DØnote 4514-CONF

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

The DØ Collaboration
URL: http://www-d0.fnal.gov

(Dated: August 2, 2004)

We present results of a search for the Flavor-Changing Neutral Current decay $B_s \to \mu^+\mu^-$ using 240 pb^{-1} of $p\bar{p}$ collision data at $\sqrt{s} = 1.96$ TeV collected by the DØ detector in Run II of the Fermilab Tevatron collider. The selection criteria were optimized using a random grid search to enhance the signal sensitivity. In a "blind" analysis, the mass regions below and above the signal region were used to determine the shape and normalization of the background. The signal region contained four events, which was consistent with the expectation of 3.7±1.1 background events. To determine the limit on the branching fraction, $B^\pm \to J/\psi\, K^\pm$ was used as the normalizing mode. In the absence of a signal an upper limit on the branching fraction, including systematic uncertainties, has been determined to be $\mathcal{B}(B_s \to \mu^+\mu^-) \leq 4.6 \cdot 10^{-7}$ at the 95% C.L.

Preliminary Results for Winter 2004 Conferences

The purely leptonic decay $B_{d,s} \to \mu^+\mu^-$ is a Flavor-Changing Neutral Current (FCNC) process [1]. In the Standard Model (SM), this decay is forbidden at the tree level and proceeds at a very low rate through higher order diagrams. The SM branching fraction ($\mathcal{B}$) for this channel was first calculated in [2] and later refined to include QCD corrections [3]. The latest SM predictions [4] are, $\mathcal{B}(B_s \to \mu^+\mu^-) = (3.42 \pm 0.54) \cdot 10^{-9}$, where the error is dominated by non-perturbative hadronic uncertainties. The corresponding leptonic branching fraction for the B_d is suppressed by an additional factor of $|V_{td}/V_{ts}|^2$ leading to a SM branching ratio of $(1.00 \pm 0.14) \cdot 10^{-10}$. Presently, the best existing experimental bound for the branching fraction of B_s (B_d) is $\mathcal{B}(B_s\,(B_d) \to \mu^+\mu^-) < 7.5 \cdot 10^{-7}\,(1.9 \cdot 10^{-7})$ at the 95% C.L. [5].

The decay amplitude of $B_{d,s} \to \mu^+\mu^-$ can be significantly enhanced in some extensions of the SM. For instance, in the type-II two Higgs Doublet Model (2HDM), all contributions from the neutral Higgs sector cancel out and the branching fraction depends only on the charged Higgs mass M_{H^+} and $\tan\beta$. The amplitude grows like $\tan^4\beta$ [6]. In the Minimal Supersymmetric Standard Model (MSSM) however, $\mathcal{B}(B_s \to \mu^+\mu^-) \propto \tan^6\beta$, leading to an enhancement of up to three orders of magnitude [7] compared to the SM, even if MSSM with minimal flavor violation (MFV) is considered, i.e., the CKM matrix is the only source of flavor violation. In minimal supergravity, an enhancement of $\mathcal{B}(B_s \to \mu^+\mu^-)$ 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 (GUT) models based on minimal SO(10). These models predict large enhancements of $\mathcal{B}(B_s \to \mu^+\mu^-)$ as well [8],[9]. Finally, FCNC decays of $B_{d,s}$ are also sensitive to supersymmetric models with non-minimal flavor structures such as the generic MSSM [10] and R_p violating SUSY [11].

In this note we report on a search for the decay $B_s \to \mu^+\mu^-$ using 240 pb^{-1} of data recorded with the DØ detector in the years 2002–2004. A limit on the branching fraction $\mathcal{B}(B_s)$ can then be computed by normalizing the upper limit of number of events in the B_s signal region to the number of reconstructed $B^\pm \to J/\psi\, K^\pm$ events [12]:

$$\mathcal{B}(B_s) \leq \frac{N_{\rm ul}}{N_{B^\pm}} \cdot \frac{\epsilon^{B^\pm}_{\mu\mu K}}{\epsilon^{B_s}_{\mu\mu}} \cdot \frac{\mathcal{B}_1(B^\pm) \cdot \mathcal{B}_2(J/\psi)}{\frac{f_{b\to B_s}}{f_{b\to B_{u,d}}} + R \cdot \frac{\epsilon^{B_d}_{\mu\mu}}{\epsilon^{B_s}_{\mu\mu}}}, \tag{1}$$

where

- $N_{\rm ul}$ is the upper limit on the number of signal events;
- $N_{B^\pm}$ is the number of observed $B^\pm \to J/\psi\, K^\pm$ events;
- $\epsilon^{B_s}_{\mu\mu}$ and $\epsilon^{B^\pm}_{\mu\mu K}$ are the efficiencies of the signal and normalization channels, obtained from Monte Carlo (MC) simulations;
- $f_{b\to B_s}/f_{b\to B_{u,d}} = 0.270 \pm 0.034$ is the fragmentation ratio of a b or $\bar{b}$ quark producing a B_s and a $B^\pm$ or B_d. This ratio has been calculated using the latest world average values [13] for the fragmentation for $B_{u,d}$ and B_s mesons respectively. The error on the ratio is calculated assuming a full anti-correlation among the individual $B_{u,d}$ and B_s fragmentation uncertainties;
- $\mathcal{B}_1 = \mathcal{B}(B^\pm \to J/\psi\, K^\pm) = (1.00 \pm 0.04) \cdot 10^{-3}$ and $\mathcal{B}_2 = \mathcal{B}(J/\psi \to \mu\mu) = (5.88 \pm 0.1)\%$ [13]; and
- $R \cdot \epsilon^{B_d}_{\mu\mu}/\epsilon^{B_s}_{\mu\mu}$ is the branching fraction ratio $\mathcal{B}(B_d)/\mathcal{B}(B_s)$ of $B_{d,s}$ mesons decaying into two muons multiplied by their total efficiency ratio [14].

To simplify the calculation of the upper limit on the branching fraction $\mathcal{B}(B_s \to \mu^+\mu^-)$ as given in Eq. 1, it is assumed that there are no contributions from $B_d \to \mu^+\mu^-$ decays (i.e., $R = 0$) in our search region due to its suppression by $|V_{td}/V_{ts}|^2$, which holds in all models with MFV. Any non-negligible contribution due to B_d decays ($R > 0$) would make the obtained limit on the branching fraction $\mathcal{B}(B_s \to \mu^+\mu^-)$ as given in Eq. 1 smaller. Our limit presented for $\mathcal{B}(B_s \to \mu^+\mu^-)$ is therefore conservative.

The DØ detector is described elsewhere [15]. The main elements, relevant for this analysis, are the central tracking and muon detector system. 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. The muon detector located outside the calorimeter 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 after the toroids, allowing for efficient detection out to pseudorapidity (η) of about 2.0.

The data selected in this analysis were triggered by four separate dimuon triggers. Trigger efficiencies for the signal and normalization samples were estimated using a (trigger) simulation software package. These efficiencies were also checked with data samples collected with unbiased or single muon triggers.

TABLE I: Number of candidate events surviving the cuts in data used in the pre-selection analysis.

Cut	Value	# candidates
Mass window (GeV/c^2)	$4.5 < m_{\mu^+\mu^-} < 7.0$	405,307
Good muon quality		234,792
$\chi^2/d.o.f$ of vertex	< 10	146,982
Muon p_T (GeV/c)	> 2.5	129,558
Muon $\lvert\eta\rvert$	< 2.0	125,679
Tracking hits	CFT$>$ 3, SMT $>$ 2	92,678
$\delta L_{x,y}$ (mm)	< 0.15	90,935
B_s candidate p_T^B (GeV/c)	> 5.0	38,167

Event (pre-)selection started with requiring two muons 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 3D-vertex with an invariant mass between 4.5 and 7.0 GeV/c^2 and a $\chi^2/d.o.f.$ of less than 10.

The transverse momentum, p_T, of each of the muons was required to be greater than 2.5 GeV/c and $|\eta|$ less than 2.0. Tracks that were matched to each muon were required to have at least three hits in the SMT and four hits in the CFT. To select well measured secondary vertices, we determined the two-dimensional decay length in the plane transverse to the beamline (L_{xy}), and required the error ($\delta L_{x,y}$) on it to be less than 150 microns. L_{xy} was calculated as $L_{xy} = \frac{\vec{l}_{Vtx}\cdot\vec{p}_T^B}{p_T^B}$. Here, $\vec{l}_{Vtx}$ represents the vector pointing from the primary vertex to the secondary vertex. The error on the transverse decay length, $\delta L_{x,y}$, 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 [16]. To ensure a similar p_T dependence of the $\mu^+\mu^-$-system in the signal as well as in the normalization channel, the transverse momentum of the candidate B_s, p_T^B, had to be greater than 5 GeV/c. After pre-selection, a total of 38,167 signal candidate events survived these base requirements. The effect of the various pre-selection criteria on data events is shown in Table I.

For the final event selection we required the candidate events to pass additional criteria. The long lifetime of the B_s mesons allowed us to reject random combinatoric background, e.g., two fake muons that formed a good vertex. We therefore used the decay length significance, $L_{xy}/\delta L_{xy}$, as one of the criteria, since it gives better discriminating power than the transverse decay length alone (where large values may be due to large errors).

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 candidate tend to be low. The second discriminant was therefore an isolation variable, I, of the muon pair, defined as:

$$I = \frac{|\vec{p}(\mu^+\mu^-)|}{|\vec{p}(\mu^+\mu^-)| + \sum\limits_{\text{track i}\neq\text{B}} p_i(\Delta\mathcal{R} < 1)}.$$

Here, $\sum\limits_{\text{track } i\neq B} p_i$, is the scalar sum over all tracks excluding the muon pair within a cone of $\Delta\mathcal{R} < 1$ around the momentum vector $\vec{p}(\mu^+\mu^-)$ of the dimuon pair where $\Delta\mathcal{R} = \sqrt{(\Delta\Phi)^2 + (\Delta\eta)^2}$. To reduce the effect of overlapping events coming from the same bunch crossing, all tracks that were counted in the sum had to satisfy the additional requirement that the distance between the track and the vertex position of the muon pair, measured along the beam direction (z-axis), had to be smaller than 5 cm.

The final discriminating variable was the pointing angle α, defined as the angle between the momentum vector $\vec{p}(\mu^+\mu^-)$ of the dimuon pair and the vector $\vec{l}_{Vtx}$ between primary and secondary vertex. This requirement ensured consistency between the direction of the decay vertex and the momentum vector of the B_s candidate.

An optimization based on these discriminating variables was done on candidate signal MC events in the mass region, $4.53 < M_{\mu^+\mu^-} < 6.15$ GeV/c^2, containing the region around the B_s, $m_{B_s^0} = 5369.6 \pm 2.4$ MeV/c^2 [13]. This region of interest is shifted downward with respect to the world average B_s mass by 30 MeV/c^2 to account for uncertainties in the momentum scale of the DØ tracking system. The 30 MeV/c^2 mass shift valid at the scale of the B-meson mass was found by a linear extrapolation of the measured mass shifts between the J/ψ and the Υ resonances with respect to their world average values [13]. This shift by 30 MeV/c^2 is smaller than our expected mass resolution for two-body decays of 90 Me/c^2V at the B_s mass.

In order to avoid biasing the optimization procedure, the signal box was kept "hidden" and events in the sideband regions around the B_s mass were used instead. The start (end) of the upper (lower) sideband was chosen such that

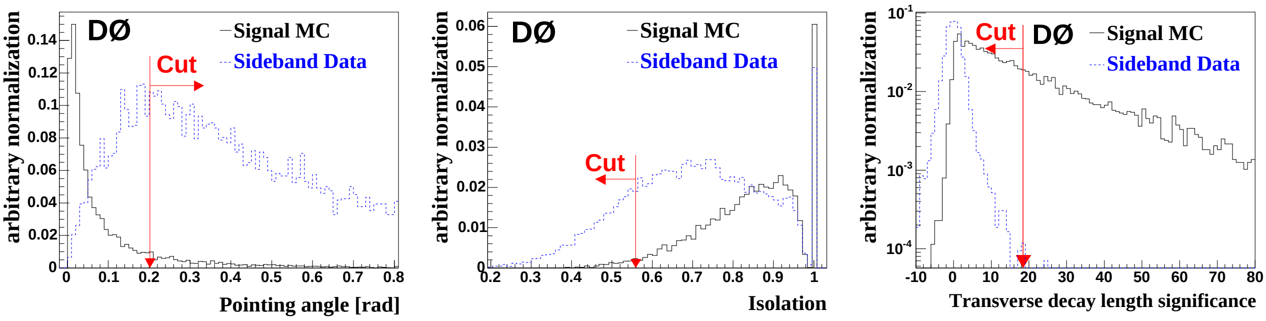

FIG. 1: Discriminating variables after the pre-selection for signal MC and data events from the sidebands. The arrows indicate the cut values that were obtained after optimization.

the interval limits were at least 3σ away from the shifted B_s mass. The width of the sidebands that were used for background estimation were chosen to be 6σ each. In the optimization phase, the size of the blind signal region was ± 270 MeV$/c^2$ around the world average value of the B_s mass (shifted by 30 MeV$/c^2$) which corresponds to $\pm 3\sigma$ of the expected mass resolution in $B_s \rightarrow \mu^+\mu^-$. To determine the final limit, we used a smaller mass region of $\pm 2\sigma$.

A Random Grid Search [17] and an optimization procedure proposed by Punzi [18] were used to find the optimal values of the discriminating variables, by maximizing the variable $P = \epsilon'^{B_s}_{\mu\mu,Pre}/(a/2 + \sqrt{N_{\rm Back}})$. Here, $\epsilon'^{B_s}_{\mu\mu}$ is the reconstruction efficiency of the signal events after the pre-selection (estimated using MC) and $N_{\rm Back}$ is the expected number of background events extrapolated from the sidebands. The constant a is the number of sigmas corresponding to the confidence level at which the signal hypothesis is tested. This number a should be defined before the statistical test and was set to 2.0, corresponding to about 95% C.L.

The optimization was performed using approximately 1/3 of the available data and on the complete set of signal MC events. Figure 1 shows the distribution of the three discriminating variables after the pre-selection for signal MC events and data in the sideband regions. After optimization, we found the following values for the cut variables and MC signal efficiencies relative to the pre-selected sample: $\alpha < 0.203$ (83.4%), $L_{xy}/\delta L_{xy} > 18.47$ (47.5%) and $I > 0.56$ (97.4%). The combined efficiency for signal events to survive these three additional selection criteria, as measured relative to pre-selection criteria, was (38.6±0.7)%, where the error is due to limited MC statistics.

A linear extrapolation of the sideband population for the whole data sample into the ($\pm 2\sigma$) signal region yielded an expected number of background events to be 3.7±1.1. This corresponds to a background rejection of approximately 99.93% due to the final event selection.

In order to calculate a branching fraction limit, $B^\pm \rightarrow J/\psi\, K^\pm$ events with $J/\psi \rightarrow \mu^+\mu^-$ were used as a normalization channel. Using this mode has the advantage that the efficiencies to detect the $\mu^+\mu^-$ system in signal and normalization events are very similar, and systematic effects tend to cancel.

A clean sample of $B^\pm \rightarrow J/\psi\, K^\pm$ was obtained by applying the following selection. The mass constrained vertex fit of the two muons to form a J/ψ was required to have a $\chi^2/d.o.f.$ of not more than 10 similar to the $\mu^+\mu^-$ vertex criterion in the B_s search. The combined vertex fit of the J/ψ and the additional $K^\pm$ had to have χ^2 less than 20 for three $d.o.f.$ The p_T of the $K^\pm$ was larger than 0.9 GeV$/c$. The cosine of the angle between the decay length vector of the $B^\pm$ and the combined momentum of J/ψ and $K^\pm$ in the transverse plane was required to be greater than 0.9. The mass spectrum of the reconstructed $B^\pm \rightarrow J/\psi\, K^\pm$ for the full data sample after all cuts is shown in Fig. 2. A fit using a Gaussian function for the signal and a second order polynomial for the background yielded 741 ± 31 ± 22 $B^\pm$ events, where the first error is statistical and the second due to systematics estimated by varying the fit range and background shape hypothesis.

The p_T distribution of the $B^\pm$ in data has a slightly harder spectrum than in MC. Therefore, MC events of the signal and normalization channel have been re-weighted accordingly. The final value for the efficiency ratio obtained from MC is then given by $\epsilon^{B^\pm}_{\mu\mu K}/\epsilon^{B_s}_{\mu\mu} = 0.229 \pm 0.008 \pm 0.014$, where the first uncertainty is due to limited MC statistics and the second uncertainty accounts for data/MC differences. These differences include the pt re-weighting of MC events, the additional kaon track reconstruction efficiency and the effects of different trigger and muon identification efficiencies. All systematic uncertainties entering into the calculation of the branching ratio limit are listed in Table II.

In the signal region, i.e., $\pm 2\sigma$ around the B_s, we found four events, whereas our estimate (from the sidebands) was 3.7±1.1. This suggests that the four data events are entirely consistent with being background. We also examined these four events in detail by studying various kinematic variables, e.g., p_T of the muons, isolation, etc., and found them to behave like background events. Figure 3 shows the remaining background events populating the lower and upper sidebands almost equally.

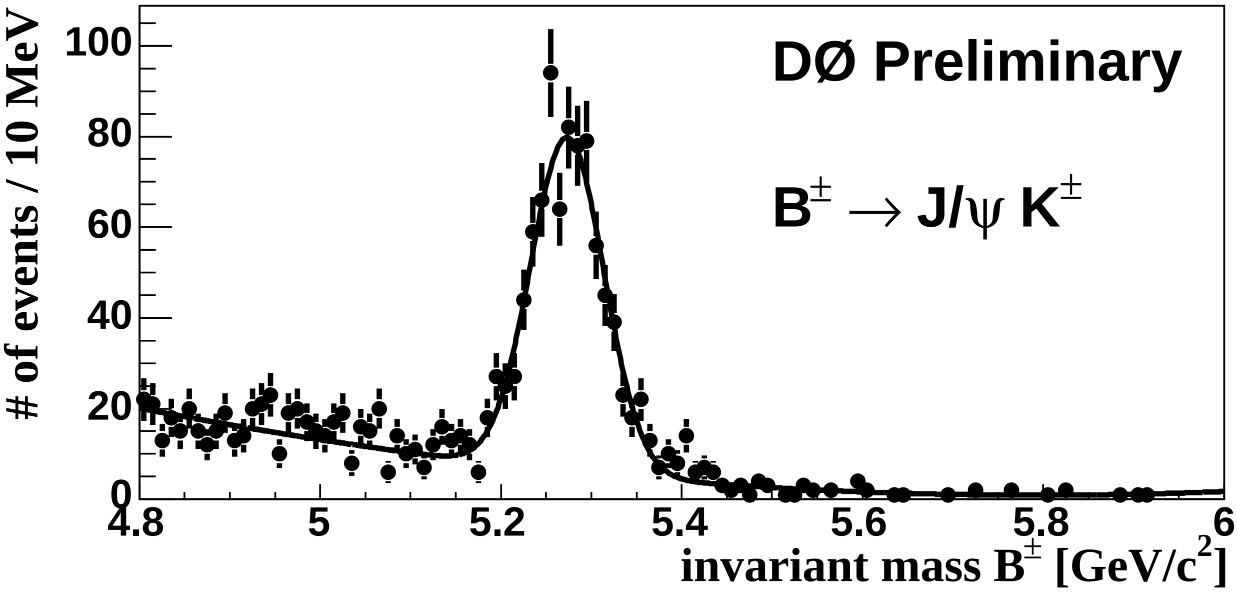

FIG. 2: Invariant mass distribution for the normalization channel $B^{\pm} \rightarrow J/\psi K^{\pm}$.

TABLE II: Relative uncertainties used in the calculation of an upper limit of $\mathcal{B}(B_s \rightarrow \mu^+\mu^-)$

Source	Relative Uncertainty [%]
$\epsilon^{B^{\pm}}_{\mu\mu K}/\epsilon^{B_s}_{\mu\mu}$	6.9
# of $B^{\pm} \rightarrow J/\psi K^{\pm}$	5.1
$\mathcal{B}(B^{\pm} \rightarrow J/\psi K^{\pm})$	4.0
$\mathcal{B}(J/\psi \rightarrow \mu\mu)$	1.7
$f_{b\rightarrow B_s}/f_{b\rightarrow B^{\pm}}$	12.7
Background uncertainty	29.7

The statistical uncertainties on the background expectation, as well as the uncertainties on the efficiencies can be included into the limit calculation of Eq. 1 by integrating over probability functions that parameterize the uncertainties. We have used a prescription [19] where we construct a frequentist confidence interval with the Feldman and Cousins [20] ordering scheme for the MC integration. The background was modeled as a Gaussian distribution with its mean value equal to the expected number of background events and its sigma equal to the background uncertainty. The uncertainty on the number of $B^{\pm}$ events as well as the uncertainties on the fragmentation ratio and measured branching ratios were added in quadrature to the efficiency uncertainties and parameterized as a Gaussian distribution.

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

$$\mathcal{B}(B_s \rightarrow \mu^+\mu^-) \leq 4.6\ \cdot 10^{-7}\ (3.8\ \cdot 10^{-7})$$

Using a Bayesian approach with flat prior and Gaussian smeared uncertainties, the limit is then given by $\mathcal{B}(B_s \rightarrow\ \mu^+\ \mu^-)\ \leq 4.7 \cdot 10^{-7}$ $(3.8 \cdot 10^{-7})$ at a 95% (90%) C.L.

This new result is presently the most stringent bound on $B_s \rightarrow \mu^+\mu^-$, improving the previously published value [5] significantly. The new limit will reduce the available parameter space for GUT inspired minimal SO(10) [9] supersymmetric models.

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

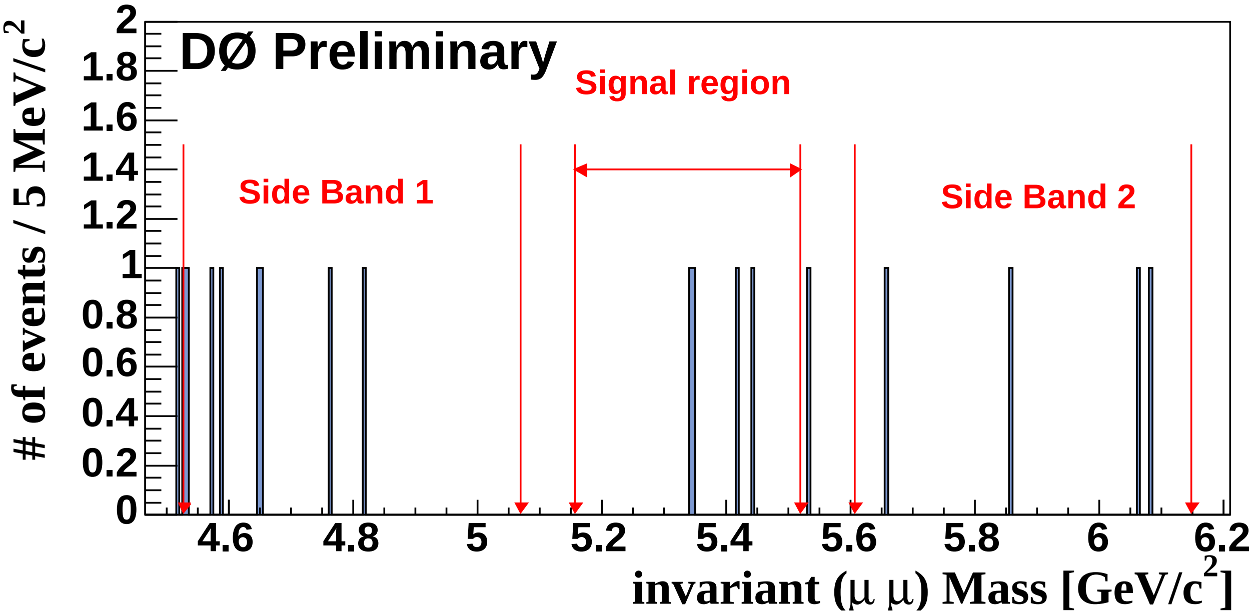

FIG. 3: Remaining background for the full data sample analyzed with our standard discriminating variables.

Research Council and WestGrid Project (Canada), BMBF (Germany), A.P. Sloan Foundation, Civilian Research and Development Foundation, Research Corporation, Texas Advanced Research Program, and the Alexander von Humboldt Foundation.

---

[*] Visitor from University of Zurich, Zurich, Switzerland.
[†] Visitor from Institute of Nuclear Physics, Krakow, Poland.

[1] Charge conjugated states are included implicitly.
[2] B. A. Campbell and P. J. O'Donnell, Phys. Rev. D **25**, 1989 (1982).
[3] G. Buchalla and A. J. Buras, Nucl. Phys B **400**, 225 (1993); M. Misiak and J. Urban, Phys. Lett. B **451**, 161 (1999); G. Buchalla and A. J. Buras, Nucl. Phys. B **548**, 309 (1999).
[4] A. J. Buras, Phys. Lett. B **566**, 115 (2003).
[5] D. Acosta *et al.* [CDF Collaboration], hep-ex/0403032, accepted by PRL.
[6] H.E. Logan and U. Nierste, Nucl. Phys. B **586**, 39 (2000).
[7] K. S. Babu and C. F. Kolda, Phys. Rev. Lett. **84**, 228 (2000); A. Dedes *et al.*, FERMILAB-PUB-02-129-T.
[8] A. Dedes *et al.*, Phys. Rev. Lett. **87**, 251804 (2001).
[9] T. Blazek *et al.*, Phys. Lett B **589**, 39 (2004); R. Dermisek *et al.*, JHEP **0304**, 37 (2003).
[10] C. Bobeth *et al.*, Phys. Rev. D **66**, 074021 (2002).
[11] R. Arnowitt *et al.* Phys. Lett. B **538**, 121 (2002).
[12] One could consider using $B_s \to J/\psi\,\phi$ as the normalizing mode, but low statistics and a large uncertainty on the branching fraction render it less attractive.
[13] S. Eidelman *et al.*, Phys. Lett. B **592**, 1 (2004).
[14] The ratio $\epsilon_{\mu\mu}^{B_d}/\epsilon_{\mu\mu}^{B_s}$ has been determined from simulation to be $(91.7 \pm 3.6)\%$, with the uncertainty due to limited MC statistics.
[15] V. Abazov, *et al.*, in preparation for submission to Nucl. Instrum. Methods Phys. Res. A, and T. LeCompte and H.T. Diehl, "The CDF and DØ Upgrades for Run II", Ann. Rev. Nucl. Part. Sci. **50**, 71 (2000).
[16] J. Abdallah *et al.* [DELPHI Collaboration], Eur. Phys. J. **C32**, 185 (2004).
[17] H. B. Prosper, Proceedings of the International Conference on Computing in High Energy Physics (CHEP'95), Rio de Janeiro, 18-22 September 1995.
[18] G. Punzi, Proceedings of the Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology (Phystat 2003), SLAC, Stanford, California, 8-11 September 2003, physics/0308063.
[19] J. Conrad *et al.*, Phys. Rev. D **67**, 012002 (2003).
[20] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998).