A Search for the Flavor-Changing Neutral Current Decay $B^0 \rightarrow \mu^+\mu^-$ in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV with the DØ 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 pp collisions at \( \sqrt{s} = 1.96 \text{ TeV} \) collected with the DØ detector in Run II of the Fermilab Tevatron collider. We find the upper limit on the branching fraction to be \( B(B^0_s \rightarrow \mu^+\mu^-) < 5.0 \times 10^{-7} \) at the 95% C.L. assuming no contributions from the decay \( B^0_s \rightarrow \mu^+\mu^- \) in the signal region. This limit is the most stringent upper bound on the branching fraction \( B^0_s \rightarrow \mu^+\mu^- \) to date.

PACS numbers: 13.20.He, 12.15.Mm, 14.40.Nd

The purely leptonic decays \( B^0_s \rightarrow \mu^+\mu^- \) [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 (\( B \)) were calculated including QCD corrections in Ref. [2]. The latest SM prediction [3] is \( B(B^0_s \rightarrow \mu^+\mu^-) = (3.47 \pm 0.64) \times 10^{-6} \), where the error is dominated by non-perturbative uncertainties. The leptonic branching fraction of the \( B^0_s \) decay is suppressed by 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 $B^0_d (B^0_s)$ is presently $B(B^0_d (B^0_s) \rightarrow \mu^+ \mu^-) < 7.5 \times 10^{-7}$ $(1.9 \times 10^{-7})$ at the 95% C.L. [4].

The decay amplitude of $B^0_d (B^0_s) \rightarrow \mu^+ \mu^-$ can be significantly enhanced in some extensions of the SM. For instance, in the type-II two-Higgs-doublet model (2HDM) the branching fraction depends only on the charged Higgs mass $M_{H^+}$ and $\tan \beta$, the ratio of the two neutral Higgs field vacuum expectation values, with the branching fraction growing like $(\tan \beta)^4$ [5]. In the minimal supersymmetric standard model (MSSM), however, $B(B^0_d \rightarrow \mu^+ \mu^-) \propto (\tan \beta)^5$, leading to an enhancement of up to three 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 $B^0_d \rightarrow \mu^+ \mu^-$ 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 $B(B^0_d \rightarrow \mu^+ \mu^-)$ 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 (GUT) based on minimal SO(10). These models predict large enhancements of $B(B^0_d \rightarrow \mu^+ \mu^-)$ as well [8, 9]. Finally, FCNC decays of $B^0_d$, are also sensitive to supersymmetric models with non-minimal 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 $B^0_d \rightarrow \mu^+ \mu^-$ 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 $B^0_d$ from $B^0_s$ lepton decays. For the final calculation of the upper limit on $B(B^0_d \rightarrow \mu^+ \mu^-)$ we assumed that there is no contribution from $B^0_s \rightarrow \mu^+ \mu^-$ decays in our search region due to its suppression by $|V_{td}/V_{ts}|^2$, which holds in all models with MFV.

The D0 detector is described in detail elsewhere [13]. The main elements, relevant for this analysis, are the central tracking and muon detector systems. The central tracking system consists of a silicon microstrip tracker (SMT) and a central fiber tracker (CFT), both located within a 2 T superconducting solenoidal magnet. Located outside the calorimeter, the muon detector consists of a layer of tracking detectors and scintillation trigger counters in front of toroidal magnets (1.8 T), followed by two more similar layers behind the toroids, allowing for efficient muon detection out to $\eta$ of about ±2, where $\eta = - \ln[\tan(\theta/2)]$ is the pseudorapidity and $\theta$ is the polar angle measured relative to the proton beam direction.

Four versions of dimuon triggers were used in the data selection of this analysis. A trigger simulation was used to estimate the trigger efficiency for the signal and normalization samples. These efficiencies were also checked with data samples collected with single muon triggers. Event preselection started by requiring two muons of opposite charge to be identified by extrapolating charged tracks reconstructed in the central tracking detectors to the muon detectors, and matching them with hits in the latter. The muons had to form a common secondary 3D-vertex with an invariant mass $m(\mu^+ \mu^-)$ between 4.5 and 7.0 GeV/c$^2$ and a $\chi^2$ per degree of freedom of $\chi^2$/d.o.f $< 10$. Each muon was required to have $p_T > 2.5$ GeV/c and $|\eta| < 2.0$. Tracks that were matched to each muon were required to have at least three hits in the SMT and at least four hits in the CFT. To select well-measured secondary vertices, we determined the two-dimensional decay length $L_{xy}$ in the plane transverse to the beamline, and required the uncertainty $\delta L_{xy}$ to be less than 0.15 mm. $L_{xy}$ was calculated as $L_{xy} = \frac{L_{tx} p_T^B}{p_T^B}$, where $p_T^B$ is the transverse momentum of the candidate $B^0_d$ and $L_{tx}$ represents the vector pointing from the primary vertex to the secondary vertex. The error on the transverse decay length, $\delta L_{xy}$, was calculated by taking into account the uncertainties in both the primary and secondary vertex positions. The primary vertex itself was found for each event using a beam-spot constrained fit as described in Ref. [14]. To ensure a similar $p_T$ dependence of the $\mu^+ \mu^-$ system in the signal and in the normalization channel, $p_T^B$ had to be greater than 5 GeV/c. A total of 38,167 events survive these preselection requirements. The effects of these criteria on the number of events are shown in Table I.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Requirement</th>
<th># Candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass (GeV/c$^2$)</td>
<td>$4.5 &lt; m_{\mu^+\mu^-} &lt; 7.0$</td>
<td>405,307</td>
</tr>
<tr>
<td>Muon quality</td>
<td></td>
<td>234,792</td>
</tr>
<tr>
<td>$\chi^2$/d.o.f of vertex</td>
<td>$&lt; 10$</td>
<td>146,982</td>
</tr>
<tr>
<td>Muon $p_T$ (GeV/c)</td>
<td>$&gt; 2.5$</td>
<td>129,558</td>
</tr>
<tr>
<td>Muon $</td>
<td>\eta</td>
<td>$</td>
</tr>
<tr>
<td>Tracking hits</td>
<td>CFT $\geq 4$, SMT $\geq 3$</td>
<td>92,678</td>
</tr>
<tr>
<td>$\delta L_{xy}$ (mm)</td>
<td>$&lt; 0.15$</td>
<td>90,935</td>
</tr>
<tr>
<td>$B^0_d$ candidate $p_T^B$ (GeV/c)</td>
<td>$&gt; 5.0$</td>
<td>38,167</td>
</tr>
</tbody>
</table>

For the final event selection, we required the candidate events to pass additional criteria. The long lifetime of the $B^0_d$ mesons allows us to reject random combinatoric 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 $B^0_d$
candidate tends to be small. The second discriminant was therefore an isolation variable, $I$, of the muon pair, defined as:

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

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 $p_{[\mu^+\mu^-]}$ 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 $\vec{p}_{[\mu^+\mu^-]}$ of the muon pair and the vector $\vec{l}_{\text{vertex}}$ between the primary and secondary vertices. This requirement ensured consistency between the direction of the decay vertex and the momentum vector of the $B_0^0$ candidate.

An optimization based on these discriminating variables was done on signal Monte Carlo (MC) events in the $B_0^0$ mass region $4.53 < M_{\mu^+\mu^-} < 6.15$ GeV/c$^2$ with $m_{B_0^0} = 5369.6 \pm 2.4$ MeV/c$^2$ [12] 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 $B_0^0$ mass by 30 MeV/c$^2$ to compensate for the shift in the momentum scale of the DØ tracking system. The mass shift was found by linear interpolation to the $B_0^0$ mass of the measured mass shifts between the $J/\psi$ and the $\Upsilon$ resonances relative to their world average values [12]. The mass shift is smaller than the MC predicted mass resolution for two-body decays of $\sigma = 90$ MeV/c$^2$ at the $B_0^0$ 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 $B_0^0$ 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 $B_0^0$ 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 $B_0^0$ mass. To determine the limit on the branching fraction, we used a smaller mass region of ±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 $P = \epsilon_{[\mu^+\mu^-]}/(a/2 + \sqrt{N_{\text{back}}})$. Here, $\epsilon_{[\mu^+\mu^-]}$ is the reconstruction efficiency of the signal events relative to the preselection (estimated using MC), 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: $L_{\text{xy}}/\delta L_{\text{xy}} > 18.5$ (47.5%), $I > 0.56$ (97.4%), and $\alpha < 0.2$ rad (83.4%). The combined efficiency for signal events to survive these three additional selection criteria, as measured relative to preselection criteria, is $(38.6\pm0.7)\%$, where the error is due to limited MC statistics.

In the absence of an apparent signal, a limit on the branching fraction $B(B_0^0 \rightarrow \mu^+\mu^-)$ can be computed by normalizing the upper limit on the number of events in the $B_0^0$ signal region to the number of reconstructed $B^\pm \rightarrow J/\psi K^\pm$ events:

$$B(B_0^0 \rightarrow \mu^+\mu^-) \leq \frac{N_{ul} \epsilon_{[\mu^+\mu^-]}^{B_{0^0}}}{N_B \epsilon_{[\mu^+\mu^-]}^{B^\pm}} \frac{B(B^\pm \rightarrow J/\psi(\mu^+\mu^-)K^\pm)}{f_{B_0 \rightarrow B_{u,d}} \epsilon_{[\mu^+\mu^-]}^{B_{u,d}}},$$

where

- $N_{ul}$ is the upper limit on the number of signal decays estimated from the number of observed events and expected background events.
- $N_B$ is the number of observed $B^\pm \rightarrow J/\psi K^\pm$ events.
- $\epsilon_{[\mu^+\mu^-]}^{B_{0^0}}$ and $\epsilon_{[\mu^+\mu^-]}^{B^\pm}$ are the efficiencies of the signal and normalization channels, obtained from MC simulations.
- $B(B^\pm \rightarrow J/\psi(\mu^+\mu^-)K^\pm)$ is the product of the branching fractions $B(B^\pm \rightarrow J/\psi K^\pm) = (1.00 \pm 0.04) \times 10^{-3}$ and $B(J/\psi \rightarrow \mu^+\mu^-) = (5.88\pm0.10) \times 10^{-2}$ [12].
- $f_{B_0 \rightarrow B_{u,d}}/f_{B \rightarrow B_{u,d}} = 0.270 \pm 0.034$ is the fragmentation ratio of a $b$ quark producing a $B_0^0$ and a $B_{u,d}$ meson. This ratio has been calculated using the latest world average fragmentation values [12] for $B_0^0$ and $B_{u,d}$ mesons, where the uncertainty on the ratio is conservatively calculated assuming a full
FIG. 1: 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.

$$\chi^2$$

FIG. 2: Invariant mass of the remaining events of the full data sample after optimized requirements on the discriminating variables.

FIG. 3: Invariant mass distribution for candidates in the normalization channel $B^\pm \rightarrow J/\psi K^\pm$.

anti-correlation among the individual $B_{u,d}$ and $B^0_s$ fragmentation uncertainties.

- $R \cdot \frac{p_{\mu^+}^2}{p_{\mu^-}^2}$ is the branching fraction ratio $R = \mathcal{B}(B^0_d) / \mathcal{B}(B^0_s)$ of $B^0_d$ mesons decaying into two muons multiplied by the total detection efficiency ratio [17]. Any non-negligible contribution due to $B^0_d$ decays ($R > 0$) would make the limit on the branching fraction $\mathcal{B}(B^0_d \rightarrow \mu^+\mu^-)$ as given in Eq. (2) smaller. Our limit presented for $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ is therefore conservative.

Using the $B^\pm \rightarrow J/\psi K^\pm$ mode [18] has the advantage that the efficiencies to detect the $\mu^+\mu^-$ system in signal and normalization events are similar, and systematic effects tend to cancel. A pure sample of $B^\pm \rightarrow J/\psi K^\pm$ events was obtained by applying the following selection criteria. The mass-constrained vertex fit of the two muons to form a $J/\psi$ was required to have a $\chi^2$/d.o.f. < 10, similar to the $\mu^+\mu^-$ vertex criterion in the $B^0_d \rightarrow \mu^+\mu^-$ search. The combined vertex fit of the $J/\psi$ and the additional $K^\pm$ ($p_T(K^\pm) > 0.9$ GeV/c) had to have $\chi^2 < 20$ for three d.o.f.. The requirements on the three discriminating variables were also applied. The mass spectrum of the reconstructed $B^\pm \rightarrow J/\psi K^\pm$ for the full data sample after all analysis requirements is shown in Fig. 3. A fit using a Gaussian for the signal and a second order polynomial for the background yields $741 \pm 31$ (stat) $\pm 22$ (sys) $B^\pm$ candidates, where the systematic uncertainty was estimated by varying the fit range, background and signal shape hypotheses.

The $p_T$ distribution of the $B^\pm$ in data has a slightly harder spectrum than that from MC. 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. 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 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 $e_{R/K} = 0.247 \pm 0.009$ (stat) $\pm 0.017$ (sys), where the first uncertainty is due to limited MC statistics and the second accounts for the $B^\pm/B^0_s$ lifetime ratio uncertainties and for uncertainties in data/MC differences. These differences include the $p_T$-dependent reweighting of MC events, signal mass width, the kaon track reconstruction efficiency and the effects of different trigger and muon identification efficiencies. All systematic uncer-
TABLE II: Relative uncertainties used in the calculation of an upper limit of \( \mathcal{B}(B_s^0 \to \mu^+ \mu^-) \).

<table>
<thead>
<tr>
<th>Source</th>
<th>Relative Uncertainty [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon_{\mu K} / \epsilon_{\mu K}^\text{n} )</td>
<td>7.7</td>
</tr>
<tr>
<td>Number of ( B^\pm \to J/\psi K^\pm ) events</td>
<td>5.1</td>
</tr>
<tr>
<td>( \mathcal{B}(B^\pm \to J/\psi K^\mp) )</td>
<td>4.0</td>
</tr>
<tr>
<td>( \mathcal{B}(J/\psi \to \mu^+ \mu^-) )</td>
<td>1.7</td>
</tr>
<tr>
<td>( f_{B^\pm} / f_{B^0} )</td>
<td>12.7</td>
</tr>
<tr>
<td>Background uncertainty</td>
<td>29.7</td>
</tr>
</tbody>
</table>

The statistical uncertainties on the background expectation, as well as the uncertainties on the efficiencies can be included into the limit calculation by integrating over probability functions that parameterize the uncertainties. We have used a prescription [19] to construct a confidence interval with the Feldman and 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 \( B^\pm \to J/\psi (\mu^+ \mu^-) K^\pm \) were added in quadrature to the efficiency uncertainties and parameterized 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

\[
\mathcal{B}(B_s^0 \to \mu^+ \mu^-) \leq 5.0 \times 10^{-7} \ (4.1 \times 10^{-7}).
\]

We also used a Bayesian approach with flat prior and Gaussian (smeared) uncertainties [22] and obtained the limit of \( \mathcal{B}(B_s^0 \to \mu^+ \mu^-) \leq 5.1 \times 10^{-7} \ (4.1 \times 10^{-7}) \) at the 95% (90%) C.L. This new result is presently the most stringent bound on \( \mathcal{B}(B_s^0 \to \mu^+ \mu^-) \), 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’Energie 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.

[1] Visitor from University of Zurich, Zurich, Switzerland.
[2] Visitor from Institute of Nuclear Physics, Krakow, Poland.
[3] Charge conjugated states are included implicitly.
[18] The ratio \( \epsilon_{\mu K} / \epsilon_{\mu K}^\text{n} \) has been determined from simulation to be 0.92±0.04, with the uncertainty due to limited MC statistics.
[19] In addition to \( B^\pm \to J/\psi K^\pm \), the other possible normalization channel is \( B_s^0 \to J/\psi \phi \). 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.
[21] This limit is derived with the world average fragmentation value [12] of \( f_{B_s^0} / f_{B_s^0}^\text{n} \to D_{u,w} = 0.270±0.034 \). A fragmentation ratio based on Tevatron data alone [21], but with larger uncertainty, gives \( f_{B_s^0} / f_{B_s^0}^\text{n} = 0.42±0.14 \) 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].