PROCEEDINGS OF SCIENCE

Measurement of $B_s^0 \to \mu^+\mu^-$ with CMS

Urs Langenegger[*†]
Paul Scherrer Institute, CH-5232 Villigen-PSI, Switzerland
E-mail: urs.langenegger@psi.ch

We present results from a search for the rare leptonic decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ in pp collisions at $\sqrt{s} = 7$ and 8 TeV, with data samples of 5 and 20 fb^{-1}, respectively, collected by the CMS experiment at the LHC. With an unbinned maximum-likelihood fit to the dimuon invariant mass distributions we determine a branching fraction $\mathscr{B}(B_s^0 \to \mu^+\mu^-) = (3.0^{+1.0}_{-0.9}) \times 10^{-9}$, where the uncertainty includes the dominant statistical and small systematic contributions. We observe an excess of $B_s^0 \to \mu^+\mu^-$ events with respect to background expectations with a significance of 4.3 standard deviations. For the decay $B^0 \to \mu^+\mu^-$ an upper limit of $\mathscr{B}(B^0 \to \mu^+\mu^-) < 1.1 \times 10^{-9}$ at the 95% confidence level is determined. Both results agree with the standard-model expectations.

The European Physical Society Conference on High Energy Physics -EPS-HEP2013
18-24 July 2013
Stockholm, Sweden

[*]Speaker.
[†]On behalf of the CMS collaboration.

PoS(EPS-HEP 2013)359

1. Introduction

The leptonic decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are expected to be very rare in the standard model (SM) of particle physics because they are mediated by effective flavor-changing neutral currents and are helicity suppressed. The predicted decay-time integrated branching fractions are $\mathscr{B}(B_s^0 \to \mu^+\mu^-) = (3.6 \pm 0.3) \times 10^{-9}$ and $\mathscr{B}(B^0 \to \mu^+\mu^-) = (1.1 \pm 0.1) \times 10^{-10}$ [1, 2]. The two decays offer very high sensitivity to models with extended Higgs-boson sectors. The search for these decays has been ongoing over the past 30 years, with results from many experiments. The ARGUS [3], UA1 [4], CLEO [5], Belle [6], *BABAR* [7], CDF [8], DØ [9], ATLAS [10], CMS [11], and LHCb [12] experiments have all published limits on these decays. In the past two years, both CDF [8] and LHCb[13] have determined central values for $\mathscr{B}(B_s^0 \to \mu^+\mu^-)$.

Here we present a complete re-analysis of the entire pp collision data sample collected in 2011 and 2012 by the CMS experiment [14] at the LHC. The integrated luminosity of the 2011 (2012) data sample amounts to 5 fb^{-1} (20 fb^{-1}) at a center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}(8\,\text{TeV})$. The improvements in this analysis with respect to last year's CMS publication [11] are: improved muon identification, improved and expanded selection variables, a multi-variate analysis, and an unbinned maximum-likelihood (UML) fit for the result extraction. The dimuon invariant mass region $5.2 < m < 5.45\,\text{GeV}$ was (re-)blinded for the selection development and the choice of the interpretation methodology. A more complete description of the analysis can be found in Ref. [15].

2. Detector

The CMS detector [14] was designed to provide excellent lepton (and photon) identification, b/τ tagging, and measurement of jets and missing transverse energy. This analysis profits from detector capabilities developed to meet the first two design criteria: excellent vertex and track reconstruction in a large and homogeneous magnetic field of 3.8 T combined with good muon trigger, reconstruction, and identification capabilities.

The silicon pixel detector, with a pixel size of $100\,\mu\text{m} \times 150\,\mu\text{m}$, provides a powerful tool [16] for measuring displaced secondary vertices from B decays[1] and minimizes effects from other pp collisions in the same bunch crossing (pileup). The silicon strip detector achieves a transverse momentum ($p_\perp$) resolution of around 1.5% for muons used in this analysis, resulting in a dimuon mass resolution of 32 MeV (for dimuon pseudorapidity $|\eta| \approx 0$) to 75 MeV (for $|\eta| > 1.8$).

The CMS muon detectors cover $|\eta| < 2.4$ and consist of drift tubes, cathode strip chambers, and resistive plate chambers to allow for precise and redundant muon track reconstruction. The baseline muon selection are "tight" muons as described in Ref. [17]. A boosted decision tree (BDT) has been developed to reduce the hadron misidentification probability by a factor two while keeping 90% of genuine muons passing the tight muon criterion. The hadron misidentification probabilities are around 10^{-3}, with variations due to the hadron species (pions, kaons, protons), transverse momentum, and pseudorapidity. The (relative) systematic uncertainty of 50% is based on differences between data and Monte Carlo (MC) simulations.

[1] We use B as abbreviation for B^0 and B_s^0.

3. Analysis

The signal is characterized by two muons from one well-reconstructed secondary B vertex, with the dimuon momentum aligned with the flight direction (from the primary vertex to the B vertex), the dimuon mass around the B mass, and isolated dimuons (because the two muons are the only decay particles of the B meson). The background has different components: the combinatorial background, estimated from data sidebands, is mainly from two separate semileptonic $b \to c\mu\bar{\nu}$ decays or from one $b \to c\mu\bar{\nu}$ decay in combination with a misidentified hadron. Furthermore, rare single B decays are studied in MC simulated events and have a peaking component from hadronic two-body B decays (e.g. $B_s^0 \to K^+K^-$) and a non-peaking component (e.g. $B^0 \to \pi^-\mu^+\nu$).

The branching fraction is determined with respect to the normalization sample $B^\pm \to J/\psi K^\pm$ ($J/\psi \to \mu^+\mu^-$) to minimize uncertainties from the B production cross section. A control sample with $B_s^0 \to J/\psi\phi$ ($J/\psi \to \mu^+\mu^-$, $\phi \to K^+K^-$) decays is reconstructed to compare B_s^0 production characteristics in MC simulation and data. Since the signal-to-background ratio is better in the central region of the detector than the forward region, the analysis is performed in two 'channels': in the 'barrel', both muons have $|\eta| < 1.4$, while in the endcap at least one muon has $1.4 < |\eta| < 2.4$.

The most powerful selection variables are based on the vertex characteristics and the isolation of the B candidate. The flight length significance is the distance between the primary vertex and the B vertex divided by its error. The B candidate impact parameter and its significance are determined with respect to the primary vertex. The pointing angle is measured between the B candidate momentum and the vector from the primary vertex to the B vertex. Four isolation variables are defined. (1) $I = {p_\perp}^{\mu\mu}/({p_\perp}^{\mu\mu} + \sum_{\rm trk} p_\perp)$, where $\sum_{\rm trk} p_\perp$ is the sum of $p_\perp$ of all tracks, other than muon candidates, satisfying $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.7$ with respect to the B candidate momentum. The sum runs over all tracks with $p_\perp > 0.9\,\text{GeV}$ that (i) are consistent with originating from the same primary vertex as the B candidate or (ii) have a distance of closest approach ($d_{\rm ca}$) with respect to the B vertex $<0.05\,\text{cm}$ and are not associated with any other primary vertex. (2) The isolation I_μ of each muon, calculated as for the B candidate. A cone size of $\Delta R = 0.5$ and tracks with $p_\perp > 0.5\,\text{GeV}$ and $d_{\rm ca} < 0.1\,\text{cm}$ from the muon are used. (3) The number of tracks $N_{\rm trk}^{\rm close}$ with $p_\perp > 0.5\,\text{GeV}$ and $d_{\rm ca}$ with respect to the B vertex less than $< 0.03\,\text{cm}$. (4) $d_{\rm ca}^0$ is defined as the smallest $d_{\rm ca}$ to the B vertex, considering all tracks in the event that are either associated with the same primary vertex as the B candidate or not associated with any primary vertex.

These variables are used to build BDTs within the TMVA framework [18]. For the training, signal events are drawn from MC simulations, while background events are from data sidebands. To avoid any possible selection bias, all samples are randomly divided into three subsets for BDT training, testing, and analysis. This procedure is applied separately to the barrel and endcap channels and for the $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ data samples, resulting in 12 separate BDTs.

A comparison of the selection variable distributions in (sideband subtracted) data and MC simulation reveals good agreement for the $B^\pm \to J/\psi K^\pm$ and $B_s^0 \to J/\psi\phi$ samples. None of the variables indicates any pileup dependence. Furthermore, pileup has no significant effect on the BDT output discriminant b distribution (mean, rms, and efficiency after application of a selection criterion).

Table 1: The signal selection efficiencies $\varepsilon_{\rm tot}$, the predicted number of SM signal events $N_{\rm signal}^{\rm exp}$, the expected number of signal and background events $N_{\rm total}^{\rm exp}$, and the number of observed events $N_{\rm obs}$ in the barrel and endcap channels for the 7 and 8 TeV data using the 1D-BDT method. The event numbers refer to the B^0 and B_s^0 signal regions, respectively, as defined in the text. The errors quoted combine the statistical and systematic uncertainties.

	$\varepsilon_{\rm tot}[10^{-2}]$	$N_{\rm signal}^{\rm exp}$	$N_{\rm total}^{\rm exp}$	$N_{\rm obs}$
	$\sqrt{s} = 7$ TeV			
B^0 Barrel	0.33 ± 0.03	0.27 ± 0.03	1.3 ± 0.8	3
B_s^0 Barrel	0.30 ± 0.04	2.97 ± 0.44	3.6 ± 0.6	4
B^0 Endcap	0.20 ± 0.02	0.11 ± 0.01	1.5 ± 0.6	1
B_s^0 Endcap	0.20 ± 0.02	1.28 ± 0.19	2.6 ± 0.5	4
	$\sqrt{s} = 8$ TeV			
B^0 Barrel	0.24 ± 0.02	1.00 ± 0.10	7.9 ± 3.0	11
B_s^0 Barrel	0.23 ± 0.03	11.46 ± 1.72	17.9 ± 2.8	16
B^0 Endcap	0.10 ± 0.01	0.30 ± 0.03	2.2 ± 0.8	3
B_s^0 Endcap	0.09 ± 0.01	3.56 ± 0.53	5.1 ± 0.7	4

4. Results

Two methodologies are used to determine the final results. (1) In the 1D-BDT method, one single minimum requirement on b per channel is used. The requirement on b is optimized for best significance. We use the 1D-BDT method for the determination of the upper limit on $\mathscr{B}(B^0 \to \mu^+\mu^-)$. In Table 1 the signal efficiencies are provided, together with the expected and observed number of events for the B^0 signal region $5.20 < m < 5.30$ GeV and the B_s^0 signal region $5.30 < m < 5.45$ GeV. A slight excess is evident in the B^0 signal regions in the barrel, while for the B_s^0 regions the observation agrees well with the expectation. (2) In the categorized-BDT method, the discriminant b is used to define twelve event categories with different signal-to-background ratios, but with the same expected signal yield in each bin. A simultaneous UML fit to the twelve mass distributions is then used to determine $\mathscr{B}(B_s^0 \to \mu^+\mu^-)$. Method (2) has higher expected sensitivity and thus provides the main methodology for the extraction of $\mathscr{B}(B_s^0 \to \mu^+\mu^-)$. As a cross check, the UML fit can also be applied to the four mass distributions of method (1).

The probability distribution functions used in the UML fit are a Crystal Ball function (plus a Gaussian with common mean) for the signal (peaking background components). Both $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are fitted for simultaneously, letting the respective normalizations free. The peaking background is constrained to the expectation normalized to the measured B^+ yield. The combinatorial background is described by a first-degree polynomial with a floating slope. Independent data samples confirm that this approach allows a good description of the high sideband data irrespective of the selected b range. The rare semileptonic background is modeled by a Gaussian kernels method of the combined MC simulation prediction. Its normalization is left floating within the large uncertainties of its predicted yield. The invariant mass distributions in the $\sqrt{s} = 8$ TeV data sample with the highest and second-highest signal-to-background ratio, together with the fit

results, are shown in Fig. 1. The candidates in the B^0 and B_s^0 signal regions have been investigated and their distributions of kinematic and vertexing variables are consistent with signal expectations based on $B_s^0 \to \mu^+\mu^-$ MC simulations.

From the UML fit in the categorized-BDT method the following branching fractions are determined

$$\mathscr{B}(B_s^0 \to \mu^+\mu^-) = (3.0^{+1.0}_{-0.9}) \times 10^{-9}$$
$$\mathscr{B}(B^0 \to \mu^+\mu^-) = (3.5^{+2.1}_{-1.8}) \times 10^{-10}$$

where the errors combine the statistical and systematic components. Separating the two errors yields $\mathscr{B}(B_s^0 \to \mu^+\mu^-) = [3.0^{+0.9}_{-0.8}(\text{stat}^{+0.6}_{-0.4}(\text{syst})] \times 10^{-9}$. For $B_s^0 \to \mu^+\mu^-$ this corresponds to an observed (expected) significance of 4.3 (4.8) standard deviations. With the 1D-BDT method, the observed (expected) significance amounts to 4.8 (4.7) standard deviations. The excess in the B^0 region has a significance of 2.0 standard deviations. In the significance determination for B_s^0 (B^0), the B^0 (B_s^0) signal is left floating and is treated as a nuisance parameter. Since the B^0 excess is not significant, we set an upper limit with the CL_s method [19, 20] at 1.1×10^{-9} at the 95% confidence level.

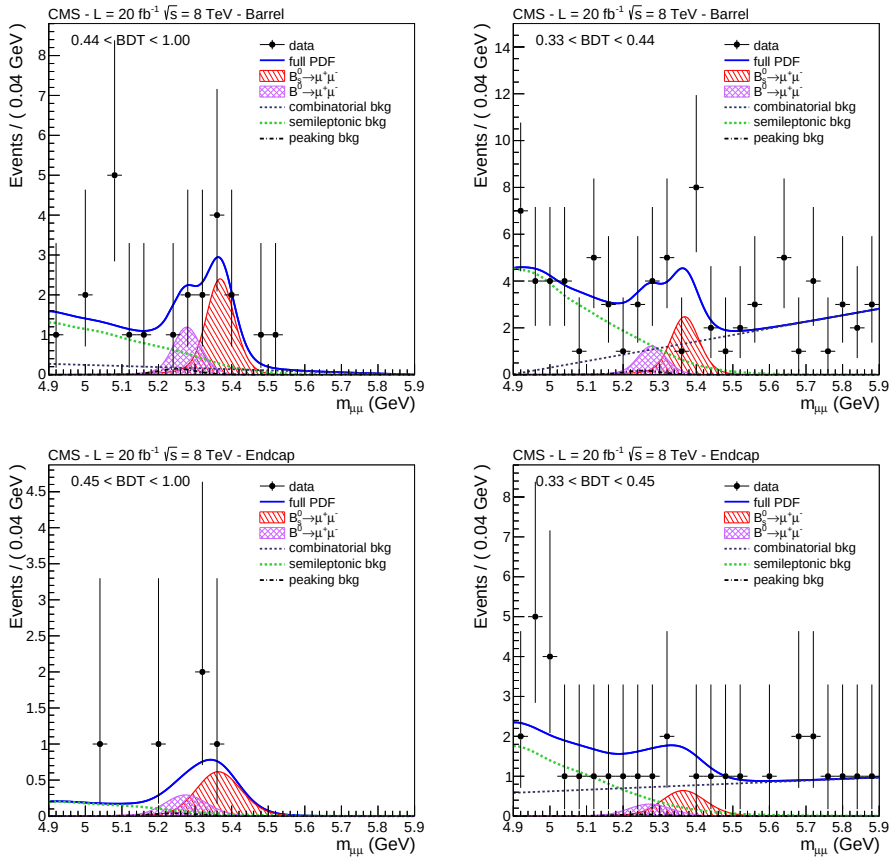

Figure 1: Results from the categorized-BDT method of the fit to the dimuon invariant mass distributions for the $\sqrt{s} = 8$ TeV data in the barrel (top) and endcap (bottom) for the BDT bins with the highest (left) and second-highest (right) signal-to-background ratio.

The external input used for these results are $\mathscr{B}(B^\pm \to J/\psi K^\pm; J/\psi \to \mu^+\mu^-) = (6.0 \pm 0.2) \times 10^{-5}$ [21] and the ratio of b-hadron fragmentation probabilities $f_s/f_u = 0.256 \pm 0.020 \pm 0.013$ as measured by LHCb [22], where the last error represents an ad-hoc uncertainty to account for a hypothetical $p_\perp$- and η-dependence of this quantity for the phase space of this analysis compared to Ref. [22]. In-situ studies comparing $B^\pm \to J/\psi K^\pm$ with $B_s^0 \to J/\psi\phi$ reveal no discernible effect within the phase space of this analysis. The uncertainties above are dominated by the statistical component. Apart from the large uncertainty on f_s/f_u for the B_s^0 result, the other systematic uncertainties include: the relative uncertainty of 50% on the hadron to muon misidentification probability; the branching fraction uncertainties of the background decay modes (dominated by the mode $\Lambda_b \to p\mu\nu$, where we use the average value between the lowest and the largest expectation in the literature [23 – 25] and assign an uncertainty of 100%); 5% uncertainty on the B^+ yield, determined from varying the fit functions and comparison with fits using a J/ψ mass constraint; muon identification and trigger uncertainties of a few percent; selection efficiency uncertainty estimated by comparing $B_s^0 \to J/\psi\phi$ efficiencies in data and MC simulation. The mass scale correction and uncertainty are measured in data and MC simulation with ψ and Υ decays to dimuons.

5. Conclusions

We have presented the first measurement of $\mathscr{B}(B_s^0 \to \mu^+\mu^-)$ with the CMS experiment, with a substantial sensitivity gain over the previous publications. The improvements are due to a BDT-based muon identification, improved and expanded selection variables, a BDT-based selection, and an UML fit for the result extraction. We observe $\mathscr{B}(B_s^0 \to \mu^+\mu^-) = (3.0^{+1.0}_{-0.9}) \times 10^{-9}$ with a significance of 4.3 standard deviations, in agreement with the SM expectations. This result constitutes the first observation of $B_s^0 \to \mu^+\mu^-$ with more than 4 standard deviations.

References

[1] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, *On the Standard Model prediction for $B^0_{s(d)} \to \mu^+\mu^-$*, *Eur. Phys. J. C* 72 (2012) 2172, [arXiv:1208.0934].

[2] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, A. Pellegrino, and N. Tuning, *Probing new physics via the $B_s^0 \to \mu^+\mu^-$ effective lifetime*, *Phys. Rev. Lett.* 109 (2012) 041801, [arXiv:1204.1737].

[3] ARGUS Collaboration, H. Albrecht *et al.*, *B meson decays into charmonium states*, *Phys. Lett. B* 199 (1987) 451.

[4] UA1 Collaboration, C. Albajar *et al.*, *A search for rare B meson decays at the CERN $Sp\overline{p}S$ collider*, *Phys. Lett. B* 262 (1991) 163.

[5] CLEO Collaboration, T. Bergfeld *et al.*, *Search for decays of B^0 mesons into pairs of leptons: $B^0 \to e^+e^-$, $B^0 \to \mu^+\mu^-$ and $B^0 \to e^\pm\mu^\mp$*, *Phys. Rev. D* 62 (2000) 091102, [hep-ex/0007042].

[6] Belle Collaboration, M. C. Chang *et al.*, *Search for $B^0 \to \ell^+\ell^-$ at Belle*, *Phys. Rev. D* 68 (2003) 111101, [hep-ex/0309069].

[7] *BABAR* Collaboration, B. Aubert *et al.*, *Search for decays of B^0 mesons into e^+e^-, $\mu^+\mu^-$, and $e^\pm\mu^\mp$ final states*, *Phys. Rev. D* 77 (2008) 032007, [arXiv:0712.1516].

[8] CDF Collaboration, T. Aaltonen *et al.*, *Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with the full CDF Run II data set*, *Phys. Rev. D* 87 (2013) 072003, [arXiv:1301.7048].

[9] DØ Collaboration, V. M. Abazov *et al.*, *Search for the rare decay $B_s^0 \to \mu^+\mu^-$*, *Phys. Rev. D* 87 (2013) 072006, [arXiv:1301.4507].

[10] ATLAS Collaboration, G. Aad *et al.*, *Search for the decay $B_s^0 \to \mu^+\mu^-$ with the ATLAS detector*, *Phys. Lett. B* 713 (2012) 387, [arXiv:1204.0735].

[11] CMS Collaboration, S. Chatrchyan *et al.*, *Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays*, *JHEP* 04 (2012) 033, [arXiv:1203.3976].

[12] LHCb Collaboration, R. Aaij *et al.*, *Search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$*, *Phys. Lett. B* 699 (2011) 330, [arXiv:1103.2465].

[13] LHCb Collaboration, R. Aaij *et al.*, *First evidence for the decay $B_s^0 \to \mu^+\mu^-$*, *Phys. Rev. Lett.* 110 (2013) 021801, [arXiv:1211.2674].

[14] CMS Collaboration, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, *JINST* 3 (2008) S08004.

[15] CMS Collaboration, S. Chatrchyan *et al.*, *Measurement of $B_s^0 \to \mu^+\mu^-$ and Search for $B^0 \to \mu^+\mu^-$ with the CMS experiment*, *Phys. Rev. Lett* 111 (2013) 101804, [arXiv:1307.5025].

[16] C. Nägeli and R. Horisberger, *Impact of pixel size and shape on physics analysis*, *acc. for publication in Nucl. Instrum. Meth. A* (2013) Proceedings of 'Pixel 2012,' 6^{th} International Workshop on Semiconductor Pixel Detectors.

[17] CMS Collaboration, S. Chatrchyan *et al.*, *Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV*, *JINST* 7 (2012) P10002, [arXiv:1206.4071].

[18] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *TMVA - toolkit for multivariate data analysis*, physics/0703039.

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

[20] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth. A* 434 (1999) 435, [hep-ex/9902006].

[21] Particle Data Group, J. Beringer, *et al.*, *Review of particle physics*, *Phys. Rev. D* 86 (2012) 010001.

[22] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, *JHEP* 04 (2013) 001, [arXiv:1301.5286].

[23] M.-Q. Huang and D.-W. Wang, *Light cone QCD sum rules for the semileptonic decay $\Lambda_b \to p\ell\bar{\nu}$*, *Phys.Rev.* D69 (2004) 094003, [hep-ph/0401094].

[24] Y.-M. Wang, Y.-L. Shen, and C.-D. Lu, *$\Lambda_b \to p, \Lambda$ transition form factors from QCD light-cone sum rules*, *Phys.Rev.* D80 (2009) 074012, [arXiv:0907.4008].

[25] K. Azizi, M. Bayar, Y. Sarac, and H. Sundu, *Semileptonic $\Lambda_{b,c}$ to Nucleon Transitions in Full QCD at Light Cone*, *Phys.Rev.* D80 (2009) 096007, [arXiv:0908.1758].

PoS(EPS-HEP 2013)359