CDF Public Note 8956

# Search for $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ Decays in $2\,\mathrm{fb}^{-1}$ of $p\overline{p}$ 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 2 fb$^{-1}$ of data collected by the CDF II detector at the Fermilab Tevatron Collider. After applying all selection requirements the observed number of $B_s^0$ and $B_d^0$ candidates is consistent with background expectations. The resulting upper limits on the branching fractions are $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 5.8 \times 10^{-8}$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 1.8 \times 10^{-8}$ 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 $|V_{ts}|^2$, while the rate of $B_d^0 \to \mu^+\mu^-$ decays is proportional to $|V_{td}|^2$. Both rates are further suppressed by helicity factors. 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 greater than one order 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 published 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 $\mathcal{B}(B_{s,d}^0 \to \mu^+\mu^-) \propto (\tan\beta)^6$, 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 [7], 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 addition, $\mathcal{B}(B_{s,d}^0 \to \mu^+\mu^-)$ has strong correlation with the the magnitude of the anomalous muon magnetic moment, dark matter nucleon scattering cross sections, and Higgs mass. This makes a measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ central to proving a large class new physics [6]. In the absence of an observation, any improvements to the limits can be used to significantly constrain many SUSY models [4–7].

In this paper, we report on a search for $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays using 2 fb$^{-1}$ 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 increasing the data sample and improving analysis technique including a likelihood based muon selection and a Neural Net based discriminant to separate signal from background. 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. Sourounding the tracking detectors are segmented electromagnetic and hadronic sampling calorimeters arranged in a projective geometry. 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 $\theta$ 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 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/$c$. Backgrounds from fake muons are suppressed by selecting muons based on a likelihood function constructed from tracker, calorimeter and muon system information including measurements of whether electromagnetic and hadronic energy is consistent with a minimum ionizing particle and differences between extrapolated track positions and muon system hits. In addition, background specifically from kaons are further suppressed by a loose selection based on specific ionization, $dE/dx$, in the drift chamber. Application of these muon identification techniques reduce combinatoric background by 25% and background from two body hadronic $B$ hadron decays to fractions of an event. 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 $L_{\rm 3D} = \vec{L} \cdot \vec{p}^{\,\mu\mu}/|\vec{p}^{\,\mu\mu}|$, 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 $\tau = M_{\mu\mu} L_{\rm 3D}/|\vec{p}^{\,\mu\mu}|$, 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 that $\lambda/\sigma_\lambda > 2$, where $\lambda = c \cdot \tau$ and $\sigma_\lambda$ is the total uncertainty on $\lambda$, $\Delta\Theta < 0.7\ rad$, where $\Delta\Theta$ is the 3D opening angle $\Delta\Theta$ between vectors $\vec{p}^{\,\mu\mu}$ and $\vec{L}$, and $B$-candidate track isolation, $I > 0.50$ [14]. There are 15594 (15072) dimuon candidates that fulfill all the above "baseline" trigger and offline reconstruction requirements in the U-U (U-X) channel (Fig. 1) . 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 and the $I$ distribution 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 discriminating variables: $M_{\mu\mu}$, $\lambda$, $\Delta\Theta$, $I$, $\vec{p}_T^{\,\mu\mu}$and the $p_T$ of the lower momentum muon candidate. Fig. 2 compares the distributions of these variables for data (which is background-dominated) to MC signal events. To enhance signal and background separation we construct a multivariate Neural Net (NN) based on the all the discriminating variables except $M_{\mu\mu}$ which is used to define signal and sideband background regions. The NN is trained using background events taken from the sideband regions and signal events from MC. Only part of the total number of background and signal events are used in order to have a sample to subsequently test the background discrimination and signal efficiency. We also conduct tests for for biases vs. mass and over-training. The search window is defined by $5.169 < M_{\mu\mu} < 5.469\,\mathrm{GeV}/c^2$, corresponding to approximately $\pm 6$ times the two-track invariant mass resolution, which is estimated to be $\sigma_M \approx 24$ 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. The resulting NN discrimination for the signal and background events are shown in Fig. 3. The NN discriminant provides approximately 25% better background rejection for the same efficiency compared to the likelihood ratio based method used in [3].

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 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

$$\mathcal{B}(B^0_s \to \mu^+\mu^-)^{95\%\mathrm{C.L.}} = \frac{N^{95\%}_{B^0_s}}{N_{B^+}} \cdot \frac{\alpha_{B^+}}{\alpha_{B^0_s}} \cdot \frac{\epsilon^{base}_{B^+}}{\epsilon^{base}_{B^0_s}} \cdot \frac{1}{\epsilon^{NN}_{B^0_s}} \cdot \frac{f_u}{f_s} \cdot \mathcal{B}(B^+ \to J/\psi K^+), \tag{1}$$

where $N^{95\%}_{B^0_s}$ is the number of $B^0_s \to \mu^+\mu^-$ decays at the 95% C.L. for $N$ observed and $N_b$ expected background events. The value of $N^{95\%}_{B^0_s}$ is estimated using the $CLs$ approach of Ref. [16] incorporating Gaussian uncertainties into the limit. Rather than conducting a simple counting experiment for events above a given NN cut in the signal region we subdivide the signal region into five equal mass bins and three NN bins delineated at 0.8, 0.95, 0.995 and 1.0 and then calculate a combined limit using all channels and bins. This set of bins is determined by optimizing the analysis based on the *a priori* expected upper limit for $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$. The expected limit for a given set of optimization requirements is computed by summing the 95% C.L. limits over all possible observations $N$, weighted by the corresponding Poisson probability when expecting $N_b$. We use the fragmentation ratio $f_u/f_s = 3.86 \pm 0.59$ [16]. The branching fraction $\mathcal{B}(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+) = (5.94 \pm 0.21) \times 10^{-5}$ is obtained from Ref. [16]. The expression for $\mathcal{B}(B^0_d \to \mu^+\mu^-)$ is derived by replacing $B^0_s$ with $B^0_d$ and the fragmentation ratio with $f_u/f_d = 1$.

The number of reconstructed $B^+ \to J/\psi K^+$ candidates, $N_{B^+}$, determined from the data using sideband subtraction, is 8117±137 for the U-U and 3270±90 for the U-X channel(Fig. 4). The parameter $\alpha_{B_s^0}$ ($\alpha_{B^+}$) is the trigger acceptance and $\epsilon^{base}_{B_s^0}$ ($\epsilon^{base}_{B^+}$) is the efficiency of the initial requirements for the signal (normalization) mode. We apply the NN requirement only to the signal mode and therefore the efficiency of the NN $\epsilon^{NN}_{B_s^0}$ appears only in the denominator of Eq. 1.

The acceptance ratio $\alpha_{B^+}/\alpha_{B^0_{s,d}}$ obtained from the signal MC is $0.297 \pm 0.020$ ($0.191 \pm 0.013$) 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 $\epsilon^{base}$ 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 $\epsilon^{trig}_{B^+}/\epsilon^{trig}_{B^0_{s,d}} = 0.9930 \pm 0.0002$ ($0.9956 \pm 0.0004$) 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. 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. Combining all effects, we find $\epsilon^{reco}_{B^+}/\epsilon^{reco}_{B^0_{s,d}} = 0.89 \pm 0.04$ ($0.88 \pm 0.04$) 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 NN requirement $\epsilon^{NN}_{B^0_{s,d}}$ is estimated from the signal MC. The efficiencies are 78%(79%) for $NN > 0.80$, 66%(67%) for $NN > 0.95$, and 44% for $NN > 0.995$ in the U-U(U-X) channels. We assign a relative systematic uncertainty of ±6% to both U-U and U-X channels based on comparisons of $B^+ \to J/\psi K^+$ MC and data samples (Fig. 5 6) and variations of the MC isolation distribution for $B_s^0 \to \mu^+\mu^-$ based on the statistical uncertainty of a study of $B_s^0 \to J/\psi\phi$ data.

Table I summarizes the acceptances and efficiencies used as inputs to Eq. 1.

| | U-U | | U-X | |
|---|---|---|---|---|
| $(\alpha_{B^+}/\alpha_{B_s^0})$ | $0.297 \pm 0.020$ | (±7%) | $0.191 \pm 0.013$ | (±7%) |
| $(\epsilon^{trig}_{B^+}/\epsilon^{trig}_{B_s^0})$ | $0.9930 \pm 0.0002$ | ( - ) | $0.9956 \pm 0.0004$ | ( - ) |
| $(\epsilon^{reco}_{B^+}/\epsilon^{reco}_{B_s^0})$ | $0.89 \pm 0.04$ | (±4%) | $0.88 \pm 0.04$ | (±4%) |
| $\epsilon^{NN}_{B_s^0}(NN > 0.80)$ | $0.779 \pm 0.047$ | (±6%) | $0.790 \pm 0.047$ | (±6%) |
| $N_{B^+}$ | $8117 \pm 137$ | (±2%) | $3270 \pm 90$ | (±3%) |
| $f_u/f_s$ | $3.86 \pm 0.59$ | (±15%) | $3.86 \pm 0.59$ | (±15%) |
| $BR(B^+)$ | $(5.94 \pm 0.21) \times 10^{-5}$ | (±4%) | $(5.94 \pm 0.21) \times 10^{-5}$ | (±4%) |
| $ses$ | $(9.5 \pm 0.7) \times 10^{-9}$ | (±18%) | $(1.5 \pm 1.0) \times 10^{-8}$ | (±18%) |

TABLE I: A summary of the inputs used in equation 1 to estimate the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The relative uncertainties are given parenthetically. The single-event-sensitivity, $ses$, corresponding to $N_{B_s^0} = 1$, is shown in the last row.

We estimate combinatoric background by extrapolating the number of events in the sideband regions passing a given cut to the signal region using a linear fit. In addition we have considered background contributions from $B^0_{s,d} \to h^+h^-$, where $h^\pm = \pi^\pm$ or $K^\pm$. Tables II and III list the background contribution in the $B_s^0$ and $B_d^0$ signal windows for the U-U and U-X channels, respectively. The expected background is calculated based on Eq. 1, with $B_s^0 \to \mu^+\mu^-$ replaced by $B \to h^+h^-$. In addition, there are two more efficiency terms in the denominator of Eq. 1 to account for the muon fake rates, measured in $D^0 \to \pi K$ data, and the fraction of misidentified $B \to h^+h^-$ events falling in either $B_s^0$ or $B_d^0$ signal window. The branching ratio for the various $B \to h^+h^-$ modes are taken from [16]. For modes with only branching ratio limits, we use the limit as the central value and assign ±100% to the systematic uncertainties. The 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.

The total background is obtained by summing the combinatoric background as estimated from the sidebands and the contribution from $B \to h^+h^-$ decays.

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

| $M_{\mu\mu}$(GeV/$c^2$) | U-U 0.800 − 0.950 | U-X 0.800 − 0.950 | U-U 0.950 − 0.995 | U-X 0.950 − 0.995 | U-U 0.995 − 1.000 | U-X 0.995 − 1.000 |
|---|---|---|---|---|---|---|
| 5.310 − 5.334 | 0.0048 | 0.0011 | 0.0091 | 0.0023 | 0.0173 | 0.0049 |
| 5.334 − 5.358 | 0.0021 | $6.036 \times 10^{-4}$ | 0.0040 | 0.0011 | 0.0078 | 0.0021 |
| 5.358 − 5.382 | $9.311 \times 10^{-4}$ | $2.841 \times 10^{-4}$ | 0.0018 | $5.534 \times 10^{-4}$ | 0.0036 | 0.0010 |
| 5.382 − 5.406 | $4.836 \times 10^{-4}$ | $1.455 \times 10^{-4}$ | $9.237 \times 10^{-4}$ | $2.949 \times 10^{-4}$ | 0.0019 | $5.372 \times 10^{-4}$ |
| 5.406 − 5.430 | $1.494 \times 10^{-4}$ | $6.849 \times 10^{-5}$ | $2.785 \times 10^{-4}$ | $1.446 \times 10^{-4}$ | $5.503 \times 10^{-4}$ | $2.771 \times 10^{-4}$ |

TABLE II: Expected background events from $B_s^0 \to$hh for $NN > 0.8$

| $M_{\mu\mu}$(GeV/$c^2$) | U-U 0.800 − 0.950 | U-X 0.800 − 0.950 | U-U 0.950 − 0.995 | U-X 0.950 − 0.995 | U-U 0.995 − 1.000 | U-X 0.995 − 1.000 |
|---|---|---|---|---|---|---|
| 5.219 − 5.243 | 0.0184 | 0.0032 | 0.0349 | 0.0067 | 0.0663 | 0.0143 |
| 5.243 − 5.267 | 0.0216 | 0.0053 | 0.0412 | 0.0096 | 0.0803 | 0.0186 |
| 5.267 − 5.291 | 0.0237 | 0.0061 | 0.0467 | 0.0119 | 0.0907 | 0.0225 |
| 5.291 − 5.315 | 0.0196 | 0.0050 | 0.0375 | 0.0101 | 0.0767 | 0.0184 |
| 5.315 − 5.339 | 0.0102 | 0.0023 | 0.0190 | 0.0050 | 0.0376 | 0.0095 |

TABLE III: Expected background events from $B_d^0 \to$hh for $NN > 0.8$

observed in the search window for a wide range of NN requirements. No statistically significant discrepancies are observed. The results are shown in Table IV.

| sample | $NN$ cut | U-U pred | U-U obsv | U-U prob(%) | U-X pred | U-X obsv | U-X prob(%) |
|---|---|---|---|---|---|---|---|
| OS- | > 0.80 | 231 ± (8) | 230 | 48 | 254 ± (9) | 278 | 10 |
| | > 0.90 | 133 ± (6) | 137 | 37 | 142 ± (7) | 163 | 6 |
| | > 0.995 | 23 ± (3) | 21 | 41 | 18 ± (2) | 29 | 2 |
| SS+ | > 0.80 | 1.2 ± (0.6) | 0 | 34 | 2.4 ± (0.8) | 1 | 34 |
| | > 0.95 | 0.6 ± (0.4) | 0 | 53 | 1.2 ± (0.6) | 1 | 66 |
| | > 0.995 | 0.3 ± (0.3) | 0 | 70 | 0.3 ± (0.3) | 0 | 70 |
| SS- | > 0.80 | 2.4 ± (0.8) | 2 | 57 | 7.5 ± (1.5) | 6 | 40 |
| | > 0.90 | 1.2 ± (0.6) | 2 | 34 | 5.1 ± (1.2) | 3 | 29 |
| | > 0.98 | 0.3 ± (0.3) | 0 | 70 | 2.1 ± (0.8) | 1 | 41 |
| FM+ | > 0.80 | 26 ± (3) | 25 | 48 | 6.3 ± (1.4) | 6 | 56 |
| | > 0.95 | 8.4 ± (1.6) | 5 | 19 | 2.1 ± (0.8) | 1 | 41 |
| | > 0.995 | 3.3 ± (1.0) | 3 | 58 | 0.6 ± (0.4) | 0 | 55 |

TABLE IV: A comparison of the predicted and observed number of events in an extended signal mass region as a function of NN 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 probabilities are calculated including the statistical uncertainties on the background prediction itself.

Using the efficiencies, background estimates and optimized selection criteria we we compute the $B_s^0$ single event $B_s^0 \to \mu^+\mu^-$ sensitivity as $(2.0 \pm 0.4) \times 10^{-8}$ [$(2.7 \pm 0.5) \times 10^{-8}$] for the U-U [U-X] channel for all mass bins and a NN selection requirement of 0.995. The expect limits combining all mass and NN bins are $6.5 \times 10^{-8}$ [$10.0 \times 10^{-8}$] for the U-U [U-X] channels and $4.9 \times 10^{-8}$ combined at 95% CL. Using these criteria for all mass bins and NN values greater than 0.995 we observe 2 [1] and 5 [1] 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(Table V).

The invariant mass distributions are shown after we apply all selection cuts in Fig. 7. Two dimensional plots of the invariant mass vs. the NN value are shown in Fig. 8.

Using Eq. 1 and combining the U-U and U-X channels taking into account the correlated uncertainties, we derive 90% (95%) C.L. limits of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 4.7 \times 10^{-8}$ $(5.8 \times 10^{-8})$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 1.5 \times 10^{-8}$ $(1.8 \times 10^{-8})$. The new limits improve the previous limits [3, 18, 19] by a factor of three and can be used to reduce the allowed parameter space of a broad spectrum of SUSY models [5–7].

| $M_{\mu\mu}$ U-U | I | II | III | IV | V |
|---|---|---|---|---|---|
| A Exp | 4.9±0.3 | 4.8±0.3 | 4.7±0.3 | 4.6±0.3 | 4.5±0.3 |
| Obs | 3 | 3 | 7 | 3 | 2 |
| B Exp | 1.6±0.2 | 1.6±0.2 | 1.5±0.2 | 1.5±0.2 | 1.5±0.2 |
| Obs | 1 | 1 | 4 | 2 | 2 |
| C Exp | 0.43±0.10 | 0.42±0.10 | 0.41±0.10 | 0.41±0.1 | 0.41±0.10 |
| Obs | 0 | 1 | 1 | 0 | 0 |
| $M_{\mu\mu}$ U-X | I | II | III | IV | V |
| A Exp | 5.4±0.4 | 5.3±0.3 | 5.2±0.3 | 5.1±0.3 | 5.0±0.3 |
| Obs | 8 | 6 | 3 | 6 | 3 |
| A Exp | 2.1±0.2 | 2.1±0.2 | 2.1±0.2 | 2.0±0.2 | 2.0±0.2 |
| Obs | 3 | 2 | 2 | 4 | 0 |
| A Exp | 0.32±0.09 | 0.31±0.09 | 0.31±0.10 | 0.31±0.10 | 0.31±0.11 |
| Obs | 0 | 0 | 0 | 0 | 1 |

| $M_{\mu\mu}$ U-U | I | II | III | IV | V |
|---|---|---|---|---|---|
| A Exp | 5.3±0.4 | 5.2±0.3 | 5.1±0.3 | 5.0±0.3 | 4.9±0.3 |
| Obs. | 7 | 11 | 5 | 7 | 2 |
| B Exp | 1.8±0.2 | 1.7±0.2 | 1.7±0.2 | 1.7±0.2 | 1.6±0.2 |
| Obs | 0 | 1 | 3 | 2 | 1 |
| C Exp | 0.47±0.10 | 0.49±0.10 | 0.50±0.10 | 0.48±0.1 | 0.45±0.10 |
| Obs | 2 | 3 | 0 | 0 | 0 |
| $M_{\mu\mu}$ U-X | I | II | III | IV | V |
| A Exp | 5.7±0.4 | 5.6±0.4 | 5.6±0.4 | 5.5±0.4 | 5.4±0.4 |
| Obs | 8 | 2 | 4 | 7 | 7 |
| B Exp | 2.2±0.2 | 2.2±0.2 | 2.2±0.2 | 2.1±0.2 | 2.1±0.2 |
| Obs | 1 | 0 | 2 | 0 | 3 |
| C Exp | 0.33±0.09 | 0.33±0.09 | 0.33±0.10 | 0.33±0.10 | 0.32±0.11 |
| Obs | 0 | 0 | 1 | 0 | 0 |

TABLE V: Expected backgrounds(Exp), including $B \to hh$, and number of observed events(Obs) in the U-U and U-X channels for the $B^0_s$ (upper) and $B^0_d$ (lower) signal windows. NN bins: A, 0.80-0.95, B, 0.95-0.995, and C, 0.995-1.0, and mass bins: I, 5.310-5.334, II, 5.334-5.358, III, 5.358-5.382, IV, 5.382-5.406, and V, 5.406-5.430(GeV/$c^2$).


[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. **09**, 029 (2005);
[6] R. Ruiz de Austri *et al.*, J. High Energy Phys. 0605, 002 (2006); S. Baek *et al.*, J. High Energy Phys. 0506, 017 (2005);
[7] R. Arnowitt *et al.*, Phys. Lett. B **538**, 121 (2002).
[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] W.M. Yao *et al.*, J. Phys. G **33**, 1 (2006).
[17] $f_x$ 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 $f_u$.
[18] D0 Collaboration, V. Abazov *et al.*, Phys. Rev. Lett. **94**, 071802 (2005).
[19] BABAR Collaboration, B. Aubert *et al.*, Phys. Rev. Lett **94**, 221803 (2005).

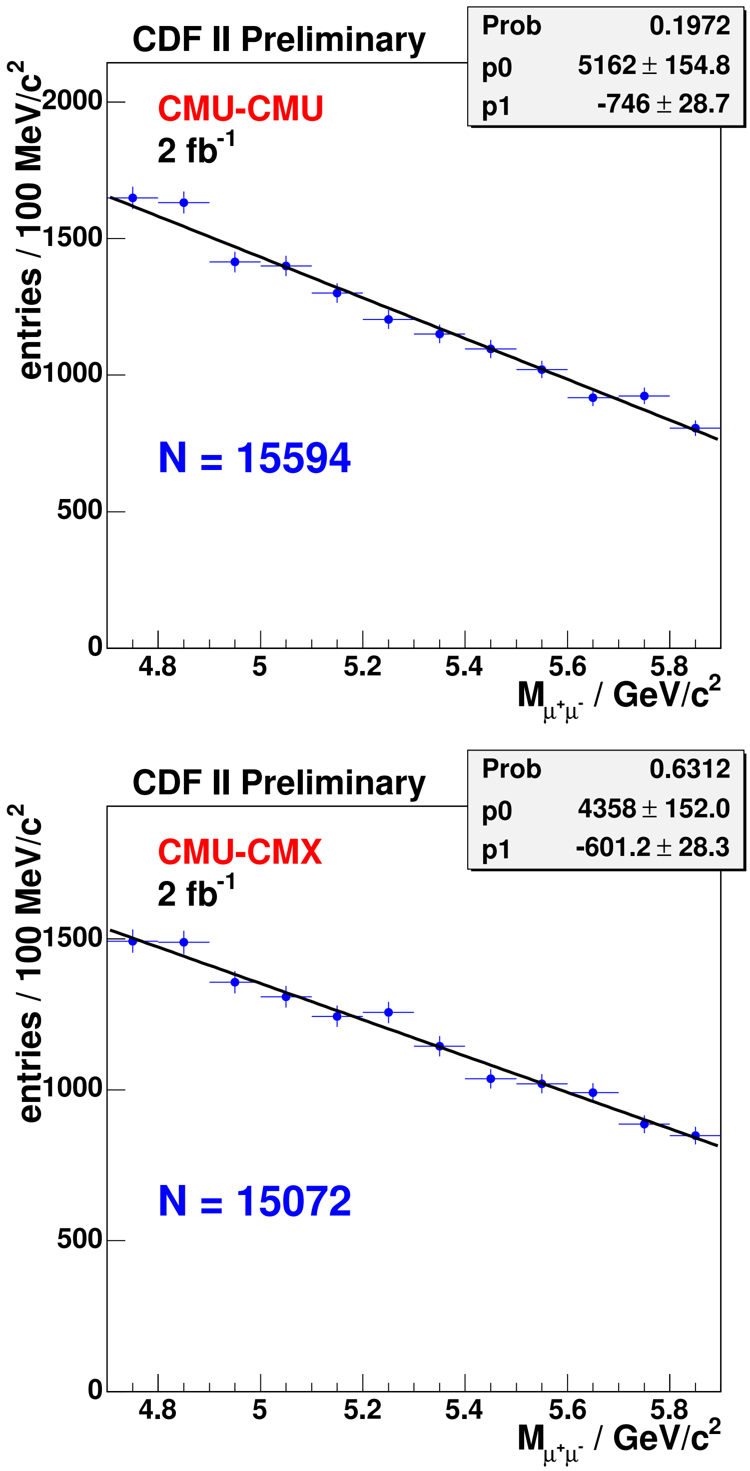



FIG. 1: Di-muon invariant mass distribution for events satisfying the baseline requirements. The results of a fit to a straight line are shown.

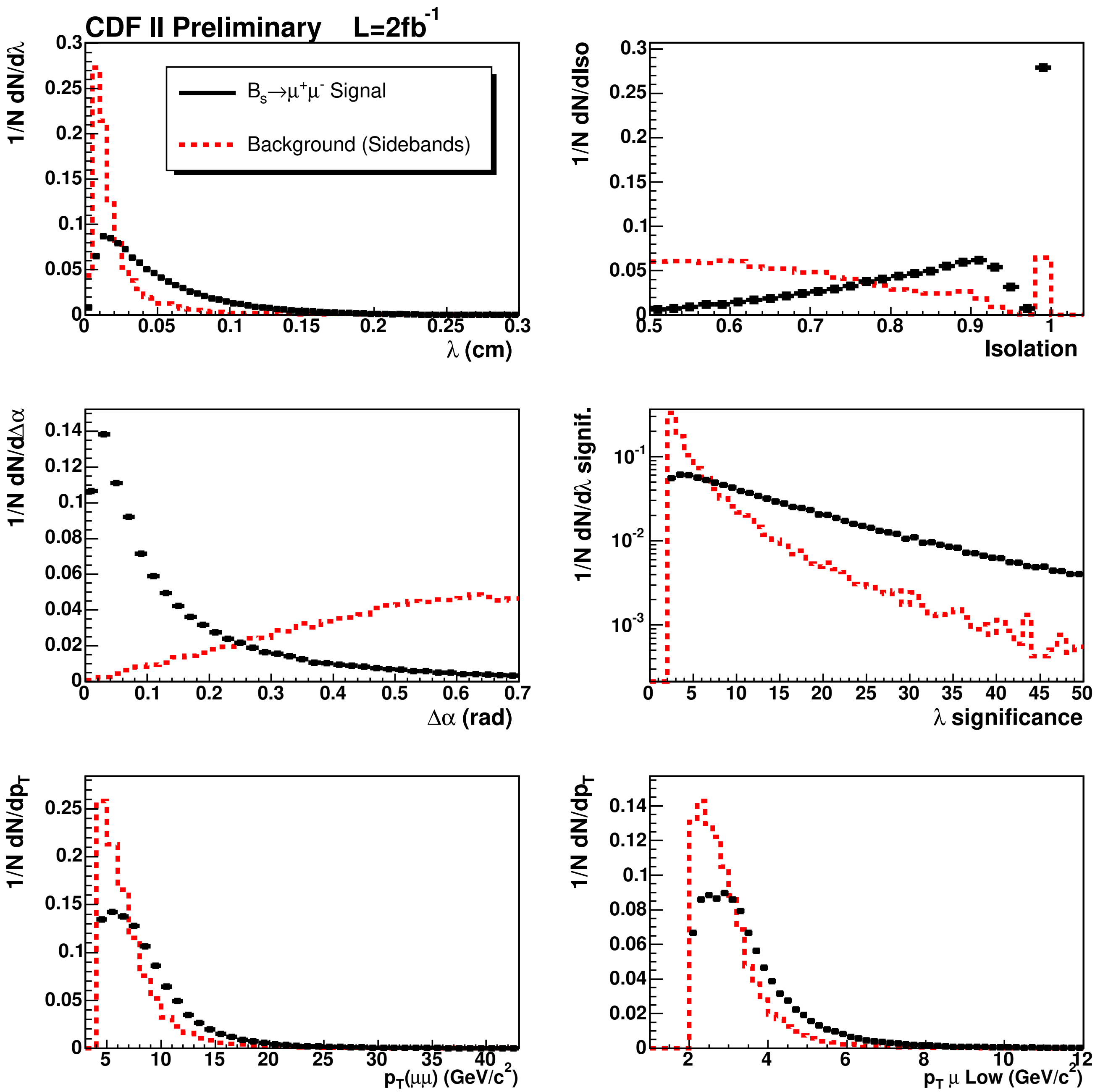



FIG. 2: Distributions of the input variables to the NN for $B_s^0 \to \mu^+\mu^-$. Data point (black) is signal MC with dashed histogram (red) being background events from data sidebands. The distributions are normalized to unit area.

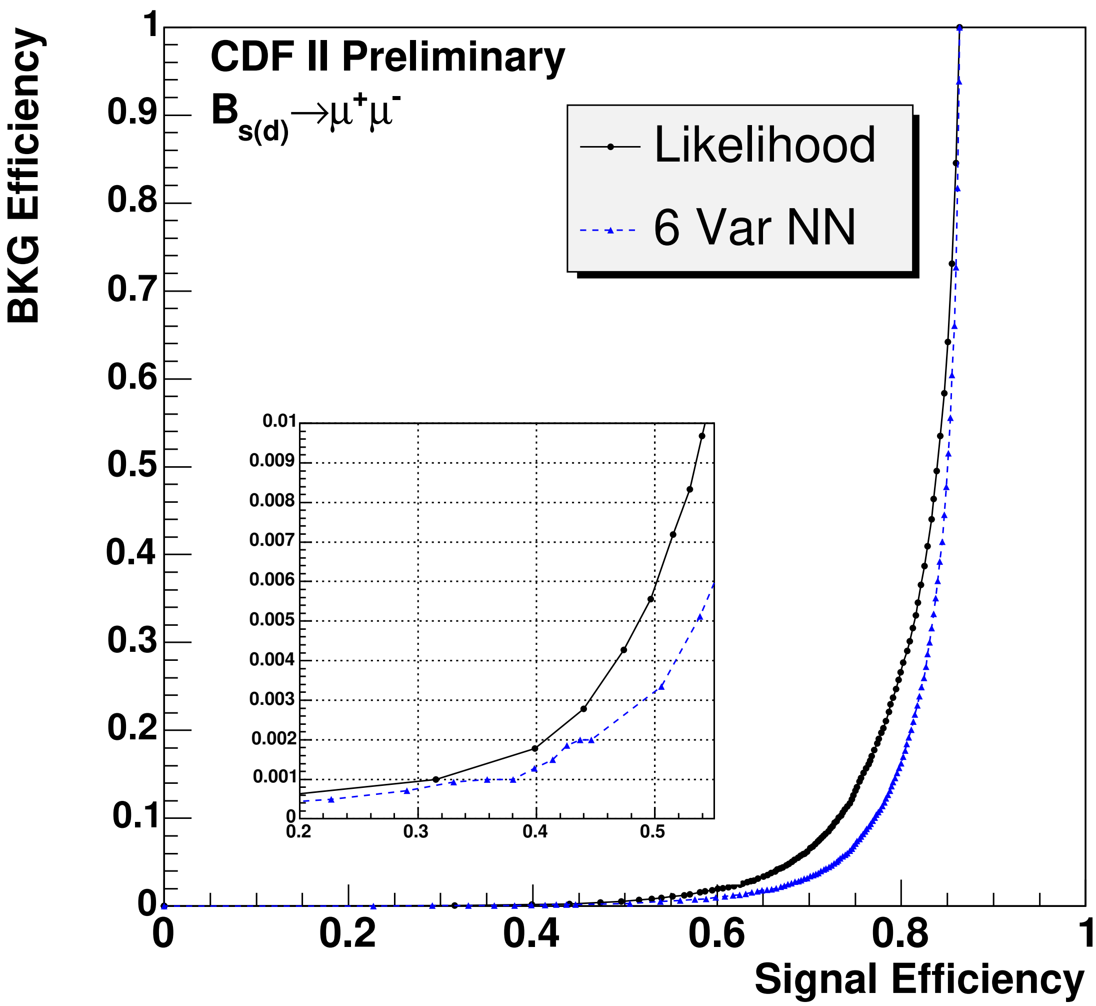



FIG. 3: Efficiency for background events as a function of signal efficiency. Likelihood ratio and NN descriminants are compared.

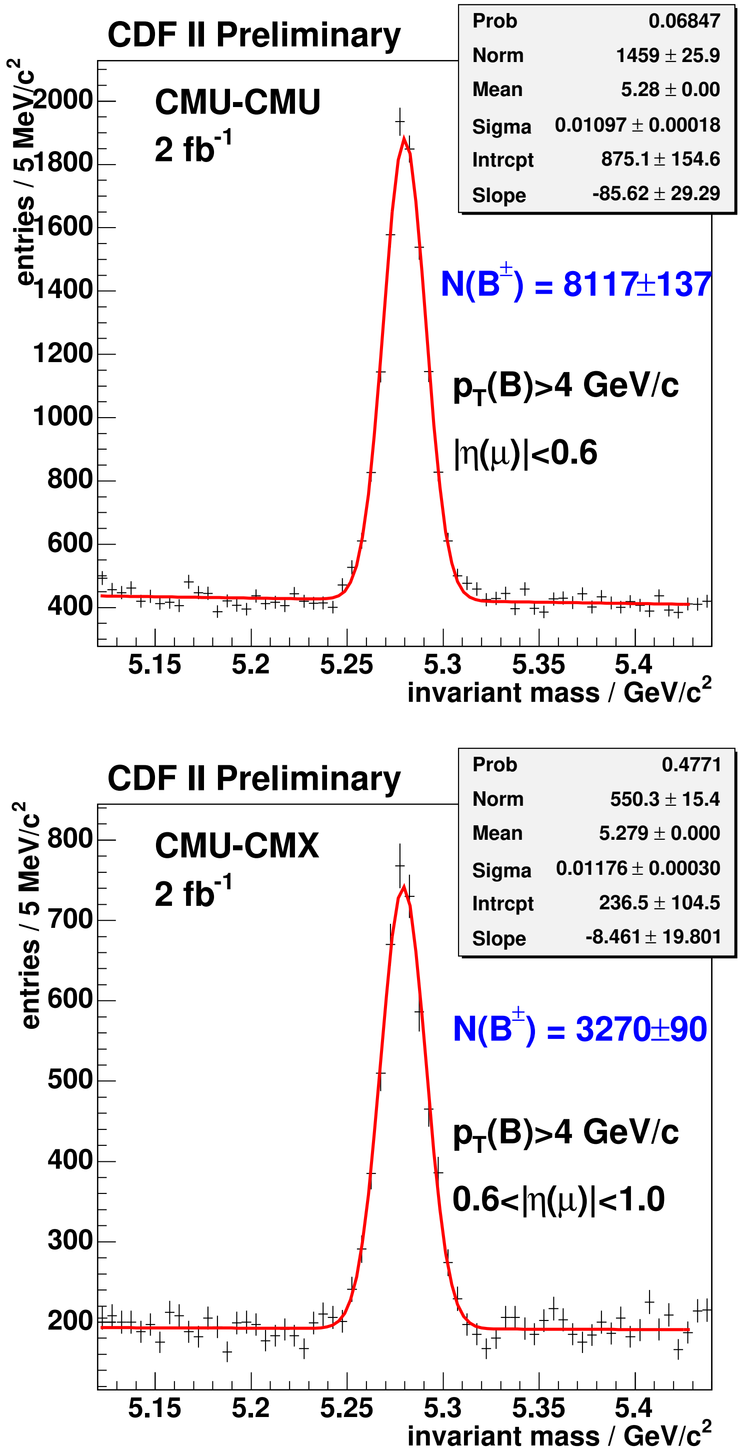



FIG. 4: The $\mu^+\mu^- K^+$ invariant mass distribution for events satisfying the baseline requirements.

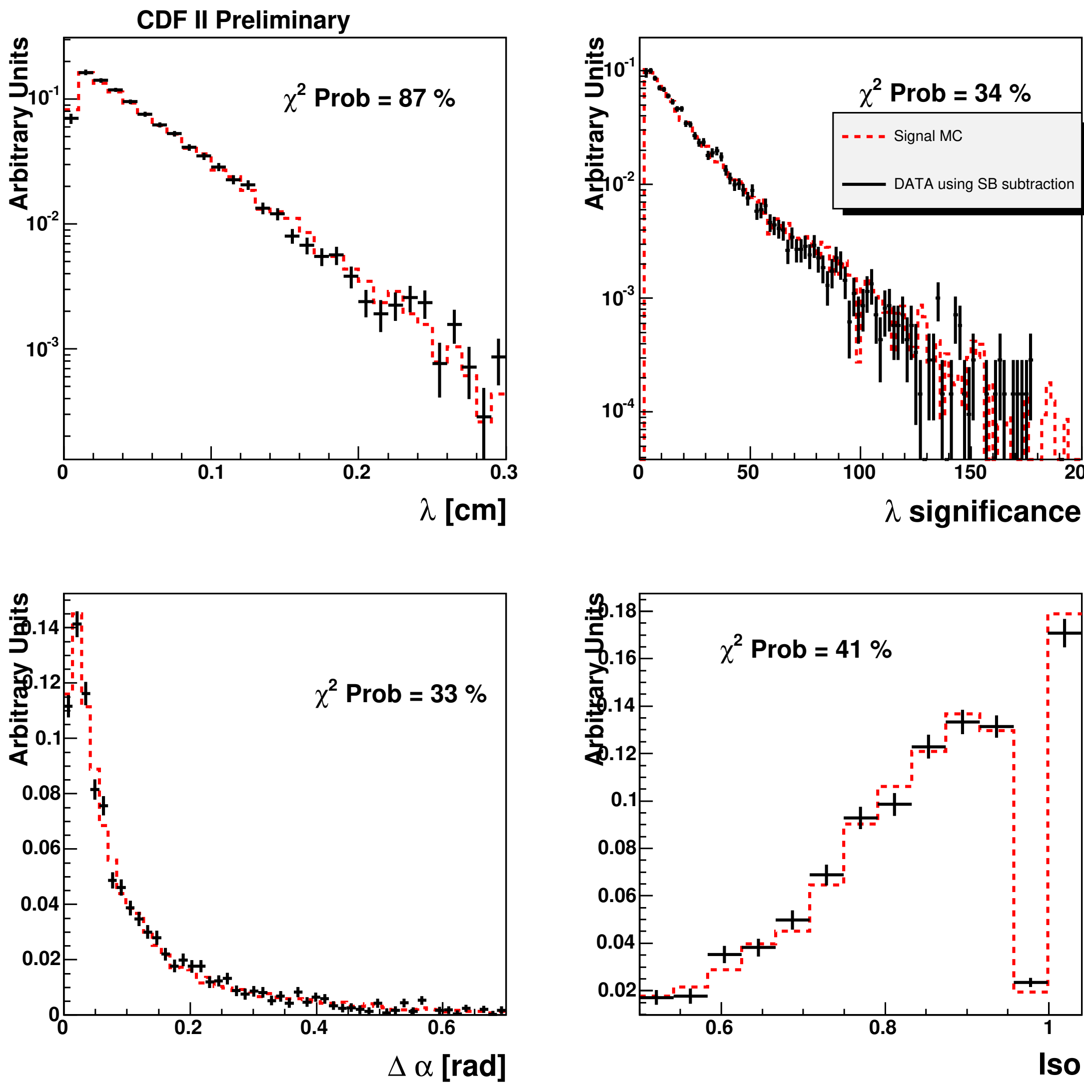



FIG. 5: Distributions of the input variables to the NN for $B^+ \rightarrow J/\psi K^+$. Red is signal MC with black being signal events from data extracted using sideband subtraction. The distributions are normalized to a area of one.

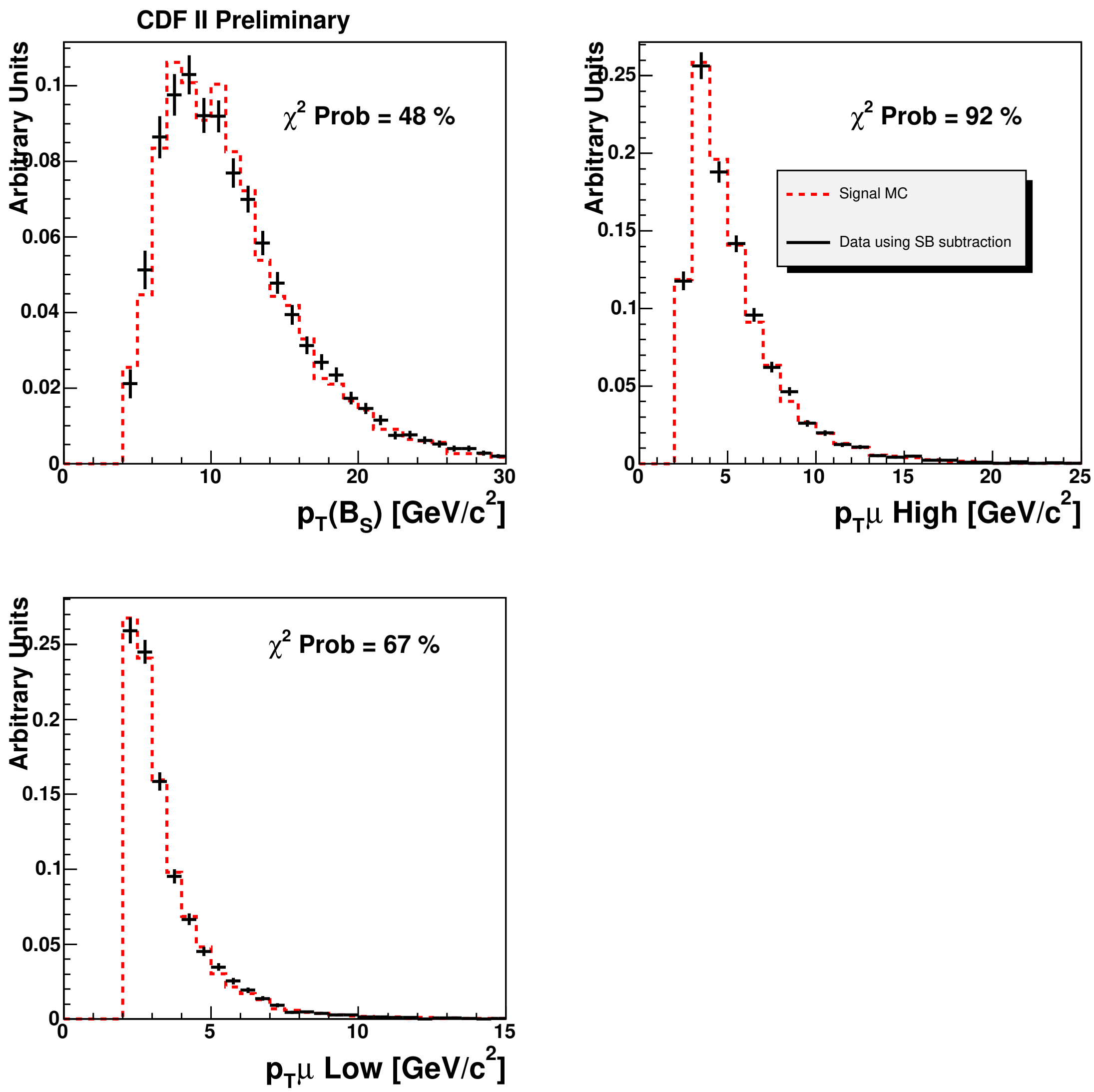



FIG. 6: Distributions of the input variables to the NN for $B^+ \to J/\psi K^+$. Red is signal MC with black being signal events from data extracted using sideband subtraction. The distributions are normalized to a area of one.

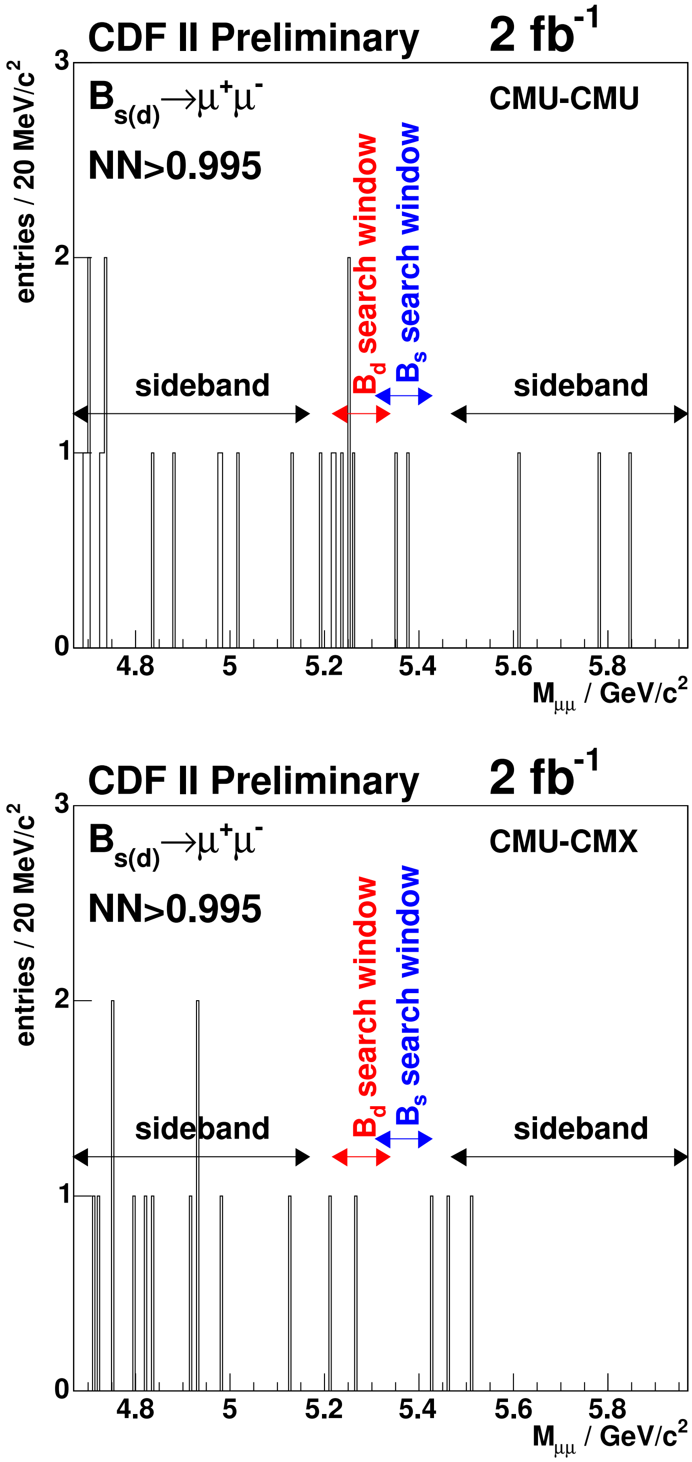
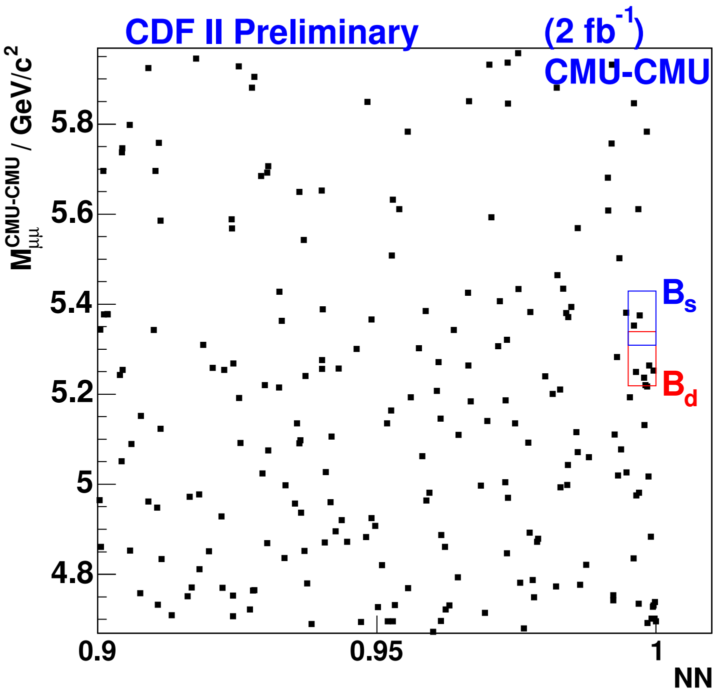



FIG. 7: The di-muon mass distribution for events with $NN > 0.995$ and satisfying all selection criteria in the U-U (top) and U-X (bottom) channels. There are seven events in the U-U channel (2 events in $B_s^0$ and 5 events in $B_d^0$ mass windows). For the U-X channel, there is one event per signal window.



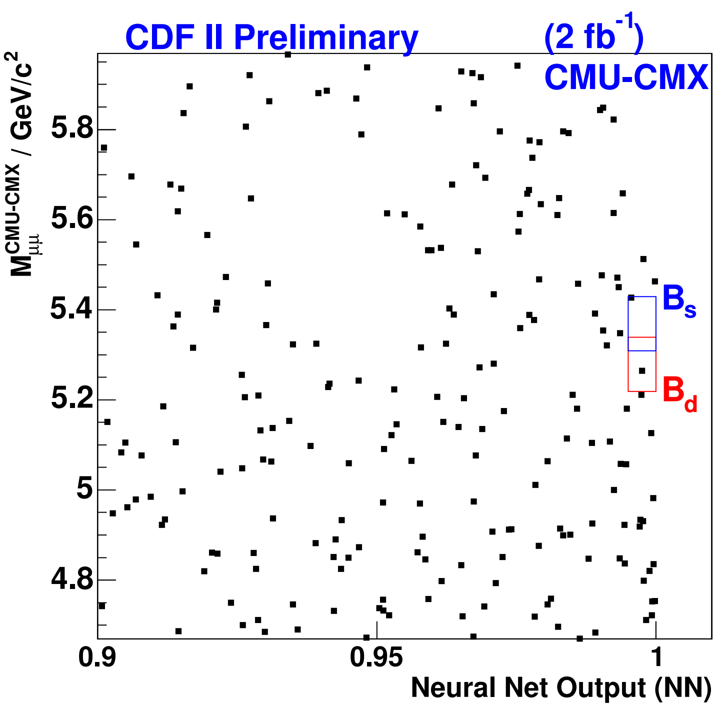




FIG. 8: The invariant mass distribution versus the NN output for events satisfying baseline in the U-U (top) and U-X (bottom) channels. Only events with $NN > 0.90$ are shown. For $NN > 0.995$, there are seven candidates in the U-U channel (2 events in $B_s^0$ and 5 events in $B_d^0$ mass windows). For the U-X channel, $B_s^0$ and $B_d^0$ mass windows have one event each.