Measurement of the $B[0 \overline{s}]^+[-]$ Branching Fraction and Search for $B[0]^+[-]$ with the CMS Experiment.


Accessed Sat Mar 02 08:53:48 EST 2019
Measurement of the $B^0 \to \mu^+ \mu^-$ Branching Fraction and Search for $B^0 \to \mu^+ \mu^-$ with the CMS Experiment

S. Chatrchyan et al.*, (CMS Collaboration)

(Received 18 July 2013; published 5 September 2013)

Results are presented from a search for the rare decays $B^0 \to \mu^+ \mu^-$ and $B^0 \to \mu^+ \mu^-$ in $p p$ collisions at $\sqrt{s} = 7$ and 8 TeV, with data samples corresponding to integrated luminosities of 5 and 20 fb$^{-1}$, respectively, collected by the CMS experiment at the LHC. An unbinned maximum-likelihood fit to the dimuon invariant mass distribution gives a branching fraction $\mathcal{B}(B^0 \to \mu^+ \mu^-) = (3.37 \pm 0.30) \times 10^{-9}$, corresponding to the decay-time integrated branching fraction, and $\mathcal{B}(B^0 \to \mu^+ \mu^-) = (1.07 \pm 0.10) \times 10^{-10}$ [1,2]. Charge conjugation is implied throughout this Letter. Several extensions of the SM, such as supersymmetric models with nonuniversal Higgs boson masses [3], specific models containing leptoquarks [4], and the minimal supersymmetric standard model with large tan$\beta$ [5,6], predict enhancements to the branching fractions for these rare decays. The decay rates can also be suppressed for specific choices of model parameters [7]. Over the past 30 years, significant progress in sensitivity has been made, with exclusion limits on the branching fractions improving by 5 orders of magnitude. The ARGUS [8], UA1 [9], CLEO [10], Belle [11], BABAR [12], CDF [13], D0 [14], ATLAS [15], CMS [16], and LHCb [17] experiments have all published limits on these decays. The LHCb experiment has subsequently shown evidence, with 3.5 standard deviation significance, for the decay $B^0 \to \mu^+ \mu^-$ with $\mathcal{B}(B^0 \to \mu^+ \mu^-) = (3.2^{+1.3}_{-1.5}) \times 10^{-9}$ [18].

This Letter reports a measurement of $\mathcal{B}(B^0 \to \mu^+ \mu^-)$ based on a simultaneous search for $B^0 \to \mu^+ \mu^-$ and $B^0 \to \mu^+ \mu^-$ decays using a data sample of $p p$ collisions corresponding to integrated luminosities of 5 fb$^{-1}$ at $\sqrt{s} = 7$ TeV and 20 fb$^{-1}$ at 8 TeV collected by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). For these data, the peak luminosity varied from $3.5 \times 10^{30}$ to $7.7 \times 10^{33}$ cm$^{-2}$ s$^{-1}$. The average number of interactions per bunch crossing (pileup) was 9 (21) at $\sqrt{s} = 7(8)$ TeV.

The search for the $B \to \mu^+ \mu^-$ signal, where $B$ denotes $B^0$ or $B_s^0$, is performed in the dimuon invariant mass regions around the $B^0$ and $B^0$ masses. To avoid possible biases, the signal region $5.20 < m_{\mu\mu} < 5.45$ GeV was kept blind until all selection criteria were established. For the 7 TeV data, this Letter reports a reapaisal of the data used in the previous result [16], where the data were rebinned. The combinatorial dimuon background, mainly from semileptonic decays of separate $B$ mesons, is evaluated by extrapolating the data in nearby mass sidebands into the signal region. Monte Carlo (MC) simulations are used to account for backgrounds from $B$ and $\Lambda_b$ decays. These background samples consist of $B \to h\mu\nu$, $B \to h\mu\mu$, and $\Lambda_b \to p\mu\nu$ decays, as well as “peaking” decays of the type $B \to hh'$, where $h$, $h'$ are charged hadrons misidentified as muons, which give a dimuon invariant mass distribution that peaks in the signal region. The MC simulation event samples are generated using PYTHIA (version 6.424 for 7 TeV, version 6.426 for 8 TeV) [19], with the underlying event simulated with the Z2 tune [20], unstable particles decayed via EVTGEN [21], and the detector response simulated with GEANT4 [22]. A normalization sample of $B^+ \to J/\psi K^+ \to \mu^+ \mu^- K^+$ decays is used to minimize uncertainties related to the $b\bar{b}$ production cross section and the integrated luminosity. A control sample of $B^0 \to J/\psi \phi \to \mu^+ \mu^- K^+ K^-$ decays is used to validate the MC simulation and to evaluate potential effects from differences in fragmentation between $B^+$ and $B^0$. The efficiencies of all samples,

---

*Full author list given at the end of the article.

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.
including detector acceptances, are determined with MC simulation studies.

A detailed description of the CMS apparatus can be found in Ref. [23]. The CMS experiment uses a right-handed coordinate system, with the origin at the nominal interaction point, the x axis pointing to the center of the LHC ring, the y axis pointing upward, and the z axis along the counterclockwise-beam direction. The polar angle \( \theta \) is measured from the positive z axis and the azimuthal angle \( \phi \) is measured in the x-y plane. The main subdetectors used in this analysis are the silicon tracker and the muon detectors. Muons are tracked within the pseudorapidity region \( |\eta| < 2.4 \), where \( \eta = - \ln(\tan(\theta/2)) \). A transverse momentum \( (p_T) \) resolution of about 1.5% is obtained for muons in this analysis [24].

The events are selected with a two-level trigger system. The first level requires only two muon candidates in the muon detectors. The high-level trigger (HLT) uses additional information from the silicon tracker to provide essentially a full event reconstruction. The dimuon invariant mass was required to satisfy \( 4.8 < m_{\mu\mu} < 6.0 \) GeV. For the 7 TeV data set, the HLT selection required two muons, each with \( p_T > 4.0 \) GeV, and a dimuon \( p_T^{\mu\mu} > 3.9 \) GeV. For events having at least one muon with \( |\eta| > 1.5 \), \( p_T^{\mu\mu} > 5.9 \) GeV was required. For the 8 TeV data set, the \( p_T \) criterion on the muon with lower \( p_T \) was loosened to \( p_T > 3.0 \) GeV, with \( p_T^{\mu\mu} > 4.9 \) GeV. For events containing at least one muon with \( |\eta| > 1.8 \), the muons were each required to have \( p_T > 4.0 \) GeV, \( p_T^{\mu\mu} > 7.0 \) GeV, and the dimuon vertex fit \( p \) value >0.5%.

For the normalization and control samples the HLT selection required the following: two muons, each with \( p_T > 4 \) GeV and \( |\eta| < 2.2 \); \( p_T^{\mu\mu} > 6.9 \) GeV; \( 2.9 < m_{\mu\mu} < 3.3 \) GeV; and the dimuon vertex fit \( p \) value >15%. Two additional requirements were imposed in the transverse plane: (i) the pointing angle \( \alpha_{xy} \) between the dimuon momentum and the vector from the average interaction point to the dimuon vertex had to fulfill \( \cos \alpha_{xy} > 0.9 \), and (ii) the flight length significance \( \ell_{xy}/\sigma(\ell_{xy}) \) must be greater than 3, where \( \ell_{xy} \) is the two-dimensional distance between the average interaction point and the dimuon vertex, and \( \sigma(\ell_{xy}) \) is its uncertainty. The signal, normalization, and control triggers required the three-dimensional (3D) distance of closest approach \( (d_{ca}) \) between the two muons to satisfy \( d_{ca} < 0.5 \) cm. The average trigger efficiency for events in the signal and normalization samples, as determined from MC simulation and calculated after all other selection criteria are applied, is in the range (39–85)%, depending on the running period and detector region. The uncertainty in the ratio of trigger efficiencies [muon identification efficiencies] for the signal and normalization samples is estimated to be (3–6)\% [(1–4)\%] by comparing simulation and data.

The \( B \rightarrow \mu^+ \mu^- \) candidates are constructed from two oppositely charged “tight” muons as described in Ref. [25]. Both muons must have \( p_T > 4 \) GeV and be consistent in direction and \( p_T \) with the muons that triggered the event. A boosted decision tree (BDT) constructed within the TMVA framework [26] is trained to further separate genuine muons from those arising from misidentified charged hadrons. The variables used in the BDT can be divided into four classes: basic kinematic quantities, silicon-tracker fit information, combined silicon and muon track fit information, and muon detector information. The BDT is trained on MC simulation samples of \( B \)-meson decays to kaons and muons. Compared to the tight muons, the BDT working point used to select muons for this analysis reduces the hadron-to-muon misidentification probability by 50% while retaining 90% of true muons. The probability to misidentify a charged hadron as a muon because of decay in flight or detector punchthrough is measured in data from samples of well-identified pions, kaons, and protons. This probability ranges from \((0.5–1.3) \times 10^{-3}, (0.8–2.2) \times 10^{-3}\), and \((0.4–1.5) \times 10^{-3}\), for pions, kaons, and protons, respectively, depending on whether the particle is in the barrel or end cap, the running period, and the momentum. Each of these probabilities is ascribed an uncertainty of 50%, based on differences between data and MC simulation.

Candidates are kept for further analysis if they have \( 4.9 < m_{\mu\mu} < 5.9 \) GeV, after constraining the tracks to a common vertex. The \( B \)-candidate momentum and vertex position are used to choose a primary vertex based on the distance of closest approach along the beam line. Since the background level and mass resolution depend significantly on \( \eta_{\mu\mu} \), where \( \eta_{\mu\mu} \) is the pseudorapidity of the \( B \)-meson candidate, the events are separated into two categories: the “barrel channel” with candidates where both muons have \( |\eta| < 1.4 \), and the “end-cap channel” containing those where at least one muon has \( |\eta| > 1.4 \). The \( m_{\mu\mu} \) resolution, as determined from simulated signal events, ranges from 32 MeV for \( \eta_{\mu\mu} = 0 \) to 75 MeV for \( \eta_{\mu\mu} > 1.8 \).

Four isolation variables are defined. (1) \( I_1 = p_T^{\mu\mu}/(p_T^{\mu\mu} + \sum_{\text{trk}} p_T) \), where \( \sum_{\text{trk}} p_T \) is the sum of \( p_T \) of all tracks, other than muon candidates, satisfying \( \Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} < 0.7 \), with \( \Delta \eta \) and \( \Delta \phi \) as the differences in \( \eta \) and \( \phi \) between a charged track and the direction of the \( B \) candidate. The sum includes all tracks with \( p_T > 0.9 \) GeV that are (i) consistent with originating from the same primary vertex as the \( B \) candidate or (ii) have a \( d_{ca} \) with respect to the \( B \) vertex <0.05 cm and are not associated with any other primary vertex. (2) \( I_2 \) is the isolation variable of each muon, calculated as for the \( B \) candidate, but with respect to the muon track. A cone size of \( \Delta R = 0.5 \) around the muon and tracks with \( p_T > 0.5 \) GeV and \( d_{ca} < 0.1 \) cm from the muon are used. (3) \( N_{\text{trk}}^{\text{close}} \) is defined as the number of tracks with \( p_T > 0.5 \) GeV and \( d_{ca} \) with respect to the \( B \) vertex less than <0.03 cm. (4) \( d_{ca} \) is defined as the smallest \( d_{ca} \) to the \( B \) vertex, considering all tracks in the event that are either...
TABLE I. The signal selection efficiencies \( \epsilon_{\text{tot}} \), the predicted number of SM signal events \( N_{\text{exp}} \), the expected number of signal and background events \( N_{\text{total}} \), and the number of observed events \( N_{\text{obs}} \) in the barrel and end-cap channels for the 7 and 8 TeV data using the 1D-BDT method. The event numbers refer to the \( B^0 \) and \( B^0 \) signal regions, respectively.

<table>
<thead>
<tr>
<th></th>
<th>( \epsilon_{\text{tot}} ) [10^{-2}]</th>
<th>( N_{\text{exp}} )</th>
<th>( N_{\text{total}} )</th>
<th>( N_{\text{obs}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 TeV</td>
<td>( B^0 ) barrel</td>
<td>0.33 ± 0.03</td>
<td>0.27 ± 0.03</td>
<td>1.3 ± 0.8</td>
</tr>
<tr>
<td></td>
<td>( B^0 ) barrel</td>
<td>0.30 ± 0.04</td>
<td>2.97 ± 0.44</td>
<td>3.6 ± 0.6</td>
</tr>
<tr>
<td></td>
<td>( B^0 ) end cap</td>
<td>0.20 ± 0.02</td>
<td>0.11 ± 0.01</td>
<td>1.5 ± 0.6</td>
</tr>
<tr>
<td></td>
<td>( B^0 ) end cap</td>
<td>0.20 ± 0.02</td>
<td>1.28 ± 0.19</td>
<td>2.6 ± 0.5</td>
</tr>
<tr>
<td></td>
<td>( B^0 ) barrel</td>
<td>0.24 ± 0.02</td>
<td>1.00 ± 0.10</td>
<td>7.9 ± 3.0</td>
</tr>
<tr>
<td></td>
<td>( B^0 ) barrel</td>
<td>0.25 ± 0.03</td>
<td>11.46 ± 1.72</td>
<td>17.9 ± 2.8</td>
</tr>
<tr>
<td>8 TeV</td>
<td>( B^0 ) end cap</td>
<td>0.10 ± 0.01</td>
<td>0.30 ± 0.03</td>
<td>2.2 ± 0.8</td>
</tr>
<tr>
<td></td>
<td>( B^0 ) end cap</td>
<td>0.09 ± 0.01</td>
<td>3.56 ± 0.53</td>
<td>5.1 ± 0.7</td>
</tr>
</tbody>
</table>

associated with the same primary vertex as the \( B \) candidate or not associated with any primary vertex.

The final selection is performed with BDTs trained to distinguish between signal and background event candidates. For the training, \( \bar{B}^0 \rightarrow \mu^+ \mu^- \) MC simulation samples are used for the signal, and candidates from the data dimuon mass sidebands after a loose preselection for the background. The preselection retains at least 10 000 events dominated by combinatorial background for each BDT. To avoid any selection bias, the data background events are randomly split into three sets, such that the training and testing of the BDT is performed on sets independent of its application. Studies with sideband events and signal MC simulation samples with shifted \( B \) mass show that the BDT response is independent of mass. Separate BDTs are trained for each of the four combinations of 7 and 8 TeV data and the barrel and end-cap regions of the detector. For each BDT, a number of variables is considered and only those found to be effective are included. Each of the following 12 variables, shown to be independent of pileup, are used in at least one of the BDTs: \( I; I_d; N_{\text{close}}; d_{\text{rel}}; p^\mu_T; n_{\mu}; \) the \( B \)-vertex fit \( \chi^2 \) per degree of freedom (dof); the \( d_{\text{rel}} \) between the two muon tracks; the 3D pointing angle \( \alpha_{3D} \); the 3D flight length significance \( \ell_{3D}/\alpha(\ell_{3D}); \) the 3D impact parameter \( \delta_{3D} \) of the \( B \) candidate; and its significance \( \delta_{3D}/\sigma(\delta_{3D}) \), where \( \sigma(\delta_{3D}) \) is the uncertainty on \( \delta_{3D} \). The last four variables are computed with respect to the primary vertex. Good agreement between data and MC simulation is observed for these variables. In total, including the division into three sets, 12 BDTs are trained.

The output discriminant \( b \) of the BDT is used in two ways for further analysis. (1) In the 1D-BDT method, a minimum requirement on \( b \) per channel is used to define the final selection. The requirement on \( b \) is optimized for best \( S/\sqrt{S+B} \) (where \( S \) is the expected signal and \( B \) the background) on statistically independent data control samples. The optimization gives \( b > 0.29 \) for both barrel and end cap in the \( \sqrt{s} = 7 \) TeV data, and \( b > 0.36 \) (0.38) in the barrel (end cap) for the \( \sqrt{s} = 8 \) TeV sample. The 1D-BDT method is used for the determination of the upper limit on \( \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) \). The signal efficiencies \( \epsilon_{\text{tot}} \) for method (1) are provided in Table I, together with the expected number of events (signal and signal plus background) for the \( B^0 \) signal region \( 5.20 < m < 5.30 \) GeV and the \( \bar{B}^0 \) signal region \( 5.30 < m < 5.45 \) GeV. (2) In the categorized-BDT method, the discriminant \( b \) is used to define 12 event categories with different signal-to-background ratios. For the \( \sqrt{s} = 7 \) TeV data in the barrel (end-cap) channel, the two categories have boundaries of \( 0.10, 0.31, 1.00 \) (0.10, 0.26, 1.00). For the \( \sqrt{s} = 8 \) TeV sample in the barrel (end-cap) channel, the corresponding boundaries for the four categories are \( 0.10, 0.23, 0.33, 0.44, 1.00 \) (0.10, 0.22, 0.33, 0.45, 1.00). This binning is chosen to give the same expected signal yield in each bin. The dimuon invariant mass distributions for the 12 categories of events are fitted simultaneously to obtain the final results. Method (2) has higher expected sensitivity and thus provides the main methodology for the extraction of \( \mathcal{B}(\bar{B}^0 \rightarrow \mu^+ \mu^-) \).

The \( B^+ \rightarrow J/\psi K^+ \rightarrow \mu^+ \mu^- K^+ \) selection requires two oppositely charged muons with \( 3.0 < m_{\mu\mu} < 3.2 \) GeV and \( p^\mu_T > 7 \) GeV, combined with one or two tracks, assumed to be kaons, fulfilling \( p_T > 0.5 \) GeV and \( |\eta| < 2.4 \) (\( |\eta| < 2.1 \) in the 8 TeV data). The distance of closest approach between all pairs among the three (four) tracks is required to be less than 0.1 cm. For \( \bar{B}^0 \rightarrow J/\psi \phi \) candidates the two assumed kaon tracks must have invariant mass \( 0.995 < m_{KK} < 1.045 \) GeV and \( \Delta R < 0.25 \). The \( B \) vertex is fitted from the three (four) tracks; a candidate is accepted if the resulting invariant mass is in the range 4.8–6.0 GeV. The final selection is achieved using the same BDT as for the signal, with the following modifications: the \( B \)-vertex \( \chi^2/dof \) is determined from the dimuon vertex fit, and for the calculation of the isolation variables all \( B \)-candidate decay tracks are neglected.

The total efficiency to reconstruct with the 1D-BDT method a \( B^+ \rightarrow J/\psi K^+ \rightarrow \mu^+ \mu^- K^+ \) decay, including the detector acceptance, is \( \epsilon_{\text{tot}} = (0.98 \pm 0.08) \times 10^{-3} \).
and \((0.36 \pm 0.04) \times 10^{-3}\), respectively, for the barrel and end-cap channels in the 7 TeV analysis, and \((0.82 \pm 0.07) \times 10^{-3}\) and \((0.21 \pm 0.03) \times 10^{-3}\) for the 8 TeV analysis, where statistical and systematic uncertainties are combined in quadrature. The distributions of \(b\) for the normalization and control samples are found to agree well between data and MC simulation, with residual differences used to estimate systematic uncertainties. No dependence of the selection efficiency on pileup is observed. The systematic uncertainty in the acceptance is estimated by comparing the values obtained with different \(b\) production mechanisms (gluon splitting, flavor excitation, and flavor creation). The uncertainty in the event selection efficiency for the \(B^+ \rightarrow J/\psi K^+\) normalization sample is evaluated from differences between measured and simulated \(B^+ \rightarrow J/\psi K^+\) events. The uncertainty in the \(B^0 \rightarrow \mu^+ \mu^-\) and \(B^0 \rightarrow \mu^+ \mu^-\) signal efficiencies \([3\pm10\%]\), depending on the channel and \(\sqrt{s}\) is evaluated using the \(B^0 \rightarrow J/\psi \phi\) control sample.

The yields for the normalization (control) sample in each category are fitted with a double (single) Gaussian function. The backgrounds under the normalization and control sample peaks are described with an exponential (plus an error function for the normalization sample). Additional functions are included, with shape templates fixed from simulation, to account for backgrounds from \(B^+ \rightarrow J/\psi \pi^+\) (Gaussian function) for the normalization sample, and \(B^0 \rightarrow J/\psi K^0\) (Landau function) for the control sample. In the 7 TeV data, the observed number of \(B^+ \rightarrow J/\psi K^+\) candidates in the barrel is \((71.2 \pm 4.1) \times 10^3\) and \((21.4 \pm 1.1) \times 10^3\) in the end-cap channel. For the 8 TeV sample the corresponding yields are \((309 \pm 16) \times 10^3\) (barrel) and \((69.3 \pm 3.5) \times 10^3\) (end cap). The uncertainties include a systematic component estimated from simulated events by considering alternative fitting functions.

The \(B^0 \rightarrow \mu^+ \mu^-\) branching fraction is measured using

\[
\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = \frac{N_S}{N^R_{\text{obs}}} \frac{f_u}{f_s} \frac{e^B_{\text{tot}}}{e^B_{\text{u}}},
\]

and analogously for the \(B^0 \rightarrow \mu^+ \mu^-\) case, where \(N_S\) \((N^R_{\text{obs}})\) is the number of reconstructed \(B^0 \rightarrow \mu^+ \mu^-\) \((B^+ \rightarrow J/\psi K^+)\) decays, \(e^B_{\text{tot}}\) \((e^B_{\text{u}})\) is the total signal \((B^+ \rightarrow J/\psi K^+)\) efficiency, \(\mathcal{B}(B^+ \rightarrow J/\psi K^+\mu^+ \mu^-) = (6.0 \pm 0.2) \times 10^{-5}\) \([27]\) is the branching fraction for \(B^+ \rightarrow J/\psi K^+\) \(\mu^+ \mu^- K^+\), and \(f_u/f_s\) is the ratio of the \(B^+ \) and \(B^0\) fragmentation fractions. The value \(f_u/f_s = 0.256 \pm 0.020\), as measured by LHCb \([28]\), is used and an additional systematic uncertainty of 5% is assigned to account for possible pseudorapidity and \(p_T\) dependence of this ratio. Studies based on the \(B^+ \rightarrow J/\psi K^+\) and \(B^0 \rightarrow J/\psi \phi\) control samples reveal no discernible pseudorapidity or \(p_T\) dependence of this ratio in the kinematic region used in the analysis.

An unblinded maximum-likelihood fit to the \(m_{\mu \mu}\) distribution is used to extract the signal and background yields. Events in the signal window can result from genuine signal, combinatorial background, background from semileptonic \(b\)-hadron decays, and the peaking background. The probability density functions (PDFs) for the signal, semileptonic, and peaking backgrounds are obtained from fits to MC simulation. The \(B^0\) and \(B^+\) signal shapes are modeled by Crystal Ball functions \([29]\). The peaking background is modeled with the sum of Gaussian and Crystal Ball functions (with a common mean). The semileptonic background is modeled with a Gaussian kernels method \([30,31]\). The PDF for the combinatorial background is modeled with a first-degree polynomial. Since the dimuon mass resolution \(\sigma\), determined on an event-by-event basis from the dimuon mass fit, varies significantly, the PDFs described above are combined as a conditional product with the PDF for the per-event mass resolution, such that the Crystal Ball function width correctly reflects the resolution on a per-event basis. To avoid any effect of the correlation between \(\sigma\) and the candidate mass, we divide the invariant mass uncertainty by the mass to obtain a “reduced” mass uncertainty, \(\sigma_r = \sigma/m_{\mu \mu}\), which is used in the fit.

The dimuon mass distributions for the four channels (barrel and end cap in 7 and 8 TeV data), further divided into categories corresponding to different bins in the BDT parameter \(b\), are fitted simultaneously. The results are illustrated for the most sensitive categories in Fig. 1. The fits for all 12 categories are shown in the Supplemental Material \([32]\) showing additional plots of the mass fits. Pseudoexperiments, done with MC simulated events, confirm the robustness and accuracy of the fitting procedure.

![Figure 1](image-url) (color online). 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 end cap (bottom) for the BDT bins with the highest (left) and second-highest (right) signal-to-background ratio.
Systematic uncertainties are constrained with Gaussian PDFs with the standard deviations of the constraints set equal to the uncertainties. Sources of systematic uncertainty arise from the hadron-to-muon misidentification probability, the branching fraction uncertainties (dominated by 100% for $\Lambda_b \rightarrow p\mu\nu$, and the normalization of the peaking background. The $B \rightarrow hh'$ and semileptonic backgrounds are estimated by normalizing to the observed $B^+ \rightarrow J/\psi K^+$ yield. The peaking background yield is constrained in the fit with log-normal PDFs with rms parameters set to the mean uncertainty. The absolute level of peaking background has been constrained in the fit with log-normal PDFs with rms parameters set to the mean uncertainty. Sources of systematic uncertainties are considered. The bottom plot of the categorized-BDT (left) and the 1D-BDT (right). For these plots, the individual categories used in the categorized-BDT method (left) and the

![Graph](image-url)

FIG. 2 (color online). Top, scan of the ratio of the joint likelihood for $B^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$. As insets, the likelihood ratio scan for each of the branching fractions when the other is profiled together with other nuisance parameters; the significance at which the background-only hypothesis is rejected is also shown. Bottom, observed and expected CL$_S$ for $B^0 \rightarrow \mu^+\mu^-$ as a function of the assumed branching fraction.

![Graph](image-url)

FIG. 3 (color online). Plots illustrating the combination of all categories used in the categorized-BDT method (left) and the 1D-BDT method (right). For these plots, the individual categories are weighted with $S/(S+B)$, where $S$ ($B$) is the signal (background) determined at the $B^0$ peak position. The overall normalization is set such that the fitted $B^0$ signal corresponds to the total yield of the individual contributions. These distributions are for illustrative purposes only and were not used in obtaining the final results.

end-cap) channel, as determined with charmonium and bottomonium decays to dimuon final states.

An excess of $B^0 \rightarrow \mu^+\mu^-$ decays is observed above the background predictions. The measured decay-time integrated branching fraction from the fit is $B(B^0_\mu \rightarrow \mu^+\mu^-) = (3.0_{-0.9}^{+1.0}) \times 10^{-9}$, where the uncertainty includes both the statistical and systematic components, but is dominated by the statistical uncertainties. The observed (expected median) significance of the excess is 4.3 (4.8) standard deviations and is determined by evaluating the ratio of the likelihood value for the hypothesis with no signal, divided by the likelihood with $B(B^0_\mu \rightarrow \mu^+\mu^-)$ floating. For this determination, $B(B^0_\mu \rightarrow \mu^+\mu^-)$ is allowed to float and is treated as a nuisance parameter in the fit (see the top plot in Fig. 2). The measured branching fraction is consistent with the expectation from the SM. With the 1D-BDT method, the observed (expected median) significance is 4.8 (4.7) standard deviations. Figure 3 shows the combined mass distributions weighted by $S/(S+B)$ for the categorized-BDT (left) and the 1D-BDT (right) methods. However, these distributions are illustrative only and were not used to obtain the final results.

No significant excess is observed for $B^0 \rightarrow \mu^+\mu^-$, and the upper limit $B(B^0 \rightarrow \mu^+\mu^-) < 1.1 \times 10^{-9}$ (9.2 $\times 10^{-10}$) at 95% (90%) confidence level (C.L.) is determined with the CL$_S$ approach [33,34], based on the observed numbers of events in the signal and sideband regions with the 1D-BDT method as summarized in Table I. The expected 95% C.L. upper limit for $B(B^0 \rightarrow \mu^+\mu^-)$ in the presence of SM signal plus background (background-only) is $6.3 \times 10^{-10}$ (5.4 $\times 10^{-10}$), where the statistical and systematic uncertainties are considered. The bottom plot in Fig. 2 shows the observed and expected CL$_S$ curves versus the assumed $B(B^0_\mu \rightarrow \mu^+\mu^-)$. From the fit, the branching fraction for this decay is determined to be...
$B(B^0 \to \mu^+ \mu^-) = (3.5^{+2.1}_{-1.8}) \times 10^{-10}$. The significance of this measurement is 2.0 standard deviations. The dimuon invariant mass distributions with the 1D-BDT method for the four channels are shown in the Supplemental Material [32].

In summary, a search for the rare decays $B^0 \to \mu^+ \mu^-$ and $B^0 \to \mu^+ \mu^-$ has been performed on a data sample of $pp$ collisions at $\sqrt{s} = 7$ and 8 TeV corresponding to integrated luminosities of 5 and 20 fb$^{-1}$, respectively. No significant evidence is observed for $B^0 \to \mu^+ \mu^-$ and an upper limit of $B(B^0 \to \mu^+ \mu^-) < 1.1 \times 10^{-9}$ is established at 95% C.L. For $B^0 \to \mu^+ \mu^-$, an excess of events with a significance of 4.3 standard deviations is observed, and a branching fraction of $B(B_s^0 \to \mu^+ \mu^-) = (3.0^{+1.0}_{-0.9}) \times 10^{-9}$ is determined, in agreement with the standard model expectations.

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: BMWF and FWF (Austria); FNRS and FWO (Belgium); CNEP, CAPES, FAPERJ, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COLCIENCIAS (Colombia); MSES (Croatia); RPF (Cyprus); MoER, SF0690030s09 and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); OTKA and NKTH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); NRF and WCU (Republic of Korea); LAS (Lithuania); CINVESTAV, CONACYT, SEP, and UASLP-FAI (Mexico); MBIE (New Zealand); NWO (Netherlands); NCN (Poland); FCT (Portugal); MOEER, SF0690030s09 and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); OTKA and NKTH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); NRF and WCU (Republic of Korea); LAS (Lithuania); CINVESTAV, CONACYT, SEP, and UASLP-FAI (Mexico); MBIE (New Zealand); FCT (Portugal); MON, RosAtom, RAS and RFBR (Russia); MESTD (Serbia); SEIDI and CPAN (Spain); Swiss Funding Agencies (Switzerland); NSC (Taipei); ThEPCenter, IPST, STAR and NSTDA (Thailand); TUBITAK and TAEK (Turkey); NASU (Ukraine); STFC (United Kingdom); DOE and NSF (U.S.).


aDeceased.
bAlso at Vienna University of Technology, Vienna, Austria.
cAlso at CERN, European Organization for Nuclear Research, Geneva, Switzerland.
dAlso at Institut Pluridisciplinaire Hubert Curien, Université de Strasbourg, Université de Haute Alsace Mulhouse, CNRS/IN2P3, Strasbourg, France.
eAlso at National Institute of Chemical Physics and Biophysics, Tallinn, Estonia.
fAlso at Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia.
gAlso at Universidade Estadual de Campinas, Campinas, Brazil.
hAlso at California Institute of Technology, Pasadena, CA, USA.
iAlso at Laboratoire Leprince-Ringuet, Ecole Polytechnique, IN2P3-CNRS, Palaiseau, France.
jAlso at Suez Canal University, Suez, Egypt.
kAlso at Zewail City of Science and Technology, Zewail, Egypt.
lAlso at Cairo University, Cairo, Egypt.
mAlso at Fayoum University, El-Fayoum, Egypt.
nAlso at British University in Egypt, Cairo, Egypt.
oPresent address: Ain Shams University, Cairo, Egypt.
pAlso at National Centre for Nuclear Research, Świerk, Poland.
qAlso at Université de Haute Alsace, Mulhouse, France.
rAlso at Joint Institute for Nuclear Research, Dubna, Russia.
sAlso at Brandenburg University of Technology, Cottbus, Germany.
tAlso at The University of Kansas, Lawrence, KS, USA.
uAlso at Institute of Nuclear Research ATOMKI, Debrecen, Hungary.
vAlso at Eötvös Loránd University, Budapest, Hungary.
wAlso at Tata Institute of Fundamental Research-EHEP, Mumbai, India.
xAlso at Tata Institute of Fundamental Research-HECR, Mumbai, India.
\(^{y}\)Present address: King Abdulaziz University, Jeddah, Saudi Arabia.
zAlso at University of Visva-Bharati, Santiniketan, India.
aaAlso at University of Ruhuna, Matara, Sri Lanka.
bbAlso at Isfahan University of Technology, Isfahan, Iran.
ccAlso at Sharif University of Technology, Tehran, Iran.
\(^{dd}\)Also at Plasma Physics Research Center, Science and Research Branch, Islamic Azad University, Tehran, Iran.
eAlso at Laboratori Nazionali di Legnaro dell’ INFN, Legnaro, Italy.
\(^{ff}\)Also at Università degli Studi di Siena, Siena, Italy.
\(^{gg}\)Also at Centre National de la Recherche Scientifique (CNRS)-IN2P3, Paris, France.
\(^{hh}\)Also at Purdue University, West Lafayette, IN, USA.
\(^{ii}\)Also at Universidad Michoacana de San Nicolas de Hidalgo, Morelia, Mexico.
\(^{jj}\)Also at Faculty of Physics, University of Belgrade, Belgrade, Serbia.
\(^{kk}\)Also at Facoltà Ingegneria, Università di Roma, Roma, Italy.
\(^{ll}\)Also at Scuola Normale e Sezione dell’INFN, Pisa, Italy.
\(^{mm}\)Also at University of Athens, Athens, Greece.
\(^{nn}\)Also at Rutherford Appleton Laboratory, Didcot, United Kingdom.
\(^{oo}\)Also at Paul Scherrer Institut, Villigen, Switzerland.
\(^{pp}\)Also at Institute for Theoretical and Experimental Physics, Moscow, Russia.
\(^{qq}\)Also at Albert Einstein Center for Fundamental Physics, Bern, Switzerland.
Also at Gaziosmanpasa University, Tokat, Turkey.

Also at Adiyaman University, Adiyaman, Turkey.

Also at Cag University, Mersin, Turkey.

Also at Mersin University, Mersin, Turkey.

Also at Izmir Institute of Technology, Izmir, Turkey.

Also at Ozyegin University, Istanbul, Turkey.

Also at Kafkas University, Kars, Turkey.

Also at Suleyman Demirel University, Isparta, Turkey.

Also at Ege University, Izmir, Turkey.

Also at Mimar Sinan University, Istanbul, Istanbul, Turkey.

Also at Kahramanmaras Sütçü Imam University, Kahramanmaras, Turkey.

Also at School of Physics and Astronomy, University of Southampton, Southampton, United Kingdom.

Also at INFN Sezione di Perugia, Università di Perugia, Perugia, Italy.

Also at Utah Valley University, Orem, UT, USA.

Also at Institute for Nuclear Research, Moscow, Russia.

Also at University of Belgrade, Faculty of Physics and Vinca Institute of Nuclear Sciences, Belgrade, Serbia.

Also at Argonne National Laboratory, Argonne, IL, USA.

Also at Erzincan University, Erzincan, Turkey.

Also at Yildiz Technical University, Istanbul, Turkey.

Also at Texas A&M University at Qatar, Doha, Qatar.

Also at Kyungpook National University, Daegu, Korea.