CDF note 7670

A Search for $B_s^0 \to \mu^+\mu^-$ Decays Using 364 pb^{-1} of Data

CDF Collaboration

Abstract

A search for the rare decays $B_s \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ is presented using $364\,\mathrm{pb}^{-1}$ of data. Opposite sign $\mu^+\mu^-$ candidates where both muons are reconstructed in the central muon system and candidates where one muons is reconstructed in the central system and one in the forward system are considered and treated as separate channels. Using selection criteria determined using an *a priori* optimization we observe no candidate events, consistent with the expected background of 0.81 ± 0.12 and 0.66 ± 0.13 events in the central-central and central-forward channels, respectively. Using candidate $B^+ \to J/\psi K^+$ events collected on the same triggers for a relative normalization, we calculate a combined limit of $\mathrm{BR}(B_s \to \mu^+\mu^-) < 1.5 \times 10^{-7}$ at 90% confidence level and $\mathrm{BR}(B_s \to \mu^+\mu^-) < 2.0 \times 10^{-7}$ at 95% confidence level. The 90% (95%) confidence level limit for the $B_d \to \mu^+\mu^-$ decay is 3.8×10^{-8} (4.9×10^{-8}).

1 Introduction

In the Standard Model, the branching ratio of $B_s \to \mu^+\mu^-$ is estimated to be $\mathrm{BR}(B_s \to \mu^+\mu^-) = 3.5 \pm 0.9 \times 10^{-9}$ [1][2]. So far, the $B_s \to \mu^+\mu^-$ final state has not been experimentally observed and the best published limit uses $240\,\mathrm{pb}^{-1}$ of D0 data, $\mathrm{BR}(B_s \to \mu^+\mu^-) < 4.1 \times 10^{-7}$ at the 90% confidence level (CL) [3]. Our most recent publication uses $171\mathrm{pb}^{-1}$ of data yielding $\mathrm{BR}(B_s \to \mu^+\mu^-) < 5.8 \times 10^{-7}$ at the 90% CL [4]. In many SUSY models, this BR can be significantly enhanced by one to three orders of magnitude [5] [6] [7] [8] [9] [10] and is thus possibly observable in Run II. For large $\tan\beta$, this is the most sensitive probe of new physics available at the Tevatron experiments.

In this analysis, candidate $B^+ \to J/\psi K^+$ events[1] collected on the same triggers are used as a relative normalization to estimate the $\mathrm{BR}(B_s \to \mu^+\mu^-)$ as:

$$\mathrm{BR}(B_s \to \mu^+\mu^-) = \frac{N_{B_s}}{\alpha_{B_s}\epsilon^{\mathrm{total}}_{B_s}} \cdot \frac{\alpha_{B^+}\epsilon^{\mathrm{total}}_{B^+}}{N_{B^+}} \cdot \frac{f_u}{f_s} \cdot BR(B^+ \to J/\psi K^+) \cdot BR(J/\psi \to \mu^+\mu^-), \quad (1)$$

where N_{B_s} is the number of candidate $B_s \to \mu^+\mu^-$ events, α_{B_s} is the geometric and kinematic acceptance of the di-muon trigger for $B_s \to \mu^+\mu^-$ decays, $\epsilon^{\mathrm{total}}_{B_s}$ is the total efficiency (including trigger, reconstruction and analysis requirements) for $B_s \to \mu^+\mu^-$ events in the acceptance, with N_{B^+}, α_{B^+}, and $\epsilon^{\mathrm{total}}_{B^+}$ similarly defined for $B^+ \to J/\psi K^+$ decays; the ratio f_u/f_s accounts for the different b-quark fragmentation probabilities and is $(0.398 \pm 0.010)/(0.104 \pm 0.015) = 3.83 \pm 0.57$ where the (anti-)correlation between the uncertainties has been accounted for [11]; the final two terms are the relevant branching ratios $BR(B^+ \to J/\psi K^+) \cdot BR(J/\psi \to \mu^+\mu^-) = (1.00 \pm 0.04) \times 10^{-3} \cdot (5.88 \pm 0.10) \times 10^{-2} = (5.88 \pm 0.26) \times 10^{-5}$ [12].

Relative to our previous analysis using 171 pb^{-1} of data [4], we have significantly increased the acceptance by including triggered muons identified in both the central muon chambers, $|\eta| < 0.6$(CMU,CMUP) and the near forward muon system $0.6 < |\eta| < 1.0$(CMX). We treat candidates including near forward muons separately dividing the analysis into a central-central(CMU-CMU) and central-forward(CMU-CMX) channel. In addition we lower the $p_T(B)$ threshold to 4 GeV compared to the previous analysis that used a $p_T(B)$ threshold of 6 GeV. By using a multivariate likelihood discriminant we significantly reduce the background while maintaining high signal efficiency. These changes are discussed in more detail below.

Sections 2 and 3 describe the Data and Monte Carlo (MC) samples used in this analysis. The likelihood discriminant is introduced in Section 4 and is used to estimate backgrounds in Section 5. The acceptance and efficiency estimates for the $B_s \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decays are detailed in Section 6 and the normalization using $B^+ \to J/\psi K^+$ events is detailed in Section 7. With all this in hand, an optimization and the corresponding sensitivity are described in Section 8. The results and conclusions are discussed in Sections 9 and 10.

[1]Throughout this note, charge conjugation is implied.

2 Data Sample

For both the $B_s \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ samples we use rare B decay di-muon trigger paths. This trigger identifies opposite sign dimuon decays and additionally requires either that one of the muons be matched to a reconstructed muon stubs in both inner and outer muon systems(CMUP) or the scaler sum p_T or the two muons be greater than 5 GeV. A minimum p_T of 1.5 GeV, 3.0 GeV and 2.0 GeV is required to trigger individual muons in the CMU, CMUP and CMX systems respectively.

For both the $B_s \to \mu^+\mu^-$ and the $B^+ \to J/\psi K^+$ samples we make the following "baseline" requirements:

trigger match: The event must have fired at least one of the di-muon triggers and have two muon legs matched to the corresponding trigger primitives from the drift chamber(COT) and muon system.

goodrun: All of the relevant systems subsystems must be operating correctly including the trigger, silicon, drift chamber and muon systems. This corresponds to approximately 364 pb^{-1} of data. When we consider the inclusion of CMX triggers we additionally require the that subsystem to be operating correctly. There is approximately 336 pb^{-1} of data which include the CMX triggers.

Track quality: Each track is required to satisfy the default quality cuts in terms of the number drift chamber and a strict requirement on the number of silicon hits associated to each track. The silicon requirement is applied to insure good vertex is available.

muon quality: Candidate muon tracks matched to muon information from the CMU and CMX must have $p_T > 2.0$ and > 2.2 GeV, respectively. Both the CMU and CMX stubs must additionally satisfy a χ^2 matching requirement between the track and muon chamber stub information of $\chi^2_{r\phi} < 9$. The muons must be of opposite charge.

Finally, we require that the B-hadron satisfy $p_T(B) > 4$ GeV and $|y(B)| < 1$, where y is the rapidity.[2] For the two-track $B_s \to \mu^+\mu^-$ sample, the B-hadron momentum, $\vec{p}(B)$, is estimated using the vector sum of the $\mu^+\mu^-$ tracks, while for the three-track $B^+ \to J/\psi K^+$ sample it is estimated using the vector sum of the $\mu^+\mu^-K^+$ tracks.

2.1 The $B_s \to \mu^+\mu^-$ Sample

For the $B_s \to \mu^+\mu^-$ sample, events surviving the baseline requirements then constrain the muons to a common 3D vertex. In analogy to the standard 2D displacement, L_{xy}, we calculate a 3D displacement, L_{3D}, as the distance between the primary vertex and the fitted secondary vertex projected along the $\vec{p}(B)$ direction. Using the invariant mass of the tracks associated with the vertex, M_{vtx}, we also calculate the proper decay length, $\lambda = c\, L_{3D}\, M_{vtx}/|\vec{p}(B)|$.

We then make the following vertex requirements:

[2] The $p_T(B)$ requirement was included as part of the optimization described in Section 8.

$\chi^2_{\mathbf{vtx}} < \mathbf{15}$

$\sigma_{\mathbf{L_{3D}}} < \mathbf{0.0150\ cm}$

$\mathbf{L_{3D}} < \mathbf{1.0\ cm}$

$\mathbf{0} < \lambda < \mathbf{0.3\ cm}$

$\lambda/\sigma(\lambda) > \mathbf{2}$

where $\sigma_{L_{3D}}$ is the associated L_{3D} uncertainty, including contributions from the primary vertex determination. Except for the last, these additional cuts were chosen to minimally impact the signal efficiency, but eliminate particularly anomalous background events. The last of these requirements minimally affects $\epsilon^{\text{total}}_{B_s}$ but dramatically reduces the correlation among the discriminating variables as discussed in Section 4.

In the data, 22459 CMU-CMU and 14305 CMU-CMX events survive these cuts with an invariant mass satisfying $4.669 < M_{\mu^+\mu^-} < 5.969$ GeV. The resulting $M_{\mu^+\mu^-}$ distributions are shown in Figure 1. These events form our $B_s \to \mu^+\mu^-$ search sample and, at this stage, are an entirely background dominated sample.

2.2 The $B^+ \to J/\psi K^+$ Sample

For the $B^+ \to J/\psi K^+$ sample, events surviving the baseline requirements must also have a di-muon invariant mass near the world average J/ψ mass, $3.017 < M_{\mu^+\mu^-} < 3.117$ GeV. Candidate $K^\pm$ tracks must have a z_0 within 5 cm of the di-muon vertexes. For each candidate kaon we simultaneously constrain the $\mu^+\mu^-K^+$ tracks to a common 3D vertex and the di-muon invariant mass to the world average J/ψ mass. The resulting fit chi-squared probability must be larger than 10^{-5}. Finally, using the two track di-muon vertex to estimate L_{3D} and $\sigma_{L_{3D}}$, and the three track vertex to estimate $\vec{p}(B)$, the $B^+ \to J/\psi K^+$ candidates are required to satisfy the same vertex requirements as discussed in Section 2.1, including $\chi^2_{(\mu\mu)vtx} < 15$.

In the data, 12121 CMU-CMU and 5353 CMU-CMX events survive these cuts with an invariant mass satisfying $5.120 < M_{\mu^+\mu^-K^+} < 5.430$ GeV. These events form our $B^+ \to J/\psi K^+$ sample and the resulting $M_{\mu^+\mu^-K^+}$ distributions are shown in Figure 2. These are fit with a Gaussian plus first order polynomial and yield a fitted mean mass that's within 1 MeV of the world average $M(B^+)$ and a mass resolution of about $\sigma(M) = 11$ MeV for both the CMU-CMU and CMU-CMX channels. We define a signal region of $\Delta M = |M_{\mu^+\mu^-K^+} - M(B^+)| < 35$ MeV and estimate the number of B^+ candidates by simple sideband subtraction. A small correction accounting for the $B^+ \to J/\psi\pi^+$ contribution to the signal region is discussed in Section 7.

3 Monte Carlo Samples

To estimate our signal efficiency, we use a sample of $B_s \to \mu^+\mu^-$ decays generated using Pythia [13] with parameters tuned to inclusive B-hadron data [14] and an `EvtGen` [15] de-

cay table which forces $B_s \to \mu^+\mu^-$ decays. The events are simulated using GEANT [16] based full detector simulation. We use a realistic silicon simulation with the parameterized charge-deposition model and a realistic beam-line simulation. We re-weight the B-hadron p_T spectrum to match that measured in Run II [17].

A $B^+ \to J/\psi K^+$ MC sample is simulated in exactly the same manner as our $B_s \to \mu^+\mu^-$ sample, except that the `EvtGen` table forces $B^+ \to J/\psi K^+$ and $J/\psi \to \mu^+\mu^-$ decays. This sample is used to estimate the efficiency of some of the $B^+ \to J/\psi K^+$ selection requirements and to cross-check the Monte Carlo modeling of the likelihood discriminant by comparing to a sample of $B^+ \to J/\psi K^+$ events in the data.

All the MC samples are reconstructed using same software used to process the data and are required to satisfy the baseline requirements described in the previous section. Figure 3 compares the $p_T(B^+)$ spectrum as observed in the data with that obtained from our Pythia MC sample for events surviving the baseline and vertex requirements described above. Figure 4 shows the resulting invariant mass distributions for $B_s \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ events surviving the baseline and vertex requirements. These are each fit to a Gaussian distribution. The $B^+ \to J/\psi K^+$ fit yields a mean fitted mass and mass resolution consistent with that observed in the data. In Figure 5 the di-muon invariant mass distribution for $J/\psi \to \mu^+\mu^-$ from $B^+ \to J/\psi K^+$ candidates[3] from data and MC is fit to a Gaussian. In each case the mean fitted mass is within a few MeV of the world average J/ψ mass while the MC mass resolution is within 10% of that observed in the data. We use the fit to the MC $B_s \to \mu^+\mu^-$ sample to estimate the corresponding mass resolution to be $\sigma(B_s \to \mu^+\mu^-) = 24$ MeV.

4 Discriminating Variables for the $B_s \to \mu^+\mu^-$ Search

Our signal is simply identified by a pair of opposite charge muons whose invariant mass is consistent with the mass of the B_s. To reduce backgrounds, it is necessary to require that the muon pair is consistent with having come from a long lived hadron, by requiring they be displaced from the primary vertex. Potential sources of background include, continuum $q\overline{q} \to \mu^+\mu^-$ Drell-Yan production, sequential semi-leptonic $b \to c \to s$ decay, double semi-leptonic $g \to b\overline{b} \to \mu^+\mu^- X$ decay, $b(c) \to \mu X +$ fake, and fake + fake events. We explored a variety of discriminating variables and identified the following as among the most promising:

$\mathbf{M}_{\mu^+\mu^-}$**:** the invariant mass of the muon pair

λ**:** the proper decay length, as defined in Section 2.1

$\mathbf{\Delta}\alpha$**:** the opening angle between the B_s flight direction, $\vec{p}(B)$, and the direction of the decay vertex - estimated as the vector originating at the primary vertex and terminating at the muon-pair vertex

Isolation: the isolation of the candidate B_s defined as, $Iso = p_T^{B_s}/(p_T^{B_s} + \sum_i p_T^i(\Delta R < 1.0))$, where the sum is over all tracks within an $\eta - \phi$ cone of radius $R = 1.0$, centered on $\vec{p}(B)$;

[3]For this comparison the requirement on the three-track χ^2 fit probability is dropped.

the tracks must satisfy the drift chamber quality requirements described in Section 2 and have a z_0 within 1 cm of the mean z_0 of the two muons.

Figure 6 compares the distributions of these variables in the signal MC and the data for events surviving the baseline and vertex requirements described in Section 2. The shapes of the signal and background-dominated data distributions are clearly very different. To further reduce the background, we make very loose requirements on the isolation and $\Delta\alpha$ variables, $Iso > 0.50$ and $\Delta\alpha < 0.70$ rads. These requirements leave 6242 and 4908 events in the CMU-CMU and CMU-CMX $B_s \to \mu^+\mu^-$ search samples, respectively. Using the $B_s \to \mu^+\mu^-$ MC sample, we estimate the efficiency for these requirements to be 92%. For the remainder of the note it should be understood that these loose Iso and $\Delta\alpha$ requirements are made for the $B_s \to \mu^+\mu^-$ samples.

The correlations among the four discriminating variables are displayed as profile histograms in Figures 7-8 for CMU-CMU and CMU-CMX data surviving the baseline and vertex cuts. We quantify the correlations by calculating the linear correlation coefficient between each pair of variables using:

$$\rho_{xy} = \frac{1}{N-1}\frac{\sum_i^N (x_i - \overline{x})(y_i - \overline{y})}{\sigma_x \sigma_y} \tag{2}$$

where N is the total number of events, $\overline{x}$ and σ_x^2 are the mean and variance of the variable x, respectively, and similarly for y. The resulting values are given in Table 1. None of the variables are significantly correlated.

	CMU-CMU				CMU-CMX			
	mass	Iso	λ	$\Delta\alpha$	mass	Iso	λ	$\Delta\alpha$
mass	1	0.012	0.037	0.045	1	0.029	0.018	0.026
Iso		1	0.003	0.028		1	0.009	0.049
λ			1	0.027			1	0.062
$\Delta\alpha$				1				1

Table 1: The linear correlation coefficients among the discriminating variables described in Section 4 for CMU-CMU (left) and CMU-CMX (right) $\mu^+\mu^-$ pairs in the data, satisfying the baseline and vertex requirements. The uncertainty is about ± 0.013 (± 0.014) for each CMU-CMU (CMU-CMX) coefficient and is estimated using the method described in reference [18].

If we were to employ the same cuts on these discriminating variables as reported in reference [4] the expected CMU-CMU background would be > 5 events for this lower $p_T(B)$ requirement with a similar expected background in the CMU-CMX channel. It would clearly be better to develop criteria which significantly suppresses the background while maintaining a high signal efficiency. In the absence of significant correlations among a set a variables, a simple multi-variate relative likelihood can be constructed which provides approximately maximal discriminating power.

4.1 A Multi-variate Relative Likelihood Discriminant

The relative likelihood is simply defined as:

$$LH = \frac{\prod_i \mathbf{P}_s(x_i)}{\prod_i \mathbf{P}_s(x_i) + \prod_j \mathbf{P}_b(x_j)} \tag{3}$$

where $i, j \in [1, N_{var}]$, N_{var} is the number of variables used to construct the likelihood, and $\mathbf{P}_{s(b)}(x_i)$ is the probability that a signal (background) event has an observed x_i. By construction, $LH \in [0, 1]$, with large LH implying that an event is signal-like. As an example, we construct a likelihood using the same (2D) variables employed in [4] for CMU-CMU events with $p_T(B) > 6$ GeV satisfying the baseline and vertex requirements. For the same signal efficiency, the LH can reduce the background by more than a factor of two. By making more stringent LH requirements, more dramatic reductions in the background are possible while still maintaining high signal efficiency.

For this analysis we construct a relative likelihood using (Iso,$\Delta\alpha$,$\mathcal{P}(\lambda)$), where $\mathcal{P}(\lambda) = \exp^{-\lambda/\lambda_{B_s}}$ ($\lambda_{B_s} = 438\ \mu m$, the world average B_s lifetime) and is the probability that a real $B_s \to \mu^+\mu^-$ decay would yield a proper decay length at least as large as that observed. The variable $\lambda_{B_s} = 438\ \mu m$ is chosen because it's expected to be a uniform distribution $\mathcal{P}(\lambda) \in [0, 1]$, unaffected by λ resolution effects. Even after the $\lambda/\sigma(\lambda) > 2$ requirement, the distribution is flat up to about $\mathcal{P}(\lambda) > 0.8$.

We use binned histograms to estimate the probability distributions, $\mathbf{P}(x_i)$. The signal and background probability distributions are compared in Figures 9 and 10 with the binning used to construct the LH variable. The probabilities are determined from events satisfying the baseline and vertex requirements and for the background are taken from the mass sidebands, $4.669 < M_{\mu^+\mu^-} < 5.169$ GeV and $5.469 < M_{\mu^+\mu^-} < 5.969$ GeV, while for the signal are taken from even numbered $B_s \to \mu^+\mu^-$ MC events. As discussed in Section 6, we will determine the signal efficiency of requiring $LH > LH_{cut}$ using the statistically independent odd-numbered $B_s \to \mu^+\mu^-$ events. The resulting likelihood curves are shown in Figure 11 for signal and background events.

5 Background Estimate

A common method for estimating backgrounds in this type of search is to apply all cuts in mass sideband regions and scale the surviving number events to get a background prediction in the signal mass region. The problem with this method is that very few events survive all cuts. This makes optimization difficult since the expected background is statistically consistent over a large fraction of the cut-parameter space. One way to improve the background estimate is to factorize the expected rejection for each cut separately. This will yield an accurate background expectation only when the correlations among the cut variables are small, which is the case here. Thus we can estimate our background as

$$N_{bgd} = N_{SB} \cdot R_{LH} \cdot R_{mass} \tag{4}$$

where N_{SB} is the number of sideband events passing the baseline and vertex requirements and having $\lambda > 0$, R_{LH} is the expected background rejection for a given LH cut, and R_{mass} is the expected number of events in the signal mass-window given a known number of surviving background events in the sideband regions. Since the mass is uncorrelated with the rest of the variables (and hence, with LH), we can evaluate R_{LH} on samples with very loose cuts, thus reducing the associated uncertainty. This method yields background estimates whose uncertainties are significantly smaller than the usual method employed.

5.1 Estimating the Mass Window Rejection

A linear fit to the $M_{\mu^+\mu^-}$ distribution, for $\mu^+\mu^-$ events in the data passing the baseline and vertex cuts, is shown in Figure 1 and yields $\chi^2/dof = 9/10$ and $4/10$ for the CMU-CMU and CMU-CMX channels, respectively. In this case, if the sidebands are chosen to be symmetric about the signal region, R_{mass} is given by the ratio of the widths of the signal to sideband regions. We define a signal region that is ± 100 MeV around the world average B_s mass, $M_{B_s} = 5369 \pm 2$ MeV. From the signal Monte Carlo sample we estimate the $M_{\mu^+\mu^-}$ resolution to be about ± 25 MeV, so this corresponds to a $\pm 4\sigma$ window. For now we keep this large window to avoid any bias in our cut optimization, when we "blind" ourselves to the signal region and use only sideband information. For the final analysis, we shrink the signal window to ± 60 MeV (corresponding to $\pm 2.5\sigma$) as discussed in Sections 6.4 and 8. Since the B_d mass is only 90 MeV lower, and our cuts will have similar efficiency for $B_d \to \mu^+\mu^-$ decays, we also include in our signal region a ± 100 MeV window around the B_d mass. Our total signal window is defined as $5.169 < M_{\mu^+\mu^-} < 5.469$ GeV. We then symmetrically define our sidebands to include an additional 0.500 GeV on either side of the signal window. The ratio of the widths is then

$$R_{mass} = \frac{\Delta M^{\mathrm{Sig}}_{\mu^+\mu^-}}{\sum_{i=lo}^{hi} \Delta M^{\mathrm{SB_i}}_{\mu^+\mu^-}} = \frac{0.3}{0.5 + 0.5} = 0.3 \tag{5}$$

For the final analysis, when we shrink the mass windows to ± 60 MeV, $R_{mass} = 0.12$. In the data, $N_{SB}^{CMU} = 4853$ and $N_{SB}^{CMX} = 3768$ for the CMU-CMU and CMU-CMX channels, respectively.

5.2 Estimating the Likelihood Rejection

We estimate R_{LH} from a likelihood curve simulated using a toy-MC method[4]. Using events surviving the baseline and vertex requirements, we form the cumulative distribution for each of the likelihood input variables, Iso, $\Delta\alpha$, and $\mathcal{P}(\lambda)$. For each event, the toy-MC throws a uniform random number for each of these variables and uses the cumulative distributions to extract the set of (Iso,$\Delta\alpha$,$\mathcal{P}(\lambda)$). This works because the correlations among these variables are so small that the toy-MC can ignore them and still produce a good estimate of the likelihood curve. By using the toy-MC, we can significantly reduce the statistical uncertainty of the R_{LH} estimate. A comparison of the likelihood curve from data sideband events and

[4] $R_{LH} \equiv \frac{\#(\text{surviving } LH>\text{cut})}{\#(\text{passing baseline+vtx cuts})}$

the likelihood curve from the toy-MC is shown in Figure 12. Using 100k toy-MC events, we estimate R_{LH} for the $B_s \to \mu^+\mu^-$ search sample for a variety of possible cut values. The results are summarized in Table 2.

	R_{LH}	
cut value	CMU-CMU	CMU-CMX
$LH > 0.85$	0.0245 ± 0.0005	0.0226 ± 0.0005
$LH > 0.90$	0.0165 ± 0.0004	0.0150 ± 0.0004
$LH > 0.92$	0.0130 ± 0.0004	0.0120 ± 0.0003
$LH > 0.95$	0.0082 ± 0.0003	0.0076 ± 0.0003
$LH > 0.98$	0.0031 ± 0.0002	0.0031 ± 0.0002
$LH > 0.99$	0.0014 ± 0.0001	0.0015 ± 0.0001

Table 2: The LH rejection factor calculated from the $\mu^+\mu^-$ toy-MC. Only the statistical uncertainties are included.

To quantify the uncertainty associated with the statistics of the data sideband sample used to estimate the cumulative distribution, we simultaneously fluctuate each bin of the cumulative distribution $\pm 1\sigma$(stat), where σ(stat) is the associated statistical uncertainty for a particular bin, and re-evaluate R_{LH}. The bin-to-bin correlations are properly accounted for. We assign half the observed difference between the two fluctuated samples as a $\pm 15\%$ and $\pm 19\%$ uncertainty to the R_{LH} estimate in the CMU-CMU and CMU-CMX channel, respectively. We also vary the manner in which we determine the cumulative distributions used by the toy-MC, the total number of toy-MC events generated, and the random number seed and compare the R_{LH} determined to that given in Table 2. All the observed differences were smaller than the assigned statistical uncertainty and no additional systematic uncertainty is thus assigned.

5.3 Residual Background Contributions from B-decays

$B \to h^+h^-$ (where $h^+ = \pi^+$ or K^+) and/or generic B-hadron decays, $B \to \mu^+X$, might anomalously contribute to the background. In particular if the $M_{\mu^+\mu^-}$ distribution were non-linear (as is clearly the case for the $B \to h^+h^-$ decays) or there were significant correlations among the discriminating variables, then the data-driven background estimate described in the previous section my not properly account for these events.

Using muon-fake rates separately determined from the data for π^+, π^-, K^+, and K^- as a function of track p_T, the expected contributions for $B_d \to \pi^+\pi^-$, $K^+\pi^-$, K^+K^-, and $B_s \to \pi^+\pi^-$, π^+K^-, K^+K^- to the signal region are estimated. $B \to h^+h^-$ decays are found to contribute at levels at least 100 times smaller than our expected sensitivity. Since these decays negligibly contribute to the background we consider them no further.

Using a generic $b\bar{b}$ MC sample, we found for generic $B \to \mu^+X$ decays, that $M_{\mu^+\mu^-}$ is linear and that no significant correlations exist among the discriminating variables. We

also demonstrated that a background estimate which exploits these assumptions (as does the method proposed in the previous section) accurately accounts for these events.

5.4 Cross-Checks in Control Samples

To help build confidence in our method for estimating the background we perform some cross-checks in several control samples. In particular, we define the following samples:

OS+: opposite-sign muon pairs, passing the baseline and vertex cuts and having $\lambda > 0$; this is our signal sample, and will not be used for cross-checks;

OS-: opposite-sign muon pairs, passing the baseline and vertex cuts and having $\lambda < 0$;

SS+: same-sign muon pairs, passing looser baseline and vertex cuts and having $\lambda > 0$;

SS-: same-sign muon pairs, passing looser baseline and vertex cuts and having $\lambda < 0$.

FM+: opposite-sign fake-muon pairs, at least one leg of which is required to *fail* the muon stub-track matching requirement, passing looser baseline and vertex cuts and having $\lambda > 0$;

For the "looser" baseline cuts, we remove the drift chamber trigger matching requirements and only demand $p_T^\mu > 1.5$ GeV. This is necessary in order to get sufficient SS and FM statistics. We can use the $\lambda < 0$ samples because of the small correlations among the discriminating variables. It should be understood in what follows that when using samples OS- or SS-, the following transformations are made: $\lambda \to -\lambda$ and $\Delta\alpha \to \pi - \Delta\alpha$.

For each of the control samples we first verify that there exist no significant correlations among four discriminating variables. We also compare the likelihood curve from the toy-MC with that observed using the data sideband events in each sample. The predicted background is compared with the observed number of events in the mass signal region for $LH > 0.50$, 0.90, and 0.99, which roughly correspond to $R_{LH} = 0.10$, 0.01, and 0.001, respectively. As demonstrated in Table 3, there is no significant discrepancy between the predicted and observed number of events in either the CMU-CMU or CMU-CMX channel.

6 Estimate of Signal Efficiency

We estimate the total acceptance times efficiency for $B_s \to \mu^+\mu^-$ decays as $\alpha_{B_s} \cdot \epsilon_{B_s}^{total} = \alpha_{B_s} \cdot \epsilon_{B_s}^{trig} \cdot \epsilon_{B_s}^{reco} \cdot \epsilon_{B_s}^{LH}$, where α_{B_s} is the geometric and kinematic acceptance of the di-muon triggers, $\epsilon_{B_s}^{trig}$ is the trigger efficiency for events within the acceptance, $\epsilon_{B_s}^{reco}$ is the di-muon reconstruction efficiency - including the baseline and vertex requirements - for events passing the trigger, and $\epsilon_{B_s}^{LH}$ is the efficiency for $B_s \to \mu^+\mu^-$ events to satisfy the likelihood requirement for events surviving the reconstruction cuts. The analogous expression for $B^+ \to J/\psi K^+$ decays is $\alpha_{B^+} \cdot \epsilon_{B^+}^{total} = \alpha_{B^+} \cdot \epsilon_{B^+}^{trig} \cdot \epsilon_{B^+}^{reco}$, where the terms are defined as for the B_s except that

sample	LH cut	CMU-CMU pred	CMU-CMU obsv	CMU-CMX pred	CMU-CMX obsv
OS-	> 0.50	236 ± 4	235	172 ± 3	168
	> 0.90	37 ± 1	32	33 ± 1	36
	> 0.99	2.8 ± 0.2	2	3.6 ± 0.2	3
SS+	> 0.50	2.3 ± 0.2	0	2.8 ± 0.3	3
	> 0.90	0.25 ± 0.03	0	0.44 ± 0.04	0
	> 0.99	< 0.1	0	< 0.1	0
SS-	> 0.50	2.7 ± 0.2	1	3.7 ± 0.3	4
	> 0.90	0.35 ± 0.03	0	0.63 ± 0.06	0
	> 0.99	< 0.1	0	< 0.1	0
FM+	> 0.50	84 ± 2	84	21 ± 1	19
	> 0.90	14.2 ± 0.4	10	3.9 ± 0.2	3
	> 0.99	1.0 ± 0.1	2	0.41 ± 0.03	0

Table 3: A comparison of the predicted and observed number of events in the signal mass region as a function of likelihood cut for the various control samples. No significant discrepancies are observed. Only the statistical uncertainties are included.

$\epsilon^{reco}_{B^+}$ includes both kaon and di-muon reconstruction efficiencies. Note that there is no $\epsilon^{LH}_{B^+}$ term since we make no likelihood cut on the $B^+ \to J/\psi K^+$ events. Using these expressions we can rewrite equation 1 as

$$\mathrm{BR}(B_s \to \mu^+\mu^-) = \frac{N_{B_s}}{N_{B^+}} \cdot \frac{\alpha_{B^+}}{\alpha_{B_s}} \cdot \frac{\epsilon^{total}_{B^+}}{\epsilon^{total}_{B_s}} \cdot \frac{1}{\epsilon^{LH}_{B_s}} \cdot \frac{f_u}{f_s} \cdot BR(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+) \qquad (6)$$

. Table 4 summarizes the inputs to equation 6. These inputs are discussed in detail in the following sections.

6.1 Acceptance

For this analysis we define our acceptance relative to those B_s and B^+ which have $p_T(B) > 4\,\mathrm{GeV}$ and $|y| < 1.0$. The acceptance is largely driven by the geometric structure of the detector and the kinematic requirements of the dimuon trigger paths.

The $B_s \to \mu^+\mu^-$ acceptance is defined as the fraction of $B_s \to \mu^+\mu^-$ events with $p_T(B_s) > 4\,\mathrm{GeV}$ and $|y(B_s)| < 1$ that satisfy the $\mu^+\mu^-$ requirements. The $B^+ \to J/\psi K^+$ acceptance is defined as the fraction of $B^+ \to J/\psi K^+$ events with $p_T(B^+) > 4\,\mathrm{GeV}$ and $|y(B^+)| < 1$ that satisfy the $\mu^+\mu^-$ and kaon requirements. We evaluate these using the Pythia samples described in Section 3 to find $\alpha_{B^+}/\alpha_{B_s} = 0.297 \pm 0.008$ and $\alpha_{B^+}/\alpha_{B_s} = 0.191 \pm 0.006$ for the CMU-CMU and CMU-CMX channels, respectively, where only the statistical uncertainties have been included.

We use a fast b hadron generator that produces b hadrons with the measured differential

	CMU-CMU		CMU-CMX	
$(\alpha_{B^+}/\alpha_{B_s})$	0.297 ± 0.020	$(\pm 7\%)$	0.191 ± 0.013	$(\pm 7\%)$
$(\epsilon^{total}_{B^+}/\epsilon^{total}_{B_s})$	0.921 ± 0.034	$(\pm 4\%)$	0.915 ± 0.034	$(\pm 4\%)$
$\epsilon^{LH}_{B_s}$	0.348 ± 0.035	$(\pm 10\%)$	0.360 ± 0.022	$(\pm 6\%)$
N_{B^+}	1785 ± 60	$(\pm 3\%)$	696 ± 39	$(\pm 6\%)$
f_u/f_s	3.83 ± 0.57	$(\pm 15\%)$	3.83 ± 0.57	$(\pm 15\%)$
$BR(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+)$	$(5.88 \pm 0.26) \times 10^{-5}$	$(\pm 4\%)$	$(5.88 \pm 0.26) \times 10^{-5}$	$(\pm 4\%)$
ses	$(1.0 \pm 0.2) \times 10^{-7}$	$(\pm 20\%)$	$(1.6 \pm 0.3) \times 10^{-7}$	$(\pm 19\%)$

Table 4: A summary of the inputs used in equation 6 to estimate the BR($B_s \to \mu^+\mu^-$)$\dot{\text{T}}$he relative uncertainties are given parenthetically. The single-event-sensitivity, *ses*, corresponding to $N_{B_s} = 1$, is shown in the last row.

cross section [17] to evaluate systematic uncertainties due to variations in the b-quark mass, fragmentation modeling, and the renormalization scale. The default Pythia and fast MC samples yield acceptances that are consistent within 1% (5%) relative for the CMU-CMU (CMU-CMX) channel. For each systematic, we calculate the ratio of $B^+ \to J/\psi K^+$ to $B_s \to \mu^+\mu^-$ acceptances for the $+1\sigma$ and -1σ samples. The resulting ratios are given in Table 5 normalized to the acceptance ratio obtained from the default samples. Note that all are statistically consistent with 1. We assign the typical statistical uncertainty of these comparisons as the associated systematic uncertainty, or $\pm 6\%$ ($\pm 7\%$) relative, for the CMU-CMU (CMU-CMX) channel.

sample		CMU-CMU $\Delta\alpha$	CMU-CMX $\Delta\alpha$
mass	$+1\sigma$	0.97 ± 0.06	0.92 ± 0.07
	-1σ	1.09 ± 0.06	1.02 ± 0.08
fragmen-	$+1\sigma$	0.94 ± 0.06	0.96 ± 0.07
tation	-1σ	1.00 ± 0.06	0.92 ± 0.07
renormali-	$+1\sigma$	0.94 ± 0.06	0.92 ± 0.07
zation	-1σ	1.07 ± 0.06	0.92 ± 0.07

Table 5: For each systematic sample generated, $\Delta\alpha$ is the ratio of $\alpha_{B^+}/\alpha_{B_s}$ for that sample normalized to the same ratio as determined from the default samples. Deviations from 1 quantify the associated systematic uncertainty. The uncertainties shown arise from the MC statistics. The CMU-CMU and CMU-CMX channels are shown separately.

6.2 Trigger Efficiencies

The trigger efficiencies are determined from data using samples of unbiased muon and J/ψ events. The single muon trigger efficiency is parameterized as a function $\epsilon_{\mu}^{muon}(run, p_T^{\mu}, |\eta^{\mu}|, \phi^{\mu}) \equiv \epsilon_{\mu}^{muon}(run, \mathbf{\Omega})$ for the CMU and CMX separately. The double-leg efficiency relevant for this analysis is estimated by the convolution of the parameterization with the $(p_T^{\mu+}, |\eta^{\mu+}|, \phi^{\mu+}, p_T^{\mu-}, |\eta^{\mu-}|, \phi^{\mu-}) = (\mathbf{\Omega}_+, \mathbf{\Omega}_-)$ spectra of $B_s \to \mu^+\mu^-$ or $B^+ \to J/\psi K^+$ MC events satisfying the acceptance criteria of Section 6.1. This is done separately for each of the relevant run ranges. The final di-muon efficiency is then the luminosity weighted average of these:

$$\epsilon_{\mu^+\mu^-}^{L1} = \sum_i w_i \int \epsilon_{\mu+}^{muon}(run_i, \mathbf{\Omega}_+) d\mathbf{\Omega}_+ \int \epsilon_{\mu-}^{muon}(run_i, \mathbf{\Omega}_-) d\mathbf{\Omega}_- \tag{7}$$

where $i \in$ [all good runs], $w_i = \frac{\mathcal{L}_i}{\sum_j \mathcal{L}_j}$, $\mathcal{L}_i$ is the integrated luminosity for the ith run, and $\epsilon_{\mu^\pm}^{L1}(run_i, \mathbf{\Omega}_\pm)$ is the parameterization of the relevant muon single-leg efficiency for the ith run. For the CMU-CMU triggers, the parameterizations are the same for each muon and are thus correlated, while for the CMU-CMX triggers they are different and thus taken as uncorrelated. The resulting ratio of efficiencies is $\epsilon_{\mu^+\mu^-}^{L1}(B^+)/\epsilon_{\mu^+\mu^-}^{L1}(B_s) = 0.9954 \pm 0.0003$ (0.9889 ± 0.0003) for the CMU-CMU (CMU-CMX) channel including statistical and systematic uncertainties, which are assumed 100% correlated in the ratio.

Once candidate single muons are identified by the trigger they are combined into opposite sign pairs and recorded for further analysis. This efficiency of this level of the trigger depends only on run range. We use luminosity and acceptance weighting to estimate the final trigger efficiencies and find them to be the same for the $B_s \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decays: $\epsilon_{CMU-CMU}^{L2} = 0.9997^{+0.0003}_{-0.0016}$, $\epsilon_{CMU-CMX}^{L2} = 0.9986^{+0.0014}_{-0.0017}$, $\epsilon_{CMU-CMU}^{L3} = 0.989 \pm 0.022$, $\epsilon_{CMU-CMX}^{L3} = 0.980 \pm 0.015$, including statistical and systematic uncertainties. The ratio of these efficiencies $\epsilon_{B^+}^{trig}/\epsilon_{B_s}^{trig}$ are equal to 1.000 with total uncertainties that are $<<$ 1% and thus negligible.

The total trigger efficiency is the product of the single muon, muon pair efficiencies. The final ratio of trigger efficiencies is then $\epsilon_{B^+}^{trig}/\epsilon_{B_s}^{trig} = 0.9954 \pm 0.0003$ and 0.9889 ± 0.0003 for the CMU-CMU and CMU-CMX channels, respectively, including statistical and systematic contributions.

6.3 Reconstruction Efficiencies

The reconstruction efficiency accounts for the efficiency of the baseline and vertex requirements of Section 2. We estimate it as the product of several factors:

$$\epsilon^{reco} = \epsilon^{COT} \epsilon^{muon} \epsilon^{silicon} \epsilon^{vtx} \tag{8}$$

For $B_s \to \mu^+\mu^-$ decays, ϵ^{COT} is the probability that the $\mu^+\mu^-$ satisfying the acceptance criteria of Section 6.1 are reconstructed in the drift chamber(COT) and survive the "drift chamber quality" requirements, ϵ^{muon} is the fraction of $\mu^+\mu^-$ satisfying the drift chamber criteria that survive the "muon quality" requirements, and $\epsilon^{silicon}$ is the fraction of $\mu^+\mu^-$ satisfying

the drift chamber and muon criteria and surviving the "silicon quality" requirements. For $B^+ \to J/\psi K^+$ decays, the $\mu^+\mu^-$ reconstruction efficiencies are defined in exactly the same manner as described for the $B_s \to \mu^+\mu^-$ decays. Additionally we define ϵ_K^{COT} as the fraction of kaons from $B^+ \to J/\psi K^+$ decays satisfying the acceptance criteria of Section 6.1 that are reconstructed in the drift chamber and survive the drift chamber quality requirements, and $\epsilon_K^{silicon}$ as the fraction of these that additionally satisfy the silicon quality requirements. In both cases ϵ^{vtx} is the fraction of fully reconstructed events which survive the relevant vertex requirements described in Sections 2.1 and 2.2. The resulting ratio of reconstruction efficiencies can then be expressed as:

$$\epsilon_{B^+}^{reco}/\epsilon_{B_s}^{reco} = \left(\frac{\epsilon_{B^+}^{COT}}{\epsilon_{B_s}^{COT}}\frac{\epsilon_{B^+}^{muon}}{\epsilon_{B_s}^{muon}}\frac{\epsilon_{B^+}^{silicon}}{\epsilon_{B_s}^{silicon}}\right)\left(\frac{\epsilon_{B^+}^{vtx}}{\epsilon_{B_s}^{vtx}}\right)\epsilon_K^{COT}\,\epsilon_K^{silicon} \tag{9}$$

$$= \left(\frac{\epsilon_{B^+}^{reco-\mu^+\mu^-}}{\epsilon_{B_s}^{reco-\mu^+\mu^-}}\right)\left(\frac{\epsilon_{B^+}^{vtx}}{\epsilon_{B_s}^{vtx}}\right)\epsilon_{B^+}^{reco-K} \tag{10}$$

The first term is the ratio of $B^+ \to J/\psi K^+$ to $B_s \to \mu^+\mu^-$ di-muon reconstruction efficiencies and is expected to be close to one. The second term is the ratio of $B^+ \to J/\psi K^+$ to $B_s \to \mu^+\mu^-$ vertex requirement efficiencies - this will be different from one due to lifetime differences and the additional vertex requirements for the $B^+ \to J/\psi K^+$ channel. The last term accounts for the reconstruction efficiency of the kaon and is unique to the $B^+ \to J/\psi K^+$ decay. We'll discuss each of these separately.

6.3.1 Di-Muon Reconstruction Efficiencies

The drift chamber and silicon efficiencies are averaged over $p_T > 2$ GeV and depend only on run number. Variations as a function of muon isolation, p_T, and di-muon opening-angle are assigned as systematic uncertainties. The resulting ratios are then $\epsilon_{B^+}^{COT}/\epsilon_{B_s}^{COT} = 1.00 \pm 0.01$ and $\epsilon_{B^+}^{silicon}/\epsilon_{B_s}^{silicon} = 1.00 \pm 0.03$ including the statistical and systematic uncertainties for both the CMU-CMU and CMU-CMX channels.

The muon reconstruction efficiencies are constants and will cancel in the ratio. The efficiency of the chi-squared track-stub matching requirements is p_T dependent. We convolute the di-muon p_T spectra for events within the acceptance that survive the trigger and that have muons which each satisfy the drift chamber quality criteria. The efficiencies are about 98% for the CMU and 99% for the CMX muons in both the $B^+ \to J/\psi K^+$ and $B_s \to \mu^+\mu^-$ decays. The resulting $B^+ \to J/\psi K^+$ to $B_s \to \mu^+\mu^-$ ratios are $\epsilon_{B^+}^{\mu\chi^2}/\epsilon_{B_s}^{\mu\chi^2} = 1.003$ and 1.002 for the CMU-CMU and CMU-CMX channels, respectively. The total associated uncertainties are $<< 1\%$ and thus negligible.

The ratio of total di-muon reconstruction efficiencies is the product of the drift chamber, $\mu\chi^2$, and silicon ratios reported above. We find for both channels that $\epsilon_{B^+}^{reco-\mu^+\mu^-}/\epsilon_{B_s}^{reco-\mu^+\mu^-} = 1.00 \pm 0.03$, including statistical and systematic uncertainties.

6.3.2 Vertex Efficiencies

We estimate the efficiency for the vertex requirements using the default MC samples. For simplicity, we also include the efficiency of the mass window requirements here too. We find $\epsilon_{B^+}^{vtx} = (75.7 \pm 1.0)\%$ and $\epsilon_{B_s}^{vtx} = (76.8 \pm 0.3)\%$ which yield a ratio of $\epsilon_{B^+}^{vtx}/\epsilon_{B_s}^{vtx} = 0.986 \pm 0.013$, for both the CMU-CMU and CMU-CMX channels and including the statistical uncertainties only. Since, by comparing the vertex efficiencies for the $J/\psi \to \mu^+\mu^-$ decays from $B^+ \to J/\psi K^+$ candidates between the MC and sideband subtracted data, we find that the MC accurately estimates these efficiencies, we assign no additional systematic uncertainty to this ratio.

6.3.3 Kaon Reconstruction Efficiencies

Using a data-MC hybrid technique where kaon tracks are embedded in data we estimate the reconstruction efficiency as a function of p_T for kaons satisfying our fiducial requirements and having $p_T(K) > 1$ GeV. We convolute the efficiency curve with the $p_T(K)$ spectrum of fiducial kaons from Monte Carlo $B^+ \to J/\psi K^+$ decays satisfying our trigger requirements to obtain $\epsilon_K^{COT} = (96.4 \pm 1.6)\%$. The uncertainty is the quadrature sum of contributions from statistics ($\pm 0.2\%$), and systematic variations due to isolation ($\pm 1.5\%$), instantaneous luminosity ($\pm 0.5\%$), and the kaon nuclear interaction cross-section ($\pm 0.3\%$). We conservatively assign a $\pm 25\%$ relative uncertainty on the kaon nuclear interaction cross-section [19].

We evaluate the kaon specific silicon reconstruction efficiency directly from the dataset used in the analysis. We remove the silicon quality requirements on the kaon leg and reconstruct a sample of $B^+ \to J/\psi K^+$ candidates. We then get the sideband subtracted efficiency for attaching > 2 silicon $r\phi$ hits to the kaon track. We find $\epsilon_K^{silicon} = 0.973 \pm 0.002$, for both the CMU-CMU and CMU-CMX channels, where only the statistical uncertainty is included. Since we evaluate this on exactly the same data set with exactly the same baseline and vertex requirements as used in the analysis, we assume the systematic uncertainties are negligible.

6.4 Efficiency of the Likelihood Requirement

The efficiency of the likelihood requirement, $\epsilon_{B_s}^{LH}$ is estimated using the odd numbered events in the $B_s \to \mu^+\mu^-$ Monte Carlo sample. The estimates are based on the sample described in Section 3 and are given in Table 6.

We cross-check the Monte Carlo efficiencies by comparing the likelihood efficiencies between $B^+ \to J/\psi K^+$ MC and sideband subtracted $B^+ \to J/\psi K^+$ distributions from the data. The ratio of data-to-MC efficiencies for $B^+ \to J/\psi K^+$ events is given in Table 7. The distributions of the three input variables are shown in Figures 13 and 14. Both channels are adequately modeled for our purposes. Based on these comparisons we assign systematic uncertainties of $\pm 10\%$ and $\pm 5\%$ (relative) to the CMU-CMU and CMU-CMX channels, respectively.

We compare the efficiency between the data and MC $B^+ \to J/\psi K^+$ samples for the $Iso > 0.50$ and $\Delta\alpha < 0.70$ rad requirements. We find that the ratio of these efficiencies between data and Monte Carlo are consistent with unity within the associated statistical uncertainty of $\pm 1.6\%$. No additional systematic uncertainty is assigned and we use the $B_s \to \mu^+\mu^-$ MC

cut	$\epsilon_{B_s}^{LH}$ CMU-CMU	CMU-CMX
$LH > 0.90$	$(70.0 \pm 0.8)\%$	$(66.3 \pm 1.0)\%$
$LH > 0.92$	$(66.5 \pm 0.8)\%$	$(64.6 \pm 1.0)\%$
$LH > 0.95$	$(61.0 \pm 0.8)\%$	$(60.1 \pm 1.0)\%$
$LH > 0.98$	$(48.1 \pm 0.9)\%$	$(48.4 \pm 1.0)\%$
$LH > 0.99$	$(37.8 \pm 0.9)\%$	$(39.1 \pm 1.0)\%$

Table 6: The $B_s \to \mu^+\mu^-$ efficiency for the likelihood requirement for the CMU-CMU (left) and CMU-CMX (right) channels. The uncertainties are the binomial statistical uncertainties.

cut	$\epsilon_{B^+}^{LH}$(data)/$\epsilon_{B^+}^{LH}$(MC) CMU-CMU	CMU-CMX
$LH > 0.70$	0.981 ± 0.026	1.015 ± 0.043
$LH > 0.80$	0.958 ± 0.027	1.028 ± 0.047
$LH > 0.90$	0.934 ± 0.030	0.989 ± 0.051
$LH > 0.94$	0.916 ± 0.032	0.966 ± 0.053
$LH > 0.98$	0.874 ± 0.038	0.947 ± 0.065

Table 7: The ratio of the $B^+ \to J/\psi K^+$ efficiency for the likelihood requirement determined from sideband subtracted data to the same efficiency determined from the Monte Carlo, including the statistical uncertainties.

sample to estimate the efficiency of these requirements to be 92%. In equation 6 we include this efficiency in the $\epsilon_{B_s}^{LH}$ term.

7 Normalization

Using the baseline and vertex requirements discussed in Section 2 we estimate the number of $B^+ \to J/\psi K^+$ candidates, $N_{B^+\to J/\psi K^+}$, using simple sideband subtraction and correcting for the small contribution of $B^+ \to J/\psi\pi^+$ decays. The signal mass window is defined as $5.240 < M_{vtx} < 5.310$ GeV while the sidebands are symmetrically defined to include an additional 120 MeV on either side of the signal region. We correct for the number of $B^+ \to J/\psi\pi^+$ events expected to fall within this mass window using this expression:

$$N_{B^+\to J/\psi K^+} = \frac{N_{K+\pi}}{1 + \left(\frac{BR(B^+\to J/\psi\pi^+)}{BR(B^+\to J/\psi K^+)} \cdot \frac{\alpha_\pi}{\alpha_K} \cdot \frac{\epsilon_\pi^{reco}}{\epsilon_K^{reco}} \cdot \frac{\epsilon_\pi^{mass}}{\epsilon_K^{mass}}\right)} \tag{11}$$

$$= \frac{N_{K+\pi}}{1 + (0.0014 \pm 0.0004)} \tag{12}$$

where $N_{K+\pi}$ is the number of ($B^+ \to J/\psi K^+ + B^+ \to J/\psi\pi^+$) decays in the $B^+ \to J/\psi K^+$ mass window determined using the sideband subtraction described above, α_π is the acceptance for $B^+ \to J/\psi\pi^+$ decays, ϵ_π^{reco} is the total reconstruction efficiency, including the trigger, COT, silicon, muon, and vertex requirements for $B^+ \to J/\psi\pi^+$ decays, and ϵ_π^{mass} is the efficiency of the $B^+ \to J/\psi K^+$ mass window requirements on the $B^+ \to J/\psi\pi^+$ sample and is equal to 0.035 ± 0.010[5]. The α_K, ϵ_K^{reco} and ϵ_K^{mass} are analogously defined for the $B^+ \to J/\psi K^+$ decays and are given in Section 6 above. The $B^+ \to J/\psi\pi^+$ terms are determined in the same manner as the $B^+ \to J/\psi K^+$ terms. The branching ratio for $B^+ \to J/\psi\pi^+$ is $BR(B^+ \to J/\psi\pi^+) = (4.0 \pm 0.5) \times 10^{-5}$ [12].

Using this expression we estimate the $B^+ \to J/\psi\pi^+$ corrected number of $B^+ \to J/\psi K^+$ events as $N^{CMU-CMU}_{B^+\to J/\psi K^+} = 1785 \pm 60$ and $N^{CMU-CMX}_{B^+\to J/\psi K^+} = 696 \pm 39$, where the uncertainties are completely dominated by the statistics of the sideband subtraction. These are used in equation 6 to estimate the BR($B_s \to \mu^+\mu^-$).

8 Optimization

For the optimization we choose as our figure-of-merit the *expected* 90% CL upper limit on the branching ratio, BR($B_s \to \mu^+\mu^-$). This is a natural choice since it's statistically rigorous and optimizes the physics result itself. We can also incorporate the effects of uncertainties into the optimization choice. We use the same methodology as described in detail in reference [4], except that we use a Bayesian integration to calculate the upper limit on the number of signal events for a given number of observed events. We assume a flat prior for the BR($B_s \to \mu^+\mu^-$) and truncated Gaussian priors for the signal efficiencies and backgrounds [20].

For the optimization we assume an integrated luminosity of 1 fb^{-1} (so that the analysis can be updated with 2005 data without re-optimizing) and vary $p_T(B)$ from 4 GeV to 6 GeV and the likelihood requirement from $0.90 - 0.99$. The CMU-CMU and CMU-CMX channels are separately optimized. For the different $p_T(B)$ requirements we re-evaluate the $B^+ \to J/\psi K^+$ yields, the acceptances, any efficiency for which there exists a p_T^μ dependence, and the likelihood curves for signal and background by remaking the probability densities, $\mathbf{P}_{s(b)}(x_i)$, used in the likelihood calculation. The background is separately estimated for each likelihood cut at each $p_T(B)$ threshold using the method of Section 5. Although we only report the optimization results using the likelihood constructed as described in Section 4.1, we include in the optimization likelihoods constructed from a variety of other variables as well, none of which outperformed the one reported here.

Because the likelihood is so effective at suppressing the background, the optimization moves the $p_T(B)$ threshold to 4 GeV and chooses $LH > 0.99$ for both the CMU-CMU and CMU-CMX channels. Moving to higher $p_T(B)$ requirements makes the expected BR limit $\sim 5\%$ worse in both channels (for the optimal LH requirement at that $p_T(B)$). Fixing the

[5]Note, the $B^+ \to J/\psi\pi^+$ mass will be shifted to higher values because the pion is assumed to have the kaon mass.

$p_T(B)$ requirement and loosening the LH cut makes the expected BR limit $\sim 10\%$ worse for both the CMU-CMU and CMU-CMX channels.

For the $364\,\mathrm{pb}^{-1}$ of CMU-CMU data these optimal requirements correspond to an expected background of $N_{bgd}^{CMU-CMU} = 0.81 \pm 0.12$ and a single-event-sensitivity of $ses^{CMU-CMU} = (1.0 \pm 0.2) \times 10^{-7}$ and yield an expected limit of $\mathrm{BR}(B_s \to \mu^+\mu^-) < 3.5 \times 10^{-7}$ at 90% CL.

For the $336\,\mathrm{pb}^{-1}$ of CMU-CMX data these optimal requirements correspond to an expected background of $N_{bgd}^{CMU-CMX} = 0.66 \pm 0.13$ and a single-event-sensitivity of $ses^{CMU-CMX} = (1.6 \pm 0.3) \times 10^{-7}$ and yield an expected limit of $\mathrm{BR}(B_s \to \mu^+\mu^-) < 5.6 \times 10^{-7}$ at 90% CL.

We can combine the CMU-CMU and CMU-CMX channels, taking into account their correlated uncertainties[21] from the f_u/f_s and $B^+ \to J/\psi K^+$ related branching ratios, to get an expected limit of $\mathrm{BR}(B_s \to \mu^+\mu^-) < 2.0 \times 10^{-7}$ at 90% CL. This is significantly better than any published result. The expected combined limit for 1 fb^{-1} of data is $\mathrm{BR}(B_s \to \mu^+\mu^-) < 1.1 \times 10^{-7}$ using the same selection criteria and assuming the single-event-sensitivity and background scale only with the luminosity.

8.1 Expectations for $B_d \to \mu^+\mu^-$

The expected limit on the $\mathrm{BR}(B_d \to \mu^+\mu^-)$ is also calculated. We estimate the acceptance, trigger, reconstruction and "final" efficiencies in the same manner as for the $B_s \to \mu^+\mu^-$ decays using a Monte Carlo sample of $B_d \to \mu^+\mu^-$ generated in the same manner as the $B_s \to \mu^+\mu^-$ sample described in Section 3. We find the ratio of $\alpha \cdot \epsilon_{total}(B_s)/\alpha \cdot \epsilon_{total}(B_d)$ is consistent with unity within the statistical uncertainties. Using the requirements optimized for the $B_s \to \mu^+\mu^-$ search yields an expected $\mathrm{BR}(B_d \to \mu^+\mu^-)$ limit of 4.9×10^{-8} at the 90% CL for the CMU-CMU and CMU-CMX channels combined using $364\,\mathrm{pb}^{-1}$ of data and including the correlations from the $B^+ \to J/\psi K^+$ related branching ratios. This is significantly better than the present best limit from the B-factories, $\mathrm{BR}(B_d \to \mu^+\mu^-) < 8.3 \times 10^{-8}$ from BaBar using 111 fb^{-1} [22].

9 Results

Figure 15 shows a two dimensional plot of $M_{\mu^+\mu^-}$ and LH for all events satisfying the baseline and vertex criteria and with $LH > 0.80$. The invariant mass distributions for events satisfying the baseline and vertex criteria of Section 2.1 and with $LH > 0.99$ are shown in Figure 16. There are no events, in either channel, which fall within the ± 60 MeV mass windows centered on the world average $B_{s(d)}$ mass, 5.369 (5.279) GeV. Taking into account the correlated systematic uncertainties from the $BR(B^+ \to J/\psi K^+) \cdot BR(J/\psi \to \mu^+\mu^-)$ and f_u/f_s we calculate a combined limit of $\mathrm{BR}(B_s \to \mu^+\mu^-) < 1.5 \times 10^{-7}$ (2.0×10^{-7}) at 90% (95%) confidence level. The resulting limit for the $B_d \to \mu^+\mu^-$ decay is $\mathrm{BR}(B_d \to \mu^+\mu^-) < 3.8 \times 10^{-8}$ (4.9×10^{-8}) at 90% (95%) confidence level.

The full likelihood distribution for events satisfying the baseline and vertex criteria and falling within the B_s search window, $5.309 < M_{\mu^+\mu^-} < 5.429$ GeV, is shown in Figure 17.

A comparison of the likelihood distribution for these signal events to the same distributions using events in the sideband region yields a KS probability of 66% and 76% for the CMU-CMU and CMU-CMX channels respectively.

As a cross-check, we compare the number of observed events to that predicted for looser LH requirements. In the CMU-CMU channel, for $LH > 0.50$ we expect 146 ± 22 and observe 136 events; while for $LH > 0.90$ we expect 24 ± 4 and observe 20 events. In the CMU-CMX channel, for $LH > 0.50$ we expect 99 ± 20 and observe 99 events; while for $LH > 0.90$ we expect 17 ± 3 and observe 9 events. The agreement for all of these is reasonable[6].

10 Conclusion

We report on an analysis which significantly improves our sensitivity to $B_s \to \mu^+\mu^-$ decays. Using approximately 360 pb^{-1} of data, we observe no candidate events in either the CMU-CMU or CMU-CMX channels while expecting 0.81 ± 0.12 and 0.66 ± 0.13 background events, respectively. Taking into account correlated systematic uncertainties we use a Bayesian integration to calculate a combined limit of $\mathrm{BR}(B_s \to \mu^+\mu^-) < 1.5 \times 10^{-7}$ (2.0×10^{-7}) at 90% (95%) confidence level. This is a significant improvement over the best published limit using 240 pb^{-1} of D0 data, 5.0×10^{-7} at 95% CL.

The same analysis is also sensitive to $B_d \to \mu^+\mu^-$ decays. We calculate a combined limit of $\mathrm{BR}(B_d \to \mu^+\mu^-) < 3.8 \times 10^{-8}$ (4.9×10^{-8}) at 90% (95%) confidence level. This is a significant improvement over the best published limit using 111 fb^{-1} of BaBar data, 8.3×10^{-8} at 90% CL.

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

[6]The Poisson probability of observing ≤ 9 while expecting 17 is approximately 2.6%, which corresponds to a fluctuation of < 2 standard deviations for a Gaussian distribution.

References

[1] G.Buchalla and A.J.Buras, Nucl. Phys. B **412** (1994) 106.

[2] M.Misiak and J.Urban, Phys. Lett. B **451** (1999) 161.

[3] D0 Collaboration, Phys. Rev. Lett. **94** (2005) 071802.
NB This result changes to $\mathrm{BR}(B_s \to \mu^+\mu^-) < 4.2 \times 10^{-7}$ at 90% CL for the updated f_u/f_s ratio used in this note and given in reference 11.

[4] CDF Collaboration, Phys. Rev. Lett. **93** (2004) 032001.

[5] A.Dedes, H.K.Dreiner, U.Nierste, and P.Richardson, `hep-ph/0207026` (2002).

[6] H.Baer, *et al.*, JHEP **0207** (2002) 050.

[7] R.Arnowitt, B.Dutta, T.Kamon, and M.Tanaka, Phys. Lett. B **538** (2002) 121.

[8] A.Dedes, H.K.Dreiner, and U.Nierste, Phys. Rev. Lett. **87** (2001) 251804.

[9] D.Auto, *et al.*, JHEP **0306** (2003) 023.

[10] R.Dermisek, S.Raby, L.Roszkowski, and R. Ruis de Austri, JHEP **0304** (2003) 037.

[11] Heavy Flavor Averaging Group, *Averages of b-hadron Properties as of Summer 2004*, `hep-ex/0412073`.

[12] S.Eidelman, *et al.*, Phys. Lett. **B 592**, (2004) 1.

[13] PYTHIA 6.2 Physics and Manual (Torbjorn Sjostrand, Leif Lonnblad, Stephen Mrenna) `hep-ph/0108264` (2001).

[14] K. Lannon, Ph.D. thesis, University of Illinois at Urbana-Champaign

[15] D. Langle, Nucl. Meth. in Phys. Res. **A 462**, (2001) 152-155

[16] S. Agostinelli, *et al.*, Nucl. Meth. in Phys. Res. **A 506**, (2003) 250-303

[17] D. Acosta, *et al.*, Phys. Rev. **D 71** 032001, (2005)

[18] B.L. van der Waerden, *Mathematical Statistics*, (1969) pp 320-323.

[19] D. Acosta, *et al.*, Phys. Rev. Lett. **94** 122001, (2005)

[20] We use the `bayes.f` algorithm provided by the statistics committee. We also used the software available as `bayesianlimit.tar`, which assumes gamma functions as priors for the signal efficiency and background, and found that it produces the same limits within 1%.

[21] We use the `bcorr.f` algorithm provided by the statistics committee, modified to accept inputs in units of 10^{-6} of $\mathrm{BR}(B_s \to \mu^+\mu^-)$. This modified version reproduces the single-channel limits of `bayes.f` and the D0 limit to within 1%.

[22] BaBar Collaboration, `hep-ex/0408096` (2004), submitted to Phys. Rev. Lett.

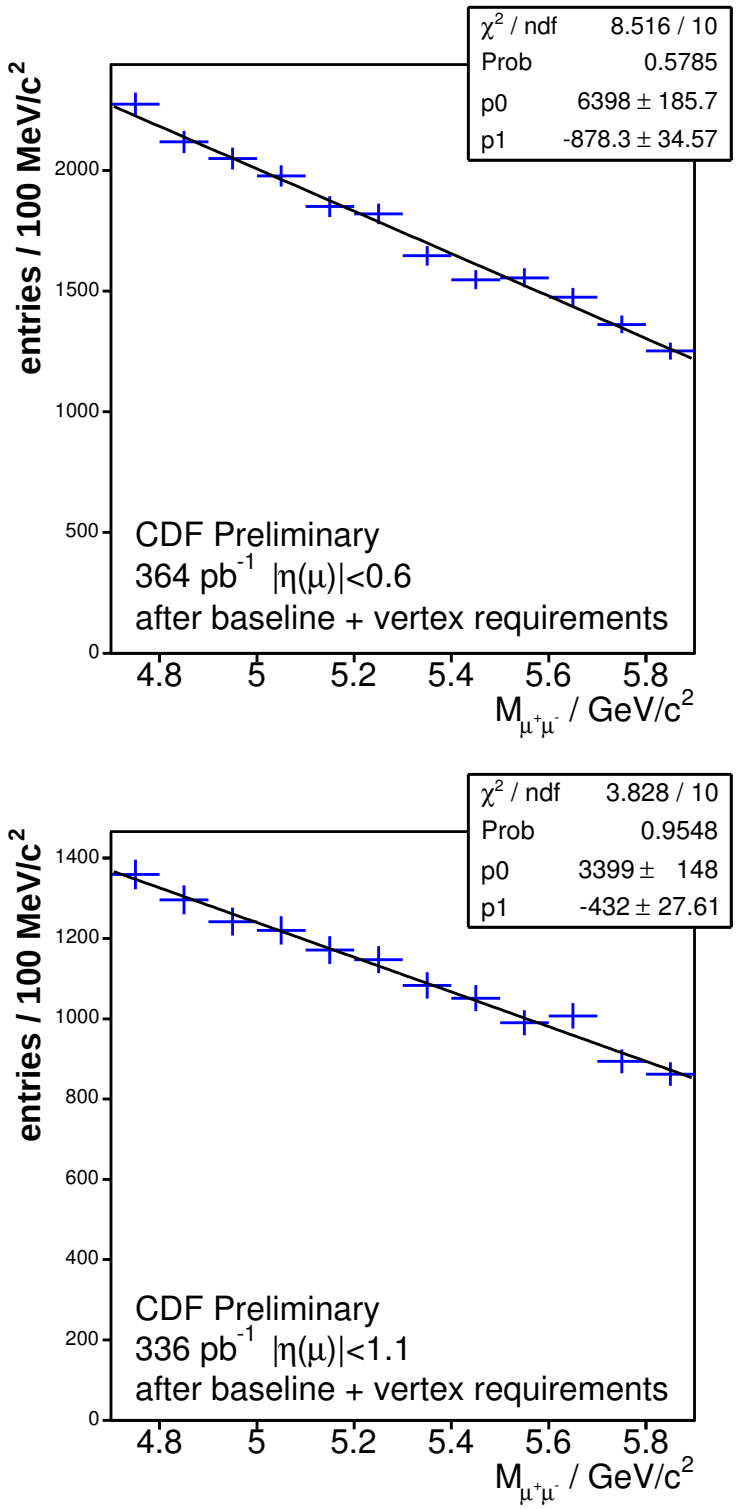

Figure 1: Di-muon invariant mass distribution for events satisfying the baseline and vertex requirements described in Section 2.1 for the $B_s \to \mu^+\mu^-$ search sample. The results of a fit to a straight line are shown.

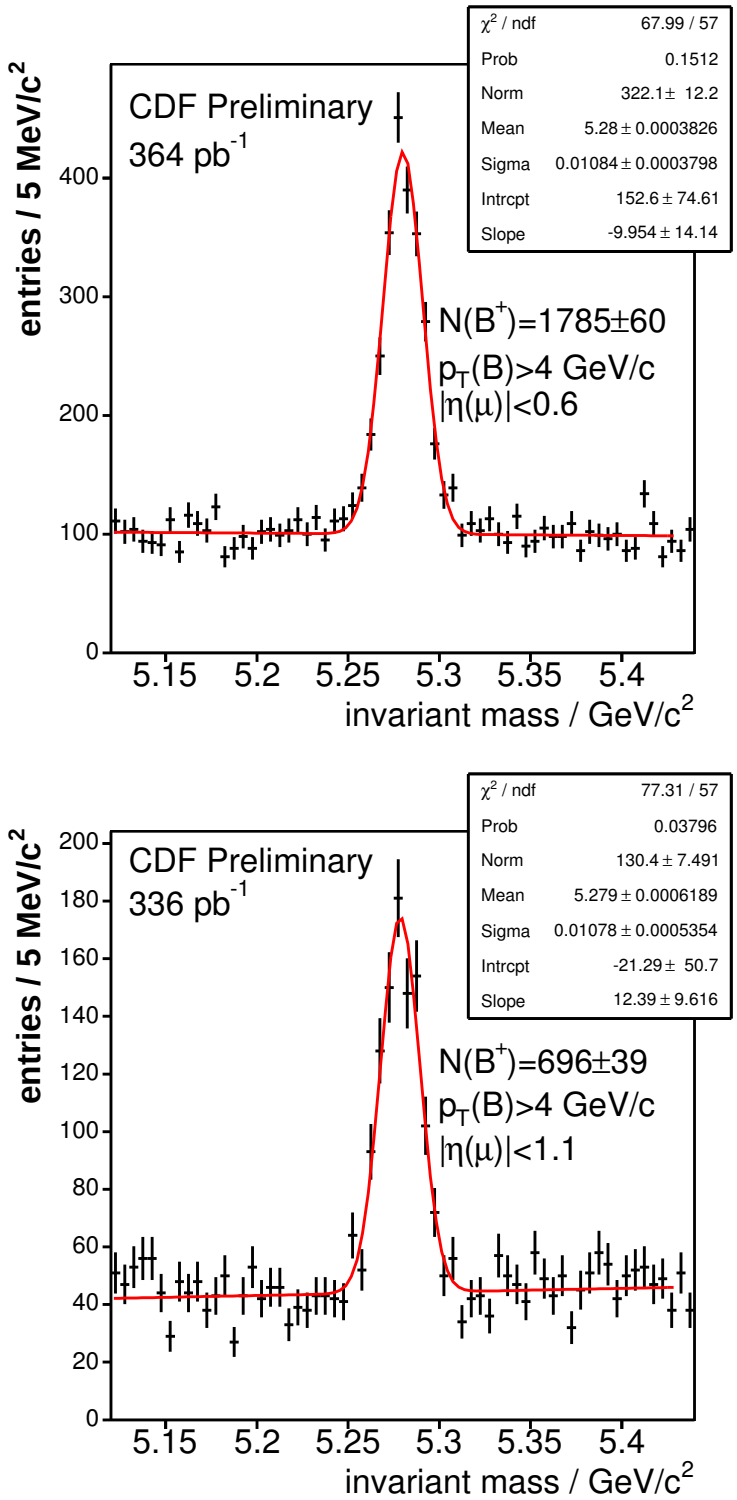

Figure 2: The $\mu^+\mu^-K^+$ invariant mass distribution for events satisfying the baseline and vertex requirements described in Section 2.2 for the $B^+ \to J/\psi K^+$ sample. The fits and estimated number of B^+ candidates are discussed in the text.

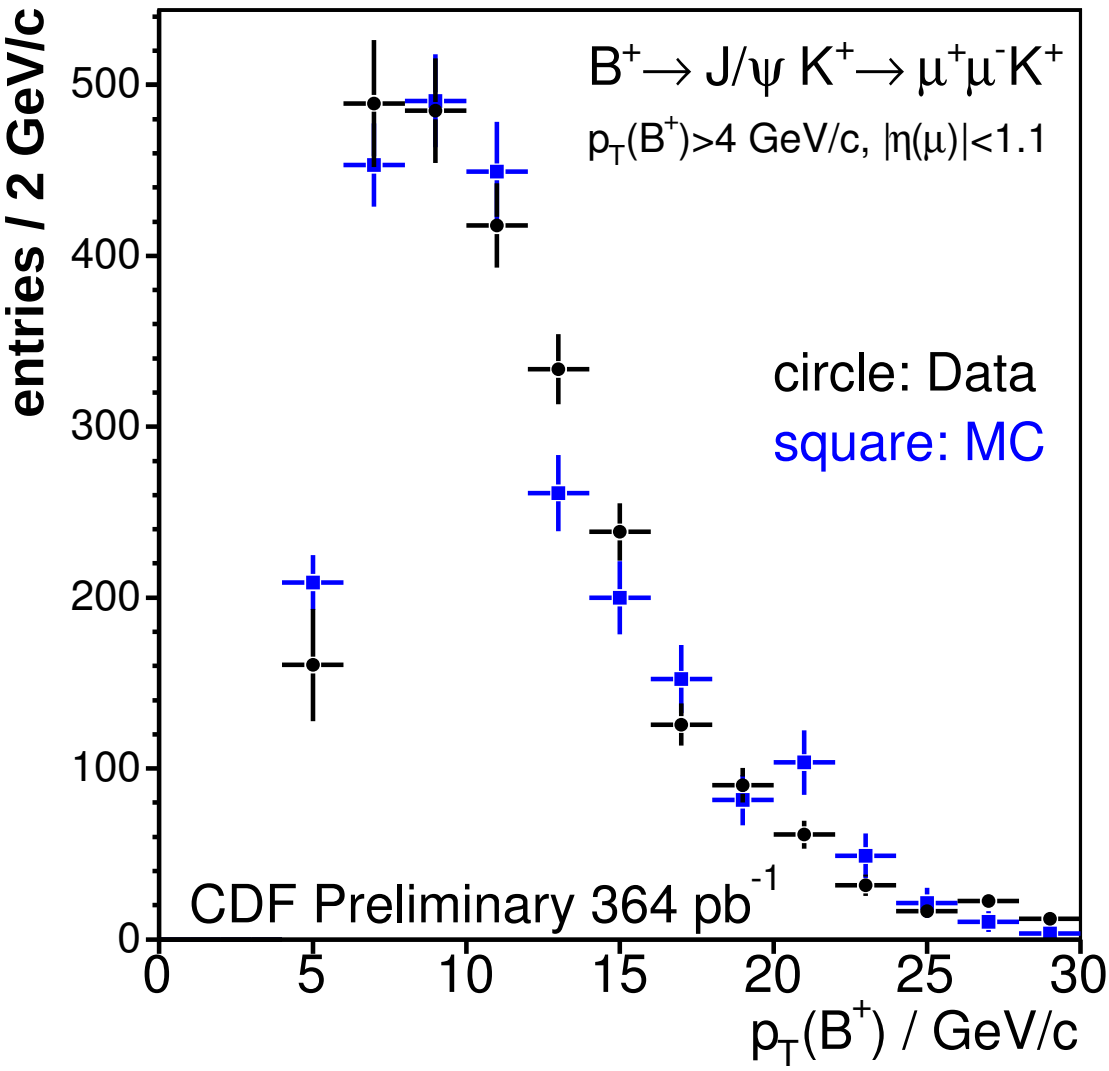

Figure 3: The $p_T(B)$ spectrum from $B^+ \to J/\psi K^+$ events satisfying the baseline and vertex requirements described in Section 2.

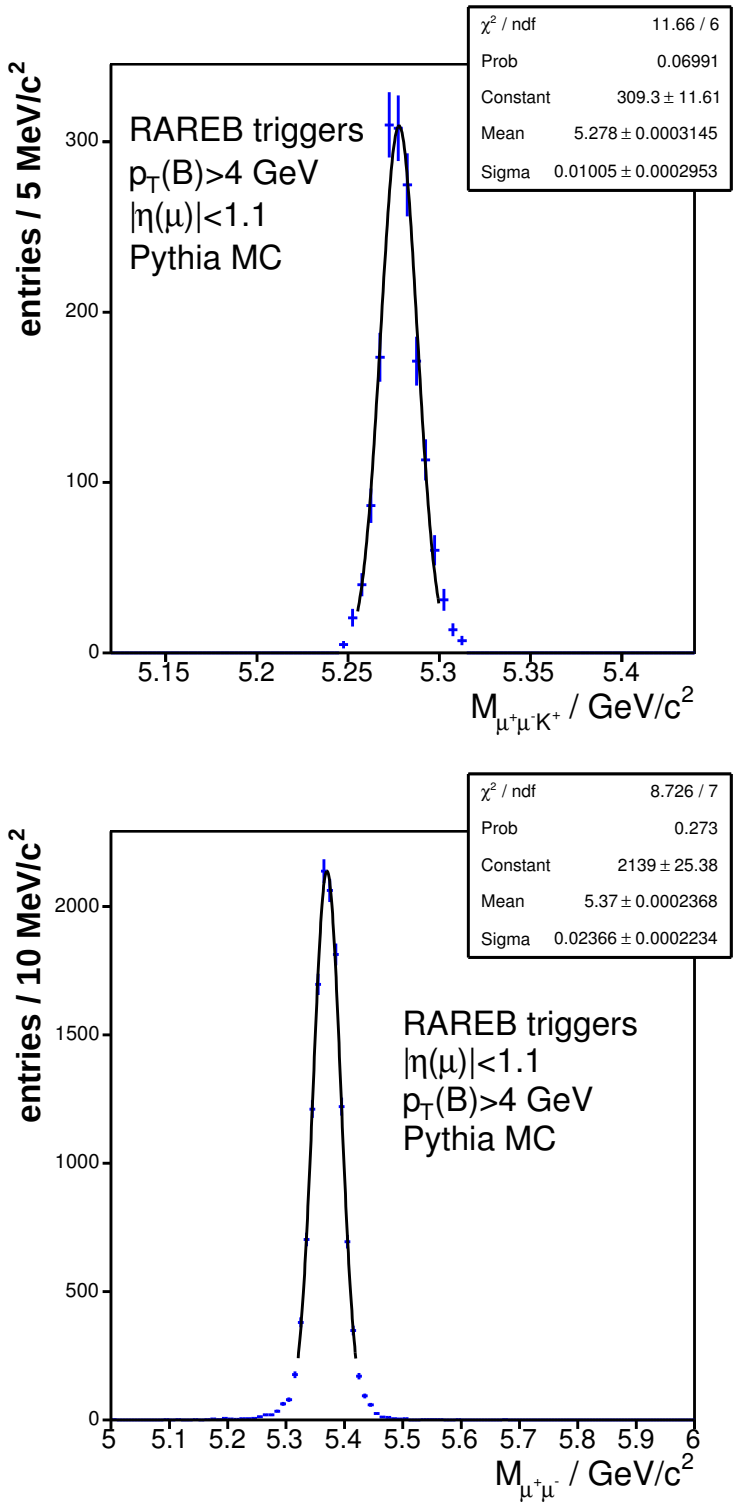

Figure 4: The invariant mass distributions for $B^+ \to J/\psi K^+$ (top) and $B_s \to \mu^+\mu^-$ (bottom) Monte Carlo events satisfying the baseline and vertex requirements described in Section 2. The results of a fit to a Gaussian are also shown.

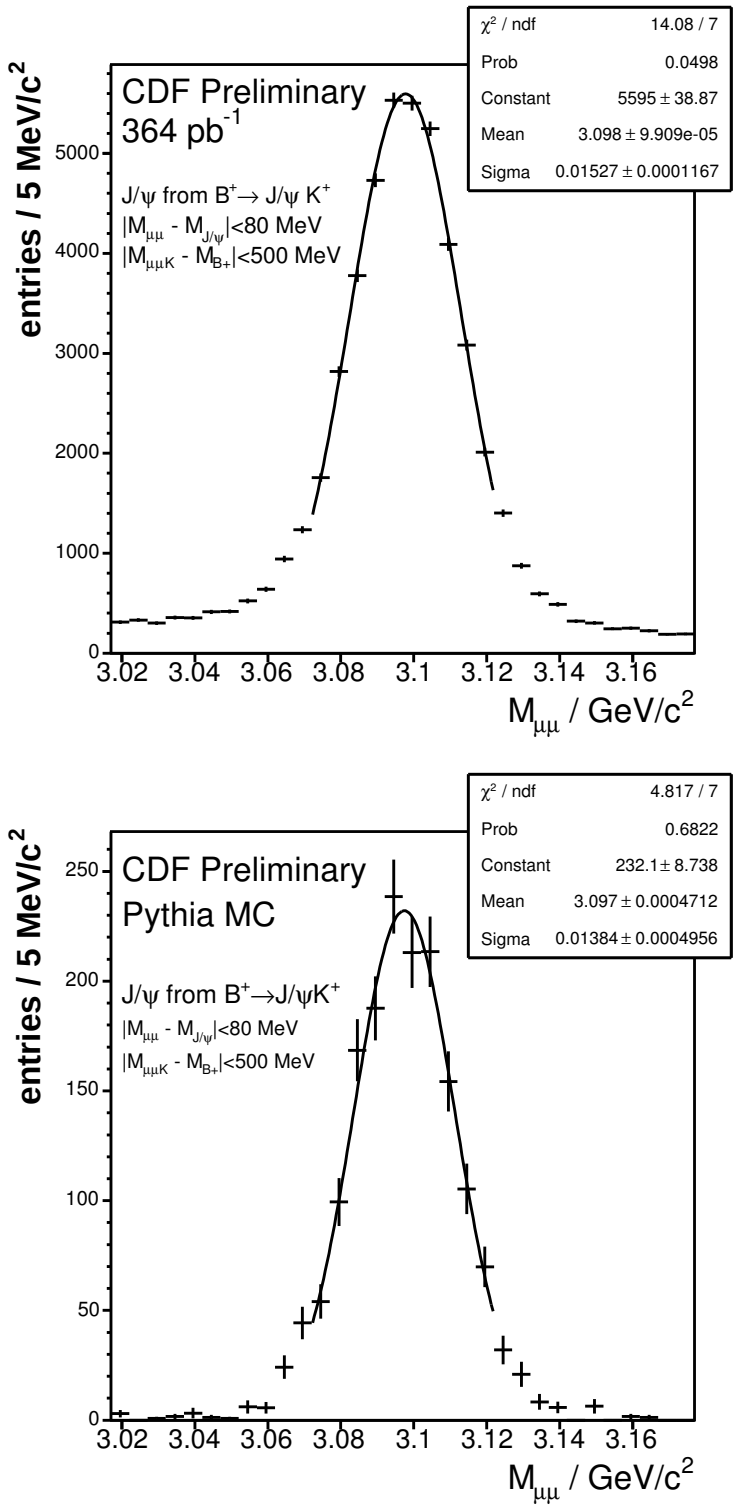

Figure 5: The invariant mass distribution for $J/\psi \to \mu^+\mu^-$ from $B^+ \to J/\psi K^+$ candidates in the data (top) and the Pythia MC (bottom). Each is fit to a Gaussian distribution.

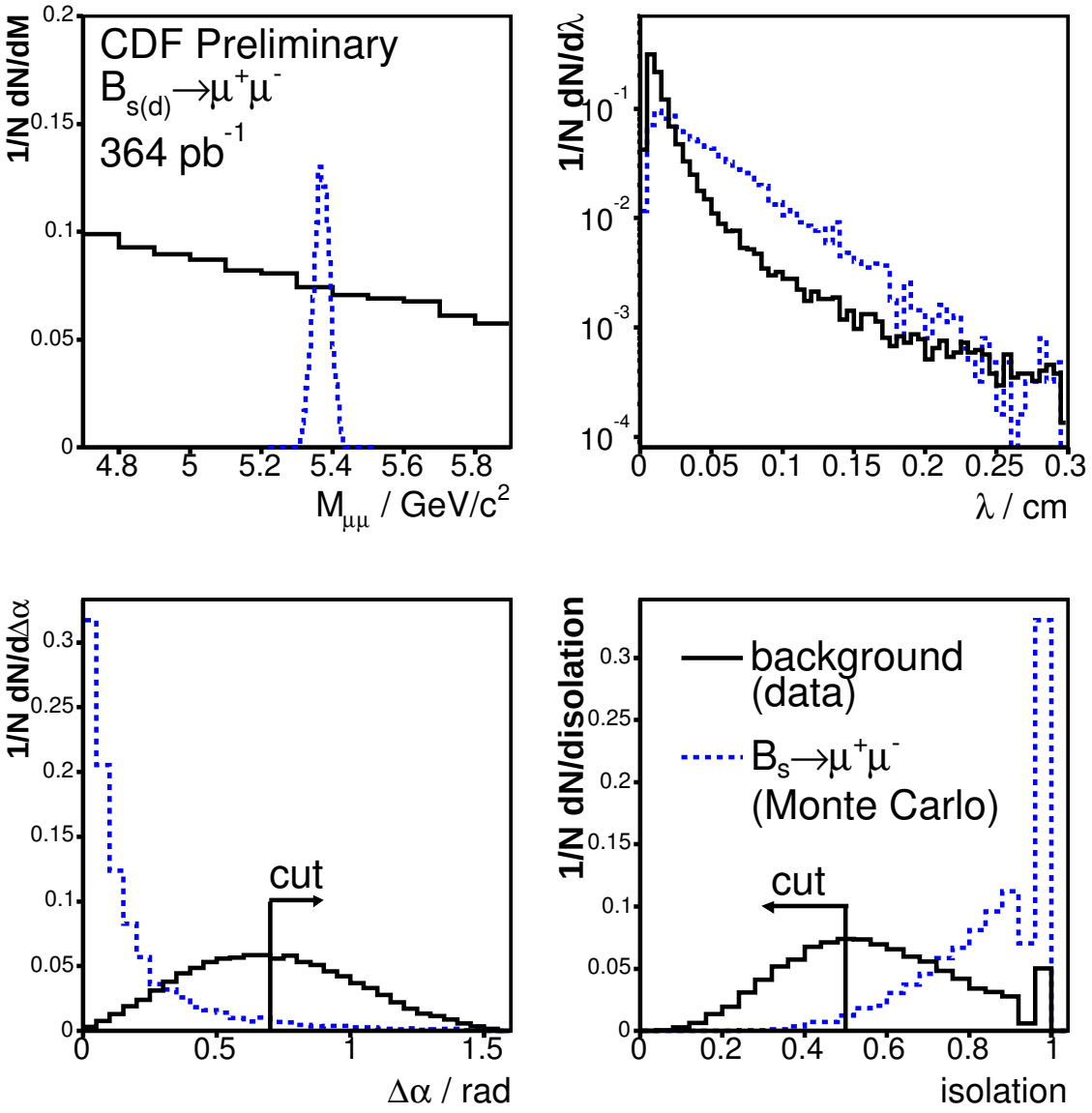

Figure 6: Distributions of various discriminating variables for Monte Carlo signal events (dashed histograms) and a background-dominated data sample (solid histograms). Only events which survive the baseline and vertex cuts and have $\lambda > 0$ are included. The histograms are all normalized to unit area. For the mass plot, different binning is used for the data and MC.

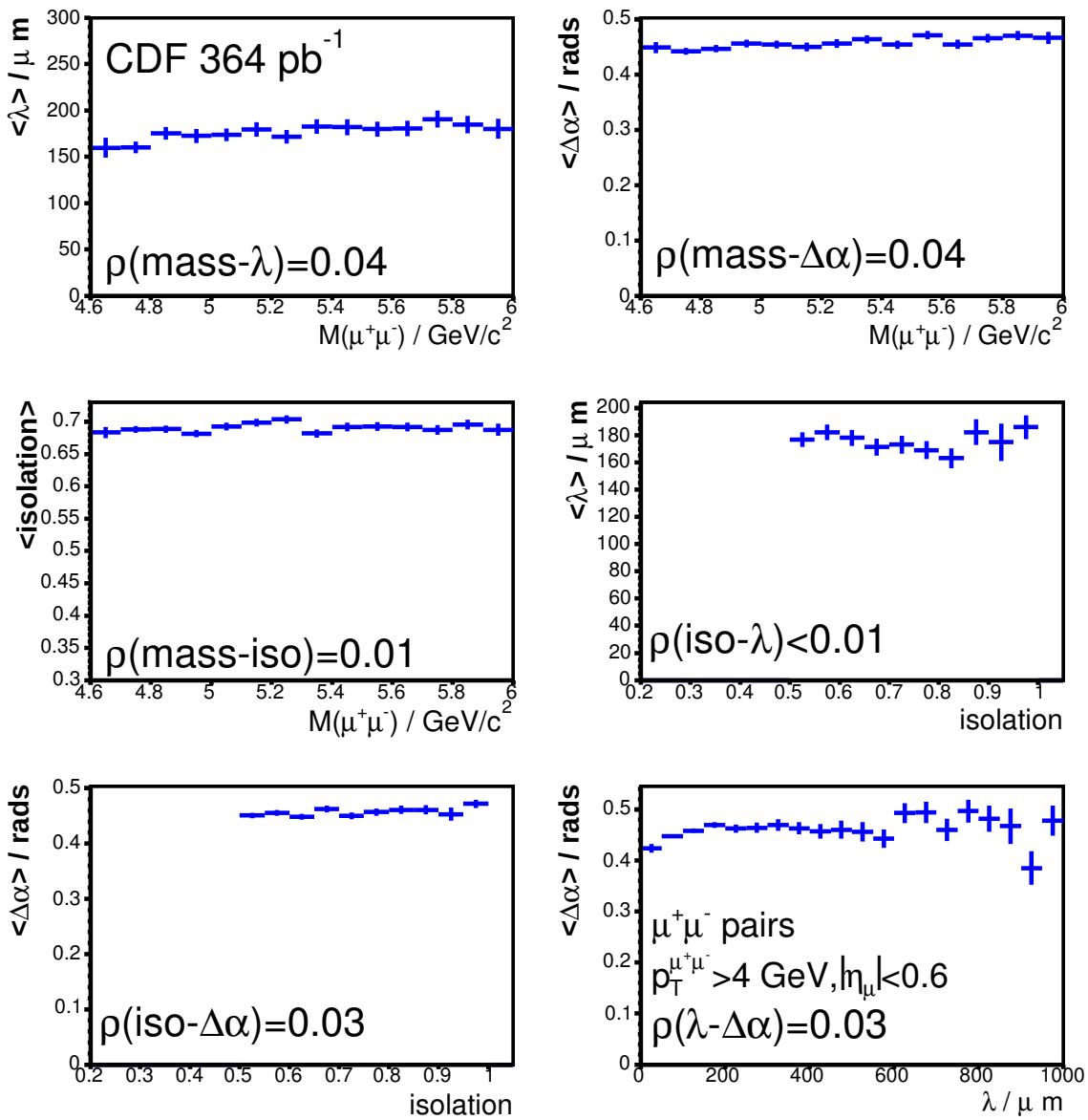

Figure 7: Profile plots showing the correlations among the four discriminating variables discussed in Section 4 for CMU-CMU $\mu^+\mu^-$ pairs satisfying the baseline and vertex requirements and having $\lambda > 0$. The y error bars are calculated as the uncertainty on the mean y value, $< y >$, in each bin of x. The linear correlation coefficients, ρ_{xy}, calculated as described in Section 4, are given for each pair of variables and have a statistical uncertainty of 0.01 each.

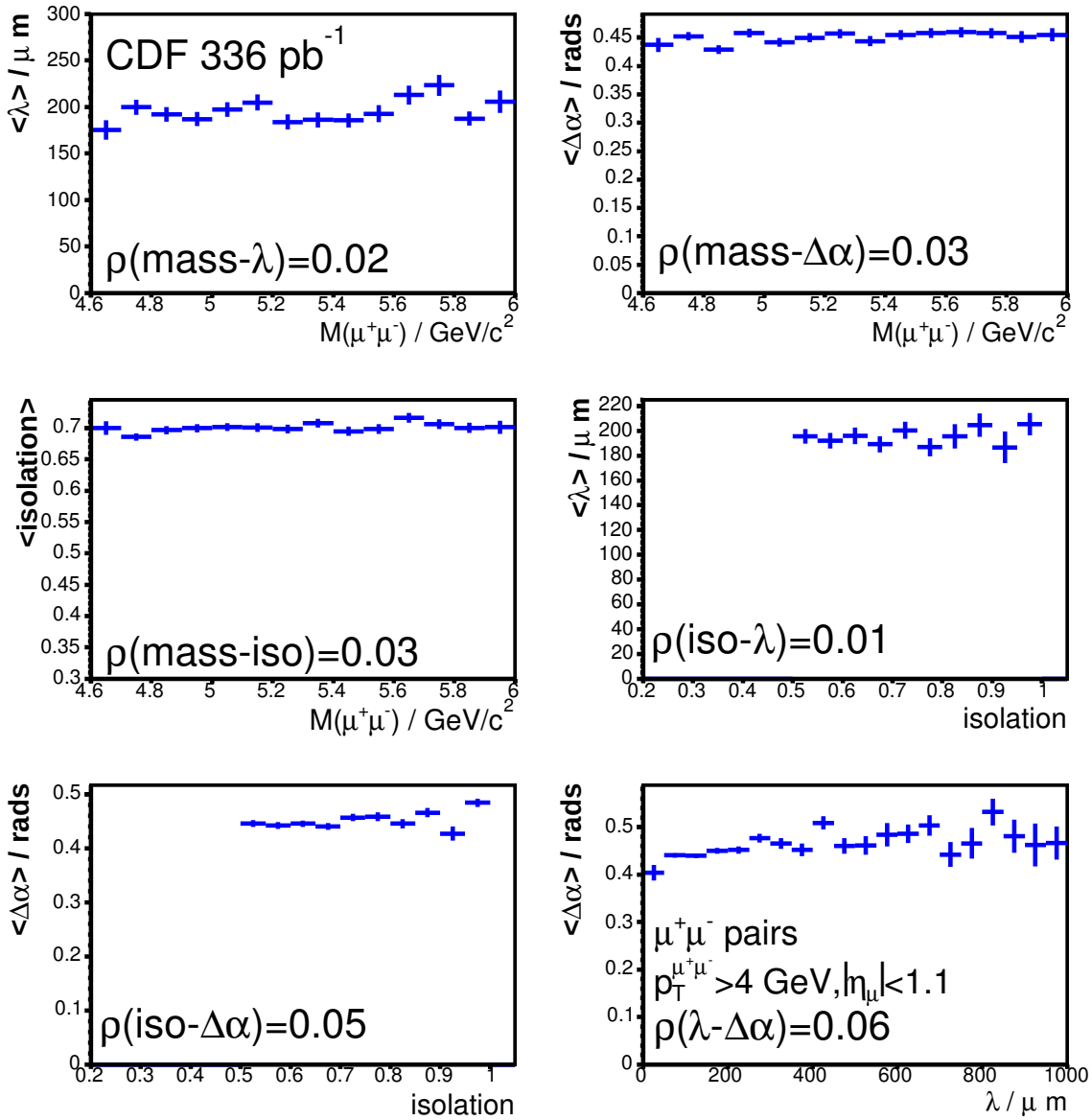

Figure 8: Profile plots showing the correlations among the four discriminating variables discussed in Section 4 for CMU-CMX $\mu^+\mu^-$ pairs satisfying the baseline and vertex requirements and having $\lambda > 0$. The y error bars are calculated as the uncertainty on the mean y value, $< y >$, in each bin of x. The linear correlation coefficients, ρ_{xy}, calculated as described in Section 4, are given for each pair of variables and have a statistical uncertainty of 0.01 each.

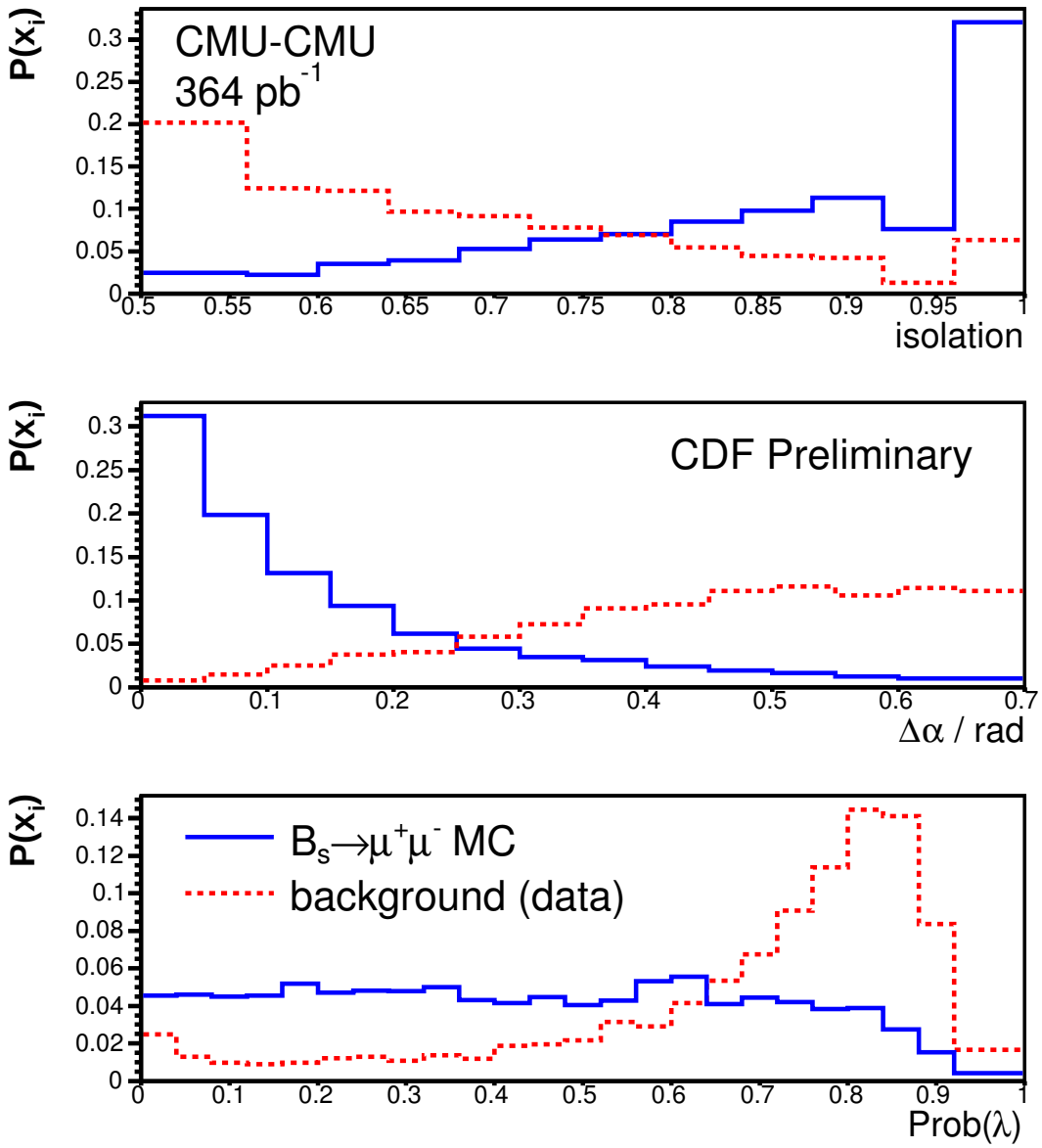

Figure 9: The Iso (top), $\Delta\alpha$ (mid), and $\mathcal{P}(\lambda)$ (bot) probability distributions used to construct the LH variable for the CMU-CMU events. Variable binning is used to properly account for regions with low statistics. Underflows and overflows are also properly accounted for.

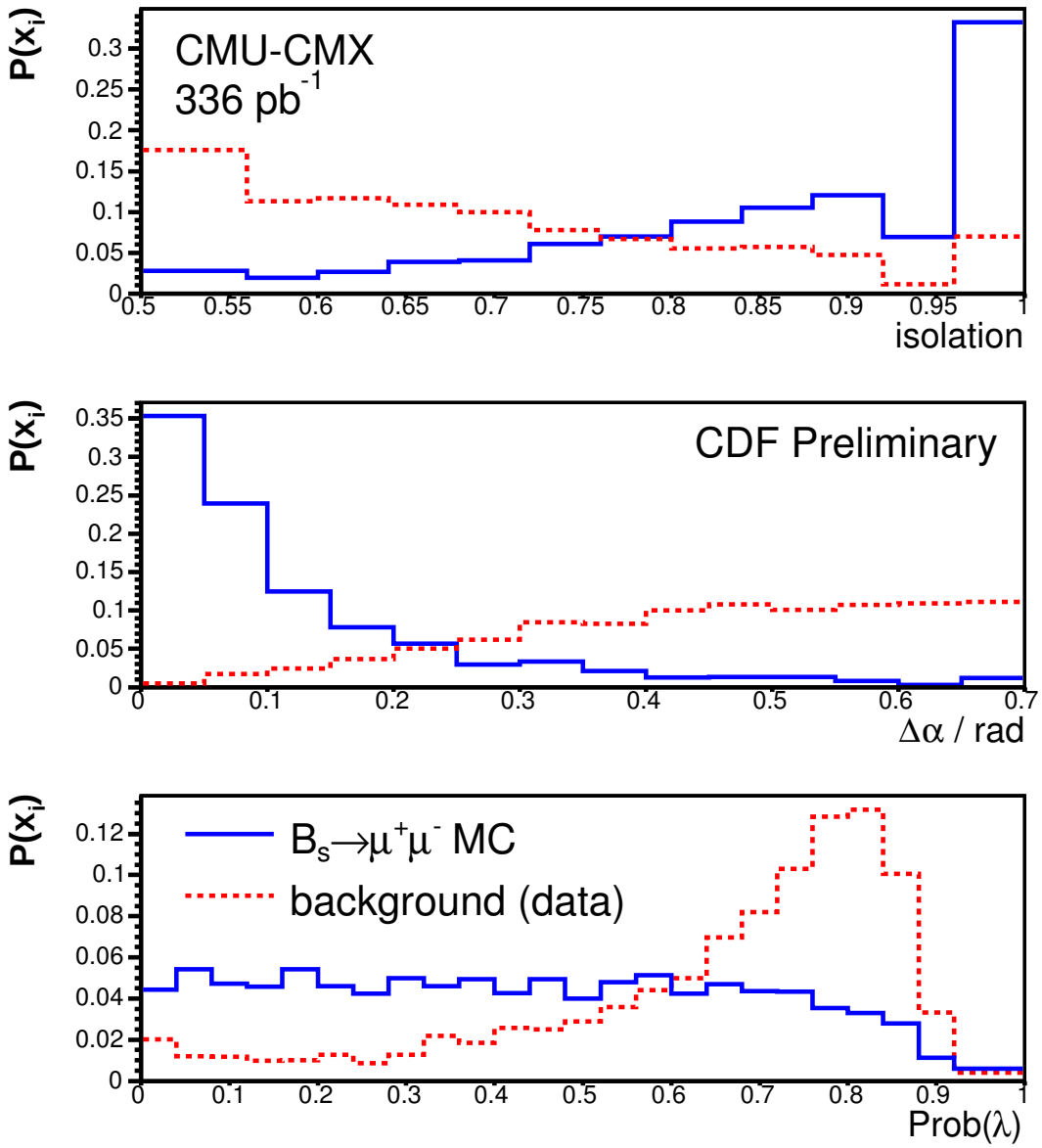

Figure 10: The Iso (top), $\Delta\alpha$ (mid), and $\mathcal{P}(\lambda)$ (bot) probability distributions used to construct the LH variable for the CMU-CMX events. Variable binning is used to properly account for regions with low statistics. Underflows and overflows are also properly accounted for.

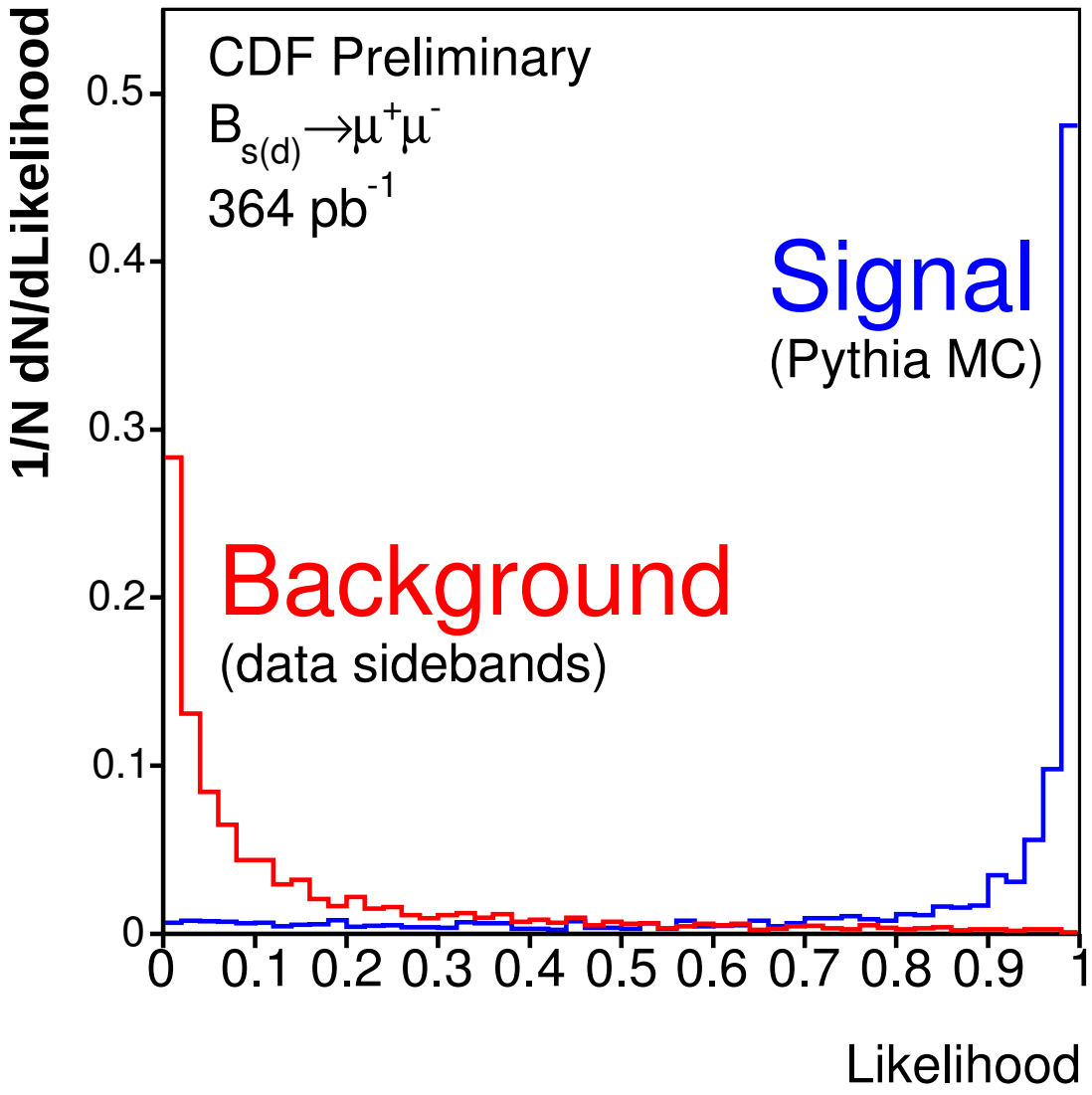

Figure 11: The resulting likelihood curves for signal and background in the CMU-CMU channel using the final choice of (Iso,$\Delta\alpha$,$\mathcal{P}(\lambda)$) variables discussed in Section 4.1.

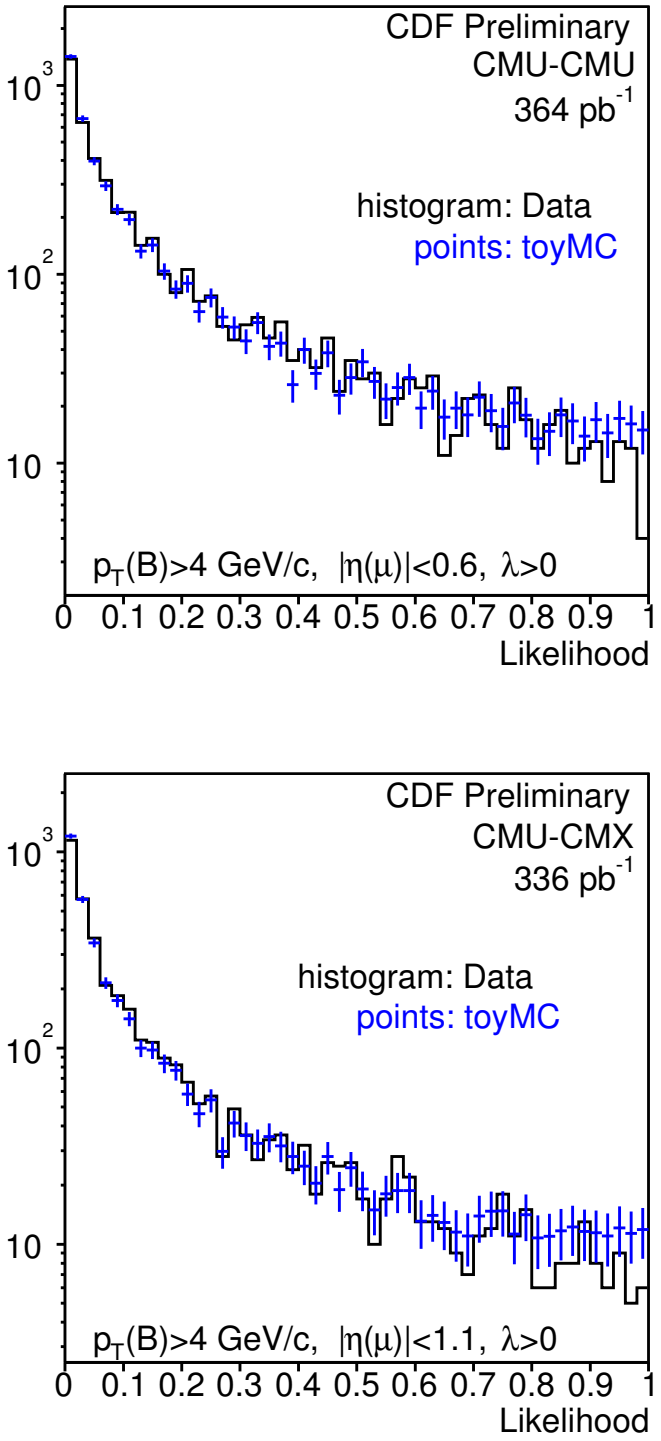

Figure 12: A comparison of the likelihood distribution from data sideband events (histogram) and from the toy-MC (points) for $\mu^+\mu^-$ events surviving the baseline and vertex requirements for the CMU-CMU (top) and CMU-CMX (bottom) channels. The KS-probability is 2% (3%) for the top (bottom).

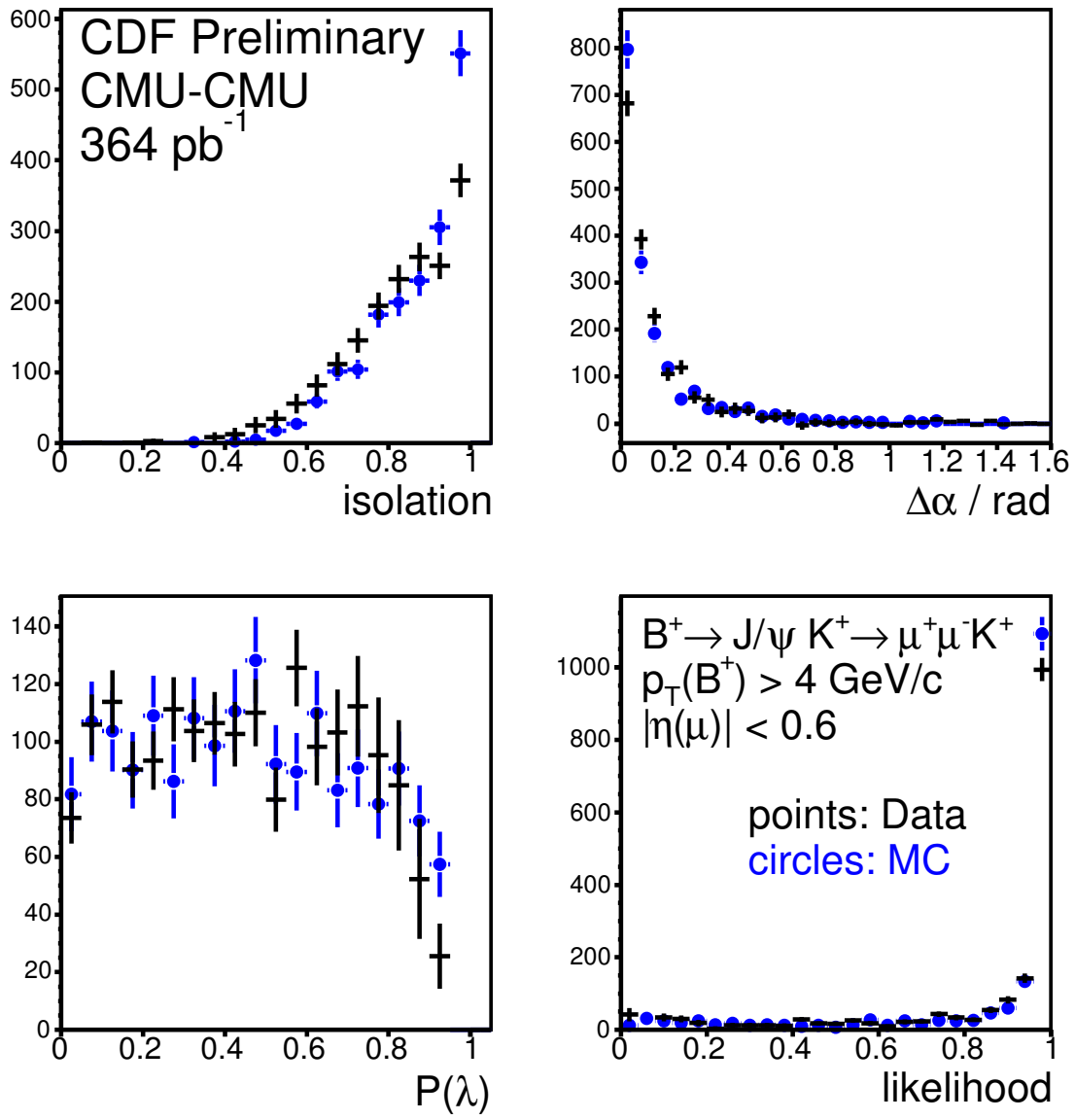

Figure 13: A comparison of the input distributions (Iso,$\Delta\alpha$, and $\mathcal{P}(\lambda)$) and the resulting likelihood distribution from sideband subtracted $B^+ \to J/\psi K^+$ data to that from the Pythia MC sample described in Section 3 for the CMU-CMU channel.

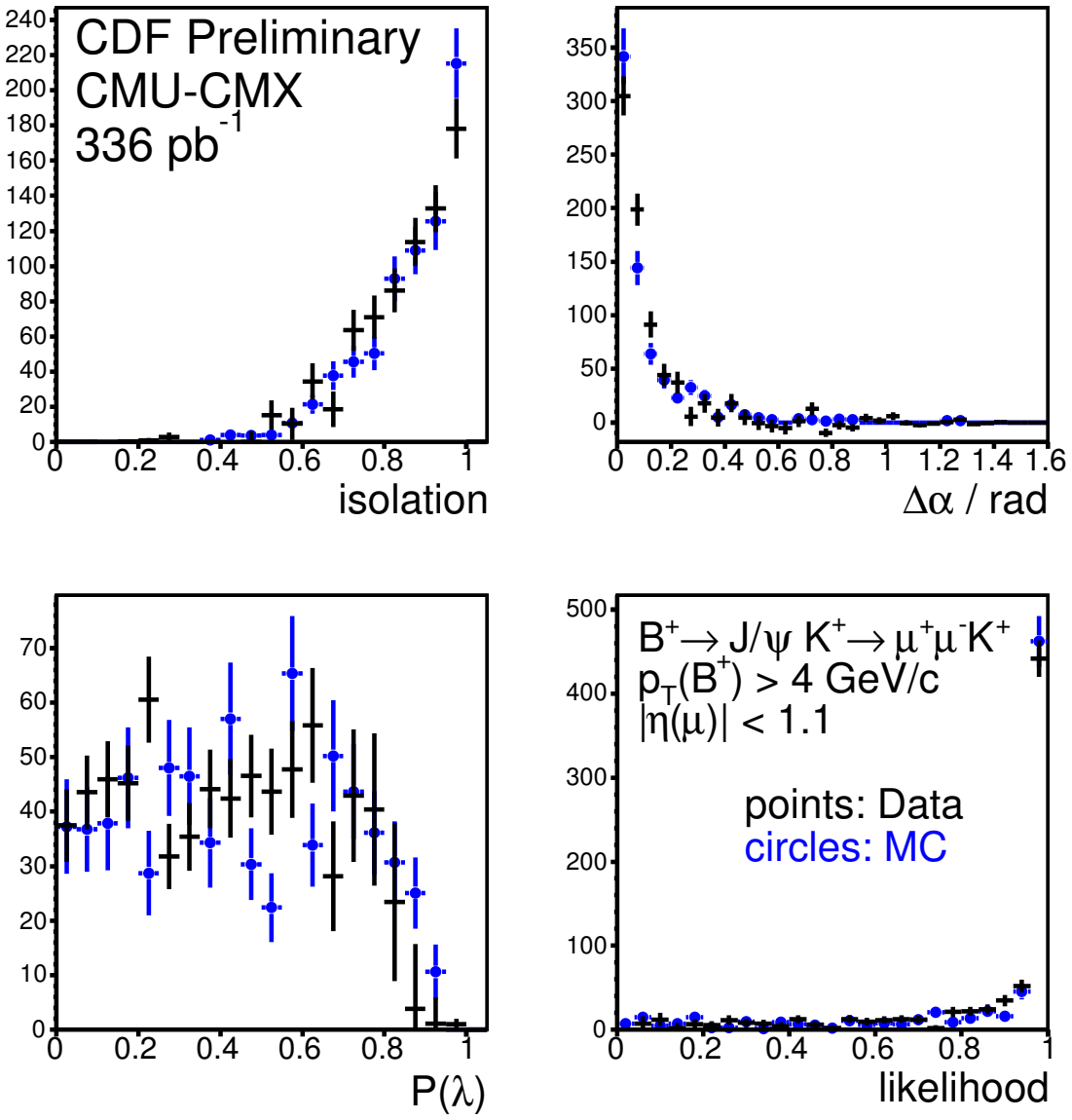

Figure 14: A comparison of the input distributions (Iso,$\Delta\alpha$, and $\mathcal{P}(\lambda)$) and the resulting likelihood distribution from sideband subtracted $B^+ \to J/\psi K^+$ data to that from the Pythia MC sample described in Section 3 for the CMU-CMX channel.

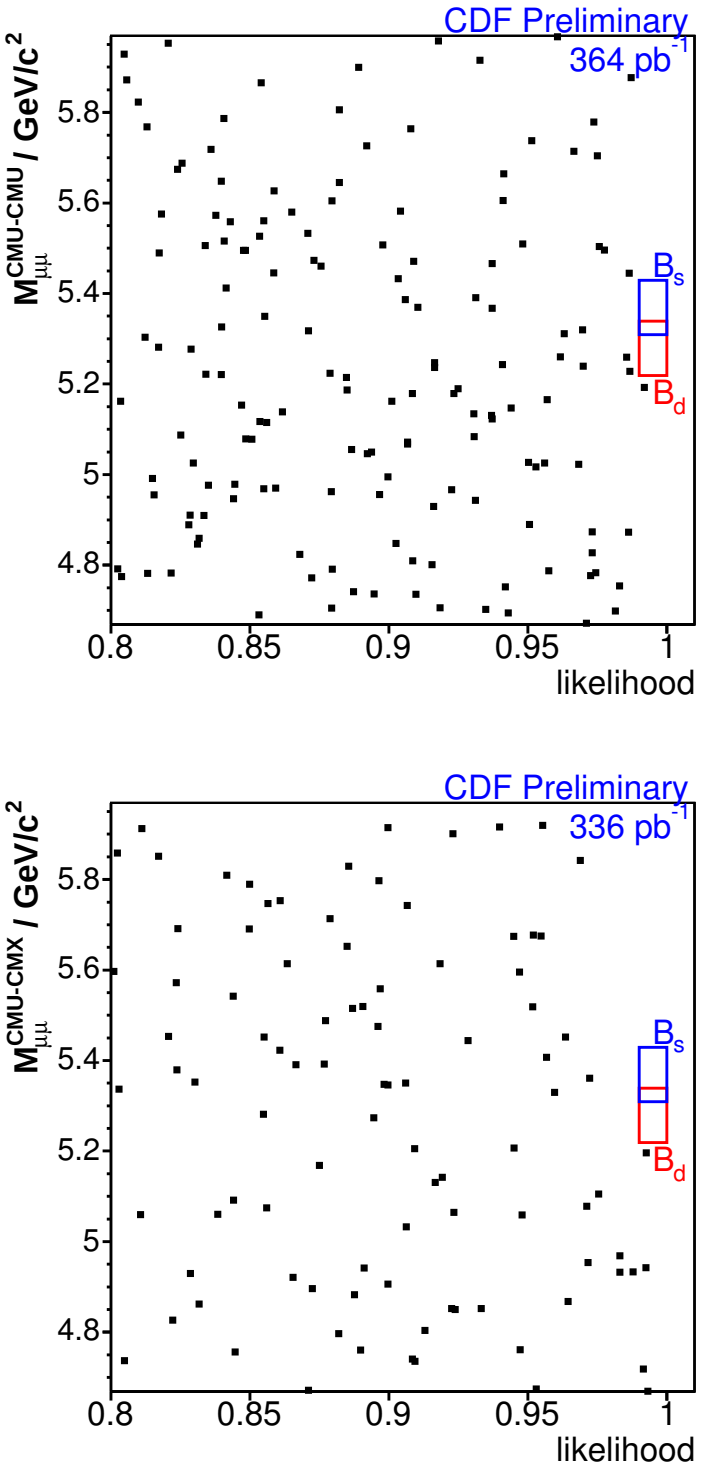

Figure 15: The invariant mass distribution versus the event likelihood for events satisfying baseline in the CMU-CMU (top) and CMU-CMX (bottom) channels. Only events with $LH >$ 0.80 are shown. No events fall within the signal box in either channel.

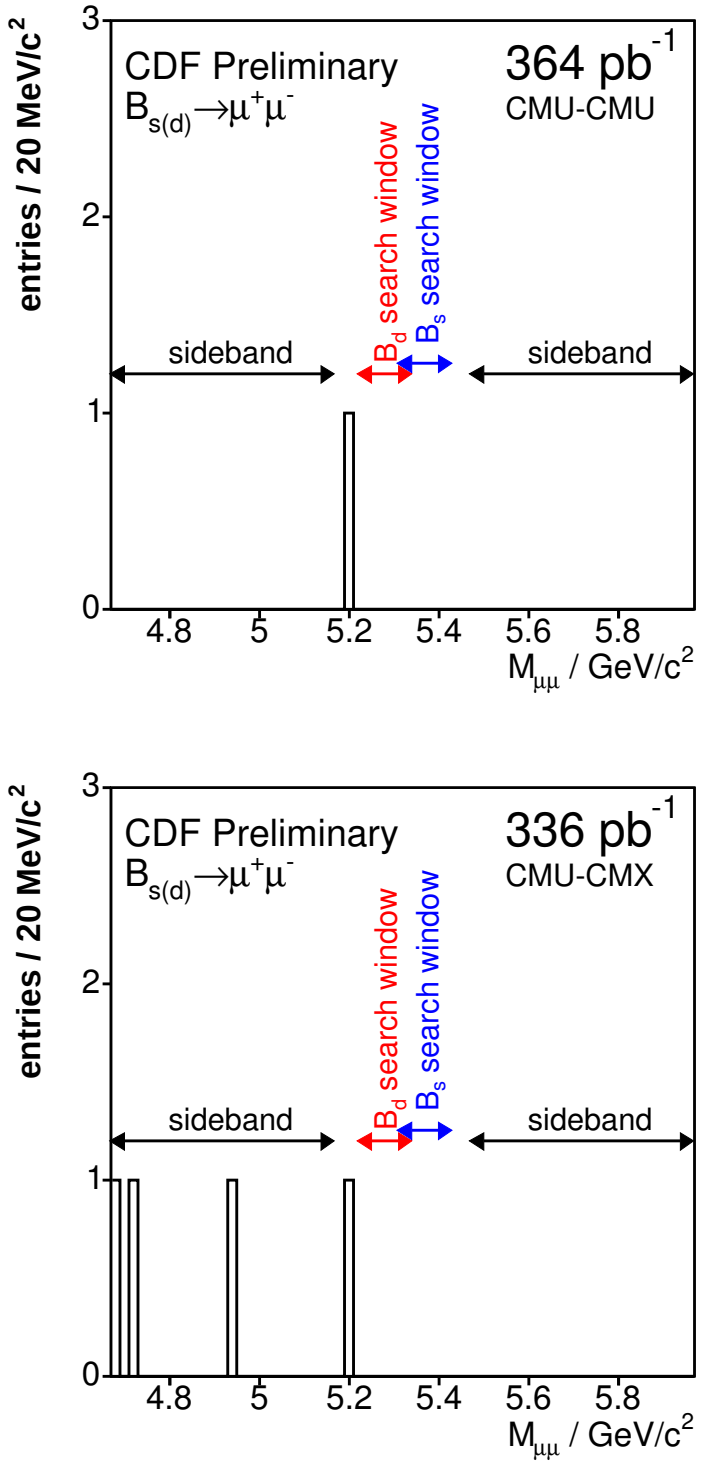

Figure 16: The invariant mass distribution for events satisfying all selection criteria in the CMU-CMU (top) and CMU-CMX (bottom) channels. No events in either channel fall within the B_s or B_d search window. The closest event has an invariant mass of 5.190 and 5.197 GeV in the CMU-CMU and CMU-CMX channel, respectively.

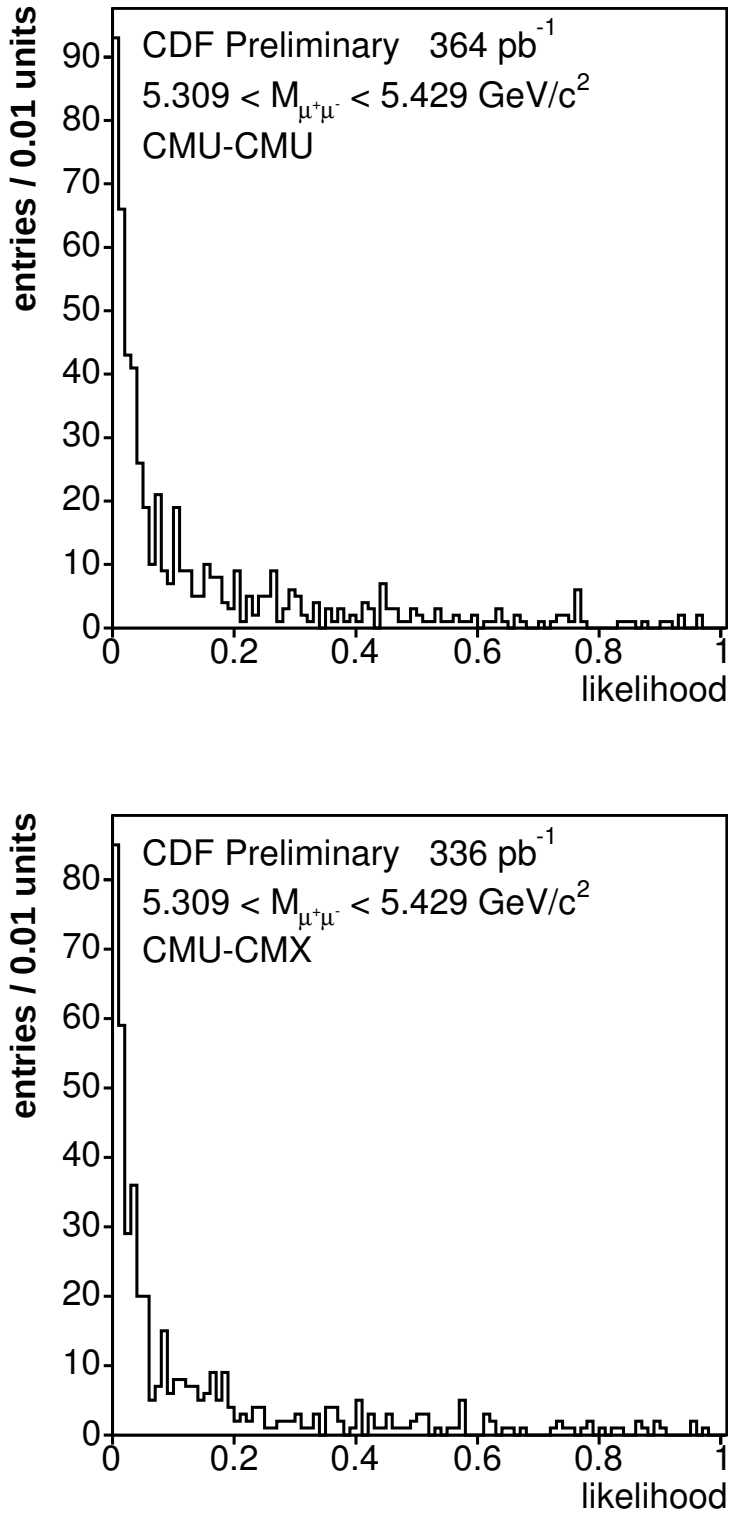

Figure 17: The likelihood distribution for events in the B_s search window for the CMU-CMU (top) and CMU-CMX (bottom) channel. There are no events with $LH > 0.99$.