CDF Public Note 8176

Search for $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ Decays in $p\overline{p}$ Collisions with CDF II

(CDF Collaboration)

We report on a search for $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV using 780 pb^{-1} of data collected by the CDF II detector at the Fermilab Tevatron Collider. After applying all selection requirements, we observe 1 and 2 candidates inside the B_s^0 and B_d^0 mass windows, consistent with background expectations. The resulting upper limits on the branching fractions are $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 1.0 \times 10^{-7}$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 3.0 \times 10^{-8}$ at 95% confidence level.

In the standard model (SM), Flavor Changing Neutral Current (FCNC) decays are highly suppressed and can only occur through higher order diagrams. The decay rate for the FCNC decay $B_s^0 \to \mu^+\mu^-$ [1] is proportional to the CKM matrix element $|V_{ts}|^2$. The rate of $B_d^0 \to \mu^+\mu^-$ decays is further suppressed by the ratio of $|V_{td}/V_{ts}|^2$. The SM expectations for these branching fractions are $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.42 \pm 0.54) \times 10^{-9}$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-) = (1.00 \pm 0.14) \times 10^{-10}$ [2], which are about two orders of magnitude smaller than the current experimental sensitivity. However, new physics contributions can significantly enhance these branching fractions. An observation of these decays at the Tevatron would be unambiguous evidence for physics beyond the SM. The best existing experimental bound on $\mathcal{B}$ is $< 2.0 \times 10^{-7}$ ($< 5.1 \times 10^{-8}$) for $B_s^0(B_d^0) \to \mu^+\mu^-$ at the 95% confidence level (C.L.) [3].

In minimal supersymmetric (SUSY) extensions of the SM, additional diagrams involving SUSY particles also contribute to FCNC decay rates and the branching fraction $\mathcal{B}(B_{s,d}^0 \to \mu^+\mu^-) \propto (\tan\beta)^6$, where $\tan\beta$ is the ratio of vacuum expectation values of the two neutral CP-even Higgs fields. Large values of $\tan\beta$ enhance the decay rate to a level observable by the Tevatron experiments [4]. For example, increases of one to three orders of magnitude are obtained in the minimal SO(10) models [5], which favor large values of $\tan\beta$. For the minimal flavor violating (MFV) models, $B_d^0 \to \mu^+\mu^-$ remains suppressed relative to $B_s^0 \to \mu^+\mu^-$ due to $|V_{td}/V_{ts}|^2$. This may not be true for non-MFV models such as R-parity violating SUSY [6], which can produce large enhancements, even for low values of $\tan\beta$, in either or both of the B_s and B_d FCNC decay rates. Thus, a simultaneous observation of $B_{s,d}^0 \to \mu^+\mu^-$ decays can be important in determining the flavor structure of the new physics. In the absence of an observation, any improvements to the limits can be used to constrain significantly many SUSY models [4–7].

In this Letter, we report on a search for $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays using 780 pb^{-1} of data collected by the upgraded Collider Detector at Fermilab (CDF II). This data set includes the 171 pb^{-1} and 364 pb^{-1} of data from our previous measurements [3, 8]. We significantly improve the sensitivity of the search over our former analysis by more than doubling our data sample. The analysis techniques are similar to those used in our previous publication [3]. The limits we present here are the most stringent to date and supersede our previous results.

The CDF II detector is described in detail in Ref. [9]. The inner tracking system is composed of a silicon microstrip detector (SVX II) [10] surrounded by an open-cell wire drift chamber (COT) [11]. These tracking detectors are immersed in a 1.4 T magnetic field and measure p_T, charged particle momentum in the plane transverse to the beamline. Four layers of planar drift chambers (CMU) [12] detect muon candidates with $p_T > 1.4$ GeV/c and provide coverage in the pseudorapidity range $|\eta| < 0.6$, where $\eta = -\ln(\tan\frac{\theta}{2})$ and θ is the angle of the track with respect to the beamline. The central muon extension (CMX) consists of conical sections of drift tubes and extends the coverage to $0.6 < |\eta| < 1.0$ for muon candidates with $p_T > 2.0$ GeV/c.

The data used in this analysis are selected by two classes of dimuon triggers: for the CMU-CMU (U-U) triggers both muon candidates are reconstructed in the CMU chambers, while for the CMU-CMX (U-X) triggers one of the muon candidates is reconstructed in the CMX chambers. The inclusion of the U-X trigger increases the signal acceptance by about 50%. The details of the trigger system and selection requirements can be found in Refs. [8, 9]. Since they have different sensitivities, we treat U-U and U-X channels separately, combining the results at the end.

The offline reconstruction begins by identifying two muon candidates of opposite charge which satisfy the online dimuon trigger requirements. To avoid regions of rapidly changing trigger efficiency, we omit CMU (CMX) muon candidates with $p_T < 2$ (2.2) GeV/c. The random combinatoric backgrounds are suppressed by requiring the vector sum of the muon transverse momenta to be $|\vec{p}_T^{\,\mu\mu}| > 4$ GeV/c. The remaining pairs of muon tracks are then refit under the constraint that they come from the same three-dimensional (3D) space point, and are required to satisfy vertex fit quality criteria. The 3D decay length is given by $L_{3D} = \vec{L} \cdot \vec{p}^{\mu\mu}/|\vec{p}^{\mu\mu}|$, where $\vec{L}$ is the displacement vector from the primary to the dimuon vertex. The primary vertex is determined using a constrained vertex fit of all tracks in the event, excluding the $\mu^+\mu^-$ pair and other secondary decay tracks. For each B-candidate we estimate the proper decay time $\tau = M_{\mu\mu} L_{3D}/|\vec{p}^{\mu\mu}|$, where $M_{\mu\mu}$ is the invariant mass and $\vec{p}^{\mu\mu}$ is the momentum vector of the dimuon system. Additional background is reduced by demanding $L_{3D} < 1.0$ cm (to remove poorly reconstructed tracks), the uncertainty on L_{3D} to be less than 150 μm, and $2\sigma_\lambda < \lambda < 0.3$ cm, where $\lambda = c \cdot \tau$ and σ_λ is the total uncertainty on λ. There are 50214 (29265) dimuon candidates that fulfill all the above trigger and offline reconstruction requirements in the U-U (U-X) channel. At this stage, the data sample is dominated by random combinatoric background.

We model the signal $B^0_{s,d} \to \mu^+\mu^-$ decays using the PYTHIA Monte Carlo (MC) program [13]. The PYTHIA events are passed through a full detector simulation and satisfy the same requirements as data. The p_T spectrum of the B-hadron in the MC is weighted to match the measurement from Ref. [9]. The signal MC samples are used primarily for analysis optimization and for estimating the efficiency of selection requirements.

For the final event selection we use the following four discriminating variables: $M_{\mu\mu}$, λ, the 3D opening angle $\Delta\Theta$ between vectors $\vec{p}^{\mu\mu}$ and $\vec{L}$, and the B-candidate track isolation (I) [14]. We exploit long lifetime of B mesons to reject prompt combinatoric background with the decay length (λ) requirement. Additional combinatoric background and partially reconstructed B hadrons are removed with the pointing angle requirement. Since b-quark fragmentation is hard, B hadrons carry most of the transverse momentum of the b quark, and thus are isolated. We use the variable I to enhance the heavy flavor content of the sample and also to reject partially reconstructed B hadrons, which are less isolated. Figure 1 compares the distributions of these variables for data (which is background-dominated) to MC signal events. Based on the observed distributions, we apply two additional loose requirements, $I > 0.5$ and $\Delta\Theta < 0.7$ rad, to further reject background while maintaining 95.4% of signal efficiency. We collectively refer to all the selection requirements applied up to this point as the "baseline" requirements. In the data, 13325 (9741) events survive the baseline requirements in the U-U (U-X) sample. The di-muon mass distribution of these events are shown in Figs 2 and 3.

To enhance signal and background separation we construct a multivariate likelihood ratio based on the input variables: I, $\Delta\Theta$, and λ probability $P(\lambda) = e^{-\lambda/c\tau_{B_{s(d)}}}$, where $\tau_{B_{s(d)}}$ is the world average $B_{s(d)}$ lifetime. We use the $P(\lambda)$ variable instead of λ in constructing the likelihood ratio because the $P(\lambda)$ distribution is nearly flat, and therefore better behaved in the likelihood. We define the likelihood ratio to be

$$L_R = \frac{\prod_i \mathbf{P}_s(x_i)}{\prod_i \mathbf{P}_s(x_i) + \prod_i \mathbf{P}_b(x_i)}, \tag{1}$$

where $x_1 = I$, $x_2 = \Delta\Theta$, $x_3 = P(\lambda)$, and $\mathbf{P}_{s(b)}(x_i)$ is the probability that a signal (background) event has an observed x_i. The probability distributions for the signal events are obtained from the signal MC and the background distributions are taken from the data sidebands (defined below) and are shown in Figs. 4 and 5. The resulting L_R distributions for the signal and background events are shown in Fig. 6.

Although a subset of the data was used previously [3, 8], we adopt an optimization strategy that only uses events in the data sidebands in order to avoid biases in our choice of final selection criteria. The search window is defined by $5.169 < M_{\mu\mu} < 5.469\,\mathrm{GeV}/c^2$, corresponding to approximately ± 6 times the two-track invariant mass resolution, which is estimated to be $\sigma_M \approx 24\ \mathrm{MeV}/c^2$. The sideband regions $4.669 < M_{\mu\mu} < 5.169\,\mathrm{GeV}/c^2$ and $5.469 < M_{\mu\mu} < 5.969\,\mathrm{GeV}/c^2$ are used to estimate the background in the signal region and for extensive analysis cross-checks.

The $B^0_s \to \mu^+\mu^-$ branching fraction is obtained by normalizing to the number of $B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+$ decays collected by the same trigger. The $B^+ \to J/\psi K^+$ mode is reconstructed using the same baseline requirements as the signal mode [15] but including an additional $p_T > 1$ GeV/c requirement on the kaon candidate. The upper limit on the branching fraction is given by

$$\mathcal{B}(B^0_s \to \mu^+\mu^-)^{90\%C.L.} = \frac{N^{90\%}_{B^0_s}}{N_{B^+}} \cdot \frac{\alpha_{B^+}}{\alpha_{B^0_s}} \cdot \frac{\epsilon^{base}_{B^+}}{\epsilon^{base}_{B^0_s}} \cdot \frac{1}{\epsilon^{L_R}_{B^0_s}} \cdot \frac{f_u}{f_s} \cdot \mathcal{B}(B^+ \to J/\psi K^+), \tag{2}$$

where $N^{90\%}_{B^0_s}$ is the number of $B^0_s \to \mu^+\mu^-$ decays at the 90% C.L. for N observed and N_b expected background events. The value of $N^{90\%}_{B^0_s}$ is estimated using the Bayesian approach of Ref. [16] assuming a flat prior for the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ and incorporating Gaussian uncertainties into the limit. The number of reconstructed $B^+ \to J/\psi K^+$ candidates, N_{B^+}, determined from the data using sideband subtraction, is 4115 ± 90 for the U-U and 1548 ± 57 for the U-X channel and are shown in figures 7 and 8. The parameter $\alpha_{B_s^0}$ (α_{B^+}) is the trigger acceptance and $\epsilon^{base}_{B_s^0}$ ($\epsilon^{base}_{B^+}$) is the efficiency of the baseline requirements for the signal (normalization) mode. We apply the likelihood ratio requirement only to the signal mode and therefore the efficiency of the likelihood ratio $\epsilon^{L_R}_{B_s^0}$ appears only in the denominator of Eq. 2. We use the fragmentation ratio $f_u/f_s = 3.71 \pm 0.44$ [18]. The branching fraction $\mathcal{B}(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+) = (5.88 \pm 0.26) \times 10^{-5}$ is obtained from Ref. [16]. The expression for $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ is derived by replacing B_s^0 with B_d^0 and the fragmentation ratio with $f_u/f_d = 1$.

The acceptance ratio $\alpha_{B^+}/\alpha_{B^0_{s,d}}$ obtained from the signal MC is 0.297 ± 0.020 (0.191 ± 0.013) for the U-U (U-X) channel. The uncertainty includes contributions from systematic variations of the modeling of the B-hadron p_T spectrum and longitudinal beam profile, and from the statistics of the MC sample.

The quantity ϵ^{base} includes the trigger and offline reconstruction efficiencies. The trigger efficiency is determined from inclusive data samples unbiased with respect to the triggers used here. The ratio of B^+ to B_s^0 trigger efficiencies is $\epsilon^{trig}_{B^+}/\epsilon^{trig}_{B^0_{s,d}} = 0.995 \pm 0.001$ (0.989 ± 0.001) for the U-U (U-X) channel. We evaluate the single track COT, SVX-II, and muon efficiencies using a data sample of inclusive $J/\psi \to \mu^+\mu^-$ decays collected on single-muon triggers. The relevant double-track efficiencies are computed by convolution with $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ MC events surviving the trigger requirements. The offline reconstruction efficiency between signal and normalization mode also largely cancels in the ratio with the exception of the kaon efficiency from the B^+ decay. Lastly, we obtain the efficiency of the remaining baseline requirements from the signal MC and cross-check the results by comparing B^+ data and MC. The comparisons of the data and MC quantities are given in figures 9 and 10 which show the transverse momentum after reweighing to the distribution observed in data and the vertex and lifetime related properties. Combining all effects, we find $\epsilon^{reco}_{B^+}/\epsilon^{reco}_{B^0_{s,d}} = 0.910 \pm 0.03$ (0.90 ± 0.03) for the U-U (U-X) channel. The uncertainty is dominated by systematic uncertainties accounting for kinematic differences between $J/\psi \to \mu^+\mu^-$ and $B^0_{s,d} \to \mu^+\mu^-$ decays.

The efficiency of the likelihood requirement $\epsilon^{L_R}_{B^0_{s,d}}$ is estimated from the signal MC. The MC isolation distribution is reweighed to match the data for the comparison. The efficiency varies from about 66% for $L_R > 0.90$ to 37% for $L_R > 0.99$. We assign a relative systematic uncertainty of $\pm 6\%$ to both U-U and U-X channels based on comparisons of $B^+ \to J/\psi K^+$ MC and data samples and variations of the MC isolation distribution for $B_s^0 \to \mu^+\mu^-$ based on the statistical error of a study of $B_s^0 \to J/\psi\phi$ data. The input distributions and the likelihood ratio comparison for $B^+ \to J/\psi K^+$ data and MC before reweighing the isolation variable are shown in figures 11 and 12 for the U-U and U-X samples. The isolation distribution comparison between $B_s^0 \to J/\psi\phi$ and MC before reweighing the isolation variable is shown in figure 13.

Table I summarizes the acceptances and efficiencies used as inputs to equation 2. We discuss the estimate of the likelihood efficiency and its associated systematic uncertainty in more detail in the next section.

	CMU-CMU		CMU-CMX	
$(\alpha_{B^+}/\alpha_{B_s^0})$	0.297 ± 0.020	$(\pm 7\%)$	0.191 ± 0.013	$(\pm 7\%)$
$(\epsilon^{trig}_{B^+}/\epsilon^{trig}_{B_s^0})$	0.995 ± 0.001	(-)	0.989 ± 0.001	(-)
$(\epsilon^{reco}_{B^+}/\epsilon^{reco}_{B_s^0})$	0.91 ± 0.03	$(\pm 3\%)$	0.90 ± 0.03	$(\pm 4\%)$
$\epsilon^{L_R}_{B_s^0}$	0.35 ± 0.02	$(\pm 7\%)$	0.39 ± 0.03	$(\pm 7\%)$
N_{B^+}	4215 ± 90	$(\pm 2\%)$	1548 ± 57	$(\pm 4\%)$
f_u/f_s	3.71 ± 0.44	$(\pm 12\%)$	3.71 ± 0.44	$(\pm 12\%)$
$BR(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+)$	$(5.88 \pm 0.26) \times 10^{-5}$	$(\pm 4\%)$	$(5.88 \pm 0.26) \times 10^{-5}$	$(\pm 4\%)$
ses	$(3.9 \pm 0.7) \times 10^{-8}$	$(\pm 16\%)$	$(6.2 \pm 1.0) \times 10^{-8}$	$(\pm 17\%)$

TABLE I: A summary of the inputs used in equation 2 to estimate the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The relative uncertainties are given parenthetically. Note that the ϵ_{L_R} includes the 95% of the loose $I > 0.5$ and $\Delta\Theta < 0.7$ requirements, which only effect the $B_s^0 \to \mu^+\mu^-$ decays. The single-event-sensitivity, ses, corresponding to $N_{B_s^0} = 1$, is shown in the last row.

The expected number of background events is estimated by extrapolating the number of sideband events passing the baseline requirements and a given L_R requirement to the signal window. The agreement of the likelihood ratio between the signal and sideband regions is shown in figures 14

We have cross-checked our background estimate procedure using control samples from the data: like sign $\mu^\pm\mu^\pm$ events, $\mu^+\mu^-$ events with $\lambda < 0$, and a fake-muon enhanced $\mu^+\mu^-$ sample in which we demand at least one muon candidate fail the muon quality requirements. We compare our background predictions to the number of events observed in the search window for a wide range of L_R requirements. No statistically significant discrepancies are observed. The results are shown in Table II.

sample	L_R cut	CMU-CMU pred	CMU-CMU obsv	CMU-CMU prob(%)	CMU-CMX pred	CMU-CMX obsv	CMU-CMX prob(%)	Combined prob(%)
OS-	> 0.50	$489 \pm (12)$	483	41	$351 \pm (10)$	338	27	28
	> 0.90	$62 \pm (4)$	73	12	$56 \pm (4)$	63	22	7
	> 0.99	$4.8 \pm (1.2)$	9	8	$3.9 \pm (1.1)$	8	7	2
SS+	> 0.50	$5.4 \pm (1.3)$	4	40	$3.3 \pm (1.0)$	2	39	27
	> 0.90	< 0.1	0	-	$0.9 \pm (0.5)$	0	43	43
	> 0.99	< 0.1	0	-	< 0.1	0	-	-
SS-	> 0.50	$6.6 \pm (1.4)$	7	49	$4.2 \pm (1.1)$	5	41	40
	> 0.90	$0.6 \pm (0.4)$	1	45	$0.3 \pm (0.3)$	0	70	57
	> 0.99	< 0.1	0	-	< 0.1	0	-	-
FM+	> 0.50	$188 \pm (8)$	159	3	$33 \pm (3)$	37	29	7
	> 0.90	$34 \pm (3)$	24	7	$6.0 \pm (1.3)$	5	46	6
	> 0.99	$4.5 \pm (1.0)$	9	6	$0.6 \pm (0.4)$	0	55	12

TABLE II: A comparison of the predicted and observed number of events in an extended signal mass region as a function of likelihood cut for the various control samples. The values given in the parentheses are the uncertainties on the mean of the background prediction. The Poisson probability for making an observation at least as large (or fewer than observed when observed is less than predicted) given the predicted background is also shown in the table. The combined probability for CMU-CMU and CMU-CMX channels are shown in the last column. The probabilities are calculated including the statistical uncertainties on the background prediction itself.

In addition we have considered background contributions from $B^0_{s,d} \to h^+h^-$, where $h^\pm = \pi^\pm$ or $K^\pm$. Tables III and IV list the background contribution in the B^0_s and B^0_d signal windows for the CMU-CMU and CMU-CMX channels, respectively. The signal windows are defined as $\pm 60 MeV/c^2$ about the world average B^0_s and B^0_d masses. The expected background is calculated based on equation 2, with $B^0_s \to \mu^+\mu^-$ replaced by $B \to h^+h^-$. In addition, there are two more efficiency terms in the denominator of equation 2 to account for the muon fake rates and the fraction of misidentified $B \to h^+h^-$ events falling in either B^0_s or B^0_d signal window. The branching ratio for the various $B \to h^+h^-$ modes are taken from [17]. For modes with only branching ratio limits, we use the limit as the central value and assign $\pm 100\%$ to the systematic uncertainties. With exception of $Br(B^0_s \to K^+\pi^-)$[21], decay modes with unmeasured branching ratios have negligible contribution to the signal windows. The uncertainties on the $B \to h^+h^-$ estimates are dominated by the fake rate and branching ratio uncertainties.

Mode	BR (10^{-6})	Fake Rate (10^{-4})	# of Evts in B^0_s Signal Window (CMU-CMU)	# of Evts in B^0_d Signal Window (CMU-CMU)
$B^0_d \to \pi\pi$	5.0 ± 0.4	1.7 ± 0.2	0.007 ± 0.002	0.081 ± 0.015
$B^0_d \to K\pi$	18.9 ± 0.7	3.4 ± 0.4	0.004 ± 0.001	0.46 ± 0.08
$B^0_d \to KK$	0.2 ± 0.2	6.0 ± 0.5	≈ 0	0.001 ± 0.001
$B^0_s \to \pi\pi$	7.0 ± 7.0	1.7 ± 0.2	0.03 ± 0.03	0.005 ± 0.004
$B^0_s \to K\pi$	7.7 ± 7.7	3.4 ± 0.4	0.049 ± 0.048	0.051 ± 0.050
$B^0_s \to KK$	35.1 ± 7.8	6.0 ± 0.5	0.073 ± 0.020	0.54 ± 0.13
Total			0.16 ± 0.06	1.14 ± 0.16

TABLE III: The estimated contribution of $B \to h^+h^-$ backgrounds in the B^0_s and B^0_d signal windows for the CMU-CMU channel.

Given that the contributions from two-body hadronic B decays are sizable in certain channels, we will include them in our total background estimates. The total background is obtained by summing the combinatoric background as estimated from the sidebands and the contribution from $B \to h^+h^-$ decays. The results with $L_R > 0.99$ are listed in Table V. With a L_R cut of 0.99, we expect to see about 1 event in the B^0_s signal window and 2 events in the B^0_d signal window.

We optimize the analysis based on the *a priori* expected 90% C.L. upper limit on $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$. The expected limit for a given set of optimization requirements is computed by summing the 90% C.L. limits over all possible observations N, weighted by the corresponding Poisson probability when expecting N_b. We scan over a range of L_R requirements and determine the optimal value to be $L_R > 0.99$. With the optimized selection requirements, the expected number of background events is the same in the B^0_s and B^0_d search windows. Within the signal region of ± 60 MeV/c^2 ($\pm 2.5\sigma_M$) about the world average B^0_s or B^0_d mass [16], N_b is 0.88 ± 0.30 [0.39 ± 0.21], and the B^0_s

Mode	BR (10^{-6})	Fake Rate (10^{-5})	# of Evts in B_s^0 Signal Window (CMU-CMX)	# of Evts in B_d^0 Signal Window (CMU-CMX)
$B_d^0 \to \pi\pi$	5.0 ± 0.4	4.5 ± 1.1	0.0013 ± 0.0004	0.014 ± 0.004
$B_d^0 \to K\pi$	18.9 ± 0.7	9.9 ± 1.9	0.0008 ± 0.0002	0.087 ± 0.019
$B_d^0 \to KK$	0.2 ± 0.2	20.5 ± 2.7	≈ 0	0.0003 ± 0.0003
$B_s^0 \to \pi\pi$	7.0 ± 7.0	4.5 ± 1.1	0.005 ± 0.005	0.0008 ± 0.0008
$B_s^0 \to K\pi$	7.7 ± 7.7	9.9 ± 1.9	0.009 ± 0.009	0.0096 ± 0.0095
$B_s^0 \to KK$	35.1 ± 7.8	20.5 ± 2.7	0.016 ± 0.005	0.12 ± 0.03
Total			0.033 ± 0.011	0.232 ± 0.038

TABLE IV: The estimated contribution of $B \to h^+h^-$ backgrounds in the B_s^0 and B_d^0 signal windows for the CMU-CMX channel.

	B_s^0 Signal Window		B_d^0 Signal Window	
Bkg Source	CMU-CMU	CMU-CMX	CMU-CMU	CMU-CMX
Combinatoric	0.72 ± 0.29	0.36 ± 0.21	0.72 ± 0.29	0.36 ± 0.21
$B \to h^+h^-$	0.16 ± 0.06	0.03 ± 0.01	1.14 ± 0.16	0.23 ± 0.04
Total	0.88 ± 0.30	0.39 ± 0.21	1.86 ± 0.34	0.59 ± 0.21

TABLE V: The total expected background events in the B_s^0 and B_d^0 signal window for $L_R > 0.99$. The total background is the sum of the combinatorial background and contribution $B \to h^+h^-$ decays, where the hadronic tracks were misidentified as muons.

single-event-sensitivity is $(3.9 \pm 0.7) \times 10^{-8}$ $[(6.2 \pm 1.0) \times 10^{-8}]$ for the U-U [U-X] channel. Using these criteria we observe 1 [0] and 2 [0] in the of U-U [U-X] channel of the B_s^0 and B_d^0 signal boxes respectively which is consistent with the background expectation. The invariant mass distributions are shown after apply all selection cut in Figs. 15. Two dimensional plots of the invariant mass vs. the likelihood ration value are shown in Figs. 16 and 17.

Using Eq. 2 and combining the U-U and U-X channels taking into account the correlated uncertainties, we derive 90% (95%) C.L. limits of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 8.0 \times 10^{-8}$ (1.0×10^{-7}) and $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 2.3 \times 10^{-8}$ (3.0×10^{-8}). The new limits improve the previous limits [3, 19, 20] by a factor of two and can be used to reduce the allowed parameter space of a broad spectrum of SUSY models [5–7].

[1] Throughout this paper inclusion of charge conjugate modes is implicit.
[2] G. Buchalla and A. J. Buras, Nucl. Phys. **B400**, 225 (1993); A.J. Buras, Phys. Lett. B **566**, 115 (2003).
[3] CDF Collaboration, D. Acosta *et al.*, Phys. Rev. Lett. **95**, 221805 (2005).
[4] S. Choudhury and N. Gaur, Phys. Lett. B **451**, 86 (1999); K.S. Babu and C. Kolda, Phys. Rev. Lett. **84**, 228 (2000).
[5] R. Dermisek *et al.*, J. High Energy Phys. **04**, 037 (2003); R. Dermisek *et al.*, J. High Energy Phys. **09**, 029 (2005); D. Auto *et al.*, J. High Energy Phys. **06**, 023 (2003); T. Blazek *et al.*, Phys. Lett. B **589**, 39 (2004).
[6] R. Arnowitt *et al.*, Phys. Lett. B **538**, 121 (2002).
[7] H. Baer *et al.*, J. High Energy Phys. **06**, 054 (2003); S. Baek, P. Ko, and W.Y. Song, Phys. Rev. Lett. **89**, 271801 (2002); H. Logan and U. Nierste, Phys. Lett. B **566**, 115 (2003); A. Dedes and B. Huffman, Phys. Lett. B **600**, 261 (2004); J. Ellis, K. Olive, and V. Spanos, Phys. Lett. B **624**, 47 (2005); G.L. Kane, C. Kolda, and J.E. Lennon, hep-ph/0310042.
[8] CDF Collaboration, D. Acosta *et al.*, Phys. Rev. Lett. **93**, 032001 (2004).
[9] CDF Collaboration, D. Acosta *et al.*, Phys. Rev. D **71**, 032001 (2005).
[10] A. Sill *et al.*, Nucl. Instrum. Meth. A **447**, 1 (2000).
[11] T. Affolder *et al.*, Nucl. Instrum. Meth. A **526**, 249 (2004).
[12] G. Ascoli *et al.*, Nucl. Instrum. Meth. A **268**, 33 (1988).
[13] T. Sjöstrand *et al.*, Comp. Phys. Commun. **135**, 238 (2001).
[14] The B-candidate isolation is defined as $I = |\vec{p}_T^{\,\mu\mu}|/(\sum_i p_T^i + |\vec{p}_T^{\,\mu\mu}|)$, where the sum is over all tracks with $\sqrt{\Delta\eta^2 + \Delta\phi^2} \leq 1$; $\Delta\phi$ and $\Delta\eta$ are the azimuthal angle and pseudorapidity of track i with respect to $\vec{p}^{\,\mu\mu}$. Also see V. Krutelyov, Ph.D. Thesis, Texas A&M University, 2005 (unpublished).
[15] We only use the two muon tracks in the B^+ vertex fit.
[16] S. Eidelman *et al.*, Phys. Lett. B **592**, 1 (2004).
[17] Heavy Flavor Averaging Group Lepton-Photon 2005 averages, `http://www.slac.stanford.edu/xorg/hfag/rare/index.html`
[18] f_x is the fraction of weakly decaying B_x hadron in b quark fragmentation. We use values from [16] and take into account

the correlations between f_s and f_u.
[19] D0 Collaboration, V. Abazov *et al.*, Phys. Rev. Lett. **94**, 071802 (2005).
[20] BABAR Collaboration, B. Aubert *et al.*, Phys. Rev. Lett **94**, 221803 (2005).
[21] As it turns out, the experimental $Br(B_s^0 \to K^+\pi^-)$ limit is near the theoretical prediction.

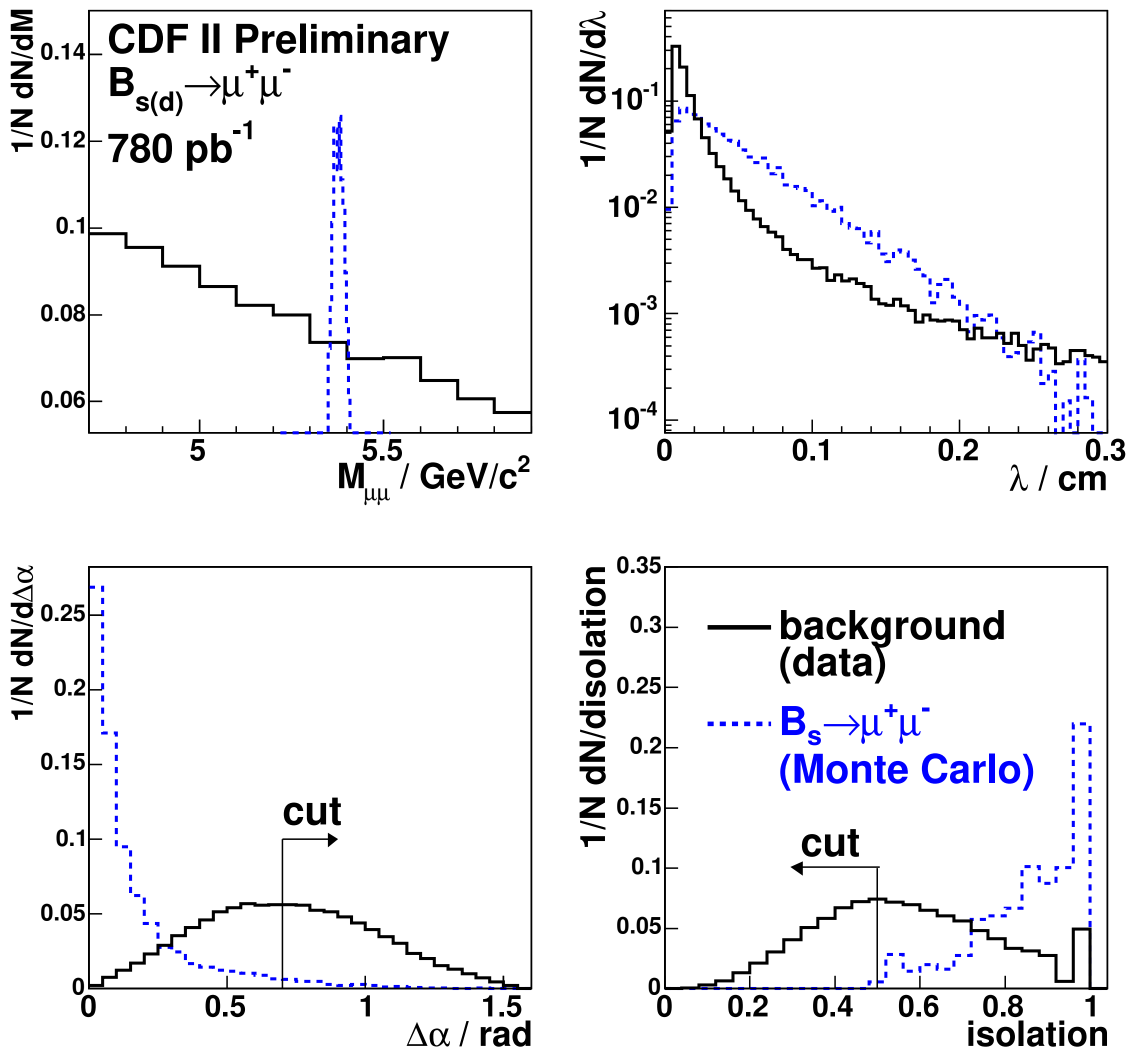

FIG. 1: Distributions of various discriminating variables for Monte Carlo signal events (dashed histograms) and a background-dominated data sample (solid histograms). Only events which survive the baseline and vertex cuts and have $\lambda > 0$ are included. The histograms are all normalized to unit area. For the mass plot, different binning is used for the data and MC.

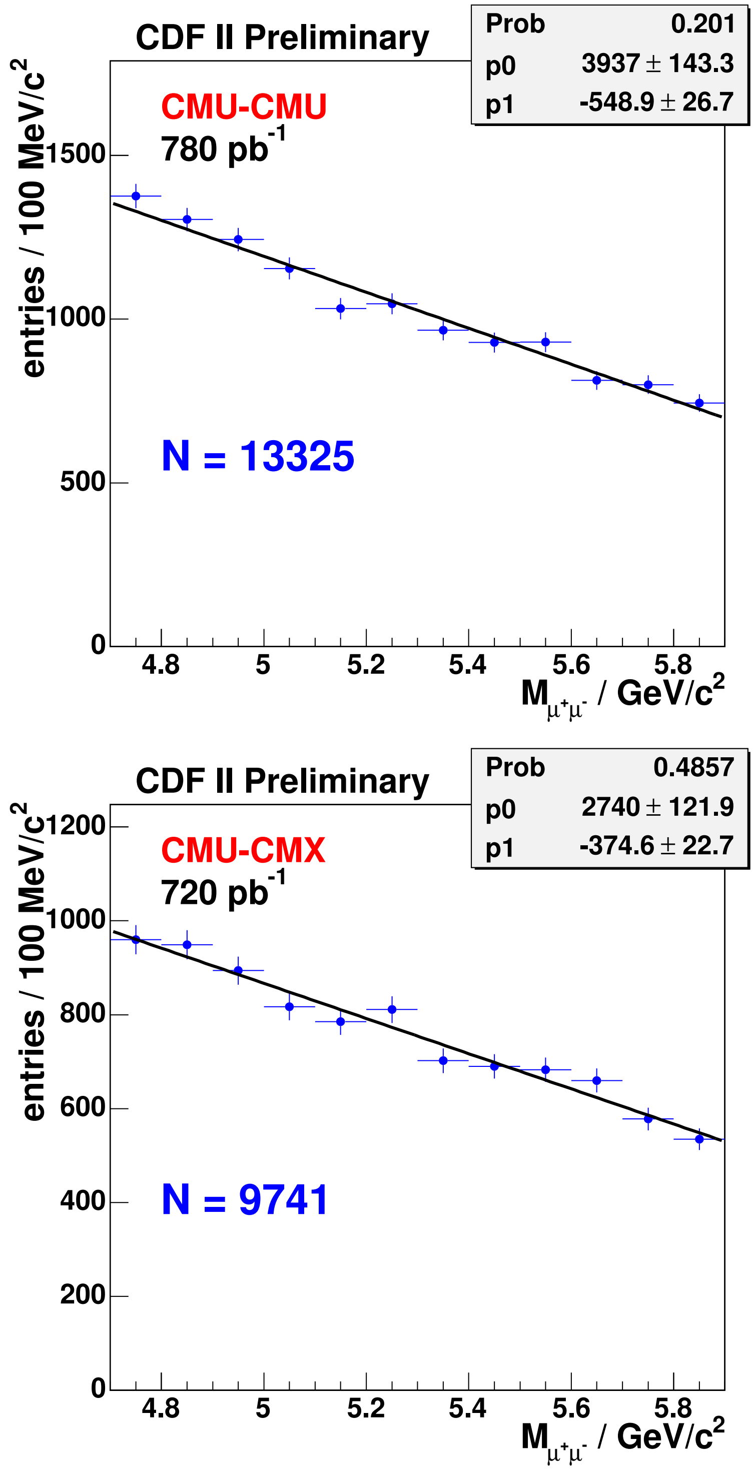

FIG. 2: Di-muon invariant mass distribution for events satisfying the baseline and vertex requirements for the $B_s^0 \to \mu^+\mu^-$ search sample for the U-U and U-X samples. The results of a fit to a straight line are shown.

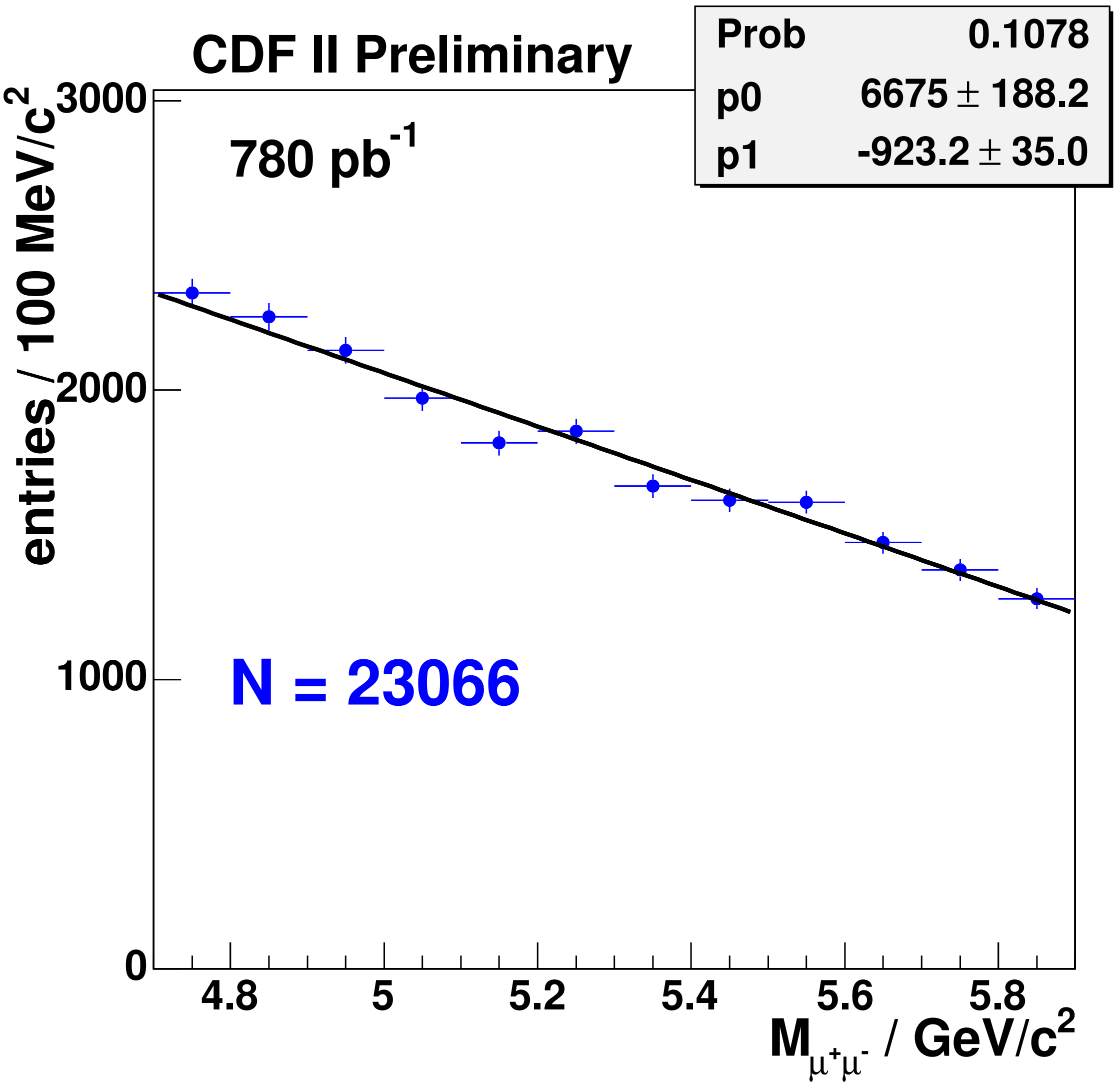

FIG. 3: Di-muon invariant mass distribution for events satisfying the baseline and vertex requirements for the $B_s^0 \to \mu^+\mu^-$ search sample for the combined U-U and U-X samples. The results of a fit to a straight line are shown.

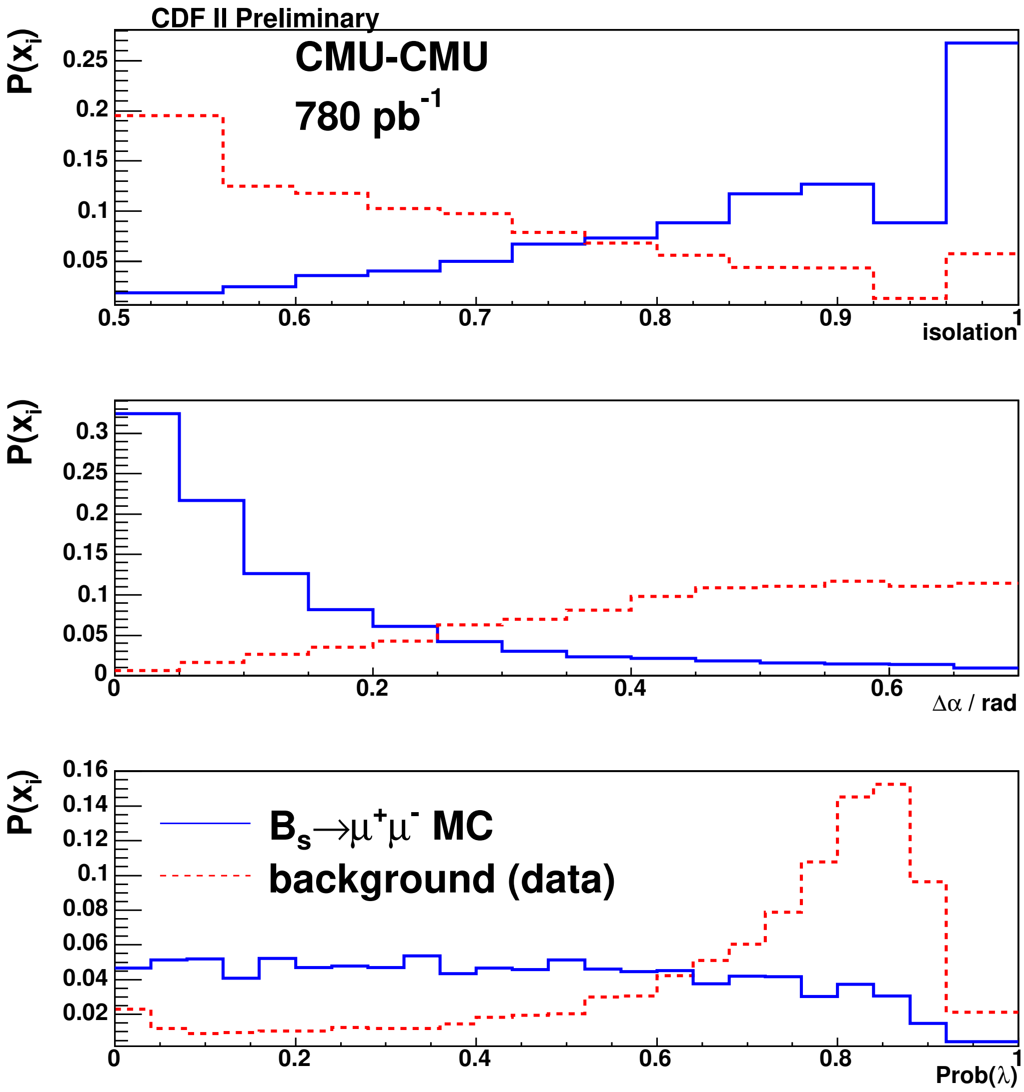

FIG. 4: Probability densities functions for the CMU-CMU events. The signal distributions are taken from MC and the background distributions are taken from data sidebands.

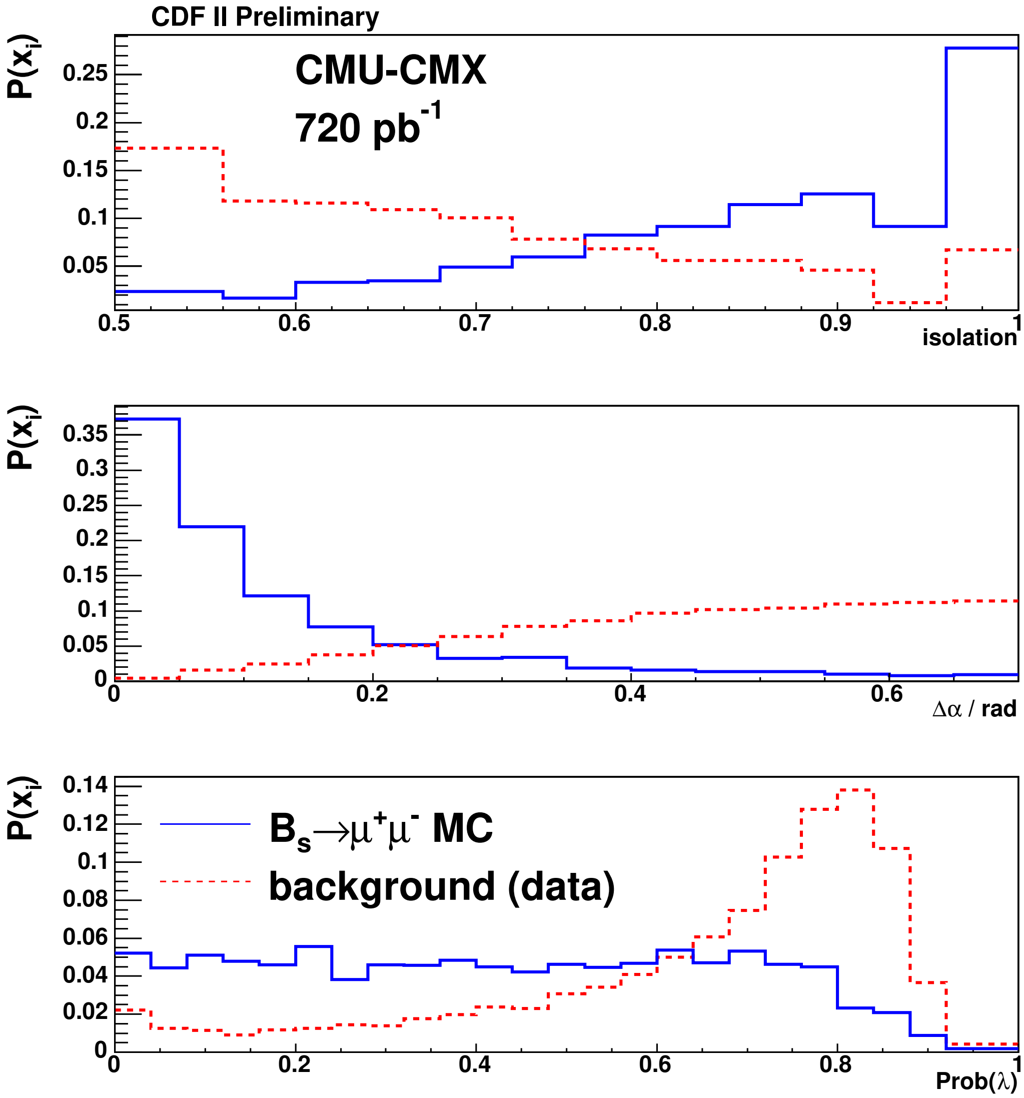

FIG. 5: Probability densities functions for the CMU-CMX events. The signal distributions are taken from MC and the background distributions are taken from data sidebands.

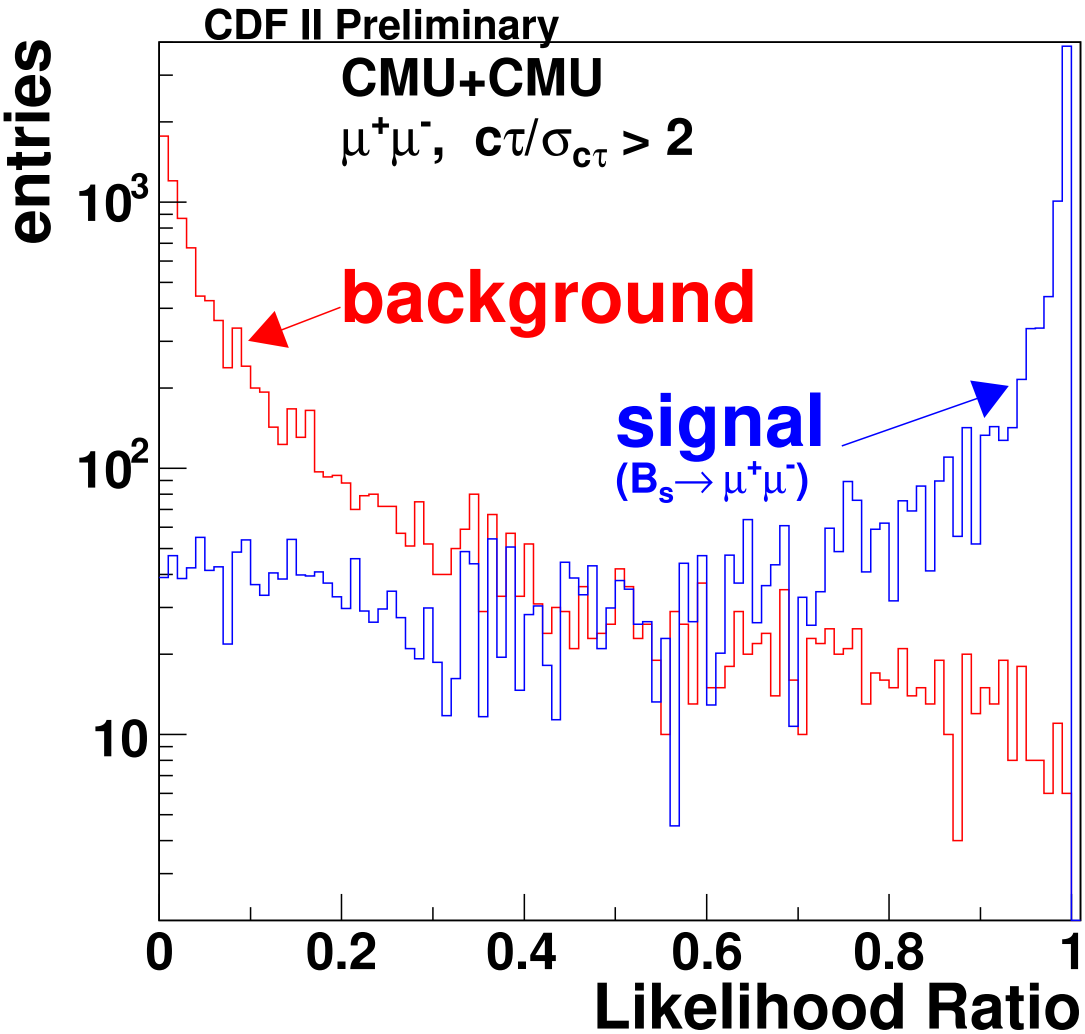

FIG. 6: The likelihood ratio distributions for signal and background in the CMU-CMU channel.

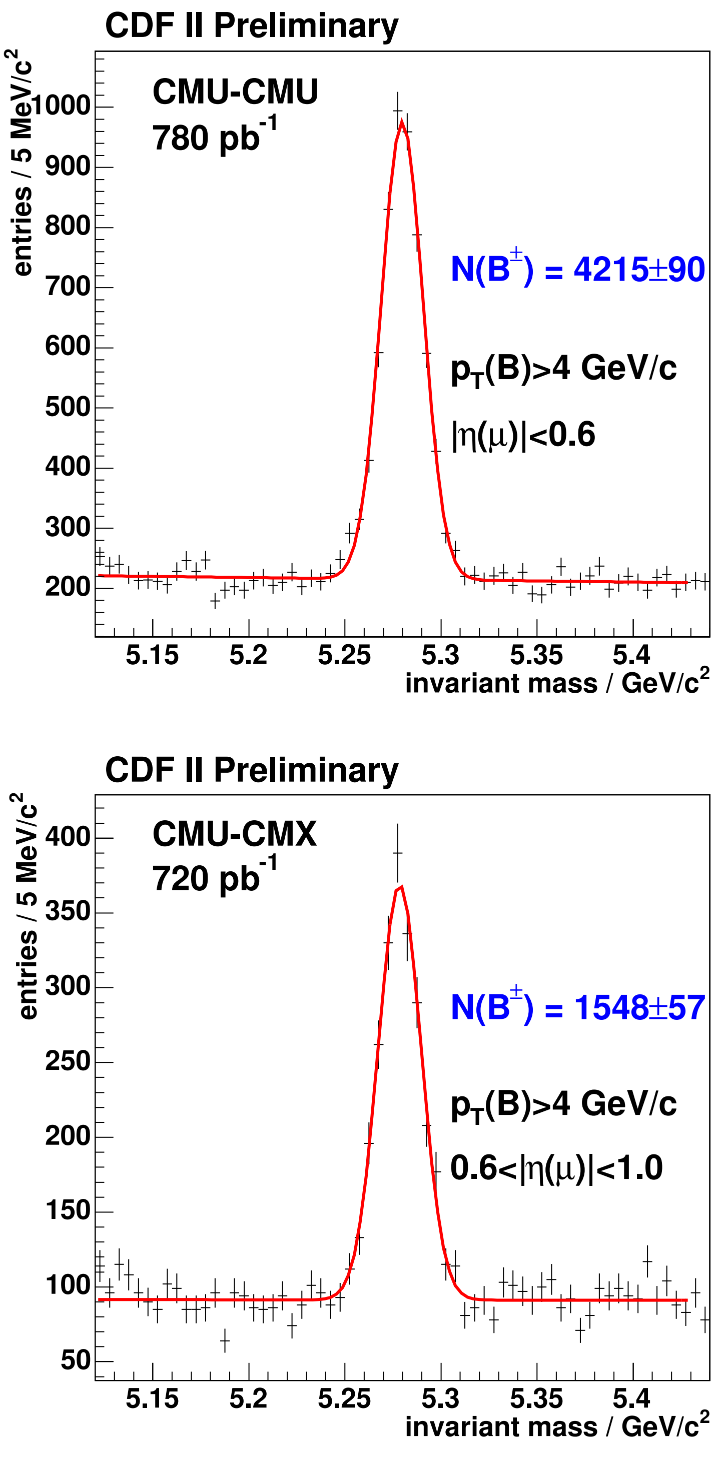

FIG. 7: The $\mu^+\mu^-K^+$ invariant mass distribution for events satisfying the baseline and vertex requirements for the U-U and U-X samples.

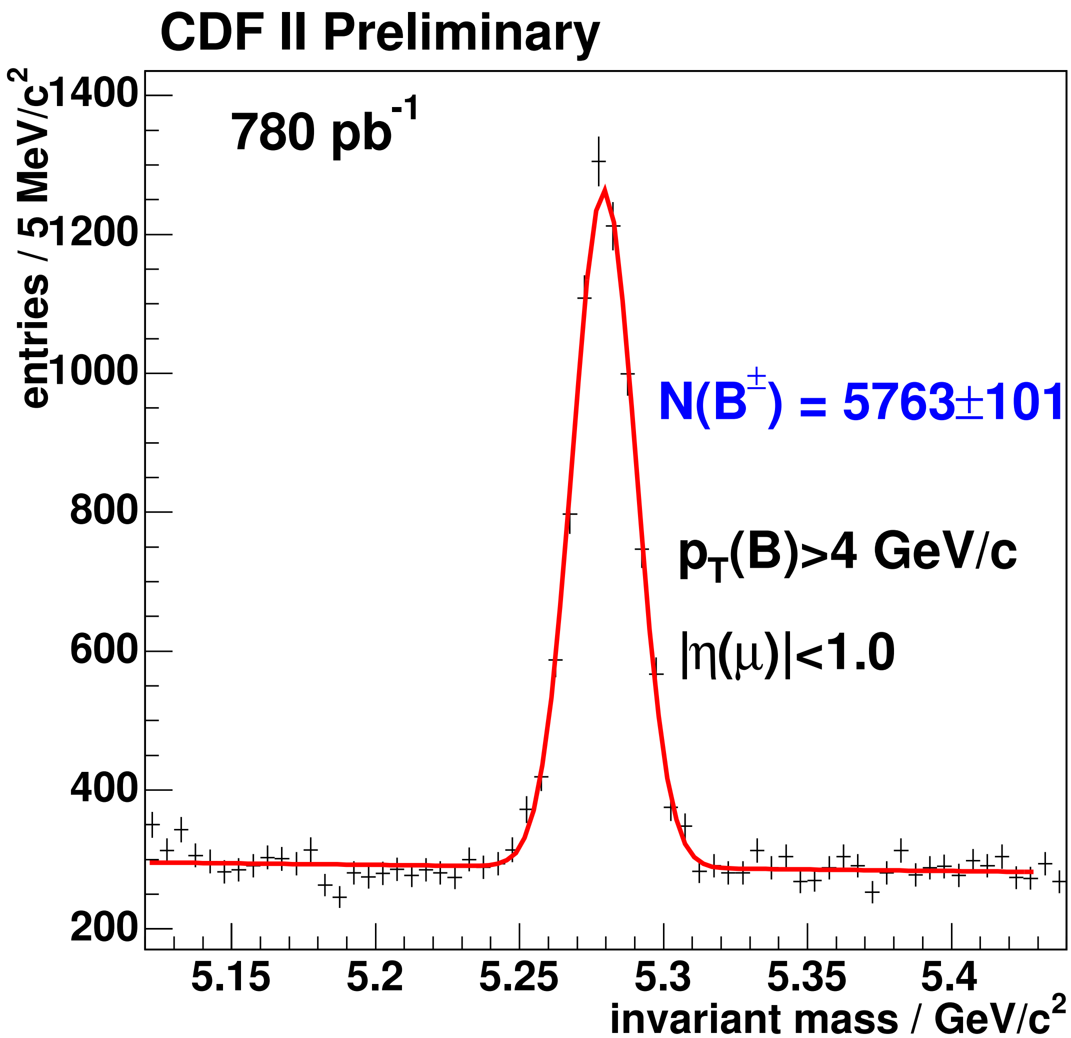

FIG. 8: The $\mu^+\mu^-K^+$ invariant mass distribution for events satisfying the baseline and vertex requirements for the combined U-U and U-X samples.

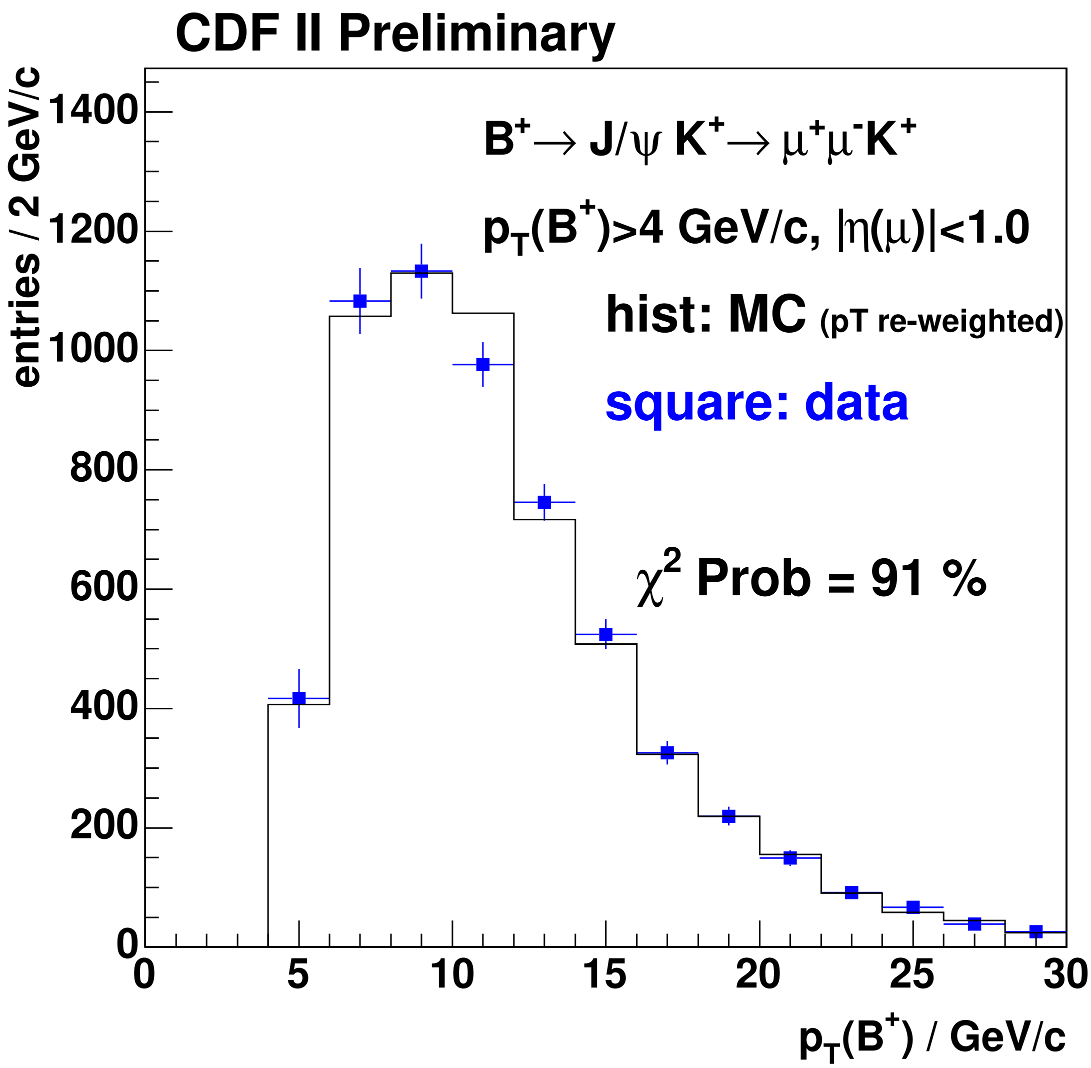

FIG. 9: The $p_T(B)$ spectrum from $B^+ \rightarrow J/\psi K^+$ events satisfying the baseline and vertex requirements. The MC distribution has been re-weighted to match the data.

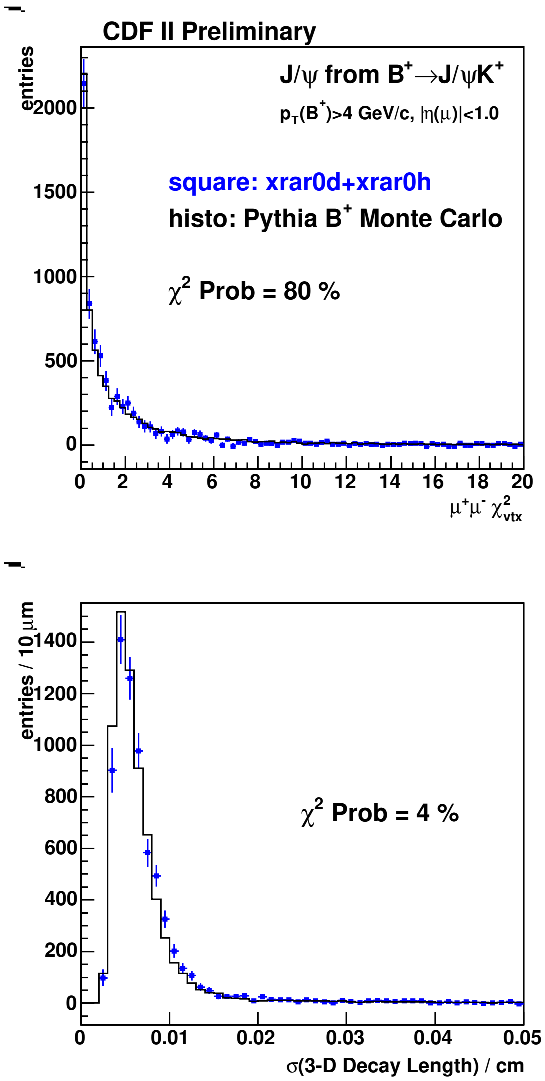

FIG. 10: The vertex χ^2, the 3D decay length, 3D decay length error, and decay length significance distributions from $B^+ \to J/\psi K^+$ using the J/ψ vertex

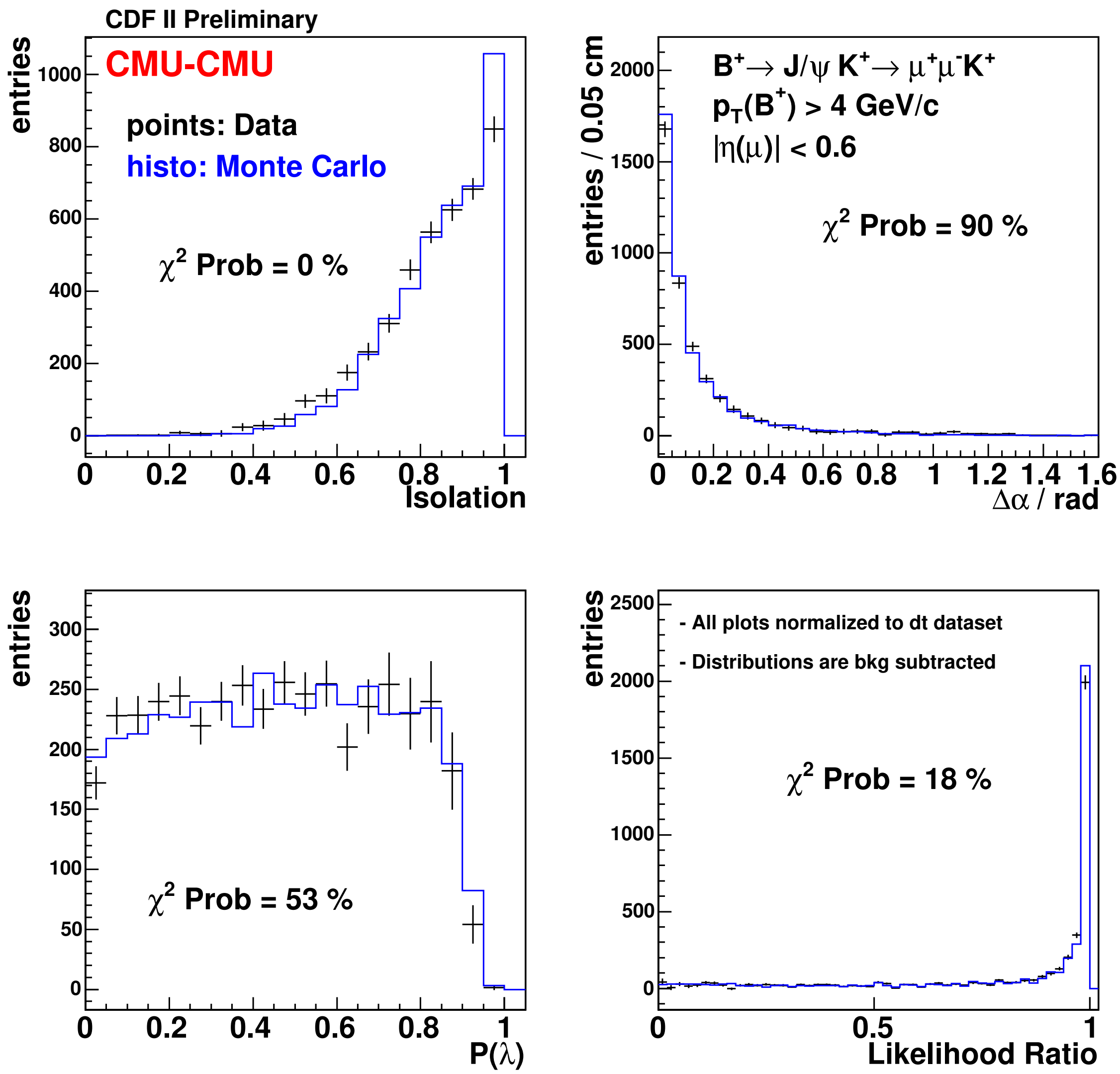

FIG. 11: A comparison of the input distributions (I,$\Delta\Theta$, and λ) and the resulting likelihood distribution from sideband subtracted $B^+ \to J/\psi K^+$ data to that from the Pythia MC sample described in the text for the CMU-CMU channel. Here, no corrections have been made to the isolation distribution.

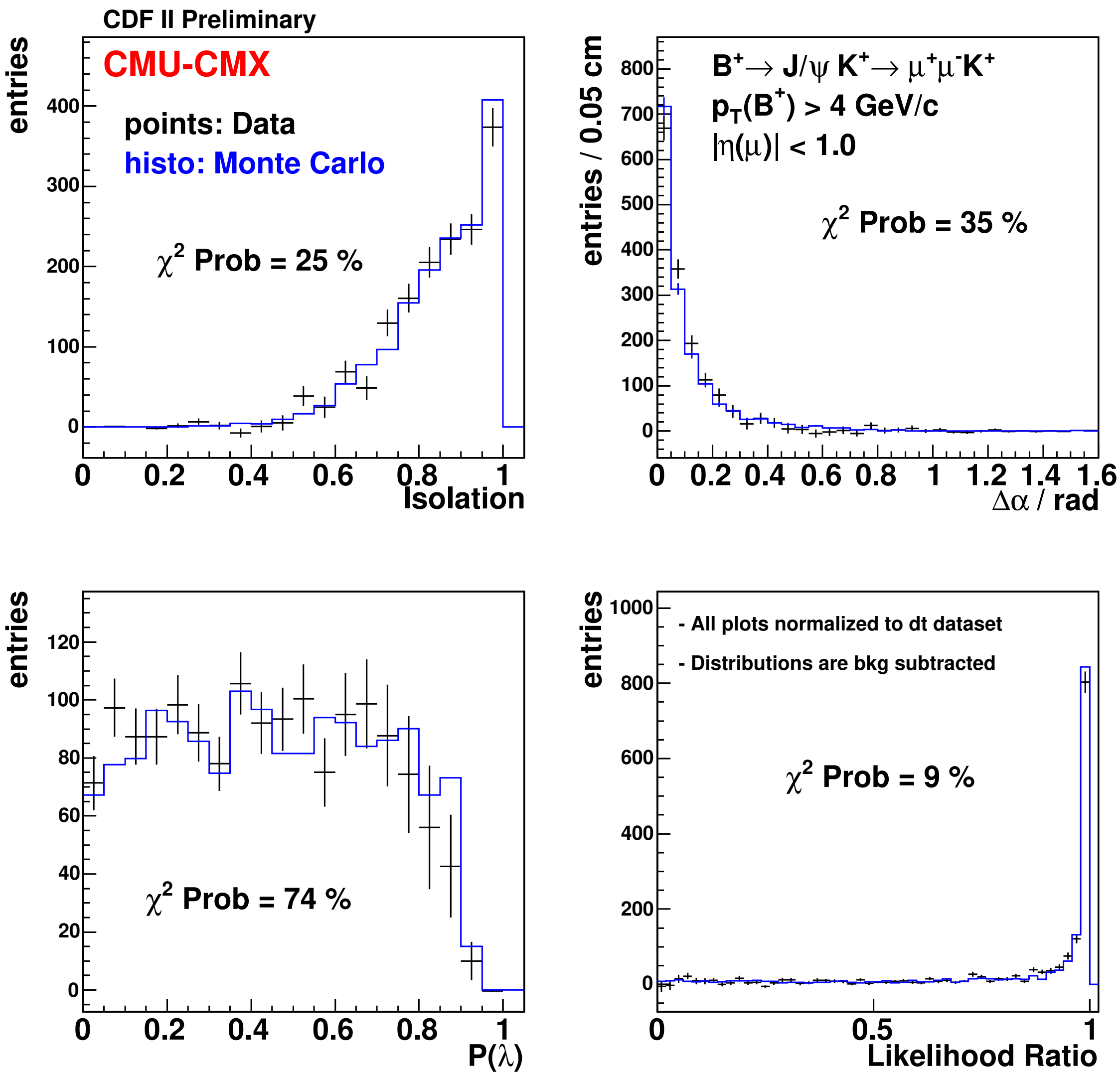

FIG. 12: A comparison of the input distributions (I,$\Delta\Theta$, and λ) and the resulting likelihood distribution from sideband subtracted $B^+ \to J/\psi K^+$ data to that from the Pythia MC sample described in the text for the CMU-CMX channel. Here, no corrections have been made to the isolation distribution.

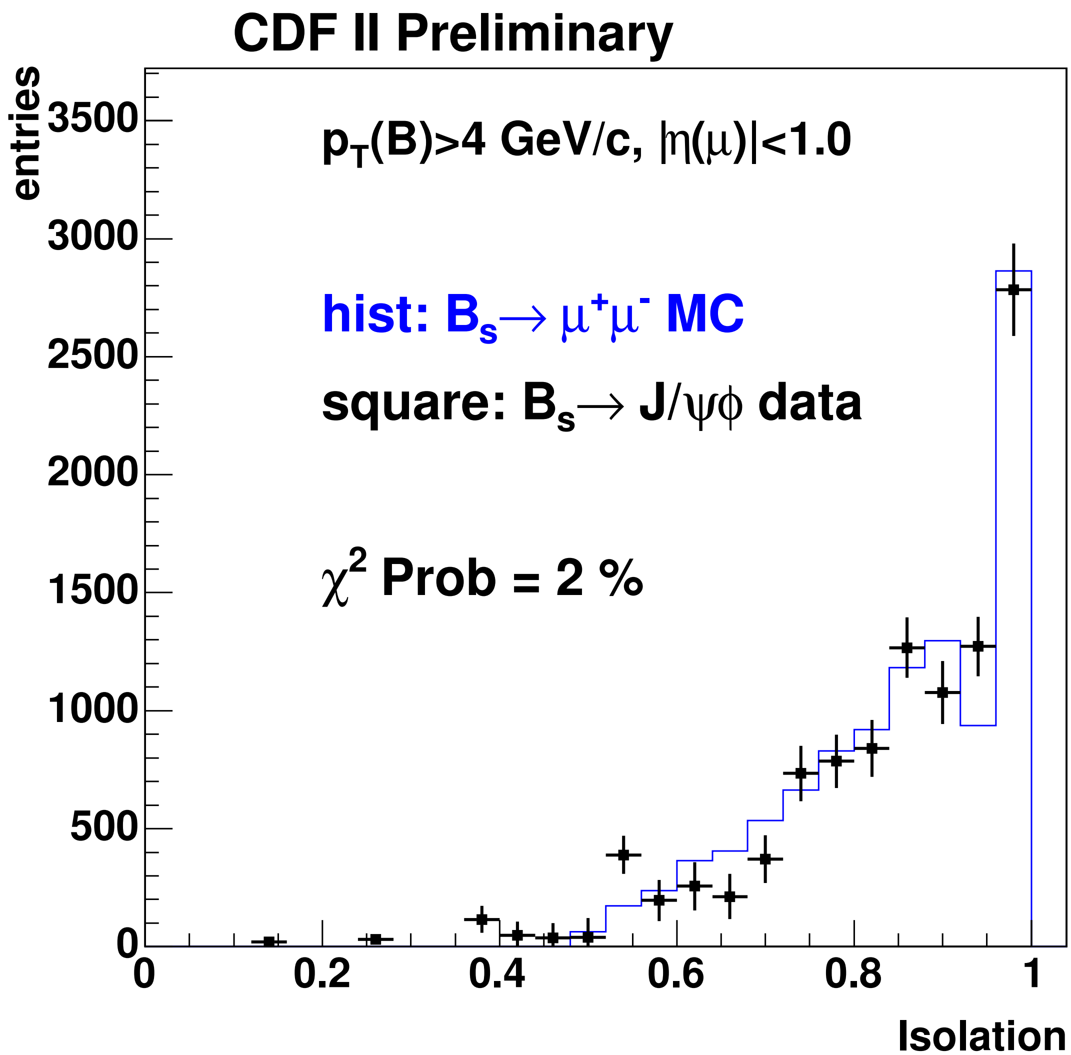

FIG. 13: A comparison of the B_s^0 isolation distribution between the $B_s^0 \rightarrow \mu^+\mu^-$ MC used in this analysis and sideband subtracted $B_s^0 \rightarrow J/\psi\phi$ candidates from the data. The data events have their $p_T(B)$ distribution re-weighted to reproduce the $B_s^0 \rightarrow \mu^+\mu^-$ MC distribution after inclusion of trigger and reconstruction effects. The MC distribution shown is before isolation re-weighting.

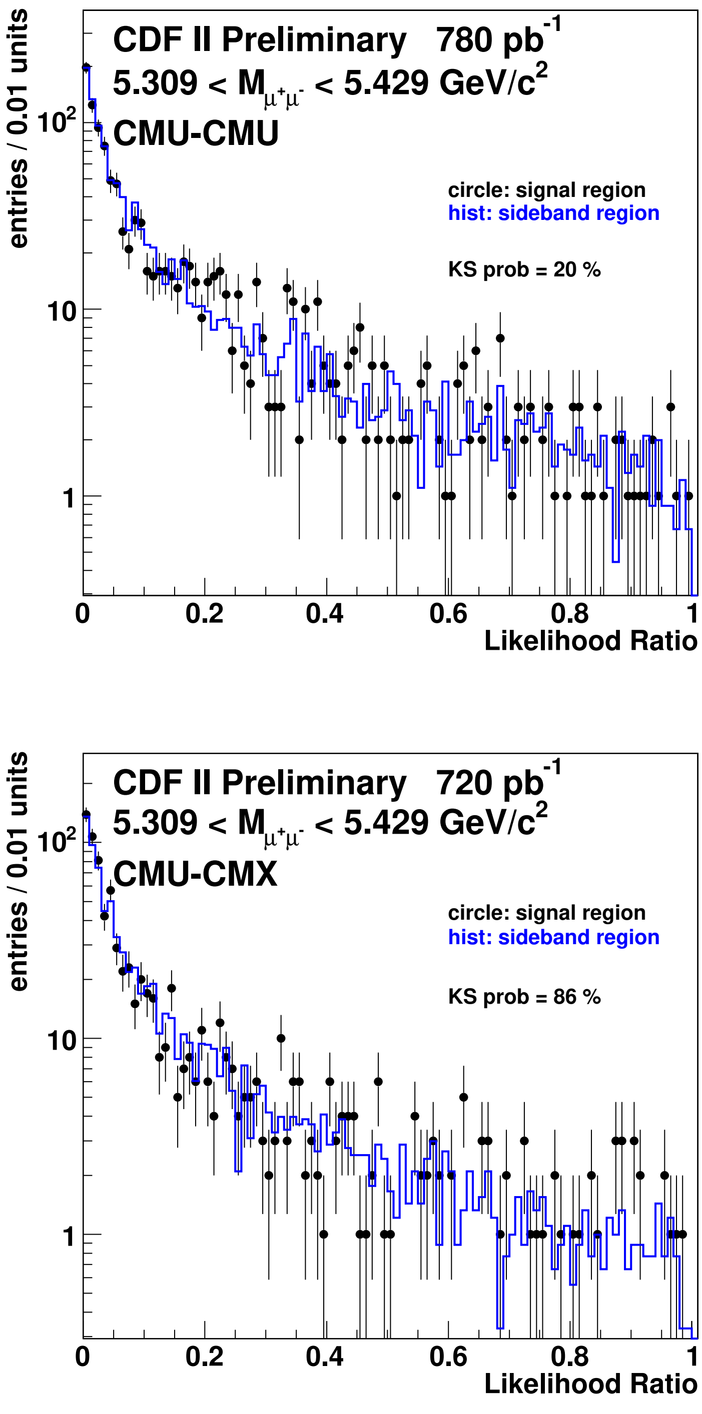

FIG. 14: The likelihood distribution for events in the B_s search window for the CMU-CMU (top) and CMU-CMX (bottom) channel. The likelihood distributions for events in the sideband regions are also shown (blue histogram).

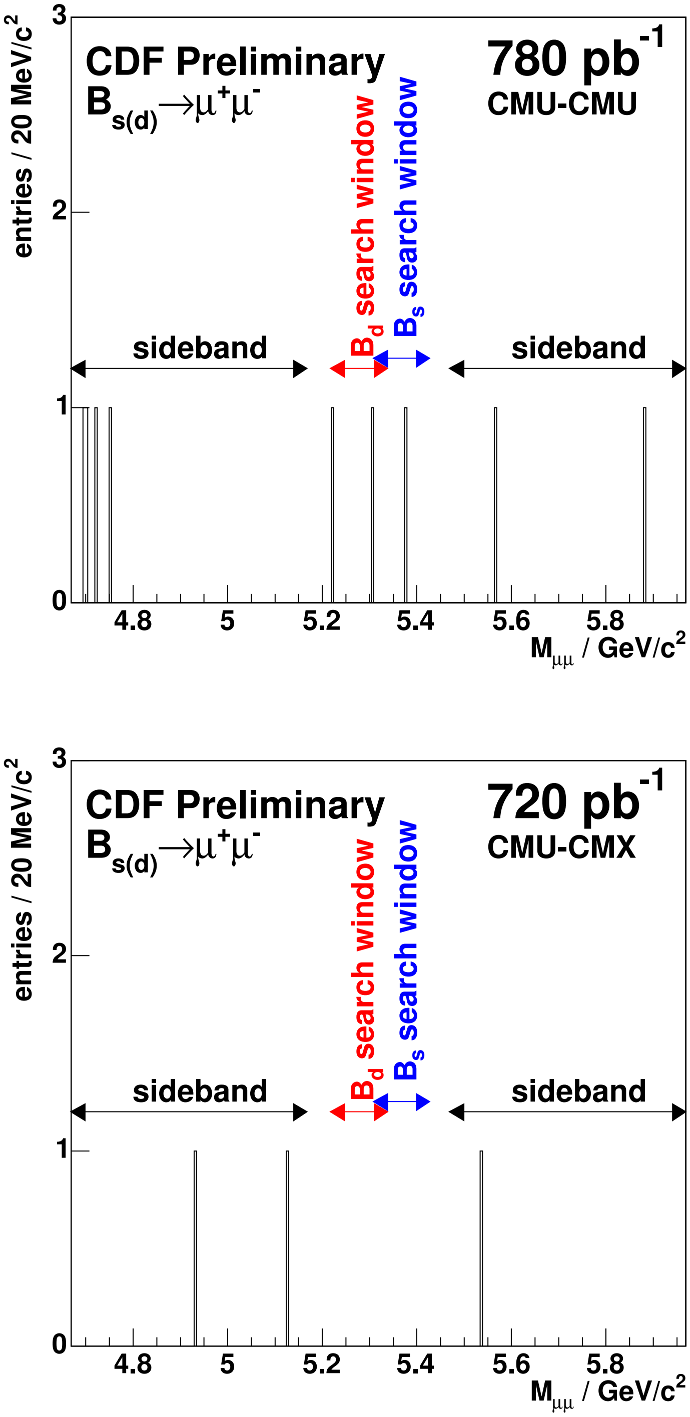

FIG. 15: The di-muon mass distribution for events satisfying all selection criteria in the CMU-CMU (top) and CMU-CMX (bottom) channels. There are three events in the CMU-CMU channel (1 event in B_s^0 and 2 events in B_d^0 mass windows). For the CMU-CMX channel, there are no events seen in either signal window.

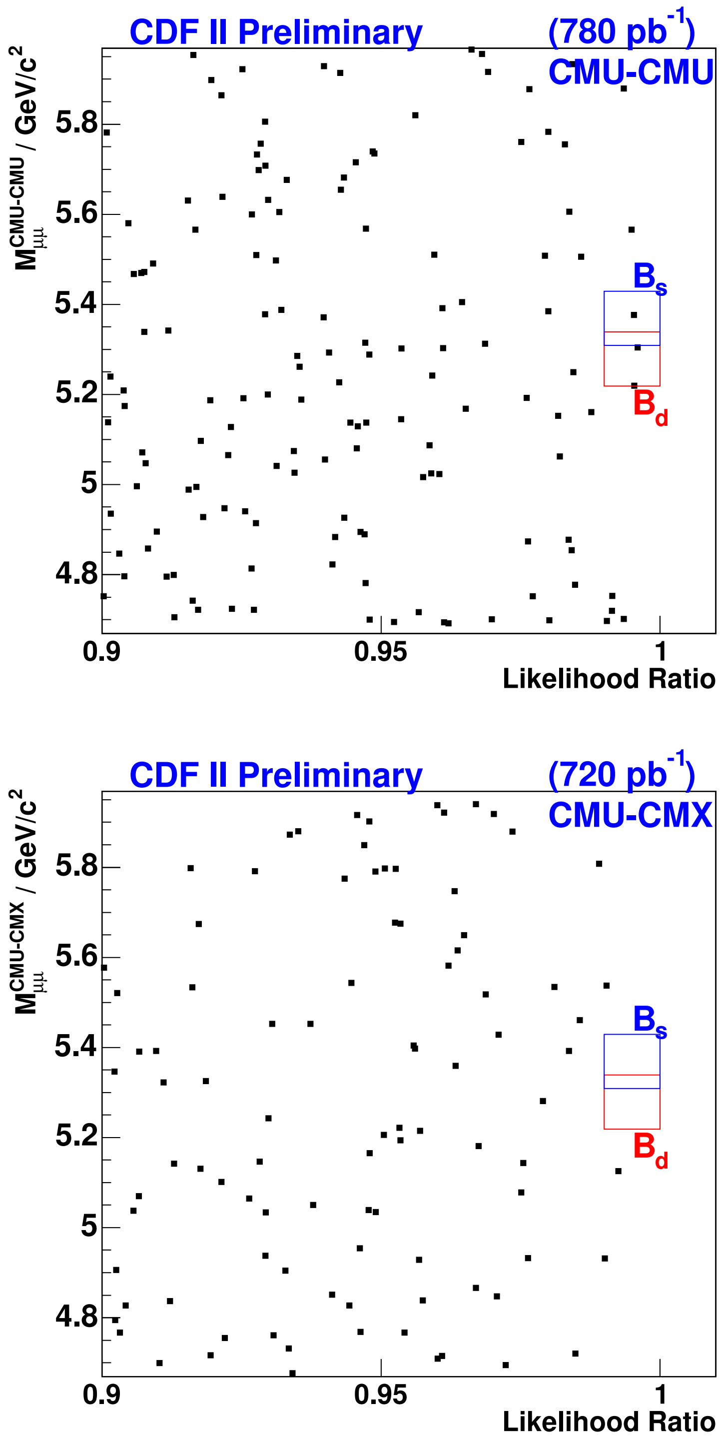

FIG. 16: The invariant mass distribution versus the event likelihood for events satisfying baseline in the CMU-CMU (top) and CMU-CMX (bottom) channels. Only events with $L_R > 0.90$ are shown. There are three events in the CMU-CMU channel (1 event in B_s^0 and 2 events in B_d^0 mass windows). For the CMU-CMX channel, there are no events seen in either signal window.

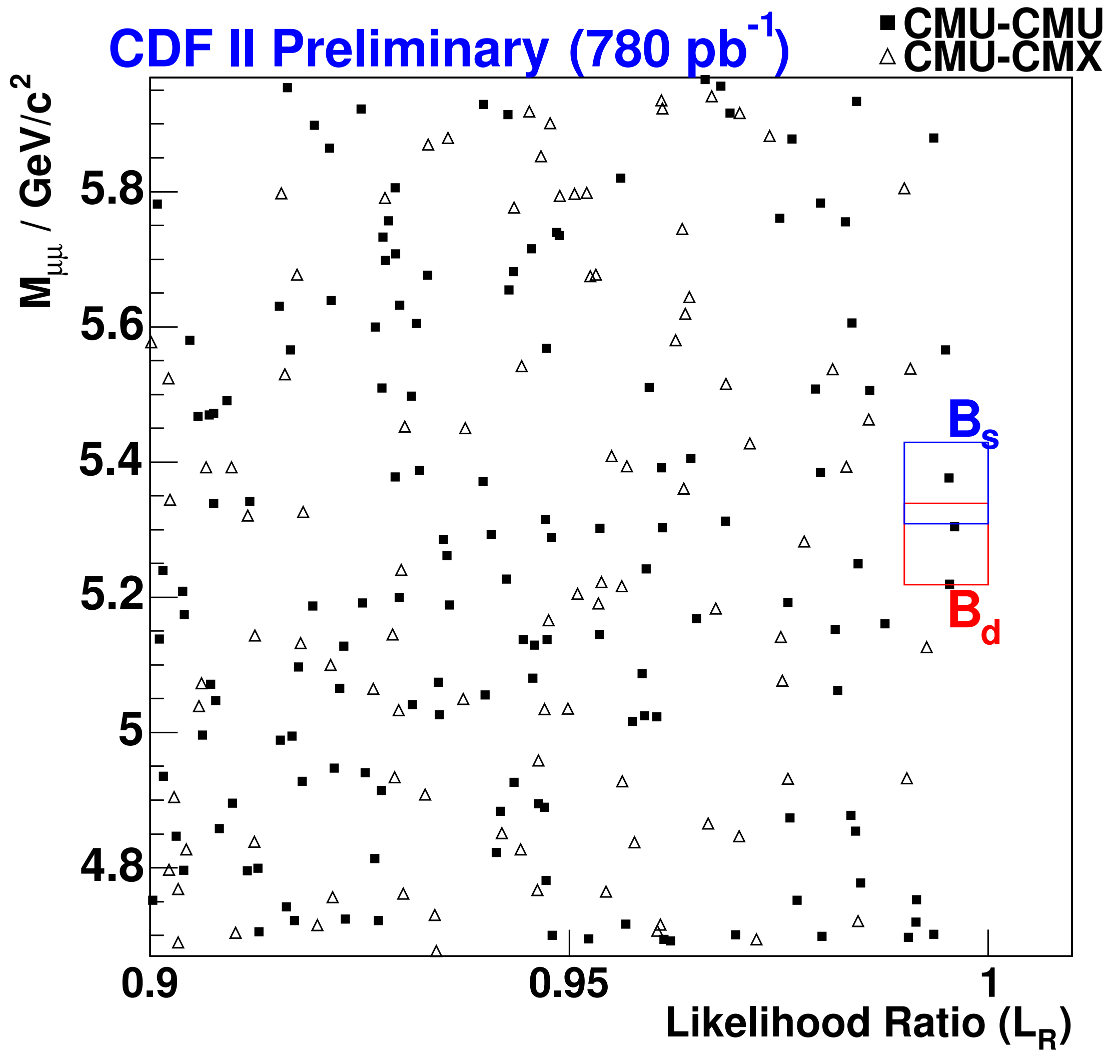

FIG. 17: The invariant mass distribution versus the event likelihood for events satisfying baseline in the CMU-CMU and CMU-CMX channels. Only events with $L_R > 0.90$ are shown.