset together with the different signal and background components in the six bins with the highest expected sensitivity is shown in Fig. 1. The bins are ranked by $S/\sqrt{S+B}$, where S and B are the number of expected signal events assuming the SM branching fractions and the number of background events under the B_s peak in each category. Two simultaneous

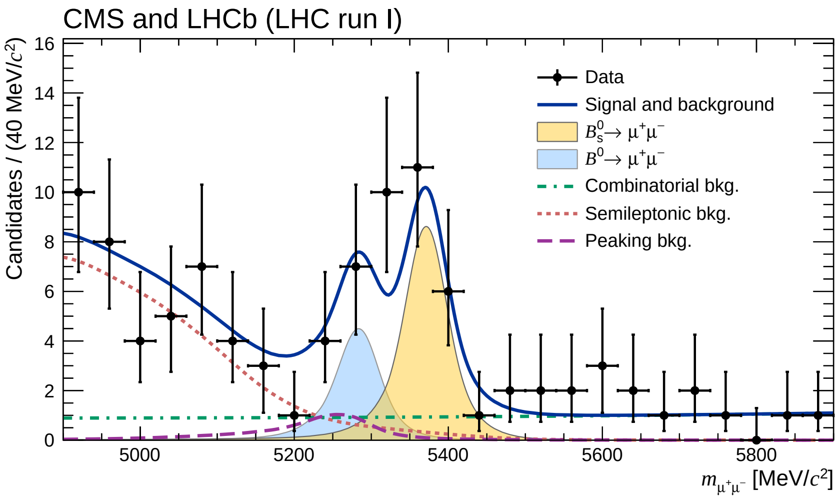

Figure 1. Dimuon invariant mass distributions of the six bins with the highest expected sensitivity according to $S/\sqrt{S+B}$. The data is shown as black points, the total signal and background model is the solid blue line, where the signal components are light blue for $B^0_s \to \mu^+\mu^-$ and yellow shaded for $B^0 \to \mu^+\mu^-$. Additionally, the different background components are the combinatorial (green dash-dotted), the semileptonic (salmon dotted) and the peaking background (violet dashed).

fits are performed to measure the signal branching fractions relative to their SM prediction $\mathcal{S}^{B^0_{(s)}}_{\rm SM}$ and the ratio $\mathcal{R}$ of the branching fractions of B^0 and B^0_s. The results are

$$\mathcal{S}^{B^0_s}_{\rm SM} = 0.76^{+0.20}_{-0.18} \quad \text{and}$$
$$\mathcal{S}^{B^0}_{\rm SM} = 3.7^{+1.6}_{-1.4} \,,$$

which are compatible with the SM within 1.2σ and 2.2σ for B^0_s and B^0, respectively. The ratio of branching fractions is measured to be $\mathcal{R} = 0.14^{+0.08}_{-0.06}$ and compatible with the SM within 2.3σ. The likelihood profile for the signal branching fractions and their ratio are shown in Fig. 2 and Fig. 3, respectively.

3. Confidence regions evaluation

In order to confirm that the approximations of Wilks' theorem hold, the confidence interval of the $B^0 \to \mu^+\mu^-$ signal is calculated using the unified frequentist method proposed by Feldman and Cousins [13]. Instead of assuming a χ^2-distribution with one degree of freedom for the test statistic $q = -2\Delta \ln L$, for each $\mathcal{B}(B^0 \to \mu^+\mu^-)$ hypothesis the actual test statistic distribution is constructed using pseudo-experiments. To generate the simulated experiments the

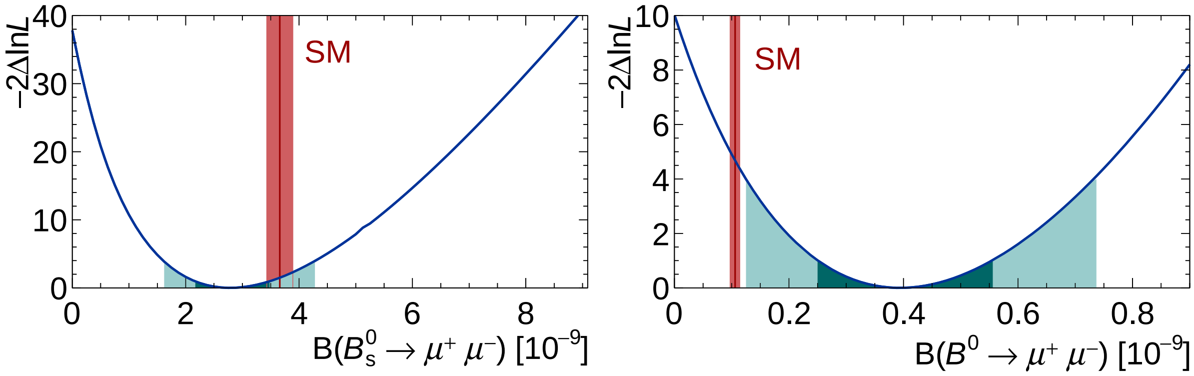

Figure 2. Likelihood profile for the branching fractions of $B_s^0 \to \mu^+\mu^-$ (left) and $B^0 \to \mu^+\mu^-$ (right). The $\pm1\sigma$ and $\pm2\sigma$ confidence regions are marked in dark and light cyan, while the SM predictions are denoted by the red bands.

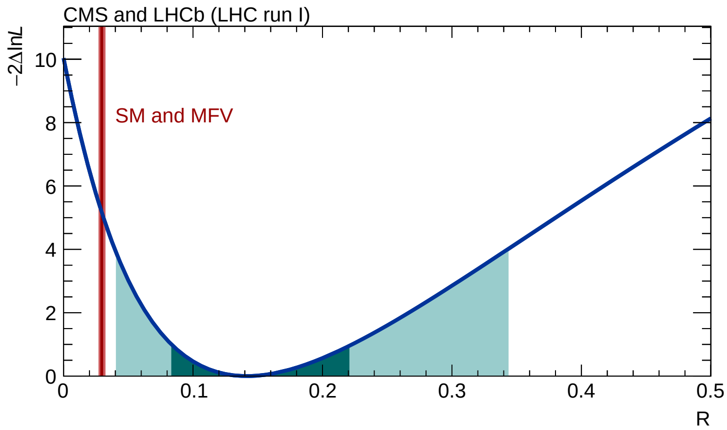

Figure 3. Likelihood profile for $\mathcal{R} \equiv \mathcal{B}(B^0 \to \mu^+\mu^-)/\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ with the $\pm1\sigma$ and $\pm2\sigma$ confidence regions highlighted in dark and light cyan. The red bands defines the predictions of the SM and MFV models.

nuisance parameters are fixed to their best value given by the result of a fit to the data with fixed $\mathcal{B}(B^0 \to \mu^+\mu^-)$. The test statistic distribution is calculated from two maximum likelihood fits, where the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ parameter is either fixed or freely floating. The mean values of the constraints on the nuisance parameters are randomised around the best-fit value used to generate each pseudo-experiment. It is found that $\mathcal{B}(B^0 \to \mu^+\mu^-)$ is in $[2.5, 5.6] \times 10^{-10}$ with 68.27% confidence level and in $[1.4, 7.4] \times 10^{-10}$ with 94.45% confidence level (CL). The outcome of this study is displayed as a $1-$CL curve in Fig. 4. The black points together with a spline interpolation in solid blue are the results of the pseudo-experiments and the approximation using Wilks' theorem is drawn as a dashed grey line. The corresponding $\pm1\sigma$ and $\pm2\sigma$ confidence intervals are indicated in dark and light cyan. In addition the significance is calculated from the same pseudo-experiments as it is defined as the $1-$CL value of the $\mathcal{B}(B^0 \to \mu^+\mu^-) = 0$ hypothesis translated into units of standard gaussian widths. For this reason $O(10^5)$ pseudo-experiments are simulated assuming no $B^0 \to \mu^+\mu^-$ signal yielding a statistical significance of 3.0σ. This confirms that the approximations in Wilks' theorem hold.

Journal of Physics: Conference Series **623** (2015) 012025 doi:10.1088/1742-6596/623/1/012025

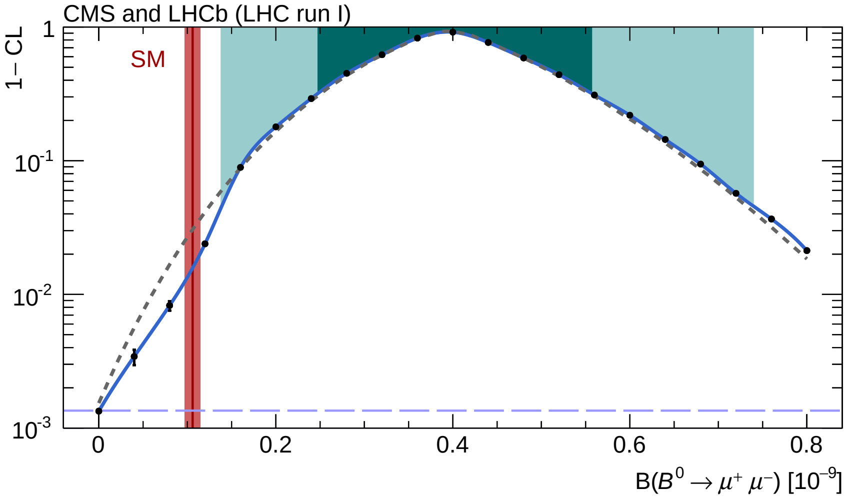

Figure 4. 1−CL curve in dependence of $\mathcal{B}(B^0 \to \mu^+\mu^-)$ hypothesis. The black points mark the 1−CL values obtained from the Feldman-Cousins procedure, while the solid blue line is their spline interpolation. The dashed grey line is the result from the evaluation of 1−CL using Wilks' theorem. The SM prediction for $\mathcal{B}(B^0 \to \mu^+\mu^-)$ is shown as a red band and the $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals are indicated in dark and light cyan.

Finally, confidence regions are evaluated in the two-dimensional $\mathcal{B}(B^0 \to \mu^+\mu^-) - \mathcal{B}(B^0_s \to \mu^+\mu^-)$ plane. Therefore, the test statistic $q = -2\Delta \ln L$ is calculated from the nominal fit and a maximum likelihood fit to the data, where both signal branching fractions are fixed to the corresponding hypotheses. The two-dimensional confidence regions are defined to have the same coverage as the corresponding one-dimensional gaussian σ intervals. The result is shown in Fig. 5.

4. Conclusions

The joint search for $B^0_{(s)} \to \mu^+\mu^-$ decays of the combined LHC-RunI dataset of the CMS and LHCb experiments is presented. For the first time the decay $B^0_s \to \mu^+\mu^-$ is observed with a statistical significance of 6.2σ. Its branching fraction is fully compatible with the SM prediction at 1.2σ.
In case of the $B^0 \to \mu^+\mu^-$ decay, an excess of events is seen over the background-only hypothesis resulting in a statistical significance of 3.2σ using Wilks' theorem and 3.0σ using pseudo-experiments. This signal is compatible with the SM within 2.2σ.
The ratio of the branching fractions of $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ is measured as $\mathcal{R} = 0.14^{+0.08}_{-0.06}$ and therefore compatible with the SM within 2.3σ.
Measuring all observables consistently with the SM predictions significantly reduces the allowed parameters space for theories beyond the SM. However, the measurement is statistically limited. During the upcoming RunII of the Large Hadron Collider both experiments will significantly increase the amount of collected B mesons, which will allow to study $B^0_{(s)} \to \mu^+\mu^-$ decays with greater precision.

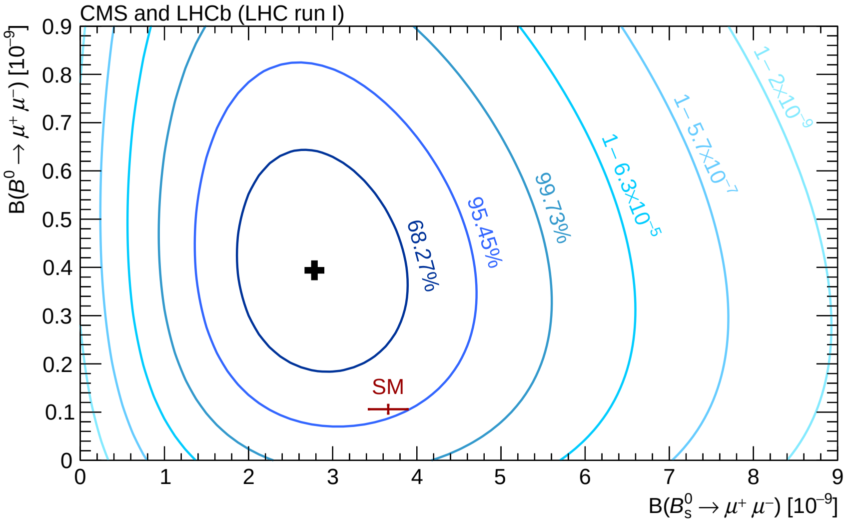

Figure 5. Two-dimensional confidence regions in the $\mathcal{B}(B^0 \to \mu^+\mu^-)-\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ plane. The black cross marks the global best-fit point. The contours indicate the region with the corresponding coverage given on the plot and the SM prediction with its uncertainties is shown as a red cross.

References

[1] Bobeth C, Gorbahn M, Hermann T, Misiak M, Stamou E *et al.* 2014 *Phys.Rev.Lett.* **112** 101801 (*Preprint* `1311.0903`)

[2] 2014 First combination of Tevatron and LHC measurements of the top-quark mass ATLAS-CONF-2014-008, CDF-NOTE-11071, CMS-PAS-TOP-13-014, D0-NOTE-6416 (*Preprint* `1403.4427`)

[3] Aaij R *et al.* (LHCb Collaboration) 2013 *Phys.Rev.Lett.* **110** 021801 (*Preprint* `1211.2674`)

[4] Aaij R *et al.* (LHCb Collaboration) 2013 *Phys.Rev.Lett.* **111** 101805 (*Preprint* `1307.5024`)

[5] Chatrchyan S *et al.* (CMS Collaboration) 2013 *Phys.Rev.Lett.* **111** 101804 (*Preprint* `1307.5025`)

[6] Khachatryan V *et al.* (CMS Collaboration, LHCb Collaboration) 2014 (*Preprint* `1411.4413`)

[7] Khodjamirian A, Klein C, Mannel T and Wang Y M 2011 *JHEP* **1109** 106 (*Preprint* `1108.2971`)

[8] Aaij R *et al.* (LHCb Collaboration) 2014 *Phys.Lett.* **B734** 122 (*Preprint* `1402.6242`)

[9] De Bruyn K, Fleischer R, Knegjens R, Koppenburg P, Merk M *et al.* 2012 *Phys.Rev.Lett.* **109** 041801 (*Preprint* `1204.1737`)

[10] DeBruyn K, Fleischer R, Knegjens R, Koppenburg P, Merk M *et al.* 2012 *Phys.Rev.* **D86** 014027 (*Preprint* `1204.1735`)

[11] Dunietz I, Fleischer R and Nierste U 2001 *Phys.Rev.* **D63** 114015 (*Preprint* `hep-ph/0012219`)

[12] Wilks S S 1938 *The Annals of Mathematical Statistics* **9** 60–62 (URL `http://www.jstor.org/stable/2957648`)

[13] Feldman G J and Cousins R D 1998 *Phys.Rev.* **D57** 3873–3889 (*Preprint* `physics/9711021`)