Search for the Flavor-Changing Neutral Current Decay $B_s^0 \rightarrow \mu^+\mu^-$ in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV with the D0 Detector

We present the results of a search for the flavor-changing neutral current decay $B^0_{s} \rightarrow \mu^+ \mu^-$ using a data set with integrated luminosity of 240 pb$^{-1}$ of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV collected with the D0 detector in run II of the Fermilab Tevatron collider. We find the upper limit on the branching fraction to be $\mathcal{B}(B^0_{s} \rightarrow \mu^+ \mu^-) \leq 5.0 \times 10^{-7}$ at the 95% C.L. assuming no contributions from the decay $B^0_{d} \rightarrow \mu^+ \mu^-$ in the signal region. This limit is the most stringent upper bound on the branching fraction $B^0_{s} \rightarrow \mu^+ \mu^-$ to date.

DOI: 10.1103/PhysRevLett.94.071802 PACS numbers: 13.20.He, 12.15.Mm, 14.40.Nd

The purely leptonic decays $B^0_{d,s} \rightarrow \mu^+ \mu^-$ [1] are flavor-changing neutral current (FCNC) processes. In the standard model (SM), these decays are forbidden at the tree level and proceed at a very low rate through higher-order diagrams. The SM leptonic branching fractions ($\mathcal{B}$) were calculated including QCD corrections in Ref. [2]. The latest SM prediction [3] is $\mathcal{B}(B^0_{s} \rightarrow \mu^+ \mu^-) = (3.42 \pm 0.54) \times 10^{-9}$, where the error is dominated by
nonperturbative uncertainties. The leptonic branching fraction of the $B^0_s$ decay is suppressed by Cabibbo-Kobayashi-Maskawa (CKM) matrix elements $|V_{td}/V_{ts}|^2$ leading to a predicted SM branching fraction of $(1.00 \pm 0.14) \times 10^{-10}$. The best existing experimental bound for the branching fraction of $B^0_s(B^0_s) \rightarrow \mu^+\mu^-$ is presently $\mathcal{B}(B^0_s(B^0_s) \rightarrow \mu^+\mu^-) < 7.5 \times 10^{-7} (1.9 \times 10^{-7})$ at the 95% C.L. [4].

The decay amplitude of $B^0_s \rightarrow \mu^+\mu^-$ can be significantly enhanced in some extensions of the SM. For instance, in the type-II two-Higgs-doublet model the branching fraction depends only on the charged Higgs boson mass $M_H^+$ and $\tan\beta$, the ratio of the two neutral Higgs field vacuum expectation values, with the branching fraction growing as $(\tan\beta)^5$ [5]. In the minimal supersymmetric standard model (MSSM), however, $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-) \propto (\tan\beta)^6$, leading to an enhancement of up to 3 orders of magnitude [6] compared to the SM, even if the MSSM with minimal flavor violation (MFV) is considered; i.e., the CKM matrix is the only source of flavor violation. An observation of $B^0_s \rightarrow \mu^+\mu^-$ would then immediately lead to an upper bound on the heaviest mass in the MSSM Higgs sector [7] if MFV applies. In minimal supergravity models, an enhancement of $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ is correlated [8] with a sizeable positive shift in $(g - 2)_\mu$ that also requires large $\tan\beta$. A large value of $\tan\beta$ is theoretically well motivated by grand unified theories based on minimal SO(10). These models predict large enhancements of $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ as well [8,9]. Finally, FCNC decays of $B^0_s$ are also sensitive to supersymmetric models with nonminimal flavor violation structures such as the generic MSSM [10] and $R$ parity violating supersymmetry [11].

In this Letter we report on a search for the decay $B^0_s \rightarrow \mu^+\mu^-$ using a data set of integrated luminosity of 240 pb$^{-1}$ recorded with the D0 detector in the years 2002–2004. Our mass resolution is not sufficient to readily separate $B^0_s$ from $B^0_s$ leptonic decays. For the final calculation of the upper limit on $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ we assumed that there is no contribution from $B^0_s \rightarrow \mu^+\mu^-$. We also required large $\tan\beta$. A large value of $\tan\beta$ is theoretically well motivated by grand unified theories based on minimal SO(10). These models predict large enhancements of $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ as well [8,9]. Finally, FCNC decays of $B^0_s$ are also sensitive to supersymmetric models with nonminimal flavor violation structures such as the generic MSSM [10] and $R$ parity violating supersymmetry [11]. For the final event selection, we required the candidate events to pass additional criteria. The long lifetime of the $B^0_s$ mesons allows us to reject random combinatorics background. We therefore used the decay length significance $L_{xy}/\delta L_{xy}$ as one of the discriminating variables, since it gives better discriminating power than the transverse decay length alone, as large values of $L_{xy}$ may originate due to large uncertainties.

The fragmentation characteristics of the $b$ quark are such that most of its momentum is carried by the $b$ hadron. Thus the number of extra tracks near the $B^0_s$ candidate tends to be small. The second discriminant was therefore an isolation variable, $I$, of the muon pair, defined as

$$I = \frac{|\vec{\rho}(\mu^+\mu^-)|}{|\vec{\rho}(\mu^+\mu^-)| + \sum_{\text{track } \neq B} p_i(\Delta R < 1)}.$$  

Here, $\sum_{\text{track } \neq B} p_i$ is the scalar sum over all tracks excluding the muon pair within a cone of $\Delta R < 1$ around the momentum vector $\vec{\rho}(\mu^+\mu^-)$ of the muon pair where $\Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2}$.

The final discriminating variable was the pointing angle $\alpha$, defined as the angle between the momentum vector $\vec{\rho}(\mu^+\mu^-)$ of the muon pair and the vector $\vec{\rho}_{vtx}$ between the primary and secondary vertices. This requirement ensured consistency between the direction of the decay vertex and the momentum vector of the $B^0_s$ candidate. An optimization based on these discriminating variables was done on signal Monte Carlo (MC) events in the $B^0_s$
mass region \(4.53 < M_{\mu^+\mu^-} < 6.15 \text{ GeV}/c^2\) with \(m_{B_s} = 5369.6 \pm 2.4 \text{ MeV}/c^2\) [14] and on data events in regions outside the signal window, i.e., in the sidebands. The mass scale throughout this analysis is shifted downward with respect to the world average \(B_s^0\) mass by \(30 \text{ MeV}/c^2\) to compensate for the shift in the momentum scale of the D0 tracking system. The mass shift was found by linear interpolation to the measured mass shifts between the \(J/\psi\) and the \(Y\) resonances relative to their world average values [14]. The mass shift is smaller than the MC predicted mass resolution for two-body decays of \(\sigma = 90 \text{ MeV}/c^2\) at the \(B_s^0\) mass.

In order to avoid biasing the optimization procedure, data candidates in the signal mass region were not examined until completion of the analysis, and events in the sideband regions around the \(B_s^0\) mass were used instead. The start (end) of the upper (lower) sideband was chosen such that they were at least \(3\sigma\) (270 MeV/c^2) away from the \(B_s^0\) mass. The widths of the sidebands used for background estimation were chosen to be 6\(\sigma\) each. The size of the blind signal region was \(3\sigma\) around the \(B_s^0\) mass. To determine the limit on the branching fraction, we used a smaller mass region of \(\pm 2\sigma\).

A random-grid search [15] and an optimization procedure [16] were used to find the optimal values of the discriminating variables, by maximizing the variable \(P = e_{\mu\mu}/(a/2 + \sqrt{N_{\text{back}}})\). Here, \(e_{\mu\mu}\) is the reconstruction efficiency of the signal events relative to the preselection (estimated using MC simulations), and \(N_{\text{back}}\) is the expected number of background events interpolated from the sidebands. The constant \(a\) is the number of standard deviations corresponding to the confidence level at which the signal hypothesis is tested. This constant \(a\) was set to 2.0, corresponding to about the 95% C.L. Figure 1 shows the distribution of the three discriminating variables after the preselection for signal MC events and data in the sideband regions. After optimization, we found the following values for the discriminating variables and MC signal efficiencies relative to the preselected sample: \(L_{xy}/\delta L_{xy} > 18.5\) (47.5\%), \(I > 0.56\) (97.4\%), and \(\alpha < 0.2\) rad (83.4\%). A linear extrapolation of the sideband population for the whole data sample into the \((\pm 180 \text{ MeV}/c^2)\) signal region yields an expected number of 3.7 \pm 1.1 background events.

Upon examining the data in all mass regions, four events are observed in the signal region, entirely consistent with the background events as estimated from sidebands. We examined the four observed events in detail by studying various kinematic variables, e.g., \(p_T\) of the muons, isolation, etc., and found them to be compatible with background events. Figure 2 shows the remaining events populating the lower and upper sidebands as well as the signal region almost equally.

In the absence of an apparent signal, a limit on the branching fraction \(\mathcal{B}(B_s^0 \to \mu^+\mu^-)\) can be computed by normalizing the upper limit on the number of events in the \(B_s^0\) signal region to the number of reconstructed \(B^0 \to J/\psi K^-\) events:

![FIG. 1 (color online). Discriminating variables after the preselection for signal MC (solid line) and data events (dashed line) from the sidebands. The arrows indicate the discriminating values that were obtained after optimization. The normalization is done on the number of signal MC and sideband data events after preselection.](image-url)
where \( N_{ul} \) is the upper limit on the number of signal decays, estimated from the number of observed events and expected background events and \( N_{B^\pm} \) is the observed number \( B^\pm \rightarrow J/\psi K^\pm \) events. The efficiencies of the signal and normalization channels obtained from MC simulations are \( \epsilon_{\mu^+\mu^-} \) and \( \epsilon_{B^\pm} \), respectively. The ratio \( B(B^\pm \rightarrow J/\psi (\mu^+\mu^-)K^\pm) = (1.00 \pm 0.04) \times 10^{-3} \) and \( B(J/\psi \rightarrow \mu^+\mu^-) = (5.88 \pm 0.10) \times 10^{-2} \). The ratio \( f_{B_d} - f_{B_d} / f_{B_d,B_{ud}} \) of a \( B_d \) meson being produced in the fragmentation compared to a \( B_{ud} \) meson is taken to be \( 0.270 \pm 0.034 \). This ratio has been calculated using the latest world average fragmentation values [14] for \( B_d \) and \( B_{ud} \) mesons, where the uncertainty on the ratio is conservatively calculated assuming a full anticorrelation among the individual \( B_{ud} \) and \( B_d \) fragmentation uncertainties.

The branching fraction ratio \( R = B(B_{d}^0) / B(B_{d}^0) \) of \( B_{d}^0 \) mesons decaying into two muons multiplied by the total detection efficiency ratio [17] is \( R \epsilon_{\mu^+\mu^-} / \epsilon_{B_d} \). Any non-negligible contribution due to \( B_{d}^0 \) decays (\( R > 0 \)) would make the limit on the branching fraction \( B(B_{d}^0 \rightarrow \mu^+\mu^-) \) as given in Eq. (2) smaller. Our limit presented for \( B_d \) decays, estimated from the number of observed events as well as the uncertainties on the branching fraction limit are listed in Table II. All systematic uncertainties entering the calculation of the branching fraction limit are listed in Table II.

We have used a prescription [19] to construct a confidence interval with the Feldman-Cousins ordering scheme. The expected background was modeled as a Gaussian distribution with its mean value equal to the expected number of background events and its standard deviation equal to the background uncertainty. The uncertainty on the number of \( B^\pm \) events as well as the uncertainties on the fragmentation ratio and branching fractions for \( B^\pm \rightarrow J/\psi (\mu^+\mu^-)K^\pm \) were added in quadrature to the efficiency uncertainties and parametrized as a Gaussian distribution.

The final corrected value for the efficiency ratio is then given by \( \epsilon_{B_d} / \epsilon_{B_{d}} = 0.247 \pm 0.009 \) (stat) \( \pm 0.017 \) (syst), where the first uncertainty is due to limited MC statistics and the second accounts for the \( B_d \) lifetime ratio uncertainties and for uncertainties in data-MC differences. All systematic uncertainties entering the calculation of the branching fraction limit are listed in Table II.

We have used a prescription [19] to construct a confidence interval with the Feldman-Cousins ordering scheme. The expected background was modeled as a Gaussian distribution with its mean value equal to the expected number of background events and its standard deviation equal to the background uncertainty. The uncertainty on the number of \( B^\pm \) events as well as the uncertainties on the fragmentation ratio and branching fractions for \( B^\pm \rightarrow J/\psi (\mu^+\mu^-)K^\pm \) were added in quadrature to the efficiency uncertainties and parametrized as a Gaussian distribution.

The resulting branching fraction limit [20] including all the statistical and systematic uncertainties at a 95% (90%) C.L. is given by

\[
\mathcal{B}(B_d^0 \rightarrow \mu^+\mu^-) \leq \frac{N_{ul}}{N_{B_d} \epsilon_{\mu^+\mu^-} \epsilon_{B_d}} \left[ \frac{B(B_d^0 \rightarrow J/\psi (\mu^+\mu^-)K^\pm)}{\frac{f_{B_d - f_{B_d,B_{ud}}}}{f_{B_d,B_{ud}}} + R \epsilon_{B^\pm} \epsilon_{\mu^+\mu^-}} \right],
\]

Gaussian for the signal and a second order polynomial for the background yields 741 \( \pm 31 \) (stat) \( \pm 22 \) (syst) \( B_d^0 \) candidates, where the systematic uncertainty was estimated by varying the fit range, background, and signal shape hypotheses.

The \( p_T \) distribution of the \( B_d^0 \) in data has a slightly harder spectrum than that from MC simulations. Therefore, MC events of the signal and normalization channels have been reweighted accordingly. In addition, the observed widths of known \( \mu^+\mu^- \) resonances [\( J/\psi \) and \( \Upsilon(1S) \)] are \( 27 \pm 4 \)% larger than predicted by MC simulations. The \( \pm 2 \sigma \) signal mass region using the MC mass resolution therefore corresponds to \( \pm 1.58 \sigma \) when the data mass resolution is considered, and the efficiency is corrected accordingly. To within errors, the MC calculation correctly reproduces the efficiency of the cuts on the discriminating variables when applied to the normalization channel.

The final corrected value for the efficiency ratio is then given by \( \epsilon_{B_d} / \epsilon_{B_{d}} = 0.247 \pm 0.009 \) (stat) \( \pm 0.017 \) (syst), where the first uncertainty is due to limited MC statistics and the second accounts for the \( B_d \) lifetime ratio uncertainties and for uncertainties in data-MC differences. All systematic uncertainties entering the calculation of the branching fraction limit are listed in Table II.

We have used a prescription [19] to construct a confidence interval with the Feldman-Cousins ordering scheme. The expected background was modeled as a Gaussian distribution with its mean value equal to the expected number of background events and its standard deviation equal to the background uncertainty. The uncertainty on the number of \( B^\pm \) events as well as the uncertainties on the fragmentation ratio and branching fractions for \( B^\pm \rightarrow J/\psi (\mu^+\mu^-)K^\pm \) were added in quadrature to the efficiency uncertainties and parametrized as a Gaussian distribution.

The resulting branching fraction limit [20] including all the statistical and systematic uncertainties at a 95% (90%) C.L. is given by

\[
\mathcal{B}(B_d^0 \rightarrow \mu^+\mu^-) \leq \frac{N_{ul}}{N_{B_d} \epsilon_{\mu^+\mu^-} \epsilon_{B_d}} \left[ \frac{B(B_d^0 \rightarrow J/\psi (\mu^+\mu^-)K^\pm)}{\frac{f_{B_d - f_{B_d,B_{ud}}}}{f_{B_d,B_{ud}}} + R \epsilon_{B^\pm} \epsilon_{\mu^+\mu^-}} \right],
\]
The ratio $\mathcal{B}(B^0_s \to J/\psi K^\pm) \leq 5.0 \times 10^{-7} (4.1 \times 10^{-7})$.

We also used a Bayesian approach with flat prior and Gaussian (smeared) uncertainties [22] and obtained the limit of $\mathcal{B}(B^0_s \to \mu^+ \mu^-) \leq 5.1 \times 10^{-7} (4.1 \times 10^{-7})$ at the 95% (90%) C.L. This new result is presently the most stringent bound on $\mathcal{B}(B^0_s \to \mu^+ \mu^-)$, improving the previously published value [4] and can be used to constrain models of new physics beyond the SM.

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the Department of Energy and National Science Foundation (USA), Commissariat à l’Énergie Atomique and CNRS/Institut National de Physique Nucléaire et de Physique des Particules (France), Ministry of Education and Science, Agency for Atomic Energy and RF President Grants Program (Russia), CAPES, CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil), Departments of Atomic Energy and Science and Technology (India), Colciencias (Colombia), CONACyT (Mexico), KRF (Korea), CONICET and UBACyT (Argentina), The Foundation for Fundamental Research on Matter (The Netherlands), PPARC (United Kingdom), Ministry of Education (Czech Republic), Natural Sciences and Engineering Research Council and WestGrid Project (Canada), BMBF and DFG (Germany), A. P. Sloan Foundation, Research Corporation, Texas Advanced Research Program, and the Alexander von Humboldt Foundation.

*Visiting from University of Zurich, Zurich, Switzerland.
†Visiting from Institute of Nuclear Physics, Krakow, Poland.

[1] Charge conjugated states are included implicitly.
[17] The ratio $\epsilon_{\mu\mu}^{B^0_s}/\epsilon_{\mu\mu}^{B^0_s}$ has been determined from simulation to be 0.92 ± 0.04, with the uncertainty due to limited MC statistics.
[18] In addition to $B^\pm \to J/\psi K^\pm$, the other possible normalization channel is $B^0 \to J/\psi \phi$. We have not used it due to low statistics, a large uncertainty on its branching fraction, and a poorly known mixture of CP even and CP odd decay modes with lifetime differences.
[20] This limit is derived with the world average fragmentation value [14] of $f_{B \to B_{s\ell}} = 0.270 \pm 0.034$. A fragmentation ratio based on Tevatron data alone [21], but with larger uncertainty, gives $f_{B \to B_{s\ell}} = 0.42 \pm 0.14$ and results in a limit of $4.1 \times 10^{-7}$ ($3.1 \times 10^{-7}$) at a 95% (90%) C.L. using the method of Ref. [19].