SEARCH FOR $B_{s,d} \to \mu^+\mu^-$ DECAYS WITH 10/FB OF $p\overline{p}$ COLLISIONS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Walter Hopkins

January 2013

SEARCH FOR $B_{s,d} \to \mu^+\mu^-$ DECAYS WITH 10/FB OF $p\overline{p}$ COLLISIONS

Walter Hopkins, Ph.D.

Cornell University 2013

A search for the rare decays of $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ is presented in $p\overline{p}$ collisions at $\sqrt{s}$=1.96 TeV corresponding to 10 fb^{-1} of integrated luminosity collected by the CDF II detector. The observed number of B_d^0 candidates is consistent with background-only expectations and yields an upper limit on the branching fraction $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ of 4.2×10^{-9} at 95% confidence level. An excess is observed for the B_s^0 signal region. The probability that background processes alone could produce such an excess or larger is 0.94%. The probability that the combination of background and the expected standard model rate $B_s^0 \to \mu^+\mu^-$ could produce such an excess or larger is 6.8%. These data are used to determine a central value of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (1.3^{+0.9}_{-0.7}) \times 10^{-8}$ as well as a two sided limit at 90% C.L. of $2.2 \times 10^{-9} < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.0 \times 10^{-8}$.

BIOGRAPHICAL SKETCH

The author was born in Cologne, Germany in 1984. He attended the Rochester Institute of Technology from 2003 to 2007 and received his Bachelor of Science in Applied Mathematics and Physics in May 2007. He continued on to research experimental high energy physics at Cornell University under the direction of Professor Julia Thom. In 2010 he received his Master of Arts in Physics from Cornell University.

This document is dedicated to my supportive parents Walter Sr. and Antje.

ACKNOWLEDGEMENTS

I would like to start by thanking my thesis adviser, Julia Thom-Levy, for her continuous support throughout the last three years. From discussions about detector hardware to theory models, I'm deeply thankful for Julia's guidance. Most of all I'm grateful for her introducing this truly interesting and educational analysis to me.

Within the $B_s^0 \to \mu^+\mu^-$ analysis group I would like to especially thank Douglas Glenzinski and Matthew Herndon for their invaluable advice and guidance. Without their vast knowledge of the CDF detector and analysis techniques this analysis wouldn't have been possible. This analysis involved many steps which all the members of the $B_s^0 \to \mu^+\mu^-$ group were a part of. I thank each member of this group for the contributions.

On a more global CDF scale, I would like to thank all the members of the CDF B meson working group. The internal review and comments to the analysis greatly improved our methods and interpretations. Special thanks goes out to the B group conveners and the $B_s^0 \to \mu^+\mu^-$ godparents who closely followed the analysis.

As a whole I would also like to thank everybody in the Cornell Physics Department. I enjoyed the fun and friendly environment in the department. From learning more about certain theory models from fellow graduate students in the high energy theory department to very informative high energy physics journal clubs. I would also like to give special thanks to my previous research advisers Jim Alexander and Ritchie Patterson who allowed me to gain experience in detector hardware both in Wilson laboratory at Cornell as well as the Compact Muon Solenoid at the Large Hadron Collider.

Last, but most definitely not least, I would like to thank my parents. These two people have been the most supportive and understanding people in my life. From my father showing me little science experiments to my mother critically thinking of every aspect in life, my parents have given me valuable lessons and skills. I would also like to thank my siblings and friends who have given me support and joy in many ways.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

During the last century a comprehensive model for all interactions and particles in particle physics has been developed and is known as the standard model (SM). This theoretical model has been extraordinarily successful in describing particle physics. The vast majority of experimental results from the past five decades have been accurately described by this model.There are, however, problems with this model that include the weak hierarchy problem, the exclusion of gravity in the model, and the use of ad-hoc parameters. The SM also fails to explain dark matter and dark energy which are the main constituents of the universe. Many new physics (NP) models have been proposed to solve these problems. Each of these models solve problems in differing ways from introducing a new symmetry to the addition of space dimensions.

Experiments have been searching for both direct and indirect indications of physics beyond the SM. Direct searches usually study particles or interactions that would directly be produced in a NP model. Indirect searches tend to measure physical quantities such as branching fractions or Cabibbo-Kobayashi-Maskawa (CKM) matrix angles that could be anomalous if any beyond the SM physics were present.

Flavor changing neutral current (FCNC) processes, such as the $B^0_{s,d} \to \mu^+\mu^-$ decays, can not occur in the SM at tree level and require higher order box and penguin diagrams. These processes are thus heavily suppressed in the SM and are a powerful probe for NP. Several NP models predict potentially significant enhancements in the decay rate of $B^0_{s,d} \to \mu^+\mu^-$.

A specific group of models, supersymmetric (SUSY) models, arose from an attempt to solve the weak hierarchy problem. These models all predict enhancements to the branching fraction $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. SUSY models are particularly popular because they not only help solve the hierarchy problem by introducing fermion/boson partners to SM boson/fermion particles, but also hypothesize very heavy particles that could serve as a dark matter candidate. One such model, the minimal supersymmetric standard model (MSSM) predicts an enhancement of $\tan^6\beta$, where $\tan\beta$ is the ratio of the vacuum expectation values (VEV's) of the two Higgs doublets in the MSSM. These models predict maximum enhancements of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ that are orders of magnitude larger than the SM predicted values.

We search for $B_{s,d}^0 \to \mu^+\mu^-$ decays in attempt to find possible enhancements to the branching fraction $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. If no enhancement is found, limits will be able to constrain the NP parameter space significantly while complementing the constraints set by direct searches.

This document is organized as follows: Chapter 2 discusses the theoretical framework and current experimental status for $B_{s,d}^0 \to \mu^+\mu^-$ decays. Chapter 3 provides an overview of both the accelerator and collider detector used for the search for $B_{s,d}^0 \to \mu^+\mu^-$ decays. This is then followed by a discussion of the analysis methods in Chapter 4. Finally the results and conclusion for this search are explored in Chapter 5.

CHAPTER 2

THEORETICAL FRAMEWORK

Fundamental particles and their interactions are very well modeled by a quantum field theory known as the standard model. This model was first developed during mid 20^{th} century by Glashow, Weinberg, and Salam. Experimental results have verified many predictions of the model since the first evidence of quarks in the 1970's. The model contains 12 fermions, 5 gauge bosons, and a Higgs boson. Gauge bosons act as the force mediators of the three fundamental interactions in the SM: electromagnetic, weak, and strong. Even though the SM is the most successful theory of physics there are many aspects of physics that the model does not describe. Even within the set of interactions the model describes there are problems that can not be solved.

2.1 Introduction to the standard model

The SM describes 12 fermions (and their charge conjugates) and their interactions through 5 gauge bosons. An additional boson, the Higgs boson, is included in the model to yield the masses of the particles in the SM.

The SM describes fermions as spin-$\frac{1}{2}$ spinor fields and divides the fermions up into quarks and leptons. Leptons only interact with the electromagnetic and weak force. There are three generations of leptons of which three are electromagnetically charged: the electron (e), the muon (μ), and the tauon (τ). The remaining leptons are neutrinos (ν) that have a very low mass (< 2 eV [1]), no charge, and are associated with a particular charged lepton. The strength with

which the lepton interacts with the electromagnetic or weak force is given by a coupling constant. For the electromagnetic force this coupling constant can be associated with the electric charge of a particle. Leptons have either an electric charge of $\pm 1e$, or 0. For the weak force the strength of interaction is described by the weak isospin which depends on the chirality of the fermion.

Quarks are divided into up and down types. Up type quarks have charge $\pm\frac{2}{3}e$ while down type quarks have charge $\mp\frac{1}{3}e$. Quarks interact with the same forces as leptons with the addition of the strong force. The strong force 'charge' can can be either red (r), green (g), or blue (b). Due to their interaction with the strong force quarks can not be observed in a free state and are confined in combinations known as hadrons. Hadrons are either baryons, a combination of three quarks, or mesons, a combination of a quark and an antiquark. Hadrons always have net color of zero. A summary of the fermions described by the SM is given in Table 2.1.

Table 2.1: Fermions in the standard model. The isospin quantum number depends on the chirality of the fermion. Right-handed particles have isospin 0 and do not interact with the weak force. The color quantum number can have three values: red, green, or blue.

	Generation			**Quantum Number**		
	I	**II**	**III**	**Electric Charge**	**Isospin**	**Color**
Leptons	e	μ	τ	$+1e$	$-1/2$	0
	ν_e	ν_μ	ν_τ	0	$+1/2$	0
Quarks	u	c	t	$+2/3e$	$+1/2$	r,g,b
	d	s	b	$-1/3e$	$-1/2$	r,g,b

The three forces are described as boson mediator particles. These three forces

each form their own symmetry gauge groups. The electromagnetic force, mathematically described by the unitary group ($U(1)$), is mediated by the massless and chargeless photon which can have spin ±1. There are three massive gauge bosons for the weak force described by a generators of the 2D special unitarity group ($SU(2)$). The $W^{\pm}$ mediate the charged current interactions while the Z mediates the neutral current weak interactions. The force mediating bosons for the strong force are the massless and chargeless gluons. These bosons are mathematically described by the generators of 3D special unitary ($SU(3)$) gauge group.

The SM unifies the electromagnetic and weak force into one electroweak force with gauge symmetry ($SU(2) \times U(1)$). The SM predicts that all particles are massless without a process known as the Higgs mechanism which spontaneously breaks the $SU(2) \times U(1)$ symmetry. This spontaneous symmetry breaking gives rise to a fifth massive boson, the Higgs boson, and gives masses to all massive particles through interactions with the Higgs boson.

Together these bosons and fermions and their couplings describe most particle physics phenomena. This document will focus on the weak interactions that can occur during the decay of diquark steady states.

2.1.1 Electroweak symmetry breaking

A key component of the SM is a process of spontaneous symmetry breaking known as electroweak symmetry breaking (EWSB). The symmetry group of the SM is $SU(3) \times SU(2) \times U(1)$ which is broken by fermion and gauge boson mass terms. To allow for masses, spontaneous breaking of the $SU(3) \times SU(2) \times U(1)$

symmetry to $SU(3) \times U(1)$ is introduced using a scalar field. A scalar complex doublet (Eq. (2.1)) is introduced to the Lagrangian with a potential known as the Mexican hat potential shown in Eq. (2.2) and Fig. 2.1.

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.1}$$

$$V(\phi) = \mu^2|\phi|^2 + \lambda|\phi|^4 \tag{2.2}$$

This yields a non zero vacuum expectation value (VEV) for $\mu^2 < 0$. The scalar field can be written as $SU(2)$ doublet expanded around the VEV:

$$vev = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{2.3}$$

A Lagrangian using the expansion around the vacuum yields terms that indicate a massive self-interacting scalar boson, the Higgs boson. In addition, the kinetic terms of the scalar field result in mass terms of the gauge boson while the fermion masses arise from the Yukawa couplings with the Higgs field. These Yukawa couplings play an important role in defining the mixing between quark generations and thus help predict the SM $B^0_{s,d} \to \mu^+\mu^-$ branching fraction.

The Higgs boson mass depends on the coupling constant, λ, and the VEV. The VEV can be determined from the gauge boson masses (γ, $W^\pm$, and Z^0) and is approximately 246 GeV/c^2. The coupling constant, however, is unknown and can only be determined from measuring interactions of known particles with the Higgs. On July 4$^{\text{th}}$ 2012 both the ATLAS and CMS collaborations announced

the discovery of a Higgs like boson with a mass of around 125 GeV/c^2. This discovery could mean that all SM particles have been observed experimentally.

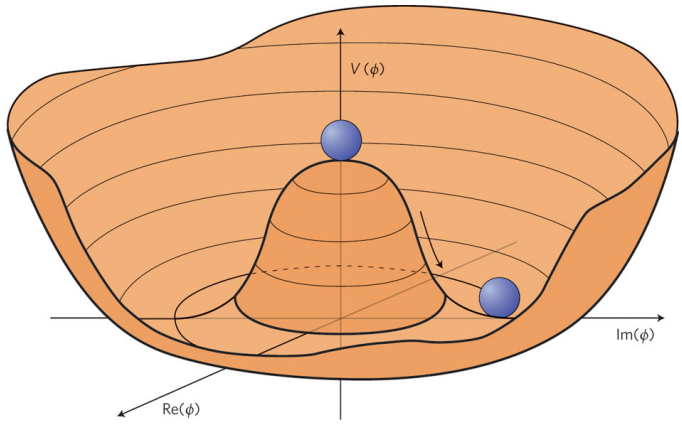

Figure 2.1: The potential of the scalar field. There is a semistable point at ϕ=0 which preserves the gauge symmetry while there is stable minimum that breaks the symmetry.

2.2 Unresolved issues with the standard model

The standard model has been a very successful model of the fundamental constituents of matter and their interactions. However, there are several problems with this model. One of the longest standing problem of the model is the "large hierarchy problem". This hierarchy problem arises from loop corrections to the masses of the fermions and bosons in the SM. For both fermions and gauge bosons the correction to their mass are proportional to the mass of the fermion or gauge boson. These loop diagrams usually are divergent and contain an energy scale cutoff. This cutoff tends to be the Planck scale, $\Lambda_{Pl} = 10^{18}$ GeV/c^2, which is the highest energy levels at which the SM could to still be valid. Beyond this energy scale, quantum gravity begins to have significant effects. Even at these scales, because the correction is proportional to the fermion or gauge bo-

son mass, the correction is greatly limited. However, for the Higgs boson, these corrections are not proportional to a Higgs mass and are of order Λ_{Pl}. This problem questions the entire SM, and many models, some of which are discussed in Sec. 2.6, have proposed methods to resolve this issue.

Besides direct problems in the SM there are also phenomena that the SM has no explanation for. One of which is the presence of dark matter and dark energy. The SM essentially describes ~4% of the universe (Fig. 2.2) while the remaining dark matter and dark energy are not described. There is strong evidence for the presence of dark matter from gravitational lensing, galactic velocity curves, the CMB power spectrum, and the abundance of light elements. Evidence for dark energy come from observation of supernovae and the CMB. Very little is known about dark matter and even less about dark energy. Some extensions of the SM, such as SUSY, propose dark matter candidates but no such candidate has yet been discovered.

In addition the SM also does not explain the matter asymmetry in our universe. Processes described by the SM do not prefer decays of matter over antimatter at the rate seen in the matter dominated universe. The SM also fails to explain the hierarchy in fermion masses. Why is the top quark mass five orders of magnitude larger than the up quark mass?

The SM is an exceptionally successful theory but there are many indications, both theoretically and experimentally, that the SM is only an effective field theory of a much more complete theory. There are many efforts in both experiment and theory to find clues of this more complete theory. The search for $B^0_{s,d} \to \mu^+\mu^-$ is one such experimental effort.

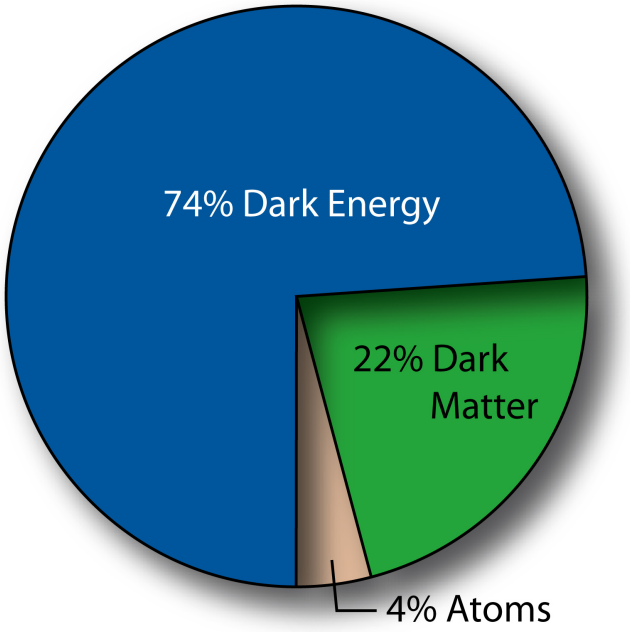

Figure 2.2: Constituents of the universe. The SM only describes small fraction of the universe.

2.3 B Mesons

Quarks can combine into unstable bound states known as mesons consisting of a quark and antiquark. These mesons are created through a process of hadronizations. In hadron colliders $b\bar{b}$ are produced by gluon fusion or light quark and antiquark annihilation. Due to quark confinement these b-quarks start to pull quarks from the vacuum to form hadrons. This process is modeled by several empirical models of which the Lund string model [2] is the most used model. In this model the quark-antiquark pair are linked by a string with a constant energy density as a function of length. This yields an increasing energy as a function of quark-antiquark ($q\bar{q}$) separation distance. When the energy reaches levels equivalent to the mass of new $q\bar{q}$ masses a new pair of quarks is produced.

The Lund string fragmentation model predicts a higher rate of light quark production during hadronization due to the energy in the string reaching levels of twice the light quark masses more easily than for heavier quarks. Additionally, the production of baryons is less likely than mesons for similar reasons. This is in agreement with experimental observations and plays a role in the $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ searches. The fragmentation ratio between a b-quark fragmenting with a s-quark and a light up or down quark need to be taken into account when measuring the branching fractions relative to a normalization decay. If the normalization decay is a decay of a B-meson with the same quark content no such fragmentation ratio enters the branching fraction estimation. However, in this analysis the $B^+ \to J/\psi K^+$ normalization mode for both the $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ searches are used, resulting in the need for an accurate estimation of the fragmentation ratio. The details of the branching fraction estimation in the $B_{s,d}^0 \to \mu^+\mu^-$ search using a normalization mode are discussed in detail in Sec. 4.3.

There are many possible $q\overline{q}$ combinations and this document will focus on B-mesons. B-mesons are a combinations of $\bar{b}$ quark and a quark with a different flavor. They are characterized by their relatively large mass (~5GeV [3]) and their long lifetimes (~440 μ m). We will limit ourselves to particular decays of the B^+, B_s^0, B_d^0, and their charge conjugates. The B^+ consists of a $\bar{b}$ and u quark, while the B_s^0 and B_d^0 contain s and d quarks along with a $\bar{b}$ quark, respectively.

2.4 The Cabibbo-Kobayashi-Maskawa Matrix

The rate at which quark flavor changes occur in the SM are summarized in a matrix known as the Cabibbo-Kobayashi-Maskawa matrix. This matrix arises naturally from the Yukawa interactions with the Higgs condensate after transformation from a weak eigenstate to a quark mass eigenstate. The CKM matrix is a 3×3 unitary complex matrix that diagonalizes mass matrices in the SM Lagrangian and is given in Equation 2.4. Each element in Equation 2.4 represents a coupling for a charged current interaction as depicted in Fig. 2.3. The approximate measured values are also shown in Equation 2.4 [3]. The diagonal terms represent mixing within one quark generation and are ~1. Interaction involving mixing of different generations are, however, significantly smaller. The smallest off diagonal interaction is from u to b and is more than 250 times smaller than the diagonal terms.

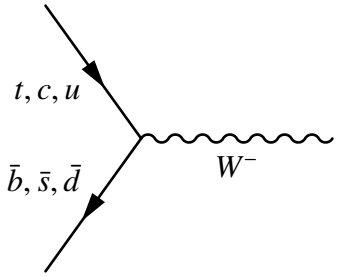

Figure 2.3: Charged current interaction. The coupling constant of the vertex, V_{CKM}^{ij}, is one of the elements of the CKM matrix.

There are several parametrization of the matrix of which the "Standard" and Wolfenstein parametrization are the most used [3]. These parametrization have four independent parameters. A complex $N \times N$ matrix has $2N^2$ real parameters, where N is the number of generations. The CKM matrix is unitary which reduces the number of parameters down to N^2. The SM Lagrangian is invari-

ant under global phase transformation which allows for the absorption of two phases into the quark fields, one for the up type and two for the down type. This means that $2N - 1$ phases can be removed from the initial N^2 phases. This yields a total of $(N-1)^2$ free parameters, which comes to four parameters in the SM.

$$V_{CKM} \equiv \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \sim \begin{pmatrix} 0.974 & 0.225 & 0.003 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix} \tag{2.4}$$

2.4.1 Flavor changing neutral currents

Flavor changing neutral current processes involve a change in flavor without a change in charge. A typical electroweak interaction either contains a change in charge or flavor. The $B^0_{s,d} \to \mu^+\mu^-$ decays are FCNC decays and occurs in the SM through loop processes. FCNC's are suppressed by the GIM-mechanism in the SM. The suppression arises from both the higher powers of the weak coupling constant (g_W), the large mass of the $W^\pm$ and Z^0 bosons, as well as subtraction of different higher order contribution. The high powers of g_W comes from the many vertices involved in the higher order diagrams. The propagator for massive bosons is given in the SM by Equation 2.5. At low energies (such as with most meson decays) where $q << m_W^2$ this can be heavily suppressed by m_W^2.

$$\frac{i(g_{\alpha\mu} + \frac{q_\alpha q_\mu}{m_W^2})}{q^2 + m_W^2} \sim \frac{ig_{\alpha\mu}}{m_W^2} \tag{2.5}$$

An example of the suppression due to addition of different contribution can

be given with the $K_L^0 \to \mu^+\mu^-$ decay. When the box diagram is considered for this example the quark in the box can be either the u, c, or t quarks. The t can be ignored due to small CKM matrix element. This leaves the u and c contributions. If these two quarks had the same mass the two contributions would cancel out since the fermion propagator is a function of fermion mass. However, the u an c quark do have different masses resulting in a non zero branching fraction. This type of calculation will be examined in more detail in the next section that discusses the SM $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ prediction in detail.

2.5 Standard model $B_{s,d}^0 \to \mu^+\mu^-$

In the SM $B_{s,d}^0 \to \mu^+\mu^-$ decays can only occur through higher order box and penguin diagrams shown in Fig. 2.4. There are no diagrams involving photon penguins ($X \to \gamma \to \mu^+\mu^-$) because the $B_{s,d}^0$ mesons are scalars with spin 0 which must be conserved. This can not happen with photons since photons can only have spin ±1. Higgs penguins are not shown in Fig. 2.4 due to their small contribution caused by the small muon masses relative to the Higgs vacuum expectation value (VEV, or v). The couplings involving the Z^0 penguin is of order $g_W^2 \sim 0.5$ while the Higgs penguin's couplings are of order $\frac{m_\mu}{v} \cdot \frac{m_{t,W}}{v} \sim O(10^{-4})$ while having similar propagators.

The branching fraction for any $B_s \to l^+l^-$ can be calculated using Equation 2.6, where B indicates either the B_s^0 or B_d^0 meson, and τ_B is the total meson lifetime. The matrix element, $\mathcal{M}$, depends on the Feynman diagram. Due to the low mass of the $B_{s,d}^0$ mesons relative to the mass of the Z^0 and $W^\pm$ bosons the matrix elements can be evaluated with an effective Lagrangian using operator

product expansion (OPE) [4]. The OPE method writes the Lagrangian as a series of effective vertices each of which depend on Wilson coefficients and local operators, that is $\mathcal{L} = \frac{G_F}{\sqrt{2}} \sum_i V^i_{CKM} C_i Q_i$, where G_F is the Fermi constant and $\frac{G_F}{\sqrt{2}} = \frac{g_W^2}{8m_W^2}$. The low energy coupling constants are represented by local operators (Q_i) while the short length scale effects are represented by Wilson coefficients (C_i). Thus the Wilson coefficients will depend on the mass of the top quark, $W^\pm$ and Z^0. For the $B^0_{s,d} \to \mu^+\mu^-$ decays the local operators will consist of a leptonic $V-A$ interactions as well as a quark $V-A$ interaction. The quark $V-A$ interaction will contain QCD corrections and is summarized as the $B^0_{s,d}$ decay function $F_{B^0_s}$. The resulting expression for the branching fraction $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ is given in Equation 2.7 [5]. In this equation $\frac{\alpha}{4\pi \sin^2\theta_W} = \frac{g_W}{16\pi}$, where α is the fine structure constant and is related to the elementary electric charge e by $\alpha = \frac{e^2}{4\pi}$. The weak mixing angle, θ_W, relates electroweak eigenstates and mass eigenstates after electroweak spontaneous symmetry breaking. The function $Y(m_t^2/m_W^2)$ correspond to the functional form of the Wilson coefficients. A more detailed description of these functions can be found in Ref. [4].

$$\mathcal{B}(B^0_{s,d} \to l^+l^-) = \frac{\tau_B}{m_B} \cdot \frac{1}{16\pi} \cdot \sqrt{1 - (\frac{2m_l}{m_B})^2} \cdot |\mathcal{M}|^2 \tag{2.6}$$

$$\mathcal{B}(B^0_s \to l^+l^-) = \frac{G_F^2}{\pi} \left(\frac{\alpha}{4\pi \sin^2\theta_W} \right)^2 \tau_{B^0_s} m_{B^0_s} m_l^2 F^2_{B^0_s} \sqrt{1 - (\frac{2m_l}{m_{B^0_s}})^2} |V^*_{tb} V_{ts}|^2 Y^2(\frac{m_t^2}{m_W^2}) \tag{2.7}$$

From Equation 2.7 one can see that the branching fraction has a m_l^2 dependence on the lepton mass. A greater lepton mass would thus result in a larger branching fraction. This would make decays into two τ's the most probably decay. However τ decays are much more challenging to detect experimentally,

leaving muons as the most experimentally attractive channel. Both the lepton and B_s^0 mass terms come from the $V - A$ interaction which yield a squared four momentum.

The final evaluation of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is given in Equation 2.8. Here $\bar{m}_t(m_t)$ represents the running top mass in the $\overline{MS}$ renormalization scheme. The greatest uncertainty in the theoretical determination of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ comes from the decay functions F_B due to low energy QCD corrections. The branching fraction for B_d^0 can be determined by replacing s with d.

$$\mathcal{B}(B_s^0 \to l^+l^-) = 3.5 \times 10^{-9} \left[\frac{\tau(B_s^0)}{1.6\text{ ps}}\right]\left[\frac{F_{B_s^0}}{210\text{ MeV}}\right]^2\left[\frac{|V_{ts}|}{0.040}\right]^2\left[\frac{\overline{m}_t(m_t)}{170\text{ GeV}}\right]^{3.12} \tag{2.8}$$

A numerical evaluation of the branching fractions yields an expected $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.2 \pm 0.2) \times 10^{-9}$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-) = (1.0 \pm 0.1) \times 10^{-10}$ [6]. The branching fractions for $B_s^0 \to \tau^+\tau^-$ and $B_s^0 \to e^+e^-$ are $O(10^{-6})$ and $O(10^{-13})$, respectively. If the decay functions for B_d^0 and B_s^0 are calculated accurately the ratio of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ should yield a measurement of $\frac{|V_{ts}|}{|V_{td}|}$, one of the input parameters of the SM, with a smaller uncertainty than the two branching fraction measurements separately.

2.6 $B_{s,d}^0 \to \mu^+\mu^-$ In new physics models

There are many NP models that predict potentials enhancements of the branching fraction $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. A summary of the maximal enhancements of several groups of models are shown in Table 2.2 [7]. The fundamentals of all these mod-

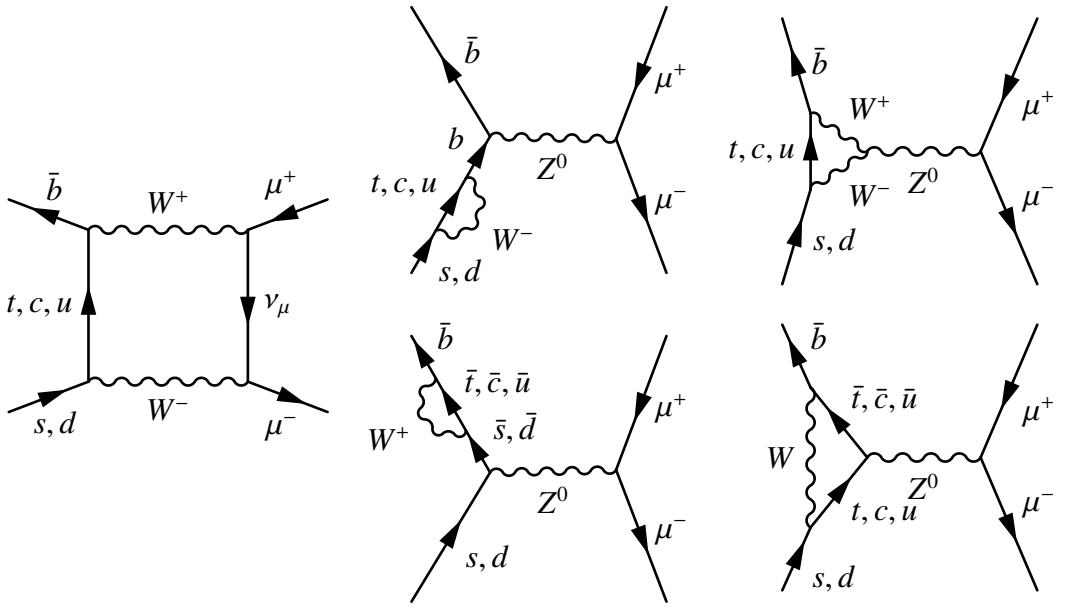

Figure 2.4: Leading order SM Feynman diagrams for $B^0_{s,d} \to \mu^+\mu^-$ decays

Table 2.2: Maximal enhancements for $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$ from different theoretical NP models. AC=Agashe, Carone; RVV=Ross, Velasco-Sevilla, Vives

Model	$\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$ Enhancement
MFV	1000%
CMFV	20%
LHT	30%
RS	10%
AC	1000%
RVV	1000%

els and the cause of the $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$ enhancement will be briefly discussed below.

Minimal flavor violation (MFV) models are a class of models for which all flavor violation is caused by the Yukawa couplings. There are many models

to which the MFV principle can be applied, such as SUSY models. There is vast array of SUSY models but all have in common that they introduce superpartners to the SM particles. The superpartners of fermions are bosons and the partners of bosons are fermions. These new particles are introduced to cancel divergences caused by loop correction to the Higgs mass. A particular example of such a model is the minimally supersymmetric standard model (MSSM). These models contains two Higgs doublets which cause significant enhancement due to new contributions to $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$ involving new neutral and charged scalar particles. The presence of two Higgs doublets causes enhancement in models other than SUSY. For most two Higgs doublet models (2HDM) the presence of the new Higgs doublet can cause enhanced contributions to $B^0_{s,d} \to \mu^+\mu^-$ decays. In these additional channels a charged W, seen in some of the diagrams in Fig. 2.4, is replaced with a charged Higgs boson. This causes very significant contribution to $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ in both 2HDM and MSSM models with MFV.

In constrained minimal flavor violation (CMFV) models all the flavor violation comes from the SM CKM matrix and only the SM interactions are relevant below the weak scales. Since $B^0_{s,d}$ meson decays are significantly below the weak scale the decays are essentially governed by SM processes. The main NP contribution to $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ originate from the $Y(m_t^2/m_W^2)$ functions, which summarize the higher energy scale effects on the branching fraction. These functions are model dependent but the maximal enhancement to $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$ is 20% for CMFV models.

A model that attempts to solve the Higgs loop correction problem without the addition of SM superpartners is the littlest Higgs with t-parity (LHT) model.

In this model, the Higgs is a composite pseudo-Goldstone boson of a $SO(5)$ broken symmetry, similar to the pion being a pseudo-Goldstone of flavor symmetry in the SM strong sector [8]. Goldstone bosons are massless when an exact symmetry is broken while pseudo-Goldstone bosons have a small mass when an approximate symmetry is broken. This breaking of a new symmetry keeps the Higgs mass light. A new set of particles with the same spin as their SM counterparts are introduced by this symmetry breaking. To be in agreement with precision electroweak measurement a new discrete symmetry is introduced: t-parity. All, but the new top quark counterpart, the T quark, have odd t-parity while the SM particles and the other new particles have even t-parity. This allows for cancellation of the Higgs mass correction divergences through the contribution of the new T quark. The advantage of this model compared to SUSY models is the relatively small number of parameters. The main contribution to $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$ comes from diagrams now including the heavier T quark which can enhance the branching fraction by up to 30%.

Only a small increase in the branching fraction $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$ is predicted by a model known as the Randall-Sundram model with custodial protections (RSc). This model introduces an new warped spacial dimension in an attempt to explain the hierarchy problem both between the Planck and electroweak scale as well as between the fermions masses and mixing angles. The first versions of these models predicted significant enhancements of FCNC's through Kaluza-Klein gluon excitations. These excitations are a result of the introduction of a new spacial dimensions. To prevent conflicts with current experimental data the gauge group is enlarged, as with the LHT model. This introduces new heavy gauge bosons that can add a 10% and 20% enhancement to $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ and $\mathcal{B}(B^0_d \to \mu^+\mu^-)$, respectively. The small enhancement is due to the large masses

of the new gauge bosons which show up in the denominator of the Y functions.

There are many SUSY models with different constraints, with some models not requiring MFV. Without MFV, SUSY models can predict significant increases in the branching fractions of $B^0_{s,d} \to \mu^+\mu^-$. Particular examples of these models are the supersymmetric flavor models which can be divided into abelian and non-abelian models. The abelian models are known as Agashe-Carone (AC) models while the non-abelian models are the Ross-Velasco-Vives (RVV) based models. The abelian models are based on a $U(1)$ flavor symmetry, while the non-abelian has a $SU(3)$ flavor symmetry. These symmetries are introduced to explain the sources of the flavor structure in the SM and to extend these to SUSY. The SM Yukawa structure would then be an effective result of the flavor symmetry breaking, similar to the mechanism for $SU(2) \times U(1)$ symmetry breaking. These models both predict significant enhancement to the branching fraction $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$. The AC model predictions is based on current measurements of other flavor physics observables.

From the discussion above one can see that there are many NP models that predict an increased $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$. There are several other flavor physics observables that together with the branching fraction $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$ measurements could yield an indication to which models are most probable. These observables and their correlations to $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$ are not discussed here but a comprehensive account is given in Ref. [7]. Experimental measurements of these parameters are imperative to be able to find the true flavor structure of nature. The following sections will discuss the search for $B^0_{s,d} \to \mu^+\mu^-$ at the Collider Detector at Fermilab.

2.7 Experimental status

There are experiments at two hadronic colliders, the Large Hadron Collider (LHC) and the Tevatron, conducting searches for the $B^0_{s,d} \to \mu^+\mu^-$ decays. At the LHC, the general purpose Compact Muon Solenoid (CMS) and the heavy flavor specific LHCb experiments have analysis efforts ongoing. The Tevatron is used by two experiments, CDF and D0, for the $B^0_{s,d} \to \mu^+\mu^-$ search. A short description of all the experiments and their results (Fig. 2.5) as of the Summer of 2011 are discussed below.

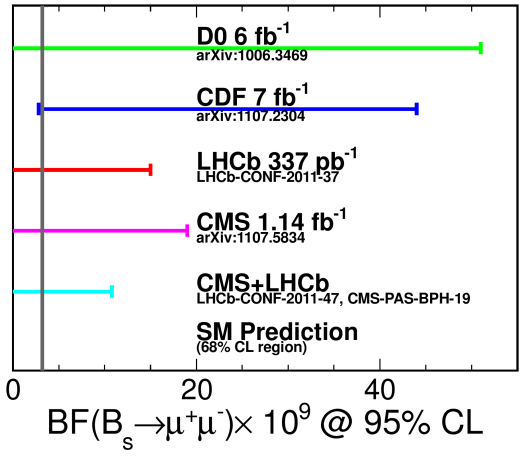

Figure 2.5: Experimental results as of the August of 2011.

2.7.1 CDF and D0

CDF and D0 are both general purpose experiments that have done several iterations of $B^0_{s,d} \to \mu^+\mu^-$ searches. Both experiments search for $B^0_{s,d} \to \mu^+\mu^-$ decays in $p\overline{p}$ collisions with $\sqrt{s} = 1.96$ TeV. Similar analysis methods were used in both experiments involving multivariate discriminants and dimuon triggers.

CDF has a dimuon mass resolution of ~25 MeV/c^2 allowing for separation of $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ while D0 can only search for the combined contribution of $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$. D0 set a limit of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 5.1 \times 10^{-8}$ at 95% confidence level (C.L.) using 6 fb^{-1} of data [9]. CDF using 7 fb^{-1} of data measured an excess in the $B_s^0 \to \mu^+\mu^-$ search while the data were consistent with a background-only hypothesis in the $B_d^0 \to \mu^+\mu^-$ search resulting in a limit of $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 6.0\times10^{-9}$ at 95% C.L. For the $B_s^0 \to \mu^+\mu^-$ search p-values of 0.27% and 1.92% were estimated using a background-only and a SM+background hypothesis, respectively. A branching fraction of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 1.8^{+1.1}_{-0.9} \times 10^{-8}$ was also determined with a fit to the data. Using the same method for the branching fraction estimate, a 95% C.L. bound was set at $2.8 \times 10^{-9} < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 4.4 \times 10^{-8}$ [10].

2.7.2 CMS and LHCb

CMS and LHCb search for $B_{s,d}^0 \to \mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 7$ TeV. CMS is a general purpose detector with a geometry similar to CDF and D0 while LHCb is a purpose built B-meson detector with a forward geometry. The first results from both CMS and LHCb were released during the Summer of 2011. Using a cut based analysis (no multivariate discriminant as in CDF and D0) and ~1 fb^{-1} of data CMS set an upper bound of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 1.9 \times 10^{-8}$ at 95% C.L. [11]. For the $B_d^0 \to \mu^+\mu^-$ search they also set an upper limit of $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 4.6 \times 10^{-9}$ at 95% C.L. LHCb used a multivariate discriminant and ~300 pb^{-1} to set an upper bound of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 1.5 \times 10^{-8}$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 5.2 \times 10^{-9}$ at 95% C.L. [12]. The CMS and LHCb results for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ were combined to yield an upper bound of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 1.1\times10^{-8}$ at 95% C.L. [13].

CHAPTER 3

EXPERIMENTAL APPARATUS

This chapter discusses the experimental apparatus used in the search for $B^0_{s,d} \to \mu^+\mu^-$. The apparatus consists of a series of accelerators that brings protons and antiprotons to an energy of 980 GeV each. These protons and antiprotons are then brought to collide while the solenoidal collider detectors detect the products of the decays.

3.1 Fermi National Accelerator Laboratory

The Fermi National Accelerator Laboratory (Fig. 3.1) was founded in 1967 and is located between Warrenville, IL and Batavia, IL, about 35 miles west of Chicago, IL. The laboratory was originally named the National Accelerator Laboratory but was renamed in 1974 to honor the Nobel prize winning physicist, Enrico Fermi. Today, Fermilab is a hub for a variety of experiments such as searches in the dark matter sector (CDMS, COUPP, DAMIC), neutrino experiments (MiniBooNE, MINERνA, MINOS, NOνA), astrophysics (Sloan Digital Sky Survey, DES), and high energy collider physics (CMS, CDF, D0). The laboratory is home to the first TeV scale particle accelerator, the Tevatron, which accelerates protons and antiprotons to a center-of-mass energy of 1.96 TeV. The experiments at Fermilab have greatly aided in the understanding of fundamental particle physics with the discovery of the bottom quark in 1977 [14], and the top quark in 1995 [15]. The first evidence of the tau during the Summer of 2000 [16] and the discovery of the Σ_b baryon [17] can also be attributed to Fermilab. Besides discoveries, Fermilab has also significantly contributed to precision measurements

such as the search for rare decays and the $W^{\pm}$ and Z^0 mass.

Figure 3.1: Areal view of the Fermi National Accelerator Laboratory.

3.2 Accelerators

The search for $B^0_{s,d} \to \mu^+\mu^-$ studies B-mesons produced from $p\overline{p}$ collisions at an energy of 1.96 TeV. A series of accelerators (Fig. 3.2) is used to produce the protons and antiprotons and to accelerate them to this high energy. The process starts at the pre-accelerator and the Linac, that produce and are the first to accelerate negative hydrogen ions (H^-), and then moves on to the booster, turning the H^- ions into proton and accelerating them further. The proton beam continues on to acceleration by the main injector while a subset of protons are sent to the antiproton source for $\overline{p}$ production. Finally, once both the protons and antiprotons are ready, they are accelerated to 980 GeV each in the Tevatron. The following sections will detail this process.

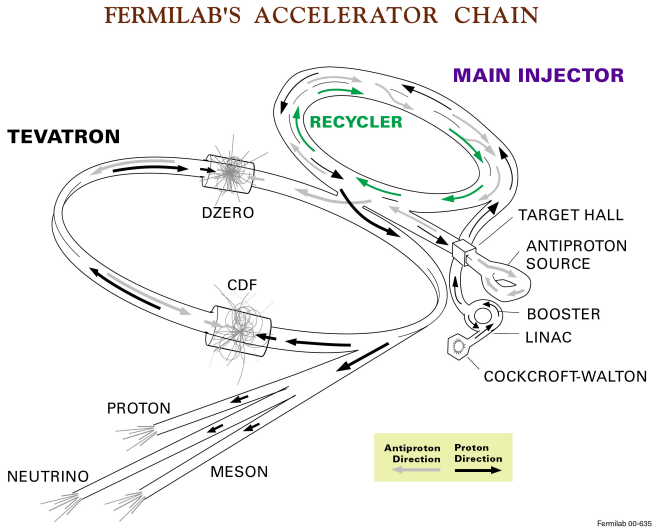

Figure 3.2: Accelerator Chain at Fermilab. The process starts at the Cockcroft-Waltron amplifier and ends at the Tevatron. The proton, neutrino, and meson lines go to the switchyard which is no longer used for proton, neutrino and meson experiments.

3.2.1 Pre-accelerator and Linac

All protons and antiprotons come from a small tank with 6.88 liters of hydrogen compressed to 5×10^{25} atmosphere. The hydrogen atoms are stripped of their electrons with an electric field and are sent to a cathode where they acquire two electrons and become H^- ions. The H^- ions are initially accelerated to 750 KeV using a Cockroft-Walton generator [18]. The Cockroft-Walton generator (Fig. 3.3) is a solid state voltage multiplier that amplifies a low AC voltage to a high DC voltage. The H^- are accelerated to 750 KeV every 66 ms and sent to the linear acceleration (Linac) for further acceleration. The energy of the hydrogen ions is increased from 750 KeV to 400 MeV in the Linac. The Linac accelerates the ion beam at the same frequency as the pre-accelerator using radio frequency (RF) cavities. Two sections accelerate the beam: the drift tube section, accelerating the beam to 116 MeV with an RF of 201 MHz, and the side-coupled cavity

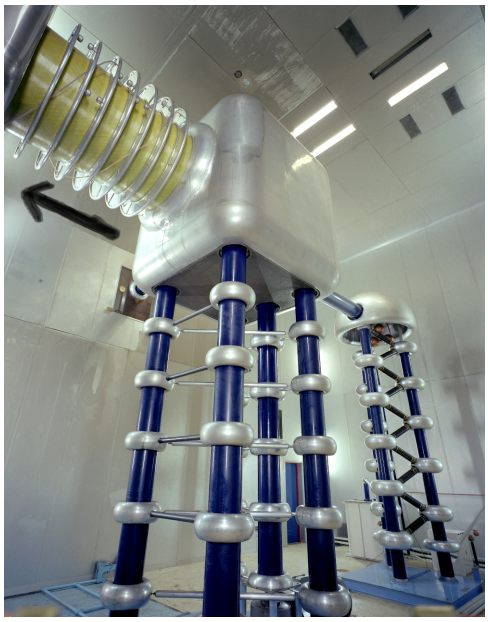

Figure 3.3: Cockroft-Walton generator. The arrow indicates the direction of acceleration of the H^- ions. The column to the right generates a –750 KV potential to accelerate the ions.

section that brings the beam to the final 400 MeV with an RF of 805 MHz. The difference in frequency is due to change in H^- ion velocity. Once the ions have reached 400 MeV they are transferred to the booster.

3.2.2 Booster

The H^- ions are brought to the booster after being accelerated by the Linac. The booster is a circular accelerator with a radius of 75 m. The ions are first stripped of their two electrons by a thin layer of carbon. The bare protons are accelerated using 17 RF cavities to an energy of 8 GeV. The resulting proton beam is sent to the main injector (MI). Alternatively, the beam can also be deflected just before it reaches the MI for the production of a neutrino beam that is used in the MiniBooNE (a booster neutrino experiment).

3.2.3 Main injector

The main injector takes the 8 GeV protons from the booster and accelerates them for injection into the Tevatron, for antiproton production, or for use in the neutrino experiments at Fermilab. The MI is a circular accelerator that, similar to the booster, uses RF cavities to accelerate the beam while using dipole magnets to keep the beam traveling in a circular path. The circumference of the MI, at ~2 miles, is about half that of the Tevatron.

Protons that are meant for use in the Tevatron are accelerated to 150 GeV. The MI also coalesces the beam into a superbunch and continues to inject these into the Tevatron until there are 36 of these proton bunches.

Some of the protons are accelerated to 120 GeV and sent to the antiproton source (discussed in the next section). When the antiproton source returns 8 GeV antiprotons, the MI accelerates them to 150 GeV and coalesced them into superbunches. These bunches are then injected into the Tevatron until there are 36 antiproton bunches.

Another subset of protons is accelerated to 120 GeV and sent to produce neutrino beams (NuMI). These neutrino beams are produced for the MINOS near and far detectors, with the far detector located in Soudan, MN, about 450 miles away. Another near neutrino experiment that uses the neutrino beam from the MI protons is MINERνA. In the near future (~2013) the NOνA experiment will also make use of the NuMI beam for both its near and far detectors.

3.2.4 Antiproton source and recycler

The antiproton source, also known as Pbar, produces antiprotons using the incoming 120 GeV protons from the MI. The Pbar is made up of a nickel source, a debuncher, and an accumulator. The protons are made to hit a nickel target producing a spray of particles. The antiprotons are selected using a series of magnetic fields which essentially selects particles within a certain mass range. The antiproton yield is relatively low, requiring 50,000 proton collisions for approximately one selected antiproton.

After the antiprotons have been selected they have a large energy spread. To reduce the spread in energy they are sent to the debuncher which is a triangular cooling ring with a radius of about 90 m. The debuncher transfers the spread in energy to a spread in time causing a debunched antiproton beam with a narrow energy spectrum at around 8 GeV.

Once the debuncher has cooled the antiproton beam the antiprotons are sent to the accumulator. The accumulator is a storage ring with a 75 m radius that collects the antiprotons until they are injected to the MI or recycler.

The recycler is a permanent magnet storage ring that stores the 8 GeV antiprotons. The permanent magnet configurations is simple and efficient by avoiding the need for cooling that is required for electromagnet storage rings. The recycler has the same radius as the MI and is actually mounted in the same tunnel, a few feet above the MI. The antiprotons are cooled and stored in the recycler for several hours until the Tevatron is ready for collider physics.

3.2.5 Tevatron

The Tevatron is a ~4 mile circumference synchrotron accelerator that accelerates protons and antiprotons from an energy of 150 GeV to 980 GeV. This is done using 8 RF cavities all located in one section of the ring. Due to the high velocities of the beams superconducting dipole magnets with magnetic fields of about 4 T are required to keep the protons and antiprotons in a circular orbit. The magnets have a constant strength because the difference in speed for 150 and 980 GeV is negligibly small. The superconducting dipole magnets require significant amounts of cooling from liquid Helium to maintain their critical temperature. In addition to the dipole magnets the Tevatron utilizes quadrupole magnets to focus the beam.

The Tevatron has 6 straight sections, A through F. The first straight, A0, connects to the switchyard and is used to dump the beam while the B0 and D0 straights are used for collider detection. The switchyard previously held meson, proton, and neutrino experiments but is no longer used. The collider detector at Fermilab (CDF) is located at the B0 straight while the D0 collider detector is located at the D0 straight. The F0 straight is where the MI connects with the Tevatron and where the accelerating RF cavities are located. The remaining straights are not currently used for a particular purpose.

The proton and antiproton beams traverse a helical orbit and pass through each other at the B0 and D0 straights. This occurs for hours at a time, known as a store, until the collision rate becomes too low at which point the store is ended and preparations are made for a new store.

3.2.6 Beam structure

The protons and antiprotons are each divided into three 12-bunch trains (Fig. 3.4) with bunches within a train separated by 396 ns. Between the trains there are 2.6 μs gaps without any proton or antiproton bunches. These gaps are known as abort gaps and are meant to allow kicker magnets enough time to ramp up to deflect the protons and antiprotons during a beam abort. The search for $B^0_{s,d} \rightarrow \mu^+\mu^-$ studies the products of these collisions at the B0 point using the CDF experiments.

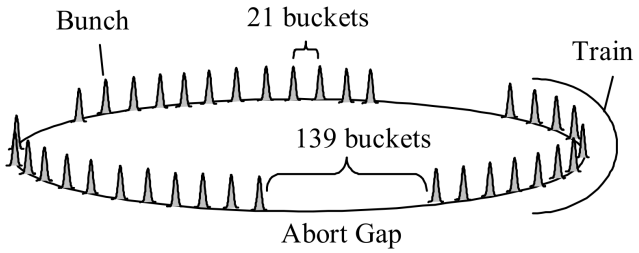

Figure 3.4: Beam structure at the Tevatron. There are three trains of 12 bunches each, separated by abort gaps. A bucket corresponds to 18.8 ns.

3.3 The CDF II detector

The CDF II detector is a general-purpose detector [19, 20, 21] with cylindrical symmetry designed to detect products of $p\overline{p}$ collisions at a center-of-mass energy of $\sqrt{s}$ = 1.96 TeV.

The detector has a cylindrical layout consisting of different sub-detector regions as can be seen in Fig. 3.5. The innermost region of the detector contains the tracking system which is comprised of a silicon vertex detector surrounded

by the central outer tracker (COT). Outside of the tracking system is a solenoidal magnet that produces a 1.4T magnetic field along the direction of the incoming proton beam. Electromagnetic and hadronic calorimeters are positioned around the magnet. The outermost detectors are muon detectors that consist of drift chambers and scintillators. Muons typically pass through the calorimeters, unlike hadrons, electrons and photons, reaching the outermost detectors.

A cylindrical coordinate system is used to describe particle trajectories. The z axis is defined as the direction of the incoming protons. Besides the familiar azimuthal angle, ϕ, radius relative to the beamline, r, and polar angle, θ, we also define a pseudorapidity: $\eta = -\ln(\tan(\theta/2))$. We define the transverse momentum, p_T, which represents the component of a particle momentum in the plane perpendicular to the beam axis, $p_T = p\sin\theta$.

The most important sub-detectors for this analysis are discussed in detail below and include the tracking system and the muon system. A more detailed description of the remaining systems, calorimeters and trigger systems can be found in Ref. [19].

3.3.1 Tracking system

The tracking system consists of a silicon vertex detector, a multi-wire open-cell drift chamber, and a solenoidal superconducting magnet. The innermost tracking system consists of three sub-systems: layer 00 (L00) [22], the silicon vertex detector (SVXII) [23], and the intermediate silicon layers (ISL) [24]. The transverse view of the silicon detectors is shown in Fig. 3.6. L00 is the closest to the interaction point and consists of 12 one-sided Rohacell foam assemblies (known

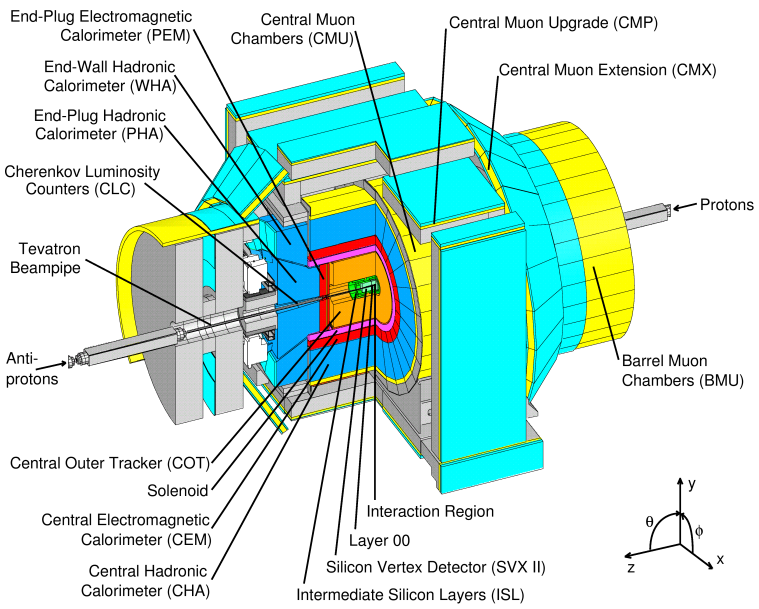

Figure 3.5: Cutaway isometric view of the CDF II detector

as ladders) supported by a carbon fiber structure. Six of these ladders are at a radius of 1.4 cm while the remaining six are at 1.6 cm. Outside the L00 is the main vertexing detector, SVXII, with five double-sided layers at radii ranging from 2.4 to 10.6 cm. Each of these layer also consists of 12 ladders. Figure 3.7 shows the L00 sub-detector together with the first two layers of the SVXII detector in the plane transverse to the beamline ($r - \phi$ plane). Three of these layers, layer 0, 1, and 3, contain $r - \phi$ modules, with strips aligned with the beam axis, on one side while having modules with strips rotated at a 90° angle relative to beam axis on the other (90° stereo layers). The remaining layers combine $r - \phi$ modules on one side with modules with strips rotated at a small angle on the reverse side. The SVXII hit resolution is 11 μm while the impact parameter resolution for tracks with a $p_T > 2.0$ GeV/c is about 40 μm of which 35 μm is due to the size of the interaction point. The impact parameter resolution improves slightly for tracks with $p_T > 2.0$ GeV/c when L00 hits are included as seen in Fig. 3.8. In addition to the five layers in the SVXII, two double-sided small-angle stereo

layers at radii 20 and 28 cm, and one layer at a radius of 22 cm provide coverage for $1 < |\eta| < 2$ and $|\eta| < 1$, respectively. These layers are collectively known as the ISL and aid in COT/SVXII hit matching. The forward two layers extend the silicon pseudorapidity coverage beyond the COT coverage into the forward section of the detector. Together these silicon detectors provide important hit information needed for secondary vertex reconstruction and pseudorapidity coverage up to $|\eta| < 2.0$.

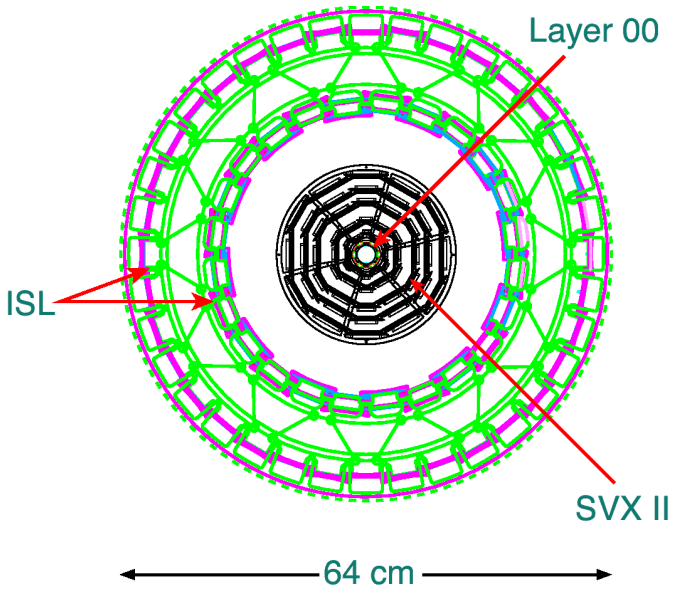

Figure 3.6: View of the silicon detectors in the $r - \phi$ plane.

Outside the silicon subsystems is the COT [25], an open-cell multi-wire drift chamber divided into 8 concentric superlayers (Fig. 3.9). The COT spans 310 cm in the z direction and from 43 to 132 cm in radius, allowing for tracking coverage for $|\eta| < 1$. The open configuration allows the ionizing gas, a 50:50 mixture of Argon and Ethane with a small admixture of isopropyl alcohol, to be shared amongst all cells. The superlayers consist of alternating axial and two-degree stereo layers. The axial layers have wires that are aligned with the beam axis while the stereo layers are rotated by a small angle. This configuration al-

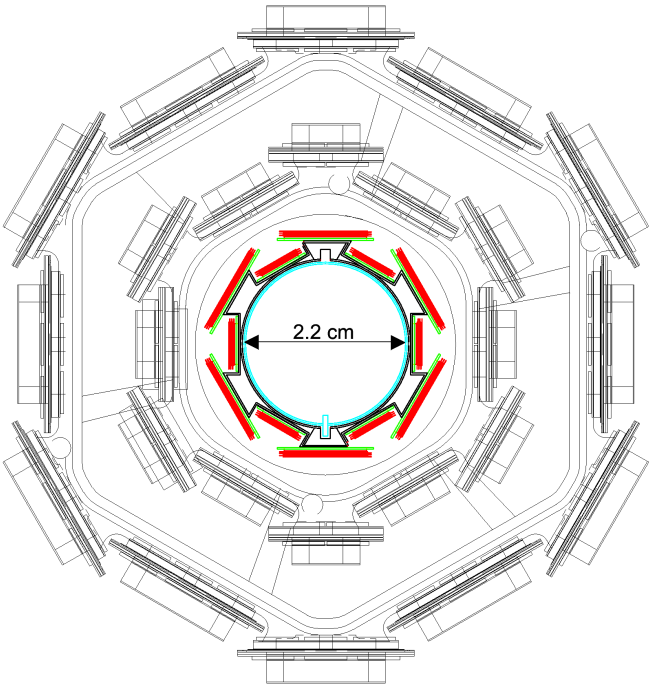

Figure 3.7: View of the L00 subsystem and two layers of the SVXII detector in the $r - \phi$ plane.

lows for a full 3D track parameter determination. The superlayers themselves are divided in ϕ into supercells that each contain 12 sense wires with a potential wire on each side (Fig. 3.10). In total the COT contains over 30,000 sense wires. The potential wires in conjunction with field panels, made of gold plated Mylar, produce a 1.9 kV/cm electric drift field. The resulting drift time is ~50 μm ns^{-1}, leading to a maximal drift time of 180 ns, which is compatible to the beam crossing interval of 396 ns. The COT also measures the ionization per unit path length, dE/dx, for particle identification. The dE/dx of a particle depends on its β, where $\beta = v/c$. For this analysis the dE/dx information is mainly used to help reject kaons.

Surrounding the COT is a superconducting solenoidal magnet producing a 1.4 T magnetic field parallel to the beam axis. The magnetic field is uniform

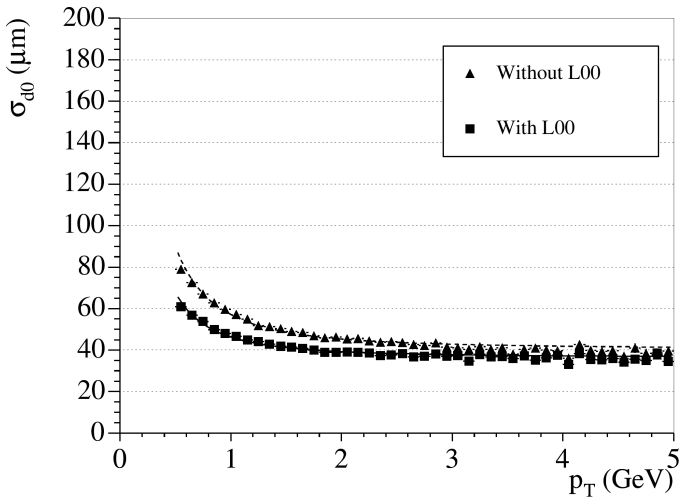

Figure 3.8: Impact parameter resolution as a function of p_T for SVXII alone and SVXII+L00.

to 0.1% within a volume for $r < 150$ cm and $|z| < 150$ cm. The transverse momentum (in GeV/c) is then determined by $p_T = 0.3qBr_c$, where q is the total charge of the particle in terms of the electron charge, B (in T) is the magnetic field strength, and r_c (in m) the radius of curvature of the reconstructed helix trajectory of the particle. The transverse momentum resolution of the COT is determined by comparing the curvature of inward and outward going tracks of cosmic ray events. The resulting resolution is $\sigma_{p_T}/p_T^2 \approx 0.15\%$ (GeV/c)$^{-1}$ [25]. The absolute momentum scale is determined using J/ψ, Υ, and Z^0 resonances, where the resonances decay into two muons [26].

In summary, the COT measures the particle trajectory inside the magnetic field of the solenoid and thus allows for momentum and charge determination. The silicon subsystems further refine the momentum measurement and provide precision vertex information for decays involving multiple charged particles.

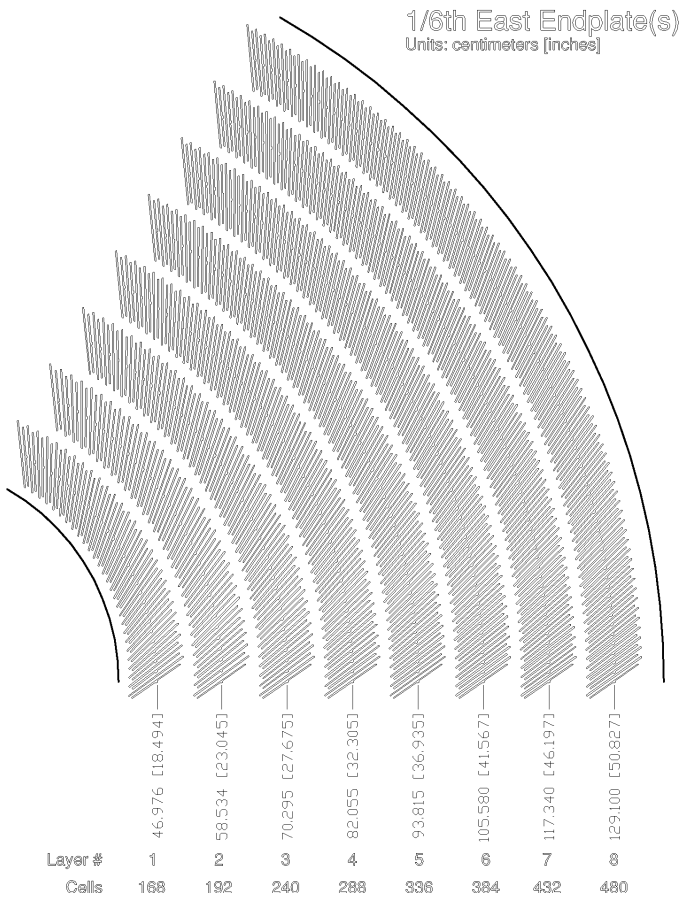

Figure 3.9: Quadrant of the COT showing the 8 superlayers.

3.3.2 Calorimeters

The CDF calorimeters are placed outside the solenoidal magnet. This placement is chosen because the CDF calorimeters use photomultipler tubes (PMT's) whose operation is problematic in magnetic fields. The calorimeters are sampling calorimeters, with absorbers interspersed between scintillators. The calorimeters can be divided into two sub-detectors known as the electromagnetic calorimeter and hadronic calorimeter which together measure electromagnetic and hadronic energy deposits. Both the calorimeters are set up in a projective tower configuration pointing towards the nominal interaction point. The combined pseudo-

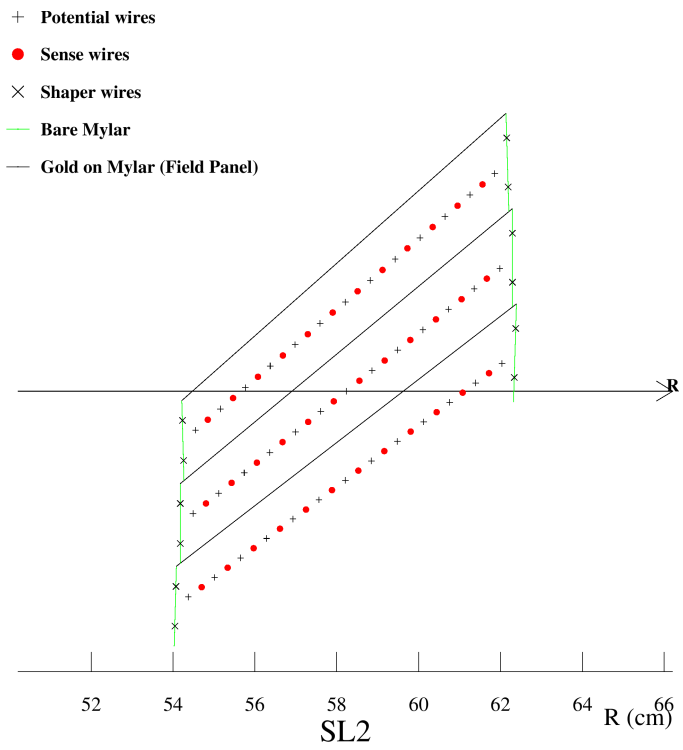

Figure 3.10: Configuration of the sense and field wires in each cell of the COT.

rapidity coverage of all the calorimeter subsystems is of $|\eta| < 3.5$.

The electromagnetic calorimeter consists of a central detector (CEM) with a η coverage of $|\eta| < 1$ [27] and a forward plug calorimeter (PEM) [28] that extends the coverage to $|\eta| < 3.6$. Both calorimeters are made of alternating layers of lead and polystyrene scintillator with photomultiplier tubes that collect the scintillation light. The CEM has a ϕ and η segmentation of 15° and 0.1, respectively. The PEM has the same ϕ segmentation for $|\eta| < 2.1$ while having a finer structure with 7.5° segments for $1 < |\eta| < 2.1$. In η the PEM has the same sectioning for $|\eta| < 1.8$ but with segments of between 0.2 and 0.6 for $|\eta| > 1.8$. The thickness of both the CEM and PEM are equivalent to one absorption length. The resolution, attained with test beam data, of the CEM is $14/\sqrt{(E_T)}$ [19] while that of the PEM is $16/\sqrt{(E)} \oplus 1$ [19], where E_T is the transverse energy given in GeV and defined

in an analogous manner to p_T, while E is the total energy given in GeV.

The energy of hadronic particles is measured by the hadronic calorimeter systems placed outside the electromagnetic calorimeters. The hadronic calorimeter system consist of a central (CHA) [27], wall (WHA), and forward (PHA) [28] calorimeter, with pseudorapidity coverage of $|\eta| < 0.6$, $0.6 < |\eta| < 1.3$ [29], and $1.3 < |\eta| < 3.6$, respectively. The CHA, WHA, and PHA have the same scintillator material as the CEM and PEM but use iron instead of lead as an absorber. Both the CHA and the WHA have the same ϕ and η segmentation as the CEM and are ~4.5 absorption lengths thick. The PHA has the same angular segmentation as the PEM and a thickness that corresponds to 7 absorption lengths. The resolution of both the CHA and the WHA is $\sim 80/\sqrt{E_T}$ while that of the PHA is $\sim 80/\sqrt{E} \oplus 5\%$.

In addition to the scintillator based sub-detectors in the electromagnetic calorimeter proportional chambers are added just before the calorimeter (CPR) and at approximately six radiation lengths (X_0) (CES for the CEM and PES for the PEM). The preshower detectors, placed immediately after the solenoid, are used to distinguish electrons from pion/photon decays from (prompt) electrons originating from the beam crossing point. These prompt electrons tend to have a softer shower profile due to interactions with the material between the interaction point and the preshower detectors. The CES detectors are used to attain position information of the incident particle within a given tower. This information aids in matching tracks with electromagnetic showers. Instead of a dedicated proportional chamber, the first layer the PEM is used as a preshower scintillator (PPR) for the plug calorimeter. This first layer is more than twice as thick as the other PEM layers and is made up of a different polyvinyltoluene-based

scintillator. The PPR has the same function as the CPR and helps discriminate non-prompt electrons.

3.3.3 Muon system

The electromagnetic and hadronic calorimeters are surrounded by the muon systems that consist of single wire drift chambers and scintillators. The drift chambers are used to reconstruct muon tracks (stubs) while the scintillators are used for timing information to match muon tracks to the correct bunch crossing. The total η and ϕ coverage of the muon sub-detectors used in this analysis is shown in Fig. 3.11.

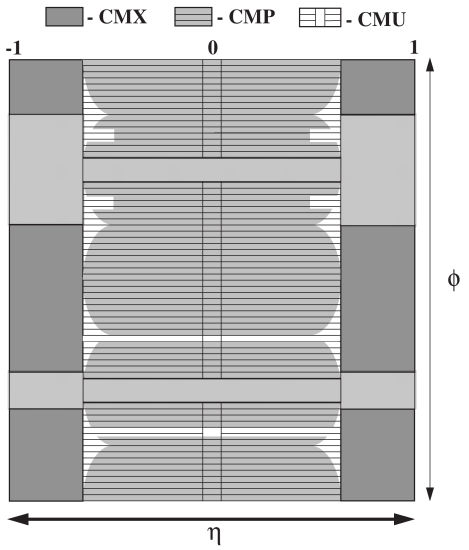

Figure 3.11: Muon system ϕ and η coverage.

In the central region the cylindrical central muon chambers (CMU) [30] provide coverage up to $|\eta| < 0.6$. Due to the large amount of material that precedes the CMU (5.5λ) a muon must have a minimum p_T of 1.4 GeV/c (range-out

threshold) to reach the CMU. The CMU is made up of 24 wedges comprised of four layers of four drift chambers. Each drift chamber is filled with the same gas mixture as in the COT and contains one sense wire that runs along the $\hat{z}$ direction. As with the COT, these chambers are run in proportional mode. This configuration results in a maximum drift time in each chamber of 800 ns. A sign ambiguity relating to which side, relative to the sense wire, of the chamber the ionizing particle traveled is resolved by shifting adjacent layers by 2 mm. Position information along the z coordinate is acquired by connecting layers 1 and 3 as well as 2 and 4 and measuring the collected charge differences in the two wires. This is done to have readout on one side of the chambers rather than one both sides. Due to mechanical constraints there are 2.4° gaps between the wedges that are uninstrumented. There are also 18 cm gaps between the two 180° arcs that each contain half the wedges.

Outside the CMU are additional central chambers, with the same η coverage, known as the central muon upgrade (CMP) [30]. Additional steel absorber with a thickness 2.3λ is placed between the CMU and CMP yielding a range-out threshold p_T of 2.2 GeV/c. The drift chambers are similar to the CMU using the same gas mixture, and having the sense wires aligned with the beam axis. The four layers of chambers, each layer being offset by a half chamber width, form a box around the cylindrical CMU. The CMP drift chambers are significantly wider resulting in an increased maximum drift time of 1.8 μs.

To extend the muon system coverage to higher pseudorapidity, $0.6 < |\eta| < 1.0$, the central muon extension (CMX) [19] was added to the muon system. The CMX consists of two arches at each end of the detector, with additional upper (keystone) and lower (miniskirts) sets of chambers. These drift chambers

are, other than being shorter, the same as in the CMP, resulting in the same maximum drift time. The chambers are configured in 8 layers with six chambers per layer. The sense wires in the drift chambers are on a conical surface that has a ~41° angle with the beamline.

The intermediate muon system (IMU) extends the η coverage further to $1.0 < |\eta| < 1.5$. This system consists of 4 layers of drift chambers known as the BMU and two scintillator sub-detectors, the BSU and TSU. The TSU covers $1.3 < |\eta| < 1.5$ and consists of two rings of counters on either side of the detector.

In addition to the drift chambers the muon system also contains scintillators. These scintillators are used for timing information and are necessary due to the long drift times of the muon chambers when compared to the bunch crossing interval. For the CMP, one layer of scintillator (CSP) is placed on the outside of the CMP. The CMX has two scintillator layers (CSX) [31]: one on each side of the 8 layers of drift chambers. In the BMU system, the scintillator (BSU) is on the outer end of the BMU, similar to the CMP.

In this analysis muons are required to have either a CMU or CMX stub and have the stub matched, in $r - \phi$, to an extrapolated track from the tracking system. Further information, such as z position and stub angle, is used in a multivariate likelihood discriminant for muon identification. The muons are paired into either a CMU-CMU (CC) or a CMU-CMX (CF) channel. The selection criteria for these two channels is discussed in Sec. 3.3.6.

3.3.4 Time of flight

Between the COT and the cryostat of the superconducting solenoid is a subdetector meant to detect the time of flight (TOF) of particles. The time of flight is the time it takes a particle to travel from the interaction point to the TOF detector. The detector sits at a radius of 138 cm and has a length, along the beam axis, of 279 cm. The TOF consists of 216 plastic, BC-408, scintillator bars in a cylindrical configuration [32]. Light produced from the scintillators is collected at the end of the bars and brought to PMT's located inside the magnetic field using wave guides. Due to the PMT's low efficiency within the magnetic field, special amplifiers with a gain of about 15 are used to amplify the PMT signal. The time of flight is determined by estimating the time of interaction from COT tracks and the time the scintillation signal is received from the TOF detectors. The time of flight information together with the track trajectory allows for the determination of the mass of particles. The COT tracks yield both a traveled distance and a momentum measurement (from the curvature of the helix). The mass can then be determined with $m = (p/c)\sqrt{\frac{ct^2}{L^2} - 1}$, where m is the particle mass, p is the particle momentum, t is the time of flight, and L is distance traveled. The mass then can be used to distinguish between pions, kaons, and protons.

3.3.5 Luminosity determination

The instantaneous luminosity of inelastic $p\overline{p}$ collisions is measured by the Cerenkov luminosity counter (CLC). The CLC consists of long cones wedged between the beamline and the forward calorimeters. There are 48 conical counters made up of two layers of aluminized mylar filled with isobutane gas at approximately at-

mospheric pressure. On the end of the CLC that is near the forward calorimeter a light collector and PMT measure the light produced from a particle traversing the detector. The CLC counts the number of particles that traverses it's volume using Cerenkov radiation. Cerenkov radiation is produced when a particles travels faster than the speed of light for the medium it is traversing. This Cerenkov light has a characteristic wavelength for a particular material. The CLC measures the luminosity by counting the number of particles that produce Cerenkov radiation per bunch crossing. The actual luminosity is estimated using the following equation: $\mu_\alpha \cdot f_{BC} = \sigma_{in} \cdot \alpha \cdot L$ [33]. Here the CLC measures μ_α, which corresponds to the average number of interactions per bunch crossing, while f_{BC} is the bunch crossing rate, σ_{in} is the inelastic $p\overline{p}$ cross section, measured in CDF Run I, α is the probability to detect a single $p\overline{p}$ interaction (the acceptance), and L is the instantaneous luminosity. The dominant systematic uncertainty for the luminosity determination comes from the $p\overline{p}$ inelastic cross section. In summary, the CLC allows for accurate instantaneous luminosity determination to within a relatively small systematic uncertainty of 3%.

3.3.6 Muon trigger

Bunch crossings occur every 396 ns (2.5 MHz) at instantaneous luminosities of $\sim 10^{32}$ cm^{-2}s^{-1}, causing of order millions of primary interactions (events) per second. The detector readout can simply not be made fast enough to read out every event at this rate. If such a readout were to exist, extremely large data sets would be produced for which there is currently no storage technology. Most interactions, however, are of little interest and involve low-momentum events. The CDF trigger system is meant to reduce the data flow from the raw $p\overline{p}$ col-

lision rate. The trigger system is divided into three consecutive triggers: level 1 (L1), level 2 (L2), and level 3 (L3). Each trigger takes the previous trigger's output data as its input. The L1 trigger uses basic track, calorimeter, and muon information to reach a decision within 5.5 μs, which corresponds to a 42 event buffer. For low-momentum events the selection criteria of the L1 trigger can still result in a too high an output rate. To reduce the rate further, a momentum threshold is introduced. The low-momentum events are treated separately from the lower-rate high-momentum events. The rate for the lower-momentum thresholds events is reduced by randomly discarding events at a variable rate. The L1 trigger employs the strongest filtering and reduces the data rate to 20–30 kHz. After the L1 trigger has reached a trigger decision the data are passed to the L2 trigger. This trigger level uses more detailed event information and if an accept decision is made passes the data to the event builder, which constructs a data structure known as an event with all the information available from all sub-detectors. The data rate after the L2 decision is 200–500 Hz. The L3 trigger then uses the full event information together with event requirements that are similar to the 'offline' requirements discussed in Sec. 4.1.1. If the L3 trigger accepts the event it is saved on mass storage facilities. The final event rate after the L3 trigger is 50–70 Hz. This entire trigger event selection process is known as the 'online' event reconstruction while all event requirements after the online reconstruction is known as the 'offline' event reconstruction. Figure 3.12 summarizes the online reconstruction procedure.

The data are collected in continuous periods of time with fixed hardware, software, and trigger configurations, known as runs. The trigger configurations can change from run to run to incorporate changes in specific requirements. Due to data runs with high instantaneous luminosities some triggers are dy-

namically prescaled. Prescaling randomly omits events even if they pass all trigger requirements to meet the maximum data output restrictions. This can be done at a rate based on the run conditions, dynamically, or at a constant rate. In this document we will only discuss a subset of muon triggers relevant to this analysis. More details on the separate levels of the muon triggers is given below.

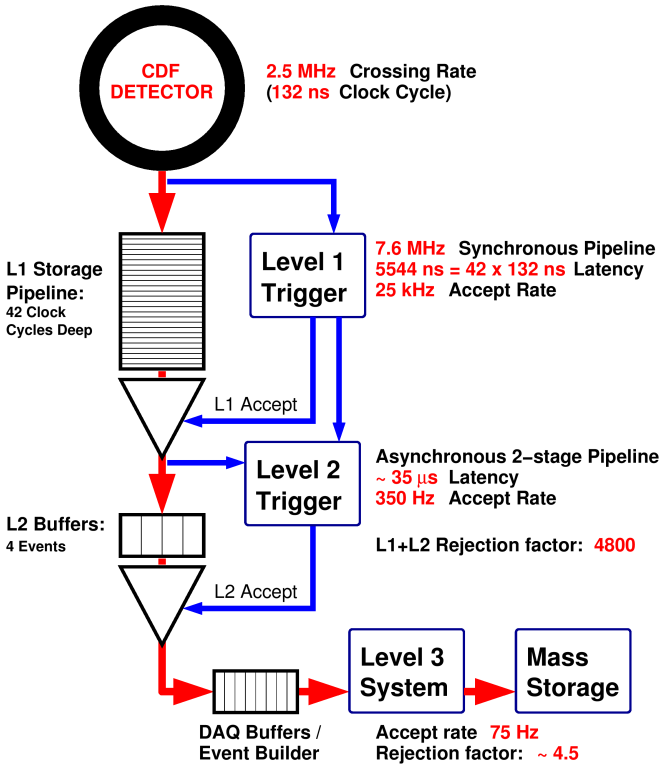

Figure 3.12: Data flow of CDF Run II data acquisition.

3.3.7 The level 1 muon trigger

The level 1 trigger uses primitives, a basic reconstruction using information from the COT and the muon chambers, to reach an accept or reject decision. Tracks are reconstructed using the extremely fast tracker (XFT) [34]. The XFT reconstructs track segments from all four of the axial COT layers. When high

instantaneous luminosities are reached additional information from stereo layers are used for segment reconstruction. In the early Run II data, there was a requirement that 10 of the 12 wires in a layer have a hit. This was changed to 11 of the 12 wires in later data due to the increased instantaneous luminosity, with only marginal effects on the XFT efficiency. The XFT track segments are linked together to form coarse tracks. The XFT only relies on ϕ and p_T information and requires a minimum track p_T of 1.5 GeV/c. The p_T and ϕ resolution of the XFT is 1.7%/(GeV/c) and $\sim 0.3^\circ$.

Once the track segments have been linked into XFT tracks the track extrapolater (XTRP) estimates the track position in the muon chambers with a ϕ resolution of 2.5°. This divides the detector into 2.5° segments known as trigger towers. The tracks are estimated as circles inside the solenoid and straight lines outside the solenoid. Deviations due to p_T dependent multiple scattering are accounted for during the matching process. Multiple scattering is dominated by multiple Coulomb scatterings of a particle off the Coulomb fields of nuclei.The L1 trigger has two separate p_T thresholds for both the CMU and CMX. For the CMU chambers the XFT has mappings for thresholds of $p_T > 1.5$ GeV/c and $p_T > 4.0$ GeV/c while the CMX has thresholds of $p_T > 2$ GeV/c and $p_T > 8$ GeV/c.

The muon chamber hits are collected, independently of the XFT, and form stub primitives. The transverse momentum of the stub is measured by comparing the drift times of chambers in adjacent layers. Due to the offsets discussed in Sec. 3.3.3 of neighboring drift chamber layers, the difference in drift times yields a measurement of the angle with which the track exits the COT. This angle, κ, is due to the deflection caused by the magnetic field inside the solenoid and thus

is dependent on p_T: $\kappa = 0.13/p_T$. This causes a p_T dependence of the drift time of $0.26/p_T$ μs, where p_T is in GeV/c. Two p_T thresholds for both the CMU and CMX are set at 1.5 GeV/c and 6.0 GeV/c. Each of these p_T threshold has a muon stub map with a granularity of 2.5°.

Additional information from the CMP and CSX muon detectors can also be used in the L1 trigger decision. If CMP hits are present, they are matched to CMU hits. A similar matching is made for CSX and CMX hits.

The XFT, XTRP, and muon chamber information is combined into 12 muon matchboxes which each process 30° in ϕ. The matchboxes produce the final L1 muon triggers of either a single muon or a dimuon trigger. If an XTRP track and a CMU stub, with the same p_T threshold, are matched a CMU trigger is reported. If in addition to the XTRP and CMU match there is also a CMP match, a CMUP trigger is reported. For a CMX trigger there must be an XTRP, CMX, and CSX match. A pair of matches is reported as a dimuon trigger for a combination of two CMU, a CMU and a CMX, a CMU and a CMUP, or a CMUP and a CMX triggers.

3.3.8 The level 2 muon trigger

The level 2 trigger works in a similar manner to the level 1 trigger. Matches are made between different subsystems to make a trigger decision. The L2 trigger, however, has a ϕ granularity of 1.25°, half that of the L1 trigger, when matching information from different sub-detectors. During the matching at this level a more detailed matching of XFT tracks and muon system hits is performed than with the XTRP. An exception is a small fraction of the data where the L2 trigger

passed every input event to the L3 trigger.

In the $B^0_{s,d} \to \mu^+\mu^-$ search a dimuon L2 trigger is required. This trigger requires two oppositely charged (opposite curvature) XFT tracks that are separated by more than 1.25° in ϕ, at the sixth COT super layer ($\Delta\phi_{SL6}$), or that are on opposite halves of the detector ($+z$ or $-z$). For a small fraction of the data the trigger also requires $\Delta\phi_{SL6} < 120°$.

3.3.9 The level 3 muon trigger and offline reconstruction

The last level of the trigger uses full event information from all sub-detectors to reach an accept or reject decision. The L3 trigger consists of Linux PC farms that have a decision output frequency of 50–70 Hz. Events are reconstructed in essentially the same method as with the offline reconstruction using the same information except for a somewhat simplified and faster tracking algorithm. The individual requirements of all the L3 triggers used are discussed in Sec. 4.1.1.

The offline reconstruction constructs physics objects that cause particular combinations of detector signals. There are several groups of objects: charged particle tracks, muons, electrons, photons, jets, and missing energy. A muon object is reconstructed when there is a charged particle track in the COT and hits in the muon chambers that match the COT track. Electrons require a charged particle track and large deposit in the electromagnetic calorimeter, while photons only have large deposits in the electromagnetic calorimeter systems. Jets are characterized by clusters of particles that leave significant energy deposits in the hadronic calorimeters. Missing energy events seem to not conserve energy in the plane transverse to the beam axis.

In this analysis we will focus on the muon reconstruction as well as the primary and secondary vertex finding algorithms. The primary vertex is defined as the interaction point of the original interaction, while the secondary vertex is the location of a secondary interaction. Good track reconstruction is essential for secondary vertex reconstruction.

A track is reconstructed as a helix due to the uniform magnetic field in the z direction inside the tracking volume. The helix is parametrized by the curvature, $C = q/2R$, where q is the charge of the particle and R is the radius of the helix, the signed impact parameter (d_0), the z coordinate (z_0), the azimuthal angle (ϕ_0), and the polar angle ($\alpha = \cot\theta$). The zero subscript indicates that these variables are measured at the point of closest approach to the beam axis. The curvatures is a function of the p_T and the strength of the magnetic field: $|C| = \frac{eBc}{2p_T}$, where B is the strength of the magnetic field. The final helix fit is done using a Kalman fitter [35]. Using the COT alone, the resolution of the helix parameters are $\sigma_{p_T}/p_T^2 \approx 0.15\%/$ GeV$/c$, $\sigma_{z_0} \approx 0.3$ cm, $\sigma_{d_0} \approx 250$ μm, $\sigma_{\phi_0} \approx 0.6$ mrad, and $\sigma_\alpha \approx 3$ mrad. Adding silicon hits reduces the impact parameter resolution by nearly an order of magnitude to $\sigma_{d_0} \approx 20$ μm. The resolution of the azimuthal and polar angle also decrease to $\sigma_{z_0} \approx 50$ μm and $\sigma_\alpha \approx 1$ mrad. The impact parameter resolutions quoted include a 30 μm beam size uncertainty and is a function of p_T. The raw impact parameter resolution is $\sigma_{d_0} = \sqrt{30^2 + (30/p_T)^2}$ without L00 hits and $\sigma_{d_0} \approx \sqrt{15^2 + (30/p_T)^2}$ with L00 hits, where p_T is in GeV$/c$.

Muons are reconstructed (for this analysis) as either CMU, CMUP, or CMX muons if hits from these respective muon systems match tracks in the COT. CMUP muons have hits in both the CMU and the CMP subsystems and have a higher p_T threshold consistent with the additional absorber placed between

the CMU and CMP. Deviations from the expected muon path are caused by multiple scattering and are proportional to $1/p_T$. The proportionality constants are roughly 15 cm×(GeV/c) for the CMU while 40 cm×(GeV/c) for all the other muon sub-detectors.

Primary and secondary vertices also need to be reconstructed in this analysis. The displaced secondary vertex requirement comes from the relatively long lifetime (~1.5 ps) of the B-meson. For the reconstruction used in this analysis the muon tracks are required to form a secondary vertex. A fit for a primary vertex is also performed. Both the primary and secondary vertex fits involve refitting track helix parameters to a common vertex. A fitting algorithm commonly used by many analyses at CDF, known as CTVMFT [36], is used for the secondary vertex fit, while a similar method is used for the primary vertex fit.

CHAPTER 4

ANALYSIS

This analysis searches for the $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays using the full 10 fb^{-1} CDF Run II data set. The same analysis methods are used for both modes with a shift in the B-meson mass search window. The branching fractions are measured relative to the $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ normalization mode. This mode together with $J/\psi \to \mu^+\mu^-$ decays are used to estimate signal efficiencies and perform several cross-checks. Monte Carlo (MC) simulations of $B_s^0 \to \mu^+\mu^-$ decays together with a full CDF detector simulation are employed to estimate many of the signal efficiencies. In addition to the $B_s^0 \to \mu^+\mu^-$ MC sample, a sample of $B^+ \to J/\psi K^+$ MC events is also produced for MC modeling cross-checks. Initially baseline requirements (described in Sec. 4.1.1) are applied to all data and MC samples. An artificial neural network (NN) is applied to enhance our expected signal over background. The entire signal selection optimization as well as the background estimates are done with the signal dimuon mass window blinded. The signal mass window is defined by a 300 MeV/c^2 window centered around the $B_{s,d}^0$ world average mass.

Major backgrounds come from Drell-Yan ($q\bar{q} \to \mu^+\mu^-$) processes through virtual γ and Z^0 states, double semileptonic decays ($b\bar{b}$ or $c\bar{c} \to \mu^+\mu^- X$), and sequential semileptonic decays ($b \to c\mu^- X \to s\mu^-\mu^+ X'$) of b or c quarks. A combination of semileptonic decays with a fake muon can also be a source of background. Additional background contributions come from $B \to h^+h'^-$ decays with two hadrons misidentified as muons. Backgrounds can be divided into a combinatorial dimuon background and a peaking background from hadronic B-meson decays, which are estimated separately. Backgrounds are studied in detail in

orthogonal control samples with various baseline requirements inverted to enhance the background contribution.

The NN binning is determined by an *a priori* optimization that uses the expected $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ limit as a figure of merit, resulting in 8 signal NN bins. The signal region is also divided into five mass bins centered on the world average B_s^0 and B_d^0 masses. This yields a total of 80 single bin counting experiments corresponding to CC and CF topologies, each with 8 NN bins, and five mass bins.

The branching fractions are measured relative to the $B^+ \to J/\psi K^+$ normalization mode using Eq. (4.1), where $\alpha_{B_s^0}$ is the geometric acceptance of the dimuon triggers, $\epsilon_{B_s^0}^{\text{trig+reco}}$ is the combined trigger and reconstruction efficiency within the acceptance, $\epsilon_{B_s^0}^{\text{NN}}$ is the NN efficiency, $\frac{f_u}{f_s}$ is the b-quark fragmentation probabilities ratio, and $\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+)$ is the normalization mode branching fraction. The equivalent efficiencies and acceptance for the normalization mode are indicated with the B^+ subscript.

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \frac{N_{B_s^0}}{N_{B^+}} \cdot \frac{\alpha_{B^+}}{\alpha_{B_s^0}} \cdot \frac{\epsilon_{B^+}^{\text{trig+reco}}}{\epsilon_{B_s^0}^{\text{trig+reco}}} \cdot \frac{1}{\epsilon_{B_s^0}^{\text{NN}}} \cdot \frac{f_u}{f_s} \cdot \mathcal{B}(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+) \quad (4.1)$$

Finally, once the signal efficiencies and the background estimates are well understood, the signal region is unblinded and a thorough statistical analysis of the result is performed. The sections below will discuss the analysis methodology and results in more detail.

4.1 Data samples

The data samples used in this analysis consist of a dimuon sample and a $B^+ \to J/\psi K^+$ sample. Both of these samples are collected with the same RAREB dimuon triggers. In addition to data samples both a $B_s^0 \to \mu^+\mu^-$ and a $B^+ \to J/\psi K^+$ MC sample are generated. This section discusses these samples and their selection criteria in detail.

4.1.1 Baseline event selection

All samples used in this analysis are required to pass basic event-selection requirements. Initially, events are required to meet the RAREB dimuon trigger requirements discussed in Sec 3.3.6. All tracks from events that pass these trigger requirements are refit during our offline reconstruction using L00 silicon hits. After the refitting, events must meet the trigger requirements again and are also required to pass requirements on secondary vertex variables.

All tracks are required to have fired a single muon L1 trigger. For CMU muons this translates to requiring $p_T > 1.5$ GeV/c with an XTRP matching of an equivalent low-p_T CMU primitive. In a small fraction of early data CMU wedge 17 is not used. CMUP muons must have a high-p_T CMU primitive as well as a CMP primitive matched to a COT track. CMX muons require an XTRP matching with a $p_T > 2.0$ GeV/c muon. Each candidate event must also have two tracks that are linked to two corresponding XFT tracks. In addition to the L1 matching, two opposite-sign tracks are required, corresponding to the L2 trigger requirements. The majority of L3 triggers require the sum of the p_T for

the two muon tracks to be greater than 5 GeV/c (known as the SUMPT trigger paths). For CMU-CMU muon events a two trigger tower separation of the two muons is required, unless each muon is on a different η side of the detector (east or west). The L3 trigger also requires that the difference in the z coordinate of the two muons at the point of closest approach to the interaction point (Δz_0) be less than 5 cm.

The following requirements are made after the trigger matching requirements have been met:

COT track quality cuts: Tracks must have $p_T > 0.5$ GeV/c and at least five hits in two of the axial layers as well as in two of the stereo layers.

Good run: Runs used for this analysis must be marked good by the data quality monitoring group [37]. Specific requirements are that the CLC, all the trigger levels, COT, offline reproduction, silicon, calorimeters, CMU, and CMX (if there are CMX hits) information is marked as being good quality. The entire CDF Run II data set is used which corresponds to approximately 9.7 fb^{-1} of data that is marked as good.

Muon quality: CMU, CMUP, and CMX tracks must have $p_T > 2.0$ GeV/c, $p_T > 3.0$ GeV/c, and $p_T > 2.2$ GeV/c, respectively. These requirements are meant to avoid rapidly changing trigger efficiency ranges near the trigger thresholds. The tracks must also have $|z_0| < 60$ cm.

XFT fiducial: Tracks are required to be within the XFT tracking volume i.e., they must have a z coordinate, at the COT exit radius ($r = 136$ cm), within 155 cm of the origin.

SVXII quality: Each muon track must have at least 3 layers of L00+SVXII with more than one associated hit.

Dimuon: A primary vertex for each event is reconstructed by refitting track helix parameters of all tracks, except for the muon candidate tracks, to a common vertex. Tracks used for the primary vertex fit are required to have $p_T > 0.5$ GeV/c and have their z_0 within 1 cm of the dimuon average z_0. If the fit fails, the beamline position at the average z_0 dimuon pair, estimated using the SVXII, is used as the primary vertex position. A well defined secondary vertex, formed by the dimuons, is also required. For tracks to be considered dimuon pair, the ϕ separation of the tracks at the sixth superlayer must be greater than 1.25°. For a small subset of early data the difference is also required to be less than 120°.

B-meson: The B-meson candidate must have $p_T > 4$ GeV/c as well as $|y| < 1$, where y is the rapidity, $y = \tanh^{-1}(\frac{p_z}{E})$. For the dimuon sample these quantities are estimated using the two muons while for the $B^+ \rightarrow J/\psi K^+$ sample the two muons and the kaon are used.

Requirements are also made on both the dE/dx and the muon likelihood of a track. dE/dx is the energy lost by the muon through ionization of the COT gas. This energy loss is a function of the β of the particle and the material that is being traversed. Thus by knowing the momentum of the particle, the mass can be determined by measuring the dE/dx. In our case this allows for kaon rejection but not pion rejection since the mass of the pion and muon are similar. The quantity $Z = log(\frac{dE/dx_{\text{Obs.}}}{dE/dx_{\text{Pred.}}})$ is used in this analysis, where the Obs. subscript indicates the observed value and the Pred. subscript signifies the predicted value based on the muon mass hypothesis. The dE/dx value used in the determination of Z is calibrated to be constant over time (run number), instantaneous luminosity, geometrical variances, and COT occupancy. The calibration was performed using a pure sample of D^*-tagged $D^0 \rightarrow K^+\pi^-$ decays and validated

with the earliest third of the data set. The average COT occupancy measured in a subset of the data and a run based luminosity approximation is used for the dE/dx calibration. This has a minimal effect on the efficiency on muons and kaon rejection while avoiding including significantly more information in the data reconstruction. A dE/dx requirement is chosen to be 98% efficient for muons from $J/\psi \to \mu^+\mu^-$ candidates. This is equivalent to a $Z > -0.083$ requirement and results in a ~50% kaon rejection.

The muon likelihood is required to be greater than 0.1 to yield similar signal efficiencies but enhanced background rejection as a previously used method. The method used in iterations of the analysis that used less than 2 fb^{-1} of data yielded a χ^2 for the muon stub and COT track matching. The muon likelihood method uses a likelihood function that incorporates muon stub matching to COT tracks and calorimeter information. The matching information comes from Δz, $\Delta\phi$, and Δx, the difference between stub and COT track z, ϕ, and x coordinate at the innermost part of the muon detector. The calorimeter deposits for real muons should be consistent with the minimum ionizing energy for muons. Signal and background distributions for these discriminating variables are determined and used as template probability distribution functions (PDF's). The final likelihood is defined as $\mathcal{L} = \frac{S}{S+B}$ where S represents the result from the fit to signal templates and $S + B$ represents the result from signal-plus-background templates.

In addition to the above requirements, secondary vertex related variables must also meet certain requirements. We define a 3D displacement length, L_{3D}, as the distance between the primary and secondary vertex. We also estimate a proper decay time, $\tau = L_{3D} M_{vtx}/|\vec{p}(B)|$, which we use to calculate the proper

decay length, $\lambda = c\tau$. The following requirements are made on these variables and their uncertainties (σ) as well as on the χ^2 returned by the secondary vertex fit:

- $\chi^2 < 15$
- $\sigma_{L_{3D}} < 0.0150$ cm
- $L_{3D} < 1.0$ cm
- $0 < \lambda < 0.3$ cm
- $\lambda/\sigma_\lambda > 2$

Finally, loose requirements on the isolation, $I > 0.5$, and the pointing angle, $\Omega < 0.7$ rads are made. Here the isolation is defined as $I = p_T(B_s^0)/(p_T(B_s^0) + \sum_i p_T(i))$, where sum goes over all tracks within an $\eta - \phi$ cone centered around the B_s^0 momentum with radius, $R = \sqrt{\eta^2 + \phi^2} < 1.0$. The pointing angle is defined as the 3D angle between the B_s^0 candidate momentum and the vector that points from primary to secondary vertex. Signal like events should be isolated ($I \to 1$) and have a small pointing angle.

4.1.2 Dimuon sample

The dimuon sample is required to have dimuon invariant mass in the range 4.669 GeV/c^2 $< M_{\mu^+\mu^-} <$ 5.969 GeV/c^2. This yields 60,842 and 64,495 events for the CC and CF channels, respectively (Fig. 4.1). After the baseline requirements have been met we blind our self to a 300 MeV/c^2 dimuon mass region from 5.169 GeV/c^2 to 5.469 GeV/c^2. We also define a $\pm$60 GeV/c^2 signal dimuon

mass region centered around the world average B_s^0 mass. The ± 60 GeV$/c^2$ is equivalent to $\pm 2.5\sigma_m$, where σ_m is the dimuon mass resolution determined by MC simulations discussed in Sec. 4.1.4. This 120 MeV$/c^2$ signal region is further divided into five dimuon mass bins with equal widths.

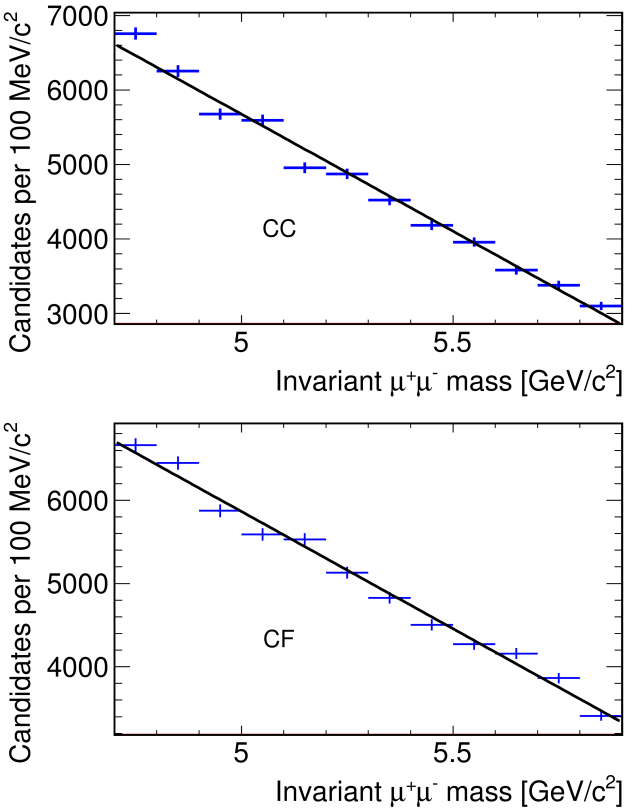

Figure 4.1: Dimuon mass distribution for the CC and CF channels after baseline requirements.

4.1.3 Normalization sample

The normalization sample consists of $B^+ \to J/\psi K^+$ candidate events collected on the same trigger as the dimuon sample. The $B^+ \to J/\psi K^+$ sample has to pass all the baseline requirements as the dimuon sample. For the secondary vertex

related selection criteria of the baseline requirements variables from the reconstructed J/ψ are used instead of from the B-meson. The kaon must have hits in at least three layers of the SVXII and L00; $p_T > 1.0$ GeV/c; z_0 within 5 cm of the dimuon vertex; and $z < 155$ cm at the radius where the track exits the COT. The fit probability of the B^+ vertex is required to be greater than 10^{-5} and the dimuon mass must be between 3.017 GeV/c^2 and 3.177 GeV/c^2. An invariant mass resolution of the two muons and the kaon is determined by fitting the invariant mass to a first order polynomial added to a Gaussian distribution. The invariant mass resolution is 11 MeV/c^2 for both the CC and CF channels. To assess the final $B^+ \to J/\psi K^+$ yield a signal region of width 35 MeV/c^2 centered around the world average B^+ mass is defined. Simple sideband subtraction, using sidebands for which $5.120 < M_{\mu^+\mu^-K^+} < 5.225$ GeV/c^2 and $5.335 < M_{\mu^+\mu^-K^+} < 5.440$ GeV/c^2, is used to attain the final $B^+ \to J/\psi K^+$ yield. A small correction of 0.14% is applied to the yield of the sideband subtraction to account for the small contribution of $B^+ \to J/\psi\pi^+$ decays. The final $B^+ \to J/\psi K^+$ yield is 28082 ± 196 and 12145 ± 1384 and is shown in Fig. 4.2 for the CC and CF channels, respectively.

4.1.4 Monte Carlo simulation sample

A $B_s^0 \to \mu^+\mu^-$ MC sample is generated using pythia [2] and EvtGen [38] with the underlying event modeling tuned to reproduce minimum bias events [39]. One of the B-mesons is forced to decay to two muons while the other is left to decay freely. The samples go through a full CDF II detector simulation to account for detector resolution effects and are reconstructed in the same manner as the data events. The MC sample must pass the same baseline requirements as the data (Sec. 4.1.1) with the exception of the muon likelihood and dE/dx requirements.

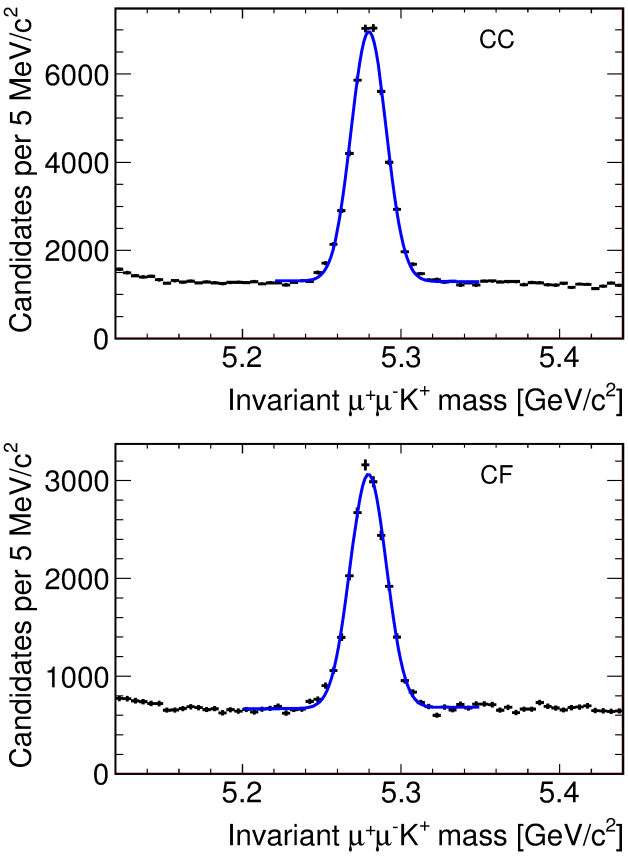

Figure 4.2: B^+ mass distribution for the CC and CF channels after baseline requirements.

These particle identification requirements are omitted because their efficiencies, which were found to be ~100%, are studied using data. The MC simulations do not accurately reproduce the B-meson p_T and isolation distribution. This is corrected by reweighting the MC using p_T distributions from $B^+ \to J/\psi K^+$ decays and isolation distributions from $B_s^0 \to J/\psi\phi$ decays.

The $B_s^0 \to \mu^+\mu^-$ MC sample is used for signal efficiency estimates of the various baseline requirements. The MC sample is also used as the signal sample for the neural network (NN) training discussed in detail in Sec. 4.2.1. The dimuon invariant mass resolution is also determined from the $B_s^0 \to \mu^+\mu^-$ MC sample at

$\sigma_M \sim 24\,\mathrm{MeV}/c^2$.

A $B^+ \to J/\psi K^+$ MC sample is produced and must pass all but the muon identification $B^+ \to J/\psi K^+$ baseline requirements. This sample is used to estimate and cross-check some of the signal efficiencies for the $B^+ \to J/\psi K^+$ normalization mode. This collection of MC events is also reweighted due to the mismodeling of the p_T and isolation distributions. The reweighting is done using $B^+ \to J/\psi K^+$ data p_T and isolation distribution. The invariant mass resolution of the $B^+ \to J/\psi K^+$ MC is compared to the mass resolution determined from $B^+ \to J/\psi K^+$ data. The simulated resolution and measured resolution are within 10% of each other which is adequate for this analysis.

An additional small sample of $B_d^0 \to \mu^+\mu^-$ MC simulations is generated for an NN mass bias check discussed in detail in Sec. 4.2.2.

4.2 Artificial neural network

To increase our discrimination power between signal and background an artificial neural network is used to combine multiple discriminating variables. In the 2 fb^{-1} iterations of the analysis a rootSNNS [40] NN package was used to combine the discriminating power of six kinematic variables. A new NN package, the neurobayes [41, 42] package, is used during the most recent iteration of the $B_{s,d}^0 \to \mu^+\mu^-$ search using 7 fb^{-1} of CDF Run II data. This analysis uses the same NN as the analysis based on 7 fb^{-1} analysis. The new NN uses 14 input variables consisting of vertex and kinematic variables and outputs a single signal-background discriminant, ν_N. Particle identification variables such as dE/dx and muon likelihood were not used in the training of the NN because

the MC simulations are unable to accurately model the distributions of these variables. The output of the NN ranges from -1 to 1 but is rescaled to range from 0 to 1. Zero corresponds to the most background like events and one signifies the most signal like events. The neurobayes NN yielded the same signal efficiency for a range of ν_N cuts as rootSNNS, but had approximately twice the background rejection (Fig. 4.3). Details on the NN training and validation are discussed in the next two sections.

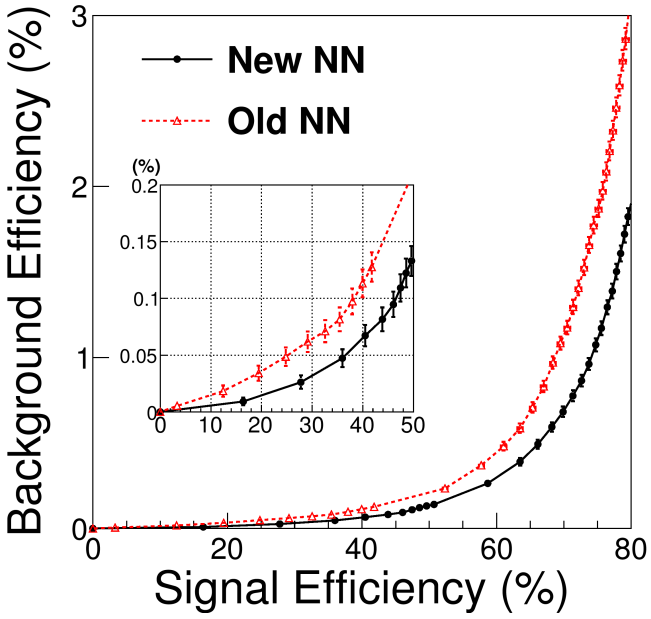

Figure 4.3: Background as a function of signal efficiency for the neurobayes (new) and rootSNNS (old) neural networks.

4.2.1 Neural net training

To train the NN a background and signal sample needs to be prepared. The background sample consists of dimuon mass sideband data that pass the baseline requirements in the first 7 fb^{-1} of data. The events are required to be within a

mass range of 4.669 GeV/c^2 < $M_{\mu^+\mu^-}$ <5.169 GeV/c^2 or 5.469 GeV/c^2 < $M_{\mu^+\mu^-}$ <5.969 GeV/c^2. This sample is representative of backgrounds that are present in the signal mass region. The data is further divided into a CC and CF sample. Because these two muon topologies have different signal/background ratios, two separate NN are trained using the CC and CF background sample. This results in a total of 36 329 CC background events and 39 657 CF events.

The MC sample discussed in Sec. 4.1.4 is used as a signal training sample. A mass window of 120 MeV/c^2, centered around the worlds average B_s^0 mass, is selected as the training sample. The signal and background samples must be the same size for NN training. To achieve this, events are randomly discarded from the MC sample to match the number of background events. When training the NN, 80% of the background and signal sample are used for training while the remaining 20% are used as a testing sample. This 80% is randomly selected from the signal and background sample. The testing sample is used to check for overtraining by the neurobayes package.

A total of 19 input variables are initially used during the NN training. These input variables consist of vertex related and kinematic variables:

***I*:** isolation of the B_s^0 candidate defined in Sec. 4.1.1.

$\Delta\Omega$: three-dimensional angle between the B_s^0 momentum and the vector pointing from primary to secondary vertex.

$\Delta\Omega_T$: angle between the B_s^0 momentum and the vector pointing from primary to secondary vertex in the plane transverse to the beamline.

Lower $p_T(\mu)$: lower transverse momentum of the muon pair.

Higher $p_T(\mu)$: higher transverse momentum of the muon pair.

Smaller $|d_0(\mu)|$: muon impact parameter of muon with the smaller value of the muon pair.

Significance of smaller $d_0(\mu)$: $d_0(\mu)/\sigma_{d_0(\mu)}$ of the muon with smaller impact parameter, where $\sigma_{d_0(\mu)}$ is the estimated uncertainty of $d_0(\mu)$.

Larger $d_0(\mu)$: impact parameter of muon with the larger value of the muon pair.

Significance of larger $d_0(\mu)$: $d_0(\mu)/\sigma_{d_0(\mu)}$ of the muon with the larger impact parameter.

$d_0(B_s^0)$: impact parameter of B_s^0 candidate.

χ^2: χ^2 of the secondary vertex fit.

L: three-dimensional vertex displacement obtained from the primary and secondary vertex fit.

L/σ_L: significance of L.

λ: three-dimensional proper decay length defined in Sec. 4.1.1.

λ/σ_λ: significance of λ.

L_T: version of L_T measured in plane transverse to the beamline.

L_T/σ_{L_T}: significance of L_T.

λ_T: version of λ in the transverse plane.

$\lambda_T/\sigma_{\lambda_T}$: significance of λ_T.

Some possible input variables were not included due to the introduction of a mass bias or MC mismodeling of the variables. The B-meson p_T and the opening angle between the two muons were excluded from the input variable list because they caused a correlation between dimuon mass and ν_N. The distance of closest approach of the two muon tracks in the z direction, Δz_0, was also omitted because the MC simulations mismodel this variable's distribution. After the

training procedure, the neurobayes package ranks the input variables in order of separation power and automatically removes variables with a significance below three [41, 42]. This results in a total of 14 input variables shown in Table 4.2.1 together with their separation significance. The distribution for data mass sideband and signal MC of the six variables with the strongest separation power are shown in Fig. 4.4.

Table 4.1: List of input parameters and their discrimination significance selected by the neurobayes package.

Rank	Variable	Significance [41, 42]
1	$\Delta\Omega$	250.9
2	Isolation	97.8
3	Larger $d_0(\mu)$	78.8
4	$d_0(B_s^0)\|$	46.9
5	L_T/σ_{L_T}	37.5
6	χ^2	24.3
7	L	23.4
8	Lower $p_T(\mu)$	14.0
9	Significance of smaller $d_0(\mu)$	9.0
10	λ/σ_λ	7.8
11	λ	6.5
12	Smaller $d_0(\mu)$	7.7
13	$\Delta\Omega_T$	4.3
14	Significance of larger $d_0(\mu)$	4.0

The final ν_N distributions for signal MC and data mass sideband after combining the CC and CF channels can be seen in Fig. 4.5.

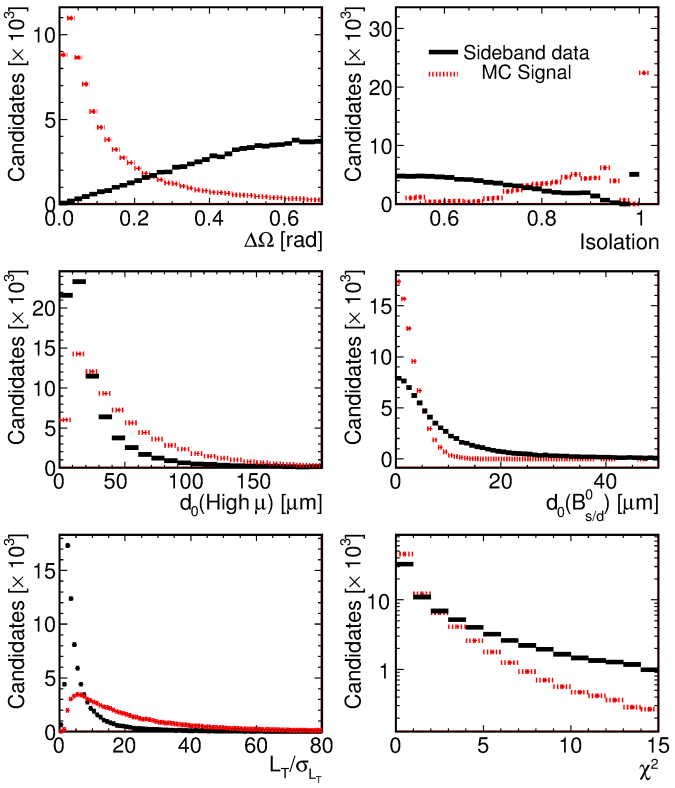

Figure 4.4: Comparison of the most powerful kinematic and lifetime parameter distributions of background from sideband (black solid) and signal MC (gray dashed) samples for the first 7 fb^{-1} of data.

4.2.2 Neural net validation: mass bias

The mass sidebands will be used to determine combinatorial background estimates making it essential that the NN does not sculpt the dimuon mass distribution. Several cross-checks are performed to ensure that the NN leaves the dimuon mass distribution unchanged. One of the cross checks compares the average ν_N with the dimuon mass for our dimuon sample after baseline requirements (Fig.4.6). The signal dimuon sample consists of two opposite-sign muons with a positive λ (OS+). In this sample, the 300 MeV/c^2 invariant dimuon mass region is still blinded. To assess possible mass sculpting of the NN in this re-

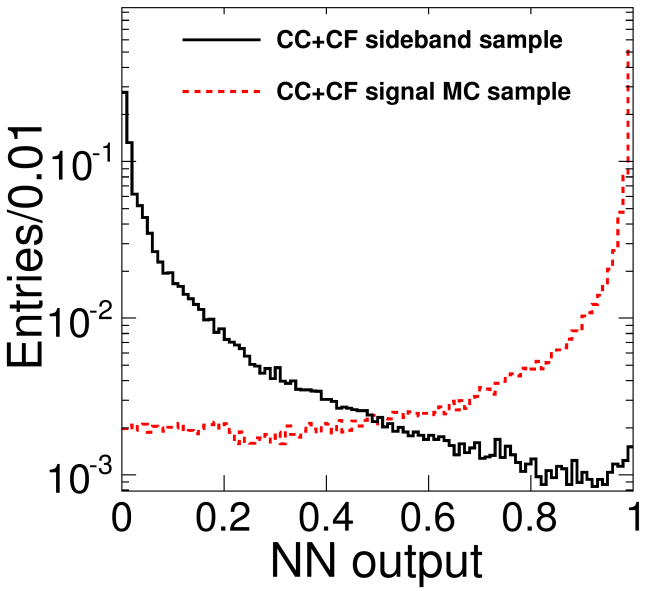

Figure 4.5: Distributions of ν_N for signal and background samples with 14 variable neurobayes NN. The background sample consists of sideband data events from the first 7 fb^{-1} of data.

gion we study the ν_N distribution as a function of dimuon mass in a sample with opposite-sign muons and negative lifetime (OS–). The resulting average NN output versus dimuon mass distribution is shown in Fig.4.6. Both distributions were fit to a constant resulting in good χ^2's and no significant indication of a correlation between NN output and dimuon mass. As a cross-check, a first order polynomial is also fit to the distribution and is described in more detail in appendix A.1. The χ^2's of the constant fits and first order polynomial fits are found to be similar.

An additional check is performed by training the NN on only sideband data. The dimuon sample, after meeting baseline requirements, is divided into an inner and an outer sideband region. The inner sideband region is defined as $5.002 < M_{\mu^+\mu^-} < 5.169$ GeV$/c^2$ while the outer is defined as $5.496 < M_{\mu^+\mu^-} < 5.636$

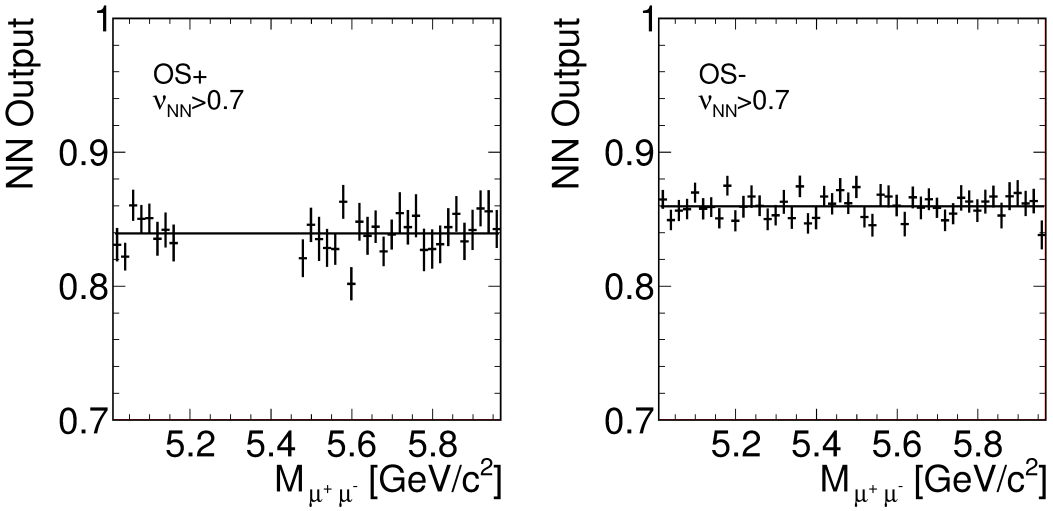

Figure 4.6: Result of the ν_N and dimuon mass correlation check using the first 7 fb^{-1} of data. Left: correlation between average ν_N and dimuon mass in OS+ sample. Right: correlation between average ν_N and dimuon mass in OS− sample.

GeV/c^2. These regions were chosen to have equal statistics for both the inner and outer sidebands. The NN is retrained using the inner sideband region as a 'signal' sample while the outer sideband regions is used as the background training sample. Since these two sideband regions are kinematically similar and only differ by invariant dimuon mass, an NN without mass bias should display similar output distributions for the two samples. The NN trainings are performed separately for the CC and CF channels. The resulting NN output distributions are shown in Fig. 4.7 and show no evidence of an NN mass bias.

To assess more subtle dimuon mass correlation of the NN due to the use of a B_s^0 MC sample for training, a small B_d^0 MC sample is produced for comparison with the B_s^0 sample. The two MC samples should produce a similar NN output distribution for an NN without any mass correlations. The two distributions for the two samples are shown in Fig. 4.8 indicating no significant mass dependence of the NN.

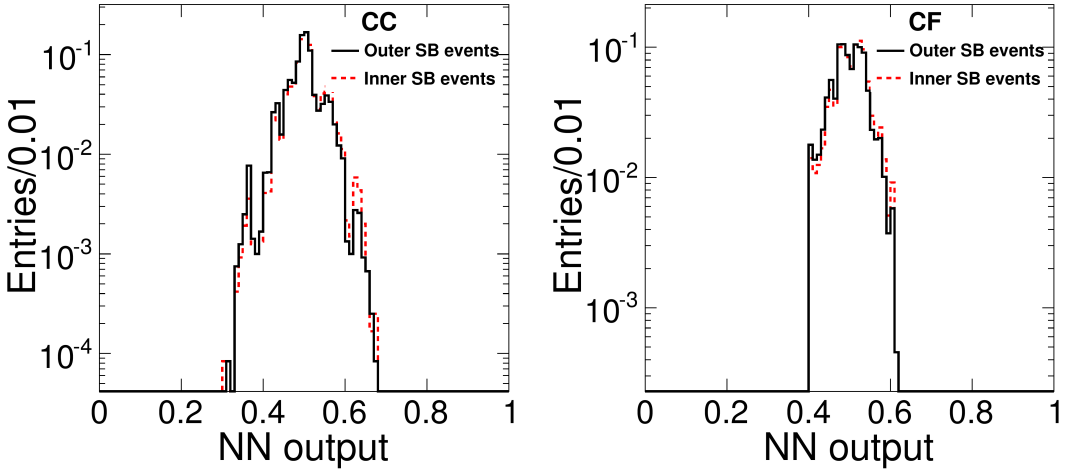

Figure 4.7: Result of the ν_N and dimuon mass correlation check using inner and outer mass sidebands from the first 7 fb^{-1} of data for separate CC and CF trainings.

4.2.3 Neural net validation: overtraining

Overtraining occurs when an NN becomes sensitive to the statistical fluctuation and properties unique to a training data set. The optimal training would be sensitive to general properties that many independent data sets of similar nature share. Overtraining can lead to an NN response in the training sample that is significantly different to the NN output in an independent sample. In this analysis 80% of the sideband data is used to train the NN. Overtraining could potentially cause an artificial suppression of the sideband yields in the 80% that is used during the NN trainings. To check for such effects, we compare the NN output distributions for the sideband data using NN's trained on 30%, 50%, and 80% of the sideband sample. The resulting distributions are shown in Fig. 4.9 and show no significant difference.

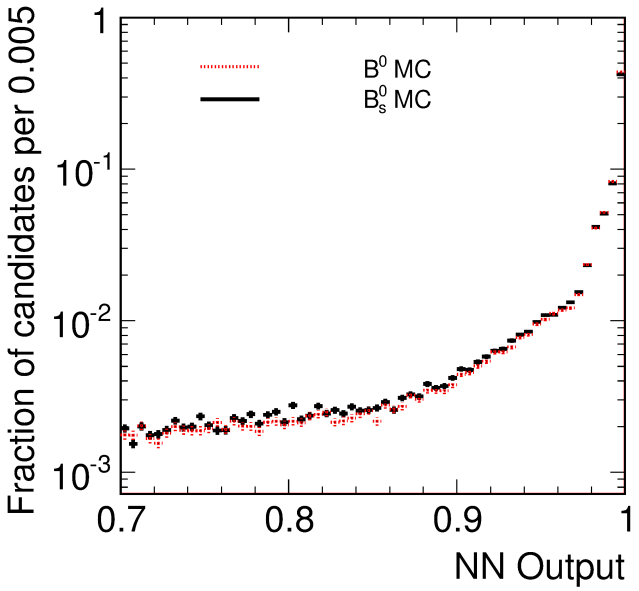

Figure 4.8: Comparison of ν_N for B_s^0 and B_d^0 MC. The distributions have been normalized to have total area of unity for the entire NN output range, $0 < \nu_N < 1$.

4.2.4 Neural net validation: MC modeling

As a check that the MC simulations model the NN input variable and some additional kinematic variable distributions correctly, we compare the distributions of $B^+ \to J/\psi K^+$ sideband-subtracted data and $B^+ \to J/\psi K^+$ MC simulations. Minor alterations were made to the input variables. Variables that are reconstructed using the B_s^0 vertex in the $B_s^0 \to \mu^+\mu^-$ sample use the J/ψ vertex in the $B^+ \to J/\psi K^+$ samples. The reconstructed B-meson p_T and isolation, however, use three-track information. The comparison of the distributions are shown in Figs. 4.10 and 4.11. The distributions are very similar for the MC sample and the sideband-subtracted data sample. Residual difference in the distributions of the isolation and the B-meson p_T distributions are assigned as systematic uncertainties to the NN efficiency and are discussed in detail in Sec. 4.3.

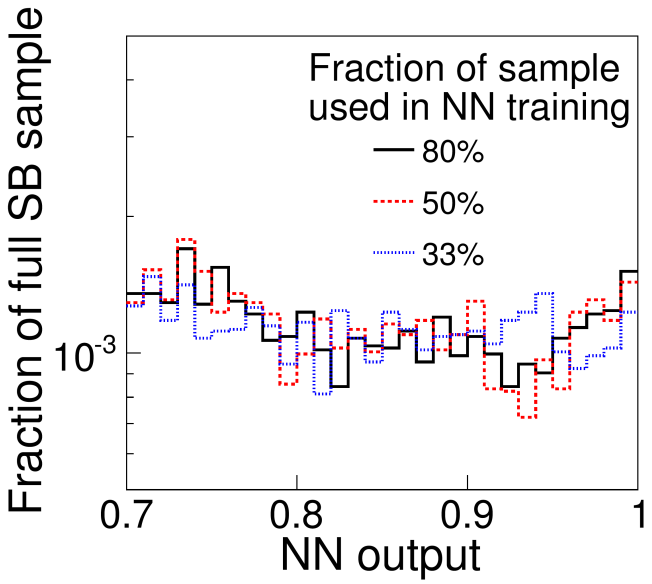

Figure 4.9: ν_N distributions for NN's trained with 33%, 50%, and 80% of the sample of sidebands events reconstructed in 7 fb^{-1}.

To ensure that our $B_s^0 \to \mu^+\mu^-$ MC simulations properly model the ν_N and NN signal efficiencies we compare the ν_N distributions for the $B^+ \to J/\psi K^+$ data and $B^+ \to J/\psi K^+$ MC sample. The NN, trained with $B_s^0 \to \mu^+\mu^-$ MC signal and $B_s^0 \to \mu^+\mu^-$ dimuon mass sideband data, is applied to the $B^+ \to J/\psi K^+$ data and MC sample. The NN output distributions for both the MC and data samples are shown in Fig. 4.12. Differences between the distributions are small and are assigned as a systematic uncertainty on the NN efficiency. The NN efficiency and its systematic uncertainty are discussed in more detail in Sec. 4.3.3.

4.3 Acceptance and efficiencies

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \frac{N_{B_s^0}}{N_{B^+}} \cdot \frac{\alpha_{B^+}}{\alpha_{B_s^0}} \cdot \frac{\epsilon_{B^+}^{\text{trig}}}{\epsilon_{B_s^0}^{\text{trig}}} \cdot \frac{\epsilon_{B^+}^{\text{reco}}}{\epsilon_{B_s^0}^{\text{reco}}} \cdot \frac{1}{\epsilon_{B_s^0}^{\text{NN}}} \cdot \frac{f_u}{f_s} \cdot \mathcal{B}(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+) \quad (4.2)$$

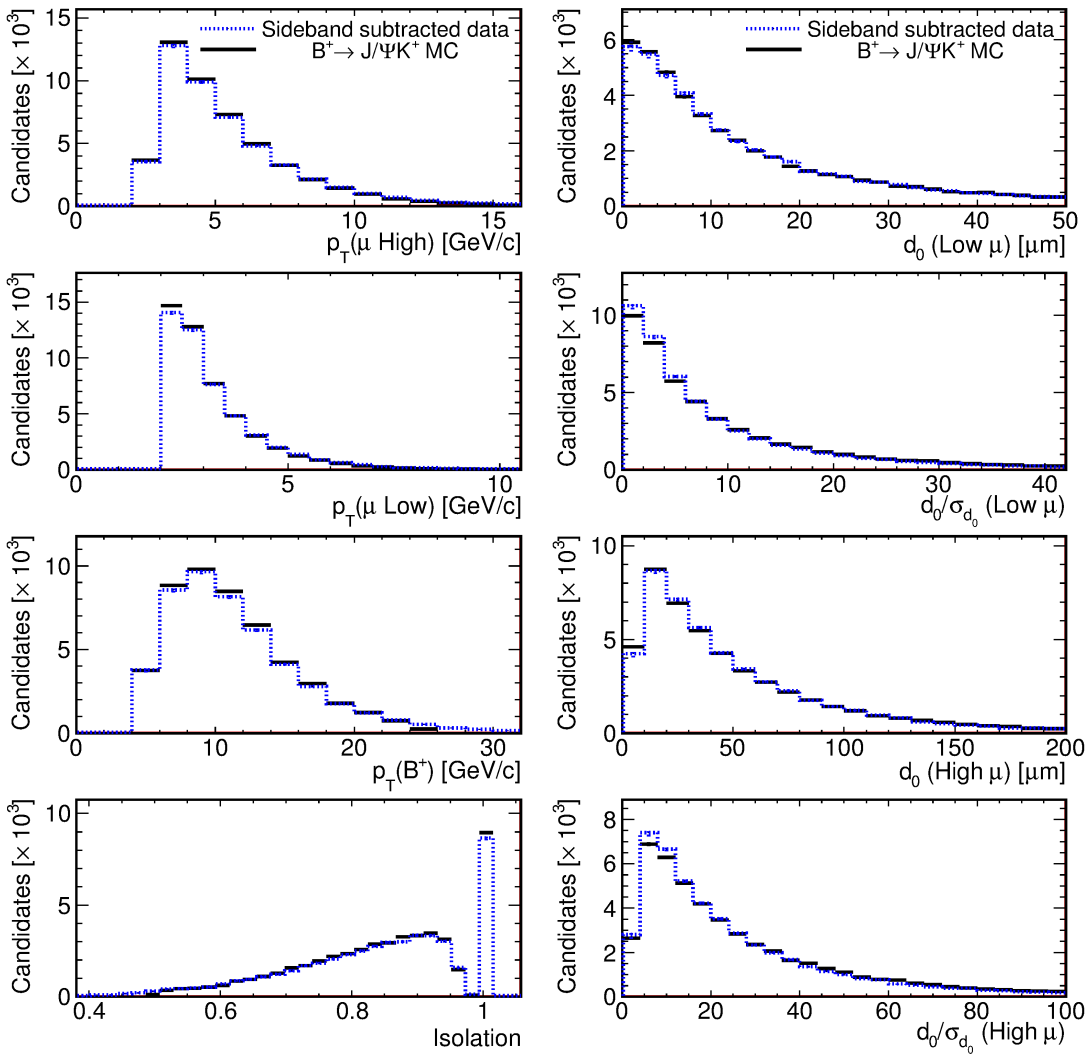

Figure 4.10: Distributions of NN input variables and additional kinematic variables for the entire 10 fb^{-1} of $B^+ \to J/\psi K^+$ sideband-subtracted data and signal MC.

The branching fraction $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$ is measured using Eq. (4.2), where $\alpha_{B^0_s}$ is the geometric acceptance of the dimuon triggers determined by the kinematic requirements on the phase space of B-meson decays considered, while $\epsilon^{\text{trig}}_{B^0_s}$ is the trigger efficiency within the acceptance, $\epsilon^{\text{reco}}_{B^0_s}$ is the efficiency of the baseline requirements for events passing the trigger requirements, $\epsilon^{\text{NN}}_{B^0_s}$ is the NN efficiency of all the NN bins combined for events satisfying the trigger and

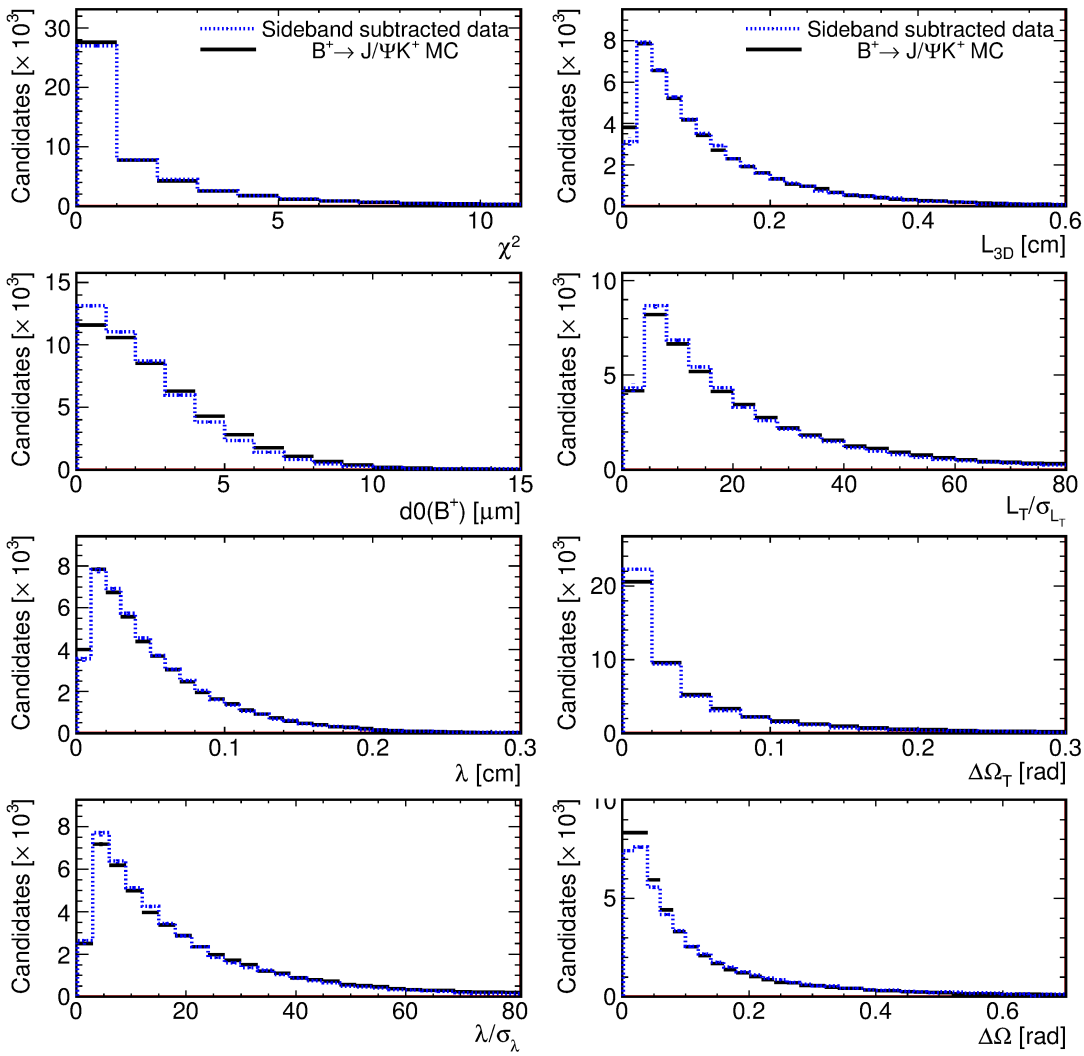

Figure 4.11: NN input variable distributions for the entire 10 fb^{-1} of $B^+ \to J/\psi K^+$ sideband-subtracted data and signal MC.

baseline requirements, $\frac{f_u}{f_s}$ is the b-quark fragmentation probabilities ratio, and $\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+)$ is the normalization mode branching fraction. The equivalent efficiencies and acceptance for the normalization mode are indicated with the B^+ subscript. No NN efficiency is indicated for the $B^+ \to J/\psi K^+$ mode because the NN is not applied to the normalization mode. The acceptance and efficiencies need to be estimated separately for the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ mode due to the kinematic differences between two and three body decays. Two

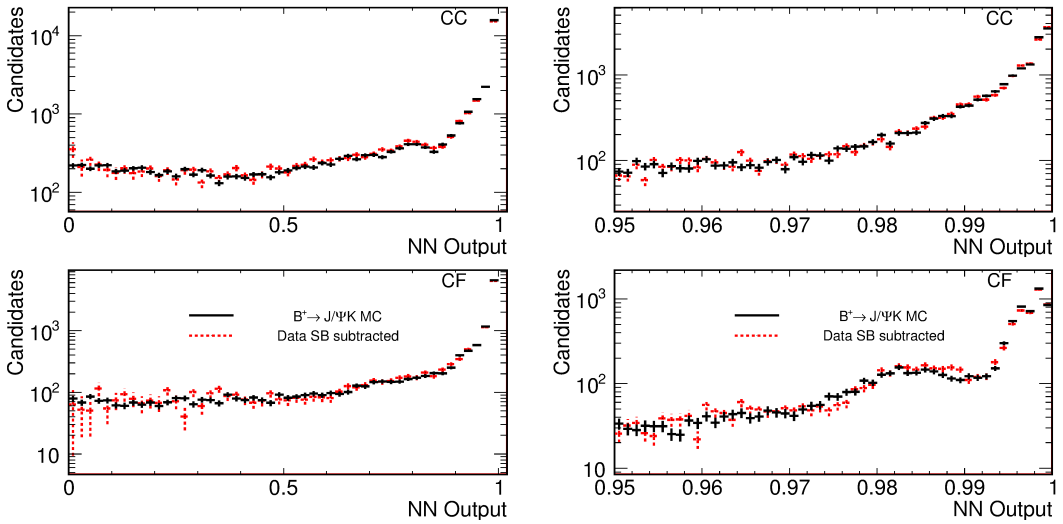

Figure 4.12: Comparison of ν_N distributions for $B^+ \to J/\psi K^+$ sideband-subtracted data and $B^+ \to J/\psi K^+$ MC simulations. The distributions in the left column show the ν_N over the entire NN output range while the distributions to the right show the ν_N for $\nu_N > 0.95$, which is our most signal sensitive range.

and three-body decays have different distributions for muon kinematic variables which affect the angular and kinematic acceptance. For the $B^+ \to J/\psi K^+$ mode the acceptance of the kaon also needs to be taken into account. Efficiencies can depend on kinematic properties and thus the different kinematic distributions of the two and three body modes can yield differing efficiencies. However, the efficiency differences for muons as a function of kinematic variables are typically not large and systematic uncertainties are substantially reduced by measuring the ratio of efficiencies.

4.3.1 Acceptance

The acceptances are determined using $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ MC simulations for B_s^0 and B^+ mesons that satisfy $|y| < 1.0$ and $p_T(B_s^0) > 4$ GeV/c, where

$y = \frac{1}{2}\ln(\frac{E+p_z}{E-p_z})$, is the rapidity. Both muons are required to satisfy the fiducial and kinematic requirements of the trigger discussed in Sec. 3.3.6. Muons are required to have $p_T > 2.0$ GeV/c for CMU muons, $p_T > 3.0$ GeV/c for CMUP, and $p_T > 2.2$ GeV/c for CMX and must extrapolate to the active fiducial volumes of the muon, COT, and L00+SVXII systems. Kaons in the normalization mode must have $p_T > 1.0$ GeV/c and extrapolate to the active fiducial volumes of the COT and L00+SVXII systems. Effects from COT tracking, multiple scattering, and stub-track matching are included in the reconstruction efficiency discussed below. Systematic uncertainties on the acceptances are assessed by varying b-quark mass, fragmentation modeling, and the renormalization scale during the MC generation. These quantities are varied by $\pm 1\sigma$, where σ is the uncertainty of the quantity, while the effect on the acceptance ratio is assessed. We also assign variations of the acceptance due to changes in the size of the beam spot as a systematic uncertainty. The observed differences are added in quadrature leading to a final acceptance ratio and systematic uncertainty given in Table 4.2.

4.3.2 Trigger efficiencies

The trigger efficiencies consist of two one-track efficiencies estimated from J/ψ dimuon events. Three separate trigger efficiencies, L1, L2, and L3, are estimated. The L1 measurement quantifies both the efficiency of the XFT to identify muon tracks as well as the efficiency of the XTRP to match these tracks to muon stubs in the muon chambers. The L1 one-track efficiency is parametrized in p_T, η, detector operation conditions, and ϕ. The p_T parametrization is meant to take into account the p_T dependence of the efficiency near the trigger p_T thresholds. This parametrization also considers the probability of muon range-out before

the muon candidate reaches the muon chambers. We parametrize the L1 efficiency in η due to the improved efficiency at higher η values due to the larger ionization path lengths in the COT. Bins in run number corresponding to detector operation conditions such as trigger configuration and COT performance are grouped into large bins with similar properties. The parametrization in detector operating conditions and ϕ is used to quantify the effects of threshold gains, gas gains in the COT, and trigger configurations on the efficiency. A tag-and-probe method is used to estimate the single-track L1 trigger efficiency. Events are collected using a single muon trigger and are reconstructed as $J/\psi \to \mu^+\mu^-$ candidates. After sideband subtraction, this yields a pure sample of $J/\psi \to \mu^+\mu^-$ events. One muon track in the sideband-subtracted J/ψ event is required to have fired an L1 muon trigger while the other passes fiducial and trigger baseline requirements described in the previous section. The numerator of the efficiency calculation consists of sideband-subtracted J/ψ events where both muon tracks are matched to an L1 trigger while the denominator consists of all sideband-subtracted J/ψ events. The dominant systematic uncertainties of the L1 efficiency comes from variations in the efficiency as a function of dimuon isolation and $\pm\eta$ asymmetries. This uncertainty is estimated separately for coarse ϕ bins and ranges from 0.3% to 2.7% for central muons. For the forward muons an additional effect due to a charge asymmetry (μ^+ or μ^-) is taken into account. The total systematic uncertainty for forward muons ranges between 0.5% and 3%. Once the one-track efficiencies are measured they are combined into a dimuon trigger efficiency by convolution with the distributions of the parameters in $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ MC. Double-track correlations are studied and are found to be small.

The L2 and L3 efficiencies quantify the ability to refine and fully reconstruct

the dimuon events. The L2 and L3 trigger efficiencies are estimated in a similar manner to the L1 efficiency but are only parametrized in bins of run numbers. Due to the changing luminosity over time, luminosity weighting is used to estimate the final L2 and L3 efficiencies. The final efficiencies for $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ are the products of L1, L2, and L3 efficiencies. The trigger efficiencies were cross-checked with $Z^0 \to \mu^+\mu^-$, for high p_T muons, and non J/ψ muon triggers. All efficiencies were consistent within statistical uncertainties. Residual variations in the L2 efficiency due to p_T dependence are assigned as systematic uncertainties, ranging between 0.14% and 3.4%. The systematic uncertainty on the L3 efficiency originates from efficiency dependence on several dimuon kinematic variables. Efficiency variations due to dimuon pseudorapidity, isolation, opening angle, and p_T are considered and result in systematic uncertainties between 1.3% and 4.3%.

The total trigger efficiency is taken as the product of the L1, L2, and L3 trigger efficiencies for the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ mode separately. The uncertainty of the efficiency ratio is estimated by treating the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ total uncertainties as 100% correlated. The final trigger efficiency ratios for the CC and CF channels and their associated uncertainties are shown in Table 4.2.

4.3.3 Reconstruction efficiencies

The reconstruction efficiency can be broken down into the COT, muon, SVXII, and vertex reconstruction efficiency of the two muons. For the normalization mode the kaon COT and SVXII efficiencies also need to be considered. An addi-

tional efficiency for the $B^+ \to J/\psi K^+$ mode is the B^+ vertex probability efficiency.

The COT track efficiency is estimated using MC track embedding. MC tracks are embedded in real COT data events to account for occupancy effects on the efficiency. The final one track efficiency is close to unity for muons with $p_T > 1.5$ GeV/c and the ratio of efficiencies for muons is taken as one. The dominant systematic uncertainties take into account the variation of the efficiency with isolation, p_T, and detector operation conditions. These parameters have different distributions for muons in $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decays. The largest systematic uncertainty comes from the absolute measurement of the kaon COT efficiency in the $B^+ \to J/\psi K^+$ decays where the full variation of the efficiency with the parameters is assigned as an uncertainty. The resulting efficiency ratio for muons is 1.00 ± 0.01 for both the CC and CF channels.

The muon stub efficiency is determined from $J/\psi \to \mu^+\mu^-$ events using a tag-and-probe method similar to the L1 trigger efficiency determination. One track in the sideband-subtracted J/ψ event is required to have matched muon stub while the other passes fiducial and baseline requirements described in Sec. 4.1.1. The numerator of the efficiency calculation is the number of events where both tracks are matched to muon stubs while the denominator consists of all sideband-subtracted events in the $J/\psi \to \mu^+\mu^-$ dimuon mass signal window. In addition to an efficiency estimated with data, an efficiency using MC events is also calculated. For the final muon stub efficiency the data efficiency is divided by the simulated efficiency to account for geometric losses already counted in the acceptance estimation. As a cross-check of the efficiency estimates the same procedure is used to determine the efficiency of high p_T muons for Z^0 decays. The difference in efficiencies is taken as a systematic uncertainty. The ratio of

the efficiencies, for both the CC and CF channels, is 1.00 ± 0.03.

We also estimate the combined efficiency of our dE/dx and muon likelihood requirements. The efficiency of the muon identification cuts are determined in a $J/\psi \to \mu^+\mu^-$ sample by comparing the signal yield with and without the requirements as a function of p_T. The efficiencies are then convoluted with the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ MC muon p_T distributions to determine a set of average efficiencies. These efficiencies are cross-checked with the $B^+ \to J/\psi K^+$ normalization mode. The difference between the efficiencies in the $B^+ \to J/\psi K^+$ MC and $B^+ \to J/\psi K^+$ sideband-subtracted data is assigned as a systematic uncertainty on the efficiency ratio. The final efficiency ratio and its uncertainty for the muon identification requirements is 1.01 ± 0.03 for both the CC channel and CF channels.

The efficiency of requiring SVXII hits on COT tracks is estimated using $J/\psi \to \mu^+\mu^-$ events in a similar manner to the muon stub and trigger efficiencies. The denominator in these calculations is the number of $J/\psi \to \mu^+\mu^-$ events that are identified as muons and matched to a COT track. This denominator includes the acceptance defined earlier in this section i.e., events must be fiducial to the silicon detectors, and accounts for dead regions of the detector. The numerator adds the requirement of hits in at least three of the SVXII and L00 layers. The efficiency is estimated as a function of muon p_T, the opening angle between the two muons in the transverse plane, and track isolation. The average over these kinematic variables is taken as the final efficiency while the variations in these variables are used for systematic uncertainties. The correlation in the silicon hit efficiency for the two muons, when both muon tracks point into the same silicon ladder, is taken into account. The efficiency ratio for adding SVXII hits,

combining all the uncertainties, is 1.00 ± 0.03 for both the CC and CF channels.

The efficiency of the vertex requirements are estimated using our $B^+ \to J/\psi K^+$ and $B_s^0 \to \mu^+\mu^-$ MC samples. The $B^+ \to J/\psi K^+$ efficiency is checked using sideband-subtracted $B^+ \to J/\psi K^+$ data events. The difference between the data and MC samples is within statistical uncertainties and no additional systematic uncertainty is assigned to the vertex requirement efficiency. The resulting efficiency ratios and its uncertainty for the vertex requirements is 0.986 ± 0.013 for the CC and CF channel.

The kaon COT efficiency is estimated using the same method as the muon COT efficiency and is 0.964 ± 0.016 for both the CC and CF channels. The systematic uncertainty of the kaon COT efficiency takes into account the probability of the kaon interacting hadronically with matter and the uncertainty on the amount of material the kaon traversed for a given path. The efficiency of requiring SVXII hits for the kaon track is directly measured from $B^+ \to J/\psi K^+$ data and thus automatically accounts for the correlation between the muons and the kaon when a combination of the three point to the same silicon ladder. The final efficiencies in the CC and CF channels for the SVXII hit requirement are 0.942 ± 0.002 and 0.948 ± 0.003, respectively. An additional requirement is made for the $B^+ \to J/\psi K^+$ mode on the $B^+ \to J/\psi K^+$ vertex probability. The efficiencies of this requirement is also directly measured in sideband-subtracted $B^+ \to J/\psi K^+$ data and are 0.938 ± 0.006 and 0.919 ± 0.010 for the CC and CF channels, respectively.

The efficiency ratios as well as the kaon and $B^+ \to J/\psi K^+$ vertex probability efficiencies are multiplied to form a combined reconstruction efficiency ratio. The resulting efficiencies for the CC and CF channel and their associated total

uncertainties are shown in Table 4.2.

Neural network efficiency

The NN efficiency for each NN bin is estimated using $B_s^0 \to \mu^+\mu^-$ signal MC. The NN efficiency is only estimated for the $B_s^0 \to \mu^+\mu^-$ mode and not the $B^+ \to J/\psi K^+$ decay because the NN is only applied to the $B_s^0 \to \mu^+\mu^-$ sample. The NN is not applied to the normalization mode because the signal/background is relatively large after the baseline requirements and sideband subtraction. Table 4.2 shows the combined efficiencies for all NN bins in the CC and CF channels. When setting limits or fitting for branching fractions the efficiencies for each NN bin are estimated separately. The highest NN bin is the most sensitive to signal and accounts for the bulk of the NN efficiency. The efficiency for the highest NN bin is approximately 46% for both the CC and CF channels. The final efficiencies for all NN bins separately are shown in Table 4.3.

The NN efficiencies are checked using the normalization mode. The same NN that is applied to the $B_s^0 \to \mu^+\mu^-$ MC is applied to sideband-subtracted $B^+ \to J/\psi K^+$ data and $B^+ \to J/\psi K^+$ MC with slight modifications of the input parameters. Here the J/ψ vertex is used for secondary vertex related input parameters. The efficiencies of various NN bins are now compared between MC and sideband-subtracted data and are shown in Table 4.4. Some difference in performance is observed between the $B^+ \to J/\psi K^+$ simulated and data samples. This difference in performance is likely due to systematic overestimation of vertex information precision in the simulation due to sensor degradation that is not simulated. For most NN bins the difference between MC and sideband-subtracted data is less than 2.5σ. Here σ is the statistical uncertainty for MC

and sideband-subtracted combined. The highest NN bin has the most significant deviation, 3.4% and 7.0% for the CC and CF, respectively.

The $B_s^0 \to \mu^+\mu^-$ MC is reweighted using $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ p_T and isolation distribution, respectively. An additional systematic of 4% is assigned based on the residual mismodeling of B-meson p_T and isolation.

Table 4.2: A summary of the inputs used in equation 4.2 to estimate the branching fraction $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The relative uncertainties are given parenthetically. The single-event-sensitivities, SES's, for the sum of all NN bins, corresponding to $N_{B_s^0} = 1$, are shown in the two last rows.

	CC		CF	
$(\alpha_{B^+}/\alpha_{B_s^0})$	0.307 ± 0.018	$(\pm 6\%)$	0.197 ± 0.014	$(\pm 7\%)$
$(\epsilon^{\text{trig}}_{B^+}/\epsilon^{\text{trig}}_{B_s^0})$	0.99935 ± 0.00012	$(-)$	0.97974 ± 0.00016	$(-)$
$(\epsilon^{\text{reco}}_{B^+}/\epsilon^{\text{reco}}_{B_s^0})$	0.86 ± 0.06	$(\pm 8\%)$	0.84 ± 0.06	$(\pm 9\%)$
$\epsilon^{\text{NN}}_{B_s^0}(\nu_N > 0.70)$	0.915 ± 0.042	$(\pm 4\%)$	0.864 ± 0.040	$(\pm 4\%)$
$\epsilon^{\text{NN}}_{B_s^0}(\nu_N > 0.995)$	0.461 ± 0.021	$(\pm 4.7\%)$	0.468 ± 0.022	$(\pm 4.6\%)$
N_{B^+}	28081 ± 219	$(\pm 1\%)$	12144 ± 153	$(\pm 1\%)$
f_u/f_s	3.55 ± 0.47	$(\pm 13\%)$	3.55 ± 0.47	$(\pm 13\%)$
$\mathcal{B}(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+)$	$(6.01 \pm 0.21) \times 10^{-5}$	$(\pm 4\%)$	$(6.01 \pm 0.21) \times 10^{-5}$	$(\pm 4\%)$
SES (All bins)	2.3×10^{-9}	$(\pm 18\%)$	3.3×10^{-9}	$(\pm 18\%)$

4.3.4 Additional systematic uncertainties

The fragmentation ratio and the $B^+ \to J/\psi K^+$ branching fraction and their associated systematics are both taken from Ref. [3].

Table 4.3: NN bin efficiencies for CC and CF channels.

	CC	CF
$0.7 < \nu_N < 0.76$	2.2%±0.1%	2.3%±0.1%
$0.76 < \nu_N < 0.85$	4.1%±0.1%	4.4±0.1%%
$0.85 < \nu_N < 0.90$	2.9%±0.1%	3.4±0.1%%
$0.90 < \nu_N < 0.94$	4.5%±0.1%	4.7±0.1%%
$0.94 < \nu_N < 0.97$	8.3%±0.1%	6.2±0.1%%
$0.97 < \nu_N < 0.987$	10.9%±0.1%	10.2%±0.1%
$0.987 < \nu_N < 0.995$	12.5%±0.1%	8.5%±0.1%
$0.995 < \nu_N < 1.0$	46.1%±0.3%	46.8%±0.3%

4.3.5 Standard model signal expectations

The expected SM signal yield for $B_s^0 \to \mu^+\mu^-$ decays is estimated for each NN bin using Eq. (4.2), the factors discussed in Sect. 4.3, and the NN bin efficiencies from Table 4.3. There is no expectation of significant SM $B_d^0 \to \mu^+\mu^-$ contribution in the B_d^0 mass window due to the expected branching fraction being a factor of 32 smaller than $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The final expected number of SM events for each NN bin is given in Table 4.5. Combining all NN bins, approximately 1.4 and 1.0 SM $B_s^0 \to \mu^+\mu^-$ events are expected in the CC and CF channels, respectively.

4.4 Background estimation

There are two types of backgrounds that need to be estimated in this analysis. The main source of background in the B_s^0 search comes from combinatorial

Table 4.4: Relative differences in NN bin efficiencies between $B^+ \to J/\psi K^+$ data and MC. A positive (negative) difference indicates that the MC efficiency was higher (lower) than the data efficiency. The differences, normalized to the standard deviation of the MC efficiency (σ), are given in parenthesis.

NN bin	CC	CF
$0.700 < \nu_N < 0.760$	–8.3% (–1.6σ)	–5.3% (–0.7σ)
$0.760 < \nu_N < 0.850$	–8.5% (–2.3σ)	–7.9% (–1.4σ)
$0.850 < \nu_N < 0.900$	4.0% (0.9σ)	–8.2% (–1.3σ)
$0.900 < \nu_N < 0.940$	–0.5% (–0.1σ)	2.4% (0.5σ)
$0.940 < \nu_N < 0.970$	0.1% (0.1σ)	–6.1% (–1.4σ)
$0.970 < \nu_N < 0.987$	2.9% (1.1σ)	0.3% (0.1σ)
$0.987 < \nu_N < 0.995$	4.4% (2.1σ)	–4.1% (–1.0σ)
$0.995 < \nu_N < 1.000$	3.4% (2.6σ)	7.0% (3.7σ)

background. In addition to the combinatorial background there is also a peaking background from $B \to h^+h'^-$ decays, where h and h' are either a pion or kaon. These backgrounds are mainly significant for the B^0_d search due to the shift in invariant mass caused by assuming a muon mass hypothesis. Both backgrounds are estimated separately for the CC and CF channels. The signal dimuon mass region is a 300 GeV/c^2 region between $5.169 < M_{\mu^+\mu^-} < 5.469$ GeV/c^2 and is blinded throughout the background estimation process. The final backgrounds are estimated for a narrower dimuon mass region that is divided into five mass bins and lies between ± 60 MeV/$c^2 \sim 2.5\sigma_m$ around the world average B^0_s (B^0_d) region.

Table 4.5: SM expected signal contribution in each NN bin for the $B_s^0 \to \mu^+\mu^-$ search.

NN bin	CC	CF
$0.700 < \nu_N < 0.760$	0.04 ± 0.01	$0.03\pm < 0.01$
$0.760 < \nu_N < 0.850$	0.07 ± 0.01	0.05 ± 0.01
$0.850 < \nu_N < 0.900$	0.05 ± 0.01	0.04 ± 0.01
$0.900 < \nu_N < 0.940$	0.07 ± 0.01	0.05 ± 0.01
$0.940 < \nu_N < 0.970$	0.10 ± 0.02	0.07 ± 0.01
$0.970 < \nu_N < 0.987$	0.13 ± 0.02	0.11 ± 0.02
$0.987 < \nu_N < 0.995$	0.20 ± 0.04	0.09 ± 0.02
$0.995 < \nu_N < 1.000$	0.75 ± 0.13	0.52 ± 0.10

4.4.1 Combinatorial backgrounds

The combinatorial background is estimated using the data mass sideband. The NN output is divided into 8 bins with the optimization process discussed in Sec. 4.6.1. For each NN bin the dimuon mass sideband regions of $5.009 < M_{\mu^+\mu^-} < 5.169$ GeV/c^2 and $5.469 < M_{\mu^+\mu^-} < 5.969$ GeV/c^2 are fitted to a first order polynomial with a fixed slope but a free floating normalization. Events below $M_{\mu^+\mu^-} = 5.009$ GeV/c^2 are not used due to $b \to \mu^+\mu^- X$ contamination. These $b \to \mu^+\mu^- X$ events tend to have a lower dimuon mass due to the missing contribution from X. A fixed slope is used because ν_N is found to have no mass correlation and thus the slope should be constant with respect to the ν_N requirement. The fixed slope is estimated using all the NN bins combined with $\nu_N > 0.7$. The final combinatorial background estimates are determined in the signal region for five mass bins centered around the B_s^0 world average mass us-

ing $B = N \times (a + b \times (x - 480))$. Here a is a constant offset determined from the sideband events for which $\nu_N > 0.7$, b is the fixed slope, 480 is the average dimuon mass in our sideband and signal window in GeV/c^2, x is the center of the mass bin considered in GeV/c^2, and B is the estimated background for that bin. The fixed slope and offset are determined separately for the CC and CF channels. We obtain $a_{CC} = 0.030 \pm 0.001$, $a_{CF} = 0.031 \pm 0.001$, $b_{CC} = -0.012 \pm 0.003$, $b_{CF} = -0.017 \pm 0.003$ from the fit shown in Fig. 4.13.

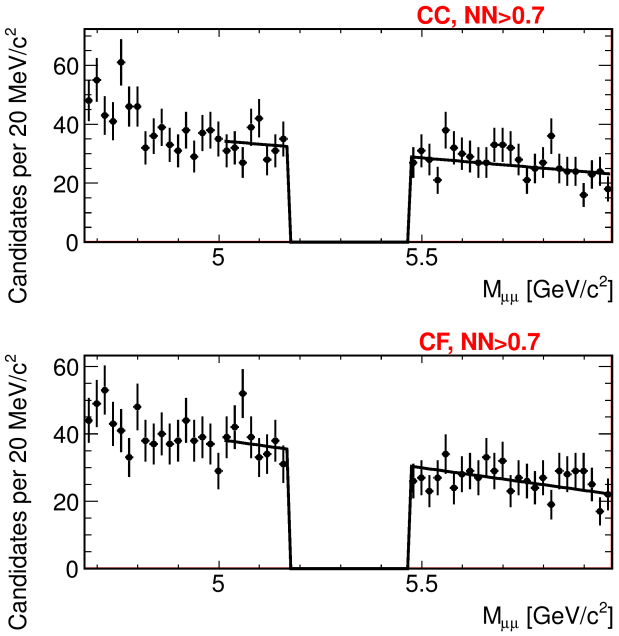

Figure 4.13: Dimuon mass distributions for $\nu_N > 0.7$ for the CC and CF channels with the signal region blinded.

The same procedure and same sidebands are used for the B_d^0 meson using five mass bins centered around the B_d^0 mass. Figures 4.14 and 4.15 show the dimuon mass distribution and the fit results for the CC and CF channel, respectively. The statistical uncertainty from the normalization and the systematic uncertainties from the slope determination are propagated into the background estimate uncertainties. The final combinatorial background estimates are given

in Tables 4.6 and 4.7.

The highest three NN bins are fitted to alternative functions to assess the systematic uncertainty based on the $M_{\mu^+\mu^-}$ distribution shape uncertainty. The highest three bins are more susceptible to contamination from partially reconstructed decays ($b \to \mu^+\mu^- X$) because these decays have similar kinematics to the $B_s^0 \to \mu^+\mu^-$ decay. This occurs when the additional hadron, X, has a low momentum causing a invariant dimuon mass close to the B_d^0 or B_s^0 mass. Contributions from these decays can cause a different dimuon mass shape in the most sensitive NN bins. A first order polynomial with free floating slope for each NN bin is used on the same dimuon mass sideband as the fixed slope fit. In addition, an exponential fit to the entire sideband mass region, $4.669 < M_{\mu^+\mu^-} < 5.169\,\mathrm{GeV}/c^2$ and $5.469 < M_{\mu^+\mu^-} < 5.969\,\mathrm{GeV}/c^2$, is performed (Figs. 4.16 and 4.17). The largest difference between the background estimates from these alternative fits and the standard fit is assigned as a systematic uncertainty across all five mass bins. The final relative uncertainties range from ±19% (±3%) to ±43% (±42%) for the CC (CF) channel.

4.4.2 Peaking backgrounds

A peaking background from two-body hadronic B-meson decays also contributes to possible backgrounds in the signal region. These $B \to h^+h'^-$ decays, where h and h' are either a $\pi^\pm$ or $K^\pm$ are a factor of 10 smaller than the combinatorial background for the B_s^0 search window while they comprise about half the total background events in the B_d^0 window. Decays involving other hadrons such as $\Lambda_b \to p\pi^-$ tend to be above the dimuon mass signal region. These hadrons also

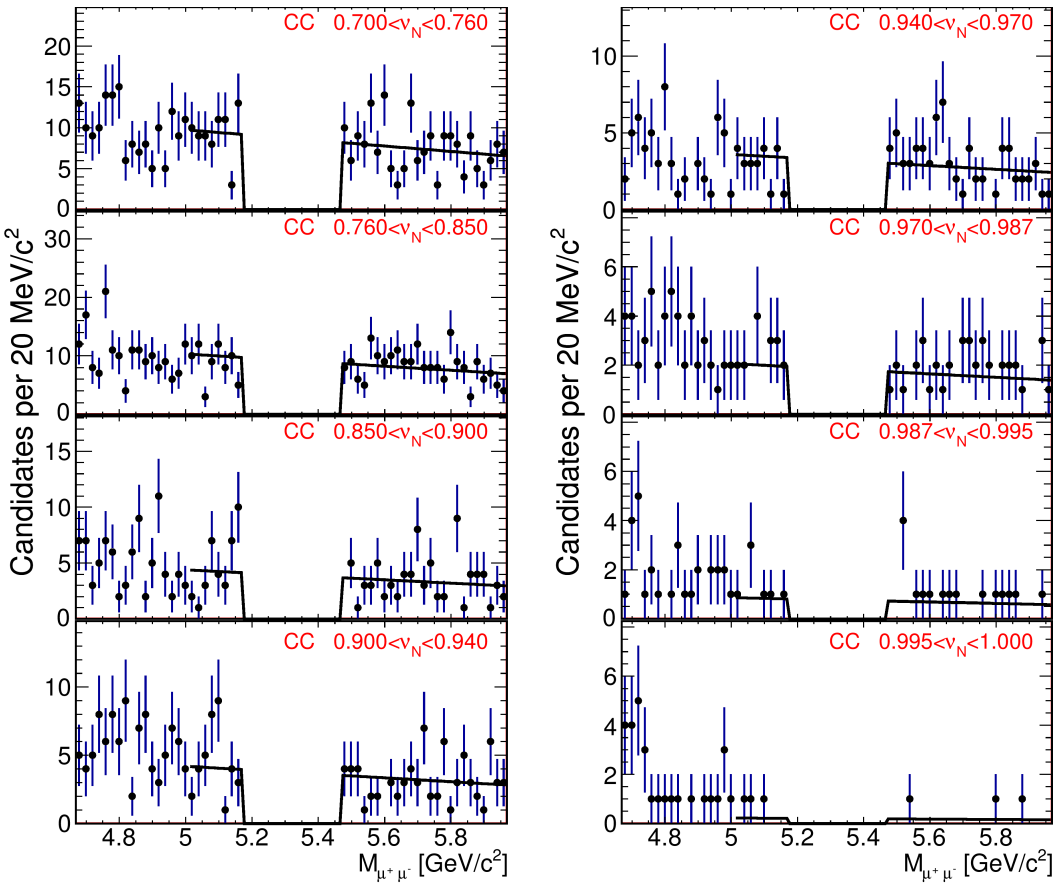

Figure 4.14: Dimuon mass distributions in all 8 NN bins for the CC channel the with extended signal region blinded.

have significantly lower productions rates than B-mesons while protons are rejected at a higher rate by the particle identification requirements than pions and kaons [43].

The rate at which pions and kaons pass all our muon requirements is extracted with a pure sample of kaons and pions from D^*-tagged $D^0 \to K^-\pi^+$ decays. These decays yield two same-sign pions, one from the $D^{*+} \to D^0\pi^+$ decay and one from the subsequent D^0 decay, and an oppositely-signed kaon. Due to the small difference between $M_{D^{*+}}$ and $M_{D^0\pi^+}$, the pion from the D^{*+} decay has a much lower p_T spectrum. The $D^0 \to K^-\pi^+$ mode is a Cabibbo-favored decay

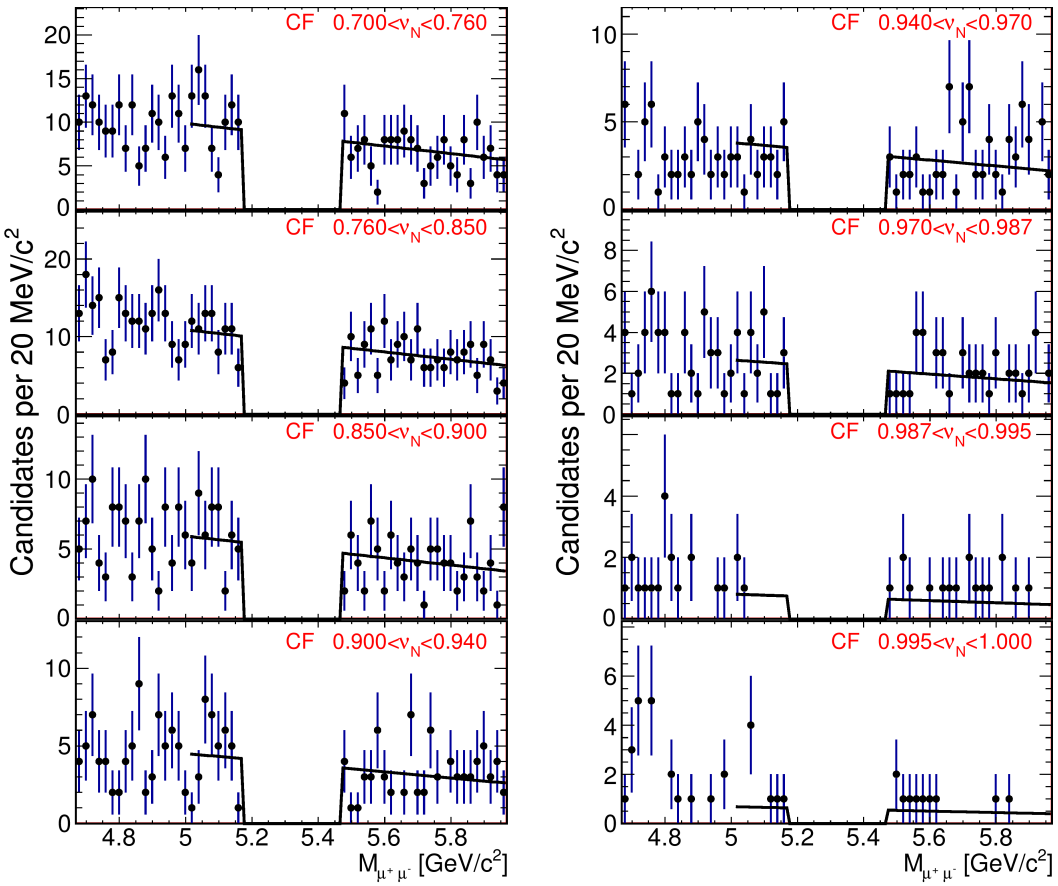

Figure 4.15: Dimuon mass distributions in all 8 NN bins for the CF channel the with extended signal region blinded.

with a branching fraction of $(3.87 \pm 0.05) \times 10^{-2}$ [3].

We use the first 7 fb^{-1} of data collected with the two track trigger (TTT), where the two triggered tracks are the two tracks from $D^0 \to K^-\pi^+$ candidates. This trigger requires two tracks with a displaced secondary vertex, by demanding the tracks have $0.1 < |d_0| < 1$ mm and $L_T > 200$ μm. The trigger also makes the following requirements: $2^\circ < \Delta\phi_0 < 90^\circ$, where $\Delta\phi_0$ is the angular separation of the two tracks; $p_T > 2.0$ GeV$/c$ for each of the two tracks; $|\Delta z_0| < 5$ cm, where Δz_0 is the difference in z coordinate of the two tracks; and $|\eta| < 1.2$ for each of the two tracks.

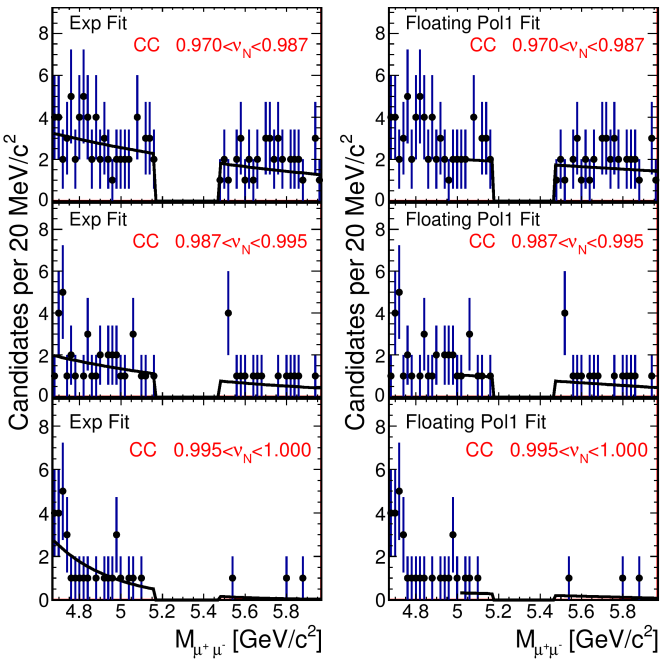

Figure 4.16: Alternative fits to the dimuon mass distributions of the three highest NN bins in the CC channel.

Additional requirements are made on events that pass the trigger requirements. The higher p_T tracks are required to have muon stubs in either the central or forward muon detectors matched to COT tracks. We demand that both tracks individually as well as the D^0 candidate, reconstructed from the pion and kaon from the secondary vertex, have $|\eta| < 1$, the χ^2 returned from the secondary vertex fit be no greater than 15, and the impact parameter of the D^0 candidate relative to the primary vertex be less than 100 μm. The third track from the primary vertex, the soft pion from the $D^{*+} \rightarrow D^0\pi^+$ decay, is required to have $p_T > 0.4$ GeV/c, $|z_0| < 1.5$ cm, $|d_0| < 600$ μm, and must have the same charge as the pion from the D^0 decay. Finally, we require $1.77 < M_{D^0} < 1.97$ GeV/c^2, where M_{D^0} is the D^0 candidate mass, and $144 < M_{D^{*+}} - M_{D^0} < 147$ MeV/c^2, where $M_{D^{*+}}$ is the reconstructed D^{*+} candidate mass. These selection criteria yield a very pure sample (> 99%) of kaons and pions from the D^0 decay which we use to estimate

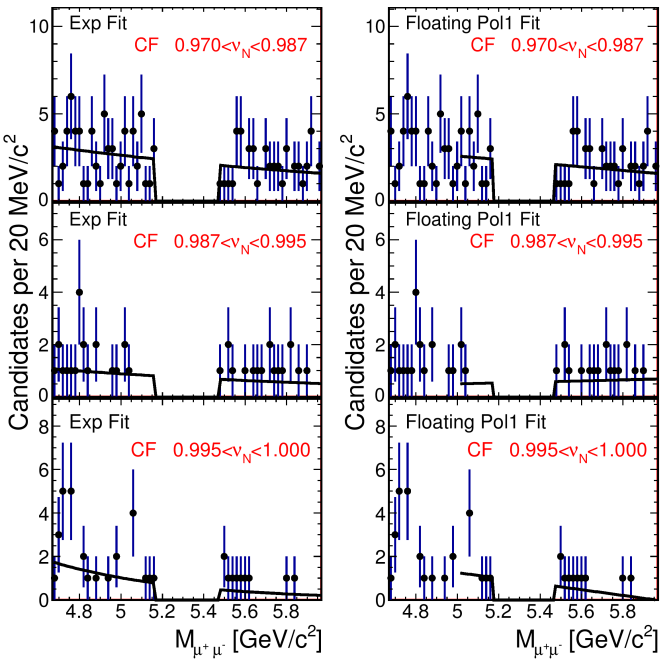

Figure 4.17: Alternative fits to the dimuon mass distributions of the three highest NN bins in the CF channel.

the efficiency of our muon identification requirements on kaons and pions, or fake rates.

The fake rates were measured for central and forward muons and separately for π^+, π^-, K^+, and K^- and binned in p_T. However, due to the near identical fake rates of π^+ and π^- the two fakes rates were combined to yield a $\pi^\pm$ fake rate. A dependence of the fake rates on instantaneous luminosity is found requiring a binning of $(0 < \mathcal{L} < 40) \times 10^{30}\text{cm}^{-2}\text{s}^{-1}$, $(40 < \mathcal{L} < 80) \times 10^{30}\text{cm}^{-2}\text{s}^{-1}$, $(80 < \mathcal{L} < 120) \times 10^{30}\text{cm}^{-2}\text{s}^{-1}$, and $(120 < \mathcal{L}) \times 10^{30}\text{cm}^{-2}\text{s}^{-1}$. The fake rates change from 20% to a factor of three due to instantaneous luminosity.

Within these p_T and luminosity bins the number of events that pass our muon identification is estimated by fitting the M_{D^0} mass distribution to a function that is the sum of a Gaussian distribution and a first order polynomial.

Table 4.6: Estimated number of combinatorial background events with statistical uncertainty for B_s^0 signal window for all NN bins.

NN bin \ Mass bins (GeV/c^2)	5.310–5.334	5.334–5.358	5.358–5.382	5.382–5.406	5.406–5.430
CC					
0.700< ν_N <0.760	10.42±0.72	10.33±0.71	10.23±0.70	10.14±0.70	10.04±0.69
0.760< ν_N <0.850	11.02±0.74	10.92±0.74	10.82±0.73	10.72±0.72	10.62±0.71
0.850< ν_N <0.900	4.69±0.46	4.65±0.45	4.61±0.45	4.56±0.44	4.52±0.44
0.900< ν_N <0.940	4.49±0.45	4.45±0.44	4.41±0.44	4.37±0.43	4.33±0.43
0.940< ν_N <0.970	3.85±0.41	3.81±0.41	3.78±0.40	3.74±0.40	3.71±0.39
0.970< ν_N <0.987	2.21±0.30	2.19±0.30	2.17±0.30	2.14±0.30	2.12±0.29
0.987< ν_N <0.995	0.92±0.19	0.91±0.19	0.91±0.19	0.90±0.19	0.89±0.19
0.995< ν_N <1.000	0.24±0.10	0.24±0.10	0.24±0.10	0.23±0.10	0.23±0.10
CF					
0.700< ν_N <0.760	10.18±0.72	10.05±0.71	9.93±0.70	9.80±0.69	9.68±0.68
0.760< ν_N <0.850	11.21±0.76	11.08±0.75	10.94±0.74	10.80±0.73	10.66±0.72
0.850< ν_N <0.900	6.11±0.54	6.03±0.53	5.96±0.52	5.88±0.52	5.81±0.51
0.900< ν_N <0.940	4.65±0.46	4.59±0.46	4.54±0.45	4.48±0.44	4.42±0.44
0.940< ν_N <0.970	3.94±0.42	3.90±0.42	3.85±0.41	3.80±0.41	3.75±0.40
0.970< ν_N <0.987	2.74±0.35	2.71±0.34	2.67±0.34	2.64±0.34	2.61±0.33
0.987< ν_N <0.995	0.83±0.19	0.82±0.18	0.81±0.18	0.80±0.18	0.79±0.18
0.995< ν_N <1.000	0.71±0.17	0.70±0.17	0.69±0.17	0.68±0.17	0.67±0.16

The number of events that fail our muon identification is taken from a fit of the M_{D^0} mass distribution to the sum of two Gaussian distributions and a first order polynomial. The muon identification requirements are identical to the ones used in the $B_s^0 \to \mu^+\mu^-$ data set and consist of a muon likelihood requirement and a dE/dx requirement. Fits with one Gaussian have the mean of the Gaussian distribution fixed to the world average of the D^0 mass while the width is set to a width determined for each p_T bin using events without the muon iden-

Table 4.7: Estimated number of combinatorial background events for B_d^0 signal window for all NN bins.

NN bin \ Mass bin (GeV/c^2)	5.219–5.243	5.243–5.267	5.267–5.291	5.291–5.315	5.315–5.339
CC					
0.700< ν_N <0.760	10.78±0.74	10.69±0.74	10.59±0.73	10.50±0.72	10.40±0.72
0.760< ν_N <0.850	11.41±0.77	11.30±0.76	11.21±0.76	11.10±0.75	11.00±0.74
0.850< ν_N <0.900	4.85±0.47	4.81±0.47	4.77±0.46	4.72±0.46	4.68±0.46
0.900< ν_N <0.940	4.64±0.46	4.60±0.46	4.56±0.45	4.52±0.45	4.48±0.44
0.940< ν_N <0.970	3.98±0.42	3.95±0.42	3.91±0.42	3.88±0.41	3.84±0.41
0.970< ν_N <0.987	2.28±0.32	2.26±0.31	2.24±0.31	2.22±0.31	2.20±0.30
0.987< ν_N <0.995	0.95±0.20	0.95±0.20	0.94±0.20	0.93±0.20	0.92±0.19
0.995< ν_N <1.000	0.25±0.10	0.25±0.10	0.24±0.10	0.24±0.10	0.24±0.10
CF					
0.700< ν_N <0.760	10.65±0.75	10.52±0.74	10.40±0.73	10.27±0.72	10.15±0.72
0.760< ν_N <0.850	11.73±0.80	11.60±0.79	11.46±0.78	11.32±0.77	11.18±0.76
0.850< ν_N <0.900	6.39±0.56	6.31±0.55	6.24±0.55	6.16±0.54	6.09±0.53
0.900< ν_N <0.940	4.87±0.48	4.81±0.48	4.75±0.47	4.70±0.47	4.64±0.46
0.940< ν_N <0.970	4.13±0.44	4.08±0.44	4.03±0.43	3.98±0.43	3.94±0.42
0.970< ν_N <0.987	2.87±0.36	2.83±0.36	2.80±0.35	2.77±0.35	2.73±0.35
0.987< ν_N <0.995	0.87±0.20	0.86±0.19	0.85±0.19	0.84±0.19	0.83±0.19
0.995< ν_N <1.000	0.74±0.18	0.73±0.18	0.72±0.18	0.71±0.17	0.70±0.17

tification requirements. The number of events is extracted from the fitted amplitude of the Gaussian for the one-Gaussian fit and the sum of amplitudes for the two-Gaussian fit. Figures 4.18 and 4.19 show the D^0 mass distributions with the accompanying fits for a lower and higher kaon p_T bin for central muons combining all luminosity bins.

As a cross-check, the fake rates, binned in p_T and luminosity, were applied to the D^*-tagged sample as weights and counted in run number bins. These sum

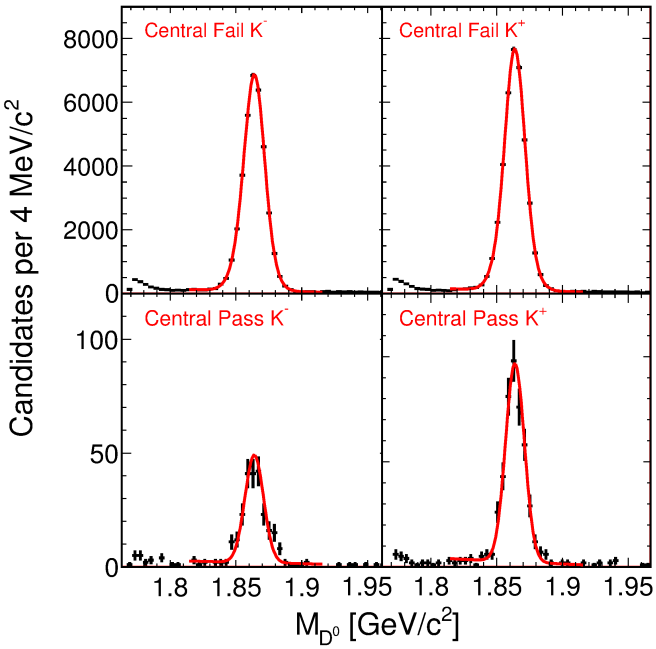

Figure 4.18: D^0 mass distributions and their associated fits for central kaons with $2.0 < p_T < 2.8$ GeV/c.

of weights were compared to the actual number of fakes in each run number bin resulting in differences of up to 20%. This 20% difference is assigned as a systematic uncertainty and accounts for the largest contribution to the fake rate uncertainty. In the final determination of the total $B \to h^+h'^-$ contribution a weighted average of fake rates is used based on the luminosity profile of the dimuon mass sideband events. The luminosity-averaged fake rates are shown in Fig. 4.20.

The B-meson mass shape and p_T distributions are modeled by MC simulations. These distributions are necessary to properly apply the fake rates and to estimate the contribution in each of the five B_s^0 (B_d^0) mass bins. The expected number of peaking background events for a specific two-body hadronic decay is given by

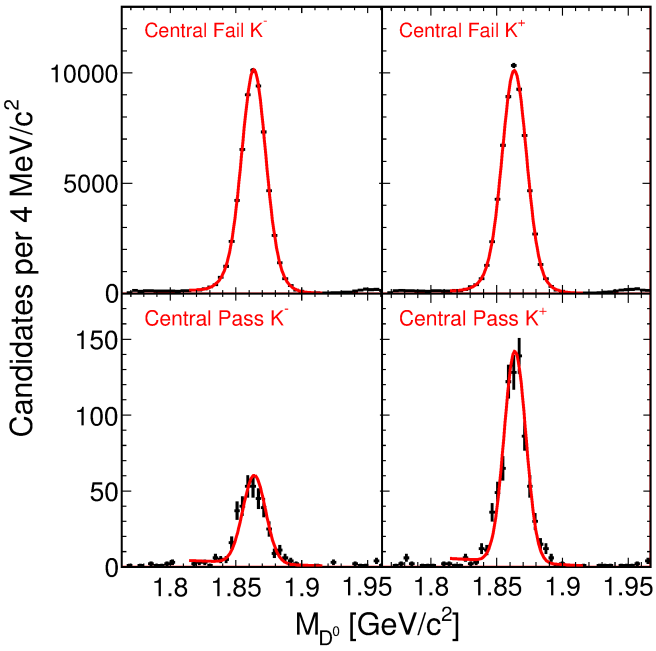

Figure 4.19: D^0 mass distributions and their associated fits for central kaons with $6.0 < p_T < 8.0$ GeV/c.

$$N_b = F_{sb} \cdot R_b \cdot \frac{\epsilon_b}{\epsilon_s} \cdot \frac{\mathcal{B}_b}{S_S},$$

where F_{sb} is the ratio of fragmentation fractions which is either unity, f_u/f_s, or f_s/f_u. If we are considering B^0_d hadronic backgrounds in the B^0_d search window this factor is equal to unity. However, if we consider B^0_d hadronic background events in the B^0_s search region $F_{sb} = f_u/f_s$, while $F_{sb} = f_s/f_u$ for the inverse. f_u/f_s is taken from Ref. [3] and has an uncertainty of 13%. R_b is the average decay mode fake rate, the rate at which both final state hadrons are misidentified as muons. This average is obtained by averaging the fake rate over the double-track p_T distribution of $B \to h^+h'^-$ MC events. An uncertainty of 35% based on the single track uncertainty of 20% is assigned to this. The efficiencies ϵ_s and ϵ_b are the mass bin efficiency, that is the efficiency of a signal $B^0_{s,d} \to \mu^+\mu^-$ event occurring in one of the five mass bins, and the efficiency for a $B \to h^+h'^-$ back-

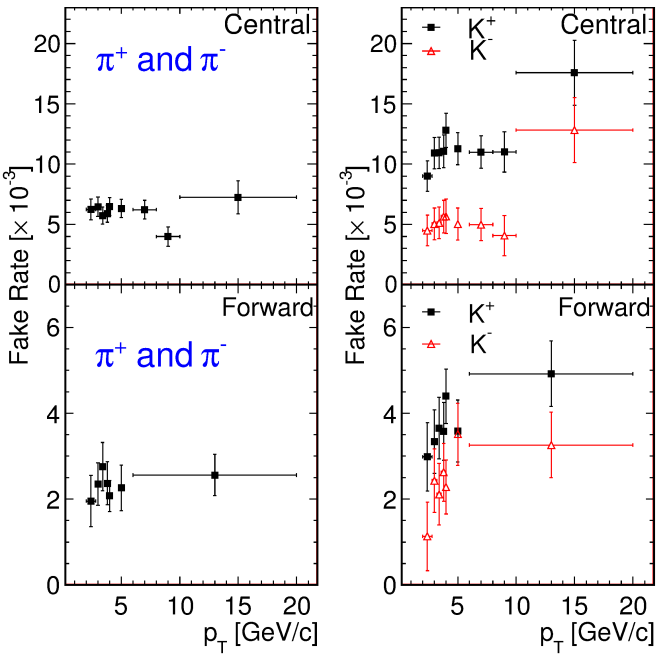

Figure 4.20: Fake rates as a function of p_T and averaged over luminosity bins for central (left two columns) and forward (right two columns) muon detectors.

ground event falling in a specific mass bin. These efficiencies are determined with MC simulations. The $\mathcal{B}_b$ is the branching fraction of a specific background mode. The value and associated uncertainty of $\mathcal{B}_b$ are current world averages from Ref. [3]. For unobserved processes, we use the current upper limit and assign a 100% uncertainty. Finally, S_s is the single-event-sensitivity which is discussed in more detail in Sec. 4.3. The final estimated peaking background contribution is given in Tables 4.8 and 4.9.

Our $B_s^0 \to \mu^+\mu^-$ MC is used to cross-check whether the p_T dependence of the fake rates could have an effect on the ν_N distribution. To test this, a ν_N distribution weighted by the fake rates is compared to an unweighted ν_N distribution yielding no significant difference in ν_N distributions.

Table 4.8: $B \rightarrow h^+h'^-$ background estimates for B_s^0 signal window for all NN bins. The contribution are negligibly small in all the lower NN bins.

Mass bin (GeV/c^2) / NN bin	5.310–5.334	5.334–5.358	5.358–5.382	5.382–5.406	5.406–5.430
CC					
0.700< ν_N <0.760	0.003± < 0.001	0.001± < 0.001	–	–	–
0.760< ν_N <0.850	0.006±0.001	0.002± < 0.001	0.001± < 0.001	–	–
0.850< ν_N <0.900	0.005±0.001	0.002± < 0.001	0.001± < 0.001	–	–
0.900< ν_N <0.940	0.007±0.001	0.003± < 0.001	0.001± < 0.001	–	–
0.940< ν_N <0.970	0.011±0.001	0.003± < 0.001	0.001± < 0.001	–	–
0.970< ν_N <0.987	0.013±0.002	0.005±0.001	0.002± < 0.001	0.001± < 0.001	–
0.987< ν_N <0.995	0.019±0.002	0.007±0.001	0.002± < 0.001	0.001± < 0.001	–
0.995< ν_N <1.000	0.074±0.010	0.026±0.003	0.009±0.001	0.003± < 0.001	0.001± < 0.001
CF					
0.700< ν_N <0.760	0.001± < 0.001	–	–	–	–
0.760< ν_N <0.850	0.002± < 0.001	0.001± < 0.001	–	–	–
0.850< ν_N <0.900	0.002± < 0.001	0.001± < 0.001	–	–	–
0.900< ν_N <0.940	0.002± < 0.001	0.001± < 0.001	–	–	–
0.940< ν_N <0.970	0.003± < 0.001	0.001± < 0.001	–	–	–
0.970< ν_N <0.987	0.004±0.001	0.002± < 0.001	0.001± < 0.001	–	–
0.987< ν_N <0.995	0.004±0.001	0.002± < 0.001	0.001± < 0.001	–	–
0.995< ν_N <1.000	0.021±0.003	0.009±0.001	0.003± < 0.001	0.002± < 0.001	–

4.4.3 Background estimate checks with control samples

Our signal data sample consists of two opposite-sign muons with $\lambda > 0$. We form four independent control samples described below.

OS–: opposite-sign muon pairs, passing the baseline requirements with $\lambda < 0$;

SS+: same-sign muon pairs, passing looser baseline requirements with $\lambda > 0$;

Table 4.9: $B \rightarrow h^+h'^-$ background estimates for B_d^0 signal window for all NN bins.

Mass bin (GeV/c^2) / NN bin	5.219–5.243	5.243–5.267	5.267–5.291	5.291–5.315	5.315–5.339
CC					
0.700< ν_N <0.760	0.015±0.002	0.013±0.001	0.011±0.001	0.006±0.001	0.002± < 0.001
0.760< ν_N <0.850	0.027±0.003	0.027±0.003	0.019±0.002	0.011±0.002	0.004±0.001
0.850< ν_N <0.900	0.022±0.002	0.019±0.002	0.014±0.002	0.008±0.001	0.003± < 0.001
0.900< ν_N <0.940	0.030±0.003	0.029±0.003	0.022±0.003	0.013±0.002	0.004±0.001
0.940< ν_N <0.970	0.047±0.005	0.039±0.004	0.031±0.004	0.016±0.002	0.005±0.001
0.970< ν_N <0.987	0.060±0.006	0.052±0.006	0.040±0.005	0.023±0.003	0.009±0.001
0.987< ν_N <0.995	0.084±0.008	0.083±0.009	0.061±0.008	0.033±0.004	0.011±0.001
0.995< ν_N <1.000	0.325±0.032	0.298±0.031	0.221±0.028	0.126±0.017	0.050±0.006
CF					
0.700< ν_N <0.760	0.004± < 0.001	0.004± < 0.001	0.003± < 0.001	0.002± < 0.001	0.001± < 0.001
0.760< ν_N <0.850	0.007±0.001	0.008±0.001	0.006±0.001	0.004±0.001	0.001± < 0.001
0.850< ν_N <0.900	0.006±0.001	0.006±0.001	0.005±0.001	0.003± < 0.001	0.001± < 0.001
0.900< ν_N <0.940	0.008±0.001	0.008±0.001	0.007±0.001	0.004±0.001	0.001± < 0.001
0.940< ν_N <0.970	0.011±0.001	0.011±0.001	0.009±0.001	0.005±0.001	0.002± < 0.001
0.970< ν_N <0.987	0.017±0.002	0.018±0.002	0.015±0.002	0.008±0.001	0.003± < 0.001
0.987< ν_N <0.995	0.014±0.001	0.015±0.002	0.012±0.002	0.007±0.001	0.002± < 0.001
0.995< ν_N <1.000	0.078±0.008	0.084±0.009	0.065±0.008	0.038±0.005	0.014±0.002

Loosening of the trigger matching is required because the dimuon trigger demands an opposite-sign dimuon pair.

SS–: same-sign muon pairs, passing looser baseline requirements with $\lambda < 0$;

FM+: opposite-sign fake-muon pairs, at least one track of which is required to *fail* the muon likelihood or dE/dx requirement, with $\lambda > 0$.

The OS– sample is enhanced with short lived decays from combinatorial backgrounds. The same-sign samples are dominated by sequential semileptonic decays and muons that are combined from independent processes. The processes that make up these control samples are representative of the backgrounds present in the OS+ sideband. We use the same background estimation process used in the signal sample for these control samples. For all but the FM+ only the combinatorial backgrounds are estimated due to the dominance of this background over the peaking background. The FM+ has an enhanced sample of $B \to h^+h'^-$ background due to the reversal of the muon identification requirements. In the FM+, the two tracks pass the standard CDF muon identification algorithm but fail our muon likelihood and dE/dx requirements. For this sample, we estimate both the combinatorial and the $B \to h^+h'^-$ background. Due to the reversed muon requirements new fake rates, shown in Fig. 4.21, are evaluated with the reversed muon identification using the same method as described in Sec. 4.4.2. The backgrounds are evaluated for the extended blinded signal mass region for each NN bin and compared to the observed number of events. A p-value is calculated using the mean and uncertainty of the background estimates. The resulting comparisons for all NN bins and control samples is shown in Table 4.10.

The comparisons in these four control samples give us confidence in our background estimation method. Overall the comparison show no large deviation between the predicted number of background events and the observed number.

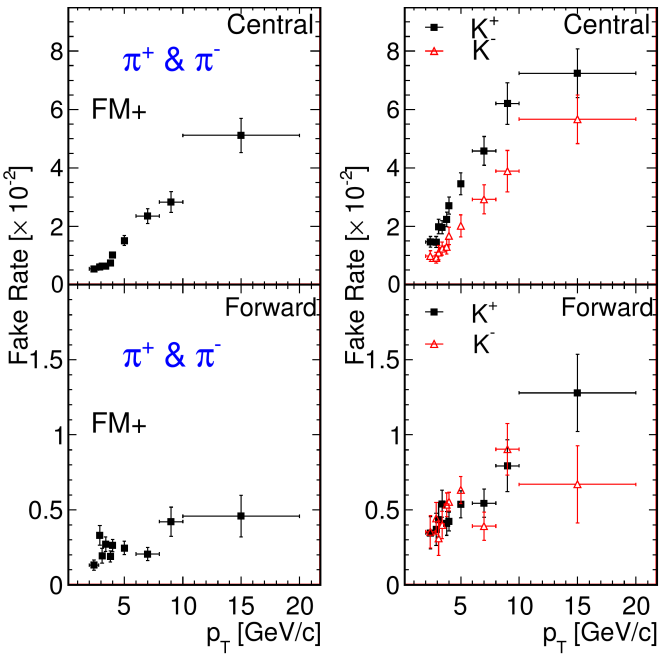

Figure 4.21: Fake rates as a function of p_T and averaged over luminosity bins for central (top) and forward (bottom) muon detectors in the FM+ control sample.

4.5 Systematic uncertainties

The systematic uncertainties in this analysis are either related to efficiencies, acceptances, normalization factors, peaking background estimates, or combinatorial background estimates. Table 4.11 summarizes all systematic uncertainties used in the search for $B^0_{s,d} \to \mu^+\mu^-$.

The dominant systematic uncertainty among the efficiencies, acceptances, and normalization factors is the uncertainty on the ratio of fragmentation fractions, f_u/f_s [3], which induces a 13% effect. The second largest systematic uncertainty is about a factor of two smaller and is due to the acceptance ratio (Sec. 4.3.1).

An additional systematic uncertainty, the $B^0_{s,d}$ mass shape uncertainty, is as-

signed based on the probability for a $B^0_{s,d}$ candidate to be reconstructed into the signal mass window. This uncertainty is based on the world average $B^0_{s,d}$ mass uncertainty, our mass scale, and our mass resolution. The final uncertainty ranges between 1% and 9% depending on mass.

The leading systematic uncertainty for the peaking background is due to the fake rate (R_b) uncertainty. The fake rates are applied to all $B \to h^+h'^-$ modes on both tracks. A final double-track 35% relative systematic uncertainty is assigned to the predicted $B \to h^+h'^-$ yield. The source of the fake rate uncertainties is discussed in detail in Sec. 4.4.2. In addition to the fake rate uncertainty, an uncertainty associated with the branching fraction of the particular $B \to h^+h'^-$ decay is taken from Ref. [3]. Branching fractions for which only upper limits have been set are assigned a 100% relative uncertainty. These decays, however, contribute a small fraction to the total $B \to h^+h'^-$ background. The leading contribution to the $B \to h^+h'^-$ background in the B^0_d signal mass window comes from $B^0_d \to K^+\pi^-$, $B^0_d \to \pi^+\pi^-$, and $B^0_s \to K^+K^-$. The branching fractions of these decays have relative uncertainties of 3%, 4%, and 16%, respectively. The $B \to h^+h'^-$ background estimates that require f_u/f_s are treated as correlated with the normalization factor since the same source of f_u/f_s is used.

The leading systematic uncertainty, up to 43%, in the combinatorial background estimates is due to the dimuon mass sideband shape uncertainty assigned to the three highest NN bins. The mass sideband shape uncertainty is discussed in Sec. 4.4.1. Another large source of systematic uncertainty, up to 42%, in the higher NN bins is due to finite sample size in the sidebands. A relatively small contribution ($\sim$ 6%) to the total combinatorial background systematic uncertainty arises from the ν_N dimuon mass slope uncertainty.

In our statistical interpretation of results all, the above systematic uncertainties are taken as nuisance parameters with Gaussian constraints. The correlations between the various systematic uncertainties are also taken into account. Combinatorial background estimates across the five mass bins are correlated because they are estimated using the same slope and offset. Peaking backgrounds are treated as correlated across all NN, mass, and muon topology bins due to the use of common fake rates. All signal efficiency ratios except for the kaon SVXII reconstruction efficiency and the B^+ vertex probability efficiency, which are measured directly in the data, are treated as correlated across NN, mass, and muon topology bins due to the use of a common MC sample.

4.6 Statistical interpretation and analysis optimization

This section will discuss the NN binning optimization as well as the methods used to statistically interpret the results.

4.6.1 Analysis optimization

The expected limit is the figure of merit for the NN optimization. The expected limit is calculated using a modified frequentist methodology, the CLs method [44], combining all NN, mass, and muon topology (CC and CF) bins while taking correlations between all the bins into account.

An ensemble of MC simulated data are generated for both a background-only, b, and a signal-plus-background, $s + b$, hypothesis by varying nuisance parameters within Gaussian constraints. Our systematic uncertainties on the

background estimates and signal efficiencies are treated as nuisance parameters. Likelihood functions for both the background-only and signal-plus-background, $\mathcal{L}(b)$ and $\mathcal{L}(s+b)$, hypotheses are constructed. The likelihoods are minimized by varying the nuisance parameters. This minimization is done for both the $s+b$ and b simulated data. The likelihood functions are defined as the product of Poisson probabilities of all 80 bins, $\mathcal{L} = \prod_{i=1}^{80} t_i^{n_i} e^{-t_i}/n_1!$, where t_i is the expected number of events according to a specific hypothesis and bin and n_i is the number of simulated experiment events for a specific bin. After the fits have been performed for each simulated experiment a log likelihood ratio (LLR) is constructed, $2\ln Q = \mathcal{L}(s+b)/\mathcal{L}(b)$, where $\mathcal{L}(s+b)$ and $\mathcal{L}(b)$ are the Poisson probabilities for the signal-plus-background and background-only hypothesis, respectively.

For each background-only LLR, the fraction of simulated experiments with a greater LLR, or p-value, is calculated for the $s+b$ and b simulated data. A ratio of p-values is constructed, $CL_s = CL_{s+b}/CL_b$, where CL_{s+b} is the p-value for the $s+b$ simulated data and CL_b is the p-value for background-only simulated data. Finally, the median value of CL_s is calculated. This process is repeated for a range of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$, which varies the expected signal in the $s+b$ hypothesis, to determine the expected 90% and 95% confidence level (C.L.) limit.

We initially chose many NN bins but combined bins that had similar expected signal-to-background. The signal contribution is estimated using our $B_s^0 \to \mu^+\mu^-$ MC sample while the background expectations are estimated using the data mass sideband and $B \to h^+h'^-$ MC simulations. The NN bin boundaries are varied to optimize the expected limits. This is done starting from the highest NN boundary down. The highest NN bin boundary is optimized for the low-

est limit while considering the statistical power and systematic uncertainties. The uncertainty on the combinatorial background estimates rise steeply as the NN boundary increases and the sideband regions become depleted of events. A boundary value of 0.995 is chosen to avoid high systematic uncertainties that could degrade the analysis performance. Once the highest NN bin is chosen the minima become relatively shallow and only vary the expected limit about 5% for the various configurations. Our final configuration results in 8 NN bins with bounds: $0.700 < \nu_N < 0.760$, $0.760 < \nu_N < 0.850$, $0.850 < \nu_N < 0.900$, $0.900 < \nu_N < 0.940$, $0.940 < \nu_N < 0.970$, $0.970 < \nu_N < 0.987$, $0.987 < \nu_N < 0.995$, and $0.995 < \nu_N < 1.000$. The final expected limits are 1.3×10^{-8} and 1.3×10^{-8} at 95% C.L. for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$, respectively.

4.6.2 Observed limits

The observed limits are calculated with the CLs methodology in a similar manner to the expected limits discussed above. As with the expected limit, simulated data are generated for the $s + b$ and b hypotheses. The same LLR is constructed with the same Poisson likelihood functions to fit to the simulated data. For the observed limit, however, the p-value is determined by also fitting the observed data to the likelihood functions. Two separate fits are performed, one for the $s + b$ and one for the b hypothesis, and are used to construct an observed LLR. The p-values for the $s + b$ and b hypotheses are taken to be the fraction of $s + b$ and b simulated data, respectively, with an LLR greater than the LLR for the observed data. As with the expected limit estimation, the two p-values are combined to form the CL_s and calculated for a range of branching fractions.

4.6.3 Hypothesis testing

We employ an ensemble of background-only simulated experiments to assess the compatibility of an observation with the predicted background as a p-value. The same Poisson likelihood functions as in the expected limit calculation are used. The likelihood functions are minimized for both the observed data and the ensemble of simulated data. Unlike in the expected limit procedure the signal yield in the $s+b$ hypothesis is left to float freely while the background yields are the same as for the background-only hypothesis. Once the fits are complete an LLR is constructed for the observed data and each simulated experiment. Finally the p-value is calculated using the number of simulated experiments with an LLR greater than the LLR for the observed data.

In the event of a significant excess, as quantified by the background-only p-value, the data will be used to access compatibility with the SM+background (SM + b) hypothesis. The SM + b hypothesis accounts for the expected standard model contribution. In this case an ensemble of SM+background simulated experiments is produced. All simulated experiments, as well as the observed data, are fit using an $s+b$ and an SM$+b$ likelihood function. As with the background-only p-value determination, the s in the $s + b$ fit is left to float freely. An LLR is constructed as $2\ln Q = \mathcal{L}(s + b)/\mathcal{L}(\text{SM} + b)$. The final p-value is taken as the ratio of simulated experiments with an LLR greater than the LLR for the observed data.

In addition to estimating an SM+background p-value in the case of a significant excess, we also perform a $\Delta\chi^2$ fit to the data to determine a central value for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. This fit is done by replacing the free-floating s in the $s+b$ hypothesis in the log-likelihood ratio with a signal yield corresponding to a specific

$\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The branching fraction is then varied to find the minimum of $-2\ln Q$ as the central value of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The 68% and 90% C.L. are taken from the interval with a change of one and 2.71 units, respectively.

Table 4.10: A comparison of the predicted and observed number of events in the extended signal mass region as a function of NN bin 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, given the predicted background, is also shown in the table. In cases where no events are observed the probability is actually the Poisson probability that exactly zero events are observed with a mean equal to the predicted mean.

		CC			CF		
Sample	NN bin	Pred.	Obs.	Prob. (%)	Pred.	Obs.	Prob. (%)
OS–	$0.700 < \nu_N < 0.760$	268.8±(14.3)	249	82.3	261.8±(13.9)	241	84.1
	$0.760 < \nu_N < 0.850$	320.8±(16.1)	282	95.1	399.0±(18.5)	397	53.2
	$0.850 < \nu_N < 0.900$	150.3±(9.9)	156	36.5	193.5±(11.4)	185	68.3
	$0.900 < \nu_N < 0.940$	146.2±(9.7)	158	23.0	177.4±(10.8)	183	37.7
	$0.940 < \nu_N < 0.970$	146.2±(9.7)	137	72.9	156.8±(10.1)	143	81.2
	$0.970 < \nu_N < 0.987$	100.4±(7.8)	98	58.3	112.6±(8.2)	110	58.3
	$0.987 < \nu_N < 0.995$	78.8±(6.8)	59	97.0	53.3±(5.4)	68	6.5
	$0.995 < \nu_N < 1.000$	41.2±(4.8)	42	47.2	48.2±(5.1)	44	70.0
SS+	$0.700 < \nu_N < 0.760$	4.8±(1.2)	3	81.8	0.9±(0.5)	3	8.9
	$0.760 < \nu_N < 0.850$	3.6±(1.0)	5	30.6	5.1±(1.2)	5	55.4
	$0.850 < \nu_N < 0.900$	2.4±(0.8)	5	12.2	0.9±(0.5)	6	0.2
	$0.900 < \nu_N < 0.940$	1.5±(0.7)	3	21.3	0.9±(0.5)	1	56.8
	$0.940 < \nu_N < 0.970$	1.5±(0.7)	1	73.3	0.9±(0.5)	1	56.8
	$0.970 < \nu_N < 0.987$	1.8±(0.7)	2	51.3	0.9±(0.5)	0	40.7
	$0.987 < \nu_N < 0.995$	0.3±(0.3)	0	74.1	0.3±(0.3)	0	74.1
	$0.995 < \nu_N < 1.000$	0.3±(0.3)	0	74.1	0.3±(0.3)	1	30.0
SS–	$0.700 < \nu_N < 0.760$	7.8±(1.5)	10	27.8	6.0±(1.3)	4	80.9
	$0.760 < \nu_N < 0.850$	10.5±(1.8)	11	47.2	7.2±(1.5)	7	55.8
	$0.850 < \nu_N < 0.900$	4.2±(1.1)	7	15.9	3.0±(0.9)	2	75.8
	$0.900 < \nu_N < 0.940$	3.6±(1.0)	4	47.2	0.9±(0.5)	7	0.1
	$0.940 < \nu_N < 0.970$	3.3±(1.0)	6	14.3	3.6±(1.0)	2	83.4
	$0.970 < \nu_N < 0.987$	3.0±(0.9)	3	55.0	2.4±(0.8)	5	12.2
	$0.987 < \nu_N < 0.995$	2.1±(0.8)	0	12.2	1.2±(0.6)	0	30.1
	$0.995 < \nu_N < 1.000$	1.2±(0.6)	1	65.9	1.8±(0.7)	0	16.5
FM+	$0.700 < \nu_N < 0.760$	152.2±(9.9)	161	29.6	66.5±(6.1)	88	2.5
	$0.760 < \nu_N < 0.850$	140.9±(9.5)	157	15.3	81.7±(6.9)	76	70.0
	$0.850 < \nu_N < 0.900$	65.2±(6.1)	50	94.4	44.7±(5.0)	34	91.6
	$0.900 < \nu_N < 0.940$	48.7±(5.2)	40	85.8	24.4±(3.6)	38	2.3
	$0.940 < \nu_N < 0.970$	27.7±(3.8)	24	73.1	12.7±(2.6)	20	7.1
	$0.970 < \nu_N < 0.987$	10.9±(2.3)	12	41.4	7.7±(2.0)	13	8.8
	$0.987 < \nu_N < 0.995$	11.0±(2.3)	4	98.3	2.7±(1.1)	3	48.3
	$0.995 < \nu_N < 1.000$	28.3±(4.1)	32	30.6	4.4±(1.6)	8	13.0

Table 4.11: Summary of systematic uncertainties for the CC and CF channels.

Category	Quantity	CC (%)	CF (%)	Source
Efficiencies, acceptance, and normalization factors	$\alpha_{B^+}/\alpha_{B_s^0}$	6	7	b mass, renorm. scale, fragmentation modeling
	$\epsilon^{\mathrm{COT}}_{B^+}/\epsilon^{\mathrm{COT}}_{B_s^0}$	1	1	Isolation, $p_T(B)$, detector effects
	$\epsilon^{\mu}_{B^+}/\epsilon^{\mu}_{B_s^0}$	3	3	$Z^0 \to \mu^+\mu^-$ and $J/\psi \to \mu^+\mu^-$ differences
	$\epsilon^{\mathrm{PID}}_{B^+}/\epsilon^{\mathrm{PID}}_{B_s^0}$	3	3	$B^+ \to J/\psi K^+$ and $J/\psi \to \mu^+\mu^-$ differences
	$\epsilon^{\mathrm{SVXII}}_{B^+}/\epsilon^{\mathrm{SVXII}}_{B_s^0}$	3	3	$p_T(\mu)$, two muon opening angle, track isolation
	$\epsilon^{\mathrm{COT}}_{K}$	1.7	1.7	Isolation, $p_T(B)$, detector effects
	ϵ^{NN}	4	4	B isolation, $p_T(B)$
	ϵ^{NN} for $\nu_N > 0.995$	3.4	7.0	Data-MC differences
	f_u/f_s	13	13	Ref. [3]
	$\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+)$	4	4	Ref. [3]
	B-meson mass shape	0.1-9	0.1-9	Mass resolution, mass scale, Ref. [3]
$B \to h^+h'^-$	R_b	20	20	Detector and luminosity effects
	$\mathcal{B}(B \to h^+h'^-)$	3-100	3-100	Ref. [3]
Comb. Bkg	Slope	6	6	Fit uncertainty
	Normalization	7-42	7-25	Sideband statistics
	Shape	10-43	3-42	Comparison of different fit functions

CHAPTER 5

RESULTS AND CONCLUSION

Summaries of the background estimates, systematic uncertainties on the background estimates, and the observed number of events are given in Fig. 5.1 and Table 5.1. The lowest five NN bins have been combined in Fig. 5.1 because the signal sensitivity is concentrated in the highest three NN bins. The observed number of events in the $B_d^0 \to \mu^+\mu^-$ signal region are consistent with background-only predictions.

The resulting p-value, assuming the background-only hypothesis, for the $B_d^0 \to \mu^+\mu^-$ search window is 41%.

Using the CLs method, we also set upper limits of 4.6×10^{-9} (3.8×10^{-9}) at 95% (90%) C.L. These observed limits are close to the expected median limits of 4.2×10^{-9} (3.4×10^{-9}) at 95% (90%) C.L. (Fig. 5.2).

In the $B_s^0 \to \mu^+\mu^-$ search window the data exceed the background-only expectations as seen in the bottom part of Fig. 5.1 and Table 5.2. The excess events populate the most sensitive muon topology bin and NN bins, and concentrate in the most central mass bins where the signal sensitivity is highest. We estimate a p-value using all NN bins ($\nu_N > 0.7$) for a background-only hypothesis as well as an SM+background hypothesis of 0.94% and 6.8%, respectively.

Since the background-only p-value corresponds to a $>2\sigma$ excess we perform a $\Delta\chi^2$ fit, that includes all our systematic uncertainties, to estimate $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ (Fig. 5.3). The central value and the associated 68% C.L. bound is $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 1.3^{+0.9}_{-0.7} \times 10^{-8}$ while the 90% C.L. bounds are $2.2 \times 10^{-9} < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.0 \times 10^{-8}$. The central value and bounds estimated from the $\Delta\chi^2$ fit were cross-

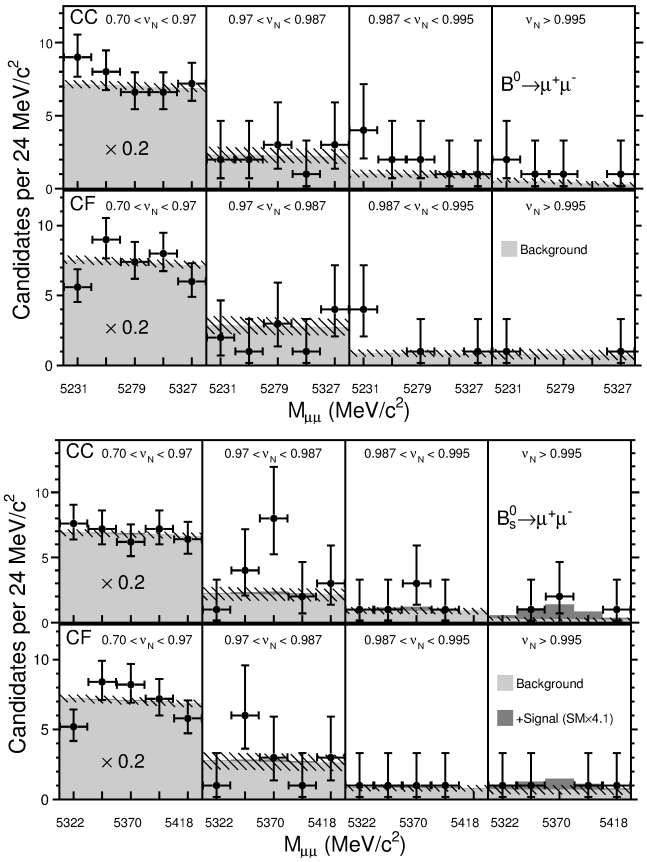

Figure 5.1: **Top:** Background estimates (light gray), systematic uncertainty on background estimates (hashed area), Poisson uncertainty on the mean (error bars on points), and data for the B^0_d signal window. **Bottom:** Similar plots for B^0_s with the addition of the 4.1×SM expectations (dark gray). The result of our $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ fit, shown in Fig. 5.3, yields a central value that is 4.1 times greater than the SM expectation. The data are divided into 8 NN bins, of which lowest 5 NN bins are combined into one bin for both figures, 5 mass bins, and two muon topologies (CC and CF).

checked using a Bayesian method using Gaussian priors for all nuisance parameters and a flat prior truncated at zero for the branching fraction. The results from the Bayesian method yielded a similar central value and bound. For comparison with our median expected limits of 1.3×10^{-8} (1.0×10^{-8}) at 95% (90%) C.L. we calculate observed upper limits using the CLs methodology of 3.1×10^{-8}

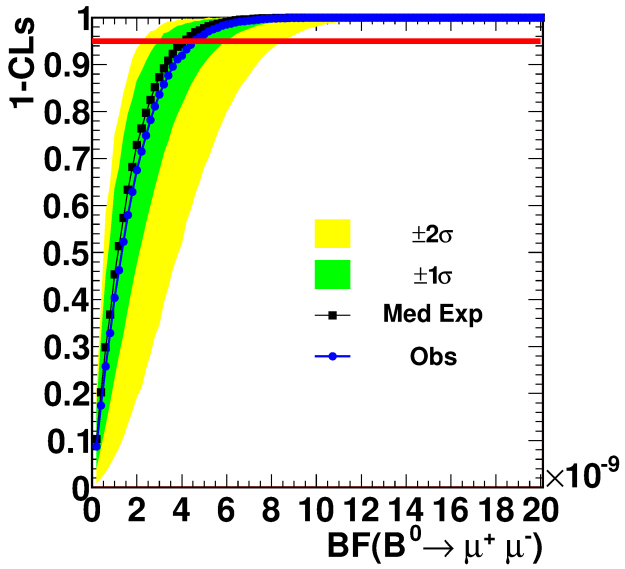

Figure 5.2: 1–CLs as a function of $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$.

(2.7×10^{-8}) at 95% (90%) C.L. (Fig. 5.4).

The excess is concentrated in the two highest NN bins of the CC channel, our most sensitive bins. The total background expectations for the $0.987 < \nu_N < 0.995$ and $0.995 < \nu_N < 1.000$ bins are 4.56 and 1.29 events while the SM expected signal yields are 0.75 and 0.20 events, respectively. However, we observe a total of 6 and 4 events, respectively, for these bins.

As a check of consistency we redo our $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ fit using only the two highest NN bins. This yields a central value of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 1.0^{+0.8}_{-0.6} \times 10^{-8}$ and a two-sided bound of $0.8 \times 10^{-9} < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 2.5 \times 10^{-8}$ at 90% C.L. Using the $\Delta\chi^2$ method we are not able to extract a 95% bound. The p-values when only using the two highest NN bins are 2.0% and 21.6% for a background-only and SM+background hypothesis, respectively.

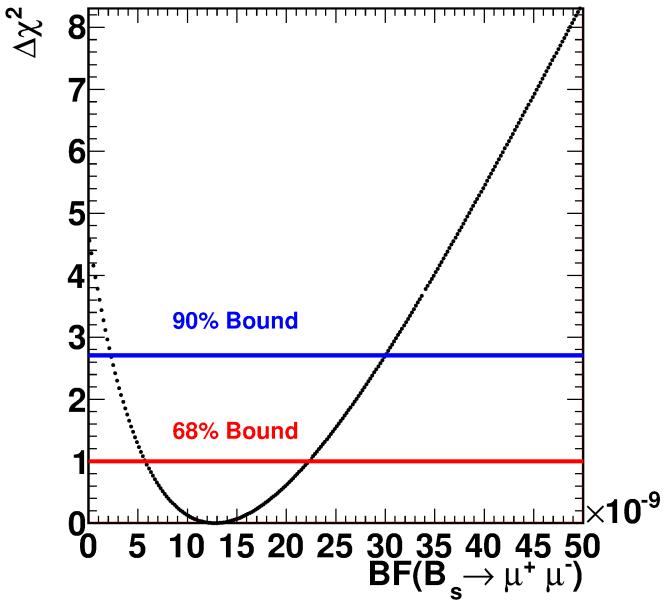

Figure 5.3: $\Delta\chi^2$ distribution as a function of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$.

We also observe a data excess in the $0.970 < \nu_N < 0.987$ bin of the CC channel, where no significant signal contribution is expected. The source of the data excess in this bin has been investigated. We employ the same background estimation methods for both the B_s^0 and B_d^0 signal windows. For the combinatorial background estimates the same sideband events are used for the background estimates in the B_d^0 and B_s^0 signal regions. The $B \to h^+h'^-$ peaking background is an order of magnitude larger in the B_d^0 region and no excess is seen in the B_d^0 signal window. We also thoroughly investigated any possible NN biases through various tests discussed in Sect. 4.2.2. We conclude that the excess in this bin is not caused by a problem with the background estimates, a NN bias, or any mismodeling of data, and is likely a statistical upward fluctuation. The analysis is repeated, discussed in detail in appendix A, using the last 3 fb^{-1} only. The results for the additional 3 fb^{-1} of data added in this analysis are shown in Fig. 5.5 and show no evidence of this fluctuation. This is consistent with our conclusion

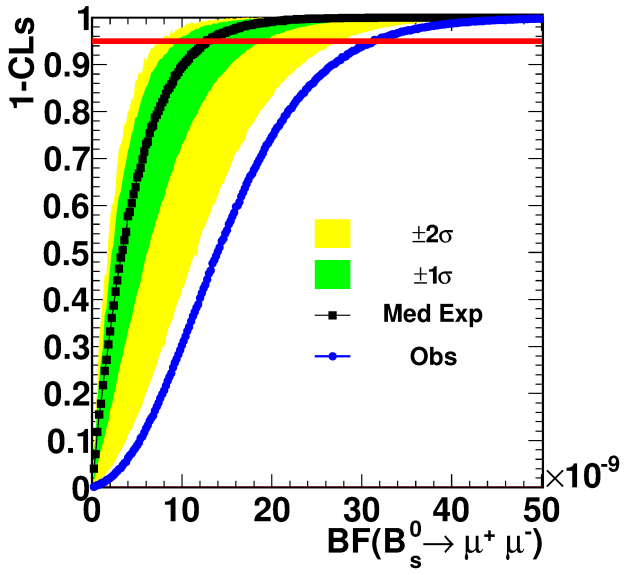

Figure 5.4: 1–CLs as a function of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$.

regarding the origin of the excess in this NN bin.

5.1 Conclusion

We report on the search for $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays using the full CDF II data set. The observed data are in agreement with background-only expectations in the $B_d^0 \to \mu^+\mu^-$ search. We set an upper bound of $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 4.6 \times 10^{-9}$ (3.8×10^{-9}) at 95% (90%) C.L.

In the $B_s^0 \to \mu^+\mu^-$ search window the observed data exceed the background expectation. We calculate the p-values for a background-only hypothesis as well as an SM+background hypothesis of 0.94% and 6.8%, respectively. A fit to the observed data yields $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 1.3^{+0.9}_{-0.7} \times 10^{-8}$. Bounds are set at $2.2 \times 10^{-9} < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.0 \times 10^{-8}$ and $0.8 \times 10^{-9} < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.4 \times 10^{-8}$ at 90% and

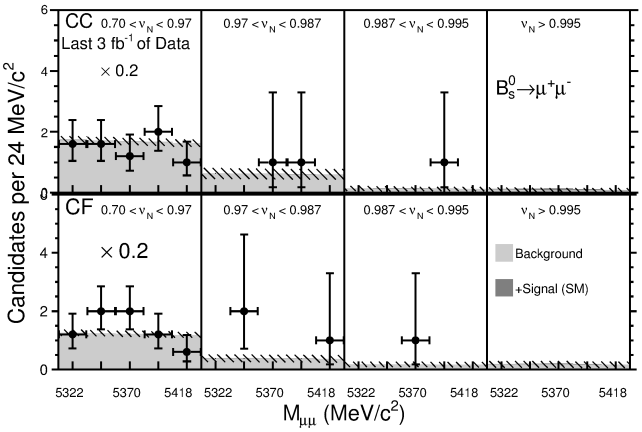

Figure 5.5: Background estimates (light gray), systematic uncertainty on background estimates (hashed area), Poisson uncertainty on the mean (error bars on points), SM expectations (dark gray), and data for the B_s^0 signal window using only the last 3 fb^{-1} of data collected. The data are divided into 8 NN bins, of which the lowest 5 NN bins are combined into one bin, 5 mass bins, and two muon topologies (CC and CF).

95% C.L., respectively.

The $B_s^0 \to \mu^+\mu^-$ results with the full CDF Run II data set is compatible with the more stringent bounds set by the LHC experiments. The bounds on $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and the measurements of the all relevant experiments are shown in Fig. 5.6. The LHCb experiment sets the strongest upper bound on $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 4.5 \times 10^{-9}$ at 95% C.L. which overlaps the CDF confidence level interval of $0.8 < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 34 \times 10^{-9}$ at 95% C.L.

5.2 Future prospects

The analysis described in this thesis is the last $B_{s,d}^0 \to \mu^+\mu^-$ analysis from CDF including the entire run II data set. As of October 2011 the Tevatron has been

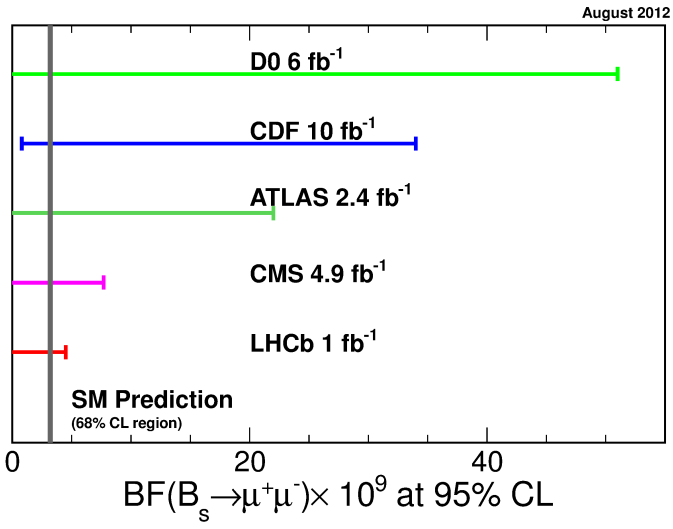

Figure 5.6: Current limits from D0 [9], CDF, ATLAS [45], CMS [46], and LHCb [47], at 95% C.L. The SM prediction is also given at 68% C.L.

decommissioned and thus no new data can be included to improve sensitivity. Many of the possible analysis improvements have been studied in several iterations of this analysis. Any improvements not implemented in this analysis would only yield very small increases in sensitivity and would not change our results. D0 could release another iteration of this analysis but the experiments mass resolution doesn't allow for separation of B_s^0 and B_d^0 mass regions. Future searches of $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ will be performed at the LHC.

The LHC is a proton-proton accelerator based in Geneva, Switzerland. The initial data taking runs were done at a center of mass energy of 7 TeV. The center-of-mass energy of the LHC was increased to $\sqrt{s} = 8$ TeV for 2012 [48]. At the end of 2012 the ATLAS and CMS expects to have collected 10–15fb^{-1} of data at which point the LHC will shutdown for 18–24 months. After this shutdown a running period from 2015 to 2017 will have the LHC running at the design $\sqrt{s} = 14$ TeV and producing 75 fb^{-1} of data. Four large experiments utilize the LHC: ALICE,

ATLAS, CMS, and LHCb. ALICE is a heavy ion experiment, while ATLAS and CMS are general purpose experiments, and LHCb is a B physics experiment.

Three experiments at the LHC already have results from early data for the $B^0_{s,d} \to \mu^+\mu^-$ searches. Two of the experiments, CMS and ATLAS, are general purpose detectors with a solenoidal geometry. The other, LHCb, is an experiment designed specifically for B physics and has a forward geometry.

As of July 2012, the current limits set by all the experiments are given in Fig. 5.6. CMS uses ~5 fb^{-1} of data and a cut based analysis, ATLAS performed the analysis using a multivariate discriminant and ~2.5 fb^{-1} of data, while LHCb performed a multivariate analysis on ~1 fb^{-1} of data. LHCb has the lowest limits and observed less than the SM+background hypothesis expected number of events at a level of 1σ. CMS and ATLAS both observe limits consistent with the SM+background hypothesis. The LHC experiments combined their result for a combined upper bound of $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 4.2 \times 10^{-9}$, which is also withing 1σ of the SM+background expectation.

All the LHC experiments are rapidly increasing the size of their data set and have great sensitivity to these decays. LHCb, in particular, has specialized in B-meson decays, having not only a forward geometry allowing for very boosted B-mesons and excellent geometric and kinematic acceptance, but also excellent vertex reconstruction, and particle identification.

The analysis strategies are in essence the same for all experiments. All use a dimuon trigger to collect $B^0_{s,d} \to \mu^+\mu^-$ candidates. The backgrounds are the same as with CDF, consisting of combinatorial and peaking $B \to h^+h'^-$ backgrounds. The first results by CMS did not use any multivariate discriminant in the selec-

tion criteria. LHCb's first result used a likelihood while their remaining results used a BDT. CMS plans on using a multivariate discriminant in future analyses.

The LHC will have a two year long shutdown at the end of 2012. By this time CMS and ATLAS would have collected approximately 10–15 fb^{-1} while LHCb would have a total of 2 fb^{-1} of data. This would give CMS a reach down to SM $B_s^0 \to \mu^+\mu^-$ 3σ evidence. This difference in integrated luminosity between CMS/ATLAS and LHCb comes from LHCb keeping their instantaneous luminosity at a constant rate. Even though CMS and ATLAS have a luminosity advantage over LHCb, trigger bandwidth allocated to B-meson searches could dampen this advantage.

Unless there are large new physics contributions to the $B_d^0 \to \mu^+\mu^-$ decay much more data is needed to reach the SM predictions. The SM prediction is a factor of 32 lower for $B_d^0 \to \mu^+\mu^-$ than for $B_s^0 \to \mu^+\mu^-$. Neither LHCb nor CMS/ATLAS will be able to reach these kind of sensitivities before the 2012 shutdown.

SUSY parameter space is now more directly affected by direct searches at the LHC than indirect searches through the $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays. The $B_{s,d}^0 \to \mu^+\mu^-$ searches, however, could still compliment any limits set by direct searches. As of November 2012 there has been no evidence of beyond the SM physics. This is an especially exciting era for particle physics that could lead to a wealth of new physics phenomena and a broader understanding of the basic constituents of the universe.

Table 5.1: Expected total backgrounds and number of observed events in the B_d^0 signal window for CC (top) and CF (bottom).

	Mass bin (GeV/c^2)	5.219–5.243	5.243–5.267	5.267–5.291	5.291–5.315	5.315–5.339	Total
CC NN bin	Exp. Bkg.	10.80±0.74	10.70±0.74	10.61±0.73	10.51±0.72	10.41±0.72	53.02
0.700–0.760	Obs.	15	14	10	7	11	57
CC NN bin	Exp. Bkg.	11.43±0.77	11.33±0.76	11.23±0.75	11.12±0.75	11.01±0.74	56.12
0.760–0.850	Obs.	12	10	7	8	9	46
CC NN bin	Exp. Bkg.	4.88±0.47	4.83±0.47	4.78±0.46	4.73±0.46	4.68±0.46	23.90
0.850–0.900	Obs.	10	3	6	6	5	30
CC NN bin	Exp. Bkg.	4.68±0.46	4.63±0.46	4.59±0.45	4.54±0.45	4.49±0.44	22.92
0.900–0.940	Obs.	6	10	6	8	6	36
CC NN bin	Exp. Bkg.	4.03±0.42	3.99±0.42	3.94±0.42	3.89±0.41	3.85±0.41	19.70
0.940–0.970	Obs.	2	3	4	4	5	18
CC NN bin	Exp. Bkg.	2.34±0.39	2.31±0.39	2.28±0.38	2.24±0.38	2.21±0.38	11.39
0.970–0.987	Obs.	2	2	3	1	3	11
CC NN bin	Exp. Bkg.	1.04±0.27	1.03±0.27	1.00±0.27	0.96±0.26	0.93±0.26	4.96
0.987–0.995	Obs.	4	2	2	1	1	10
CC NN bin	Exp. Bkg.	0.57±0.19	0.54±0.17	0.47±0.16	0.37±0.15	0.29±0.14	2.24
0.995–1.000	Obs.	2	1	1	0	1	5
CF NN bin	Exp. Bkg.	10.65±0.75	10.53±0.74	10.40±0.73	10.28±0.73	10.15±0.72	52.01
0.700–0.760	Obs.	8	13	12	16	10	59
CF NN bin	Exp. Bkg.	11.74±0.80	11.61±0.79	11.47±0.78	11.33±0.77	11.19±0.76	57.33
0.760–0.850	Obs.	9	13	13	13	12	60
CF NN bin	Exp. Bkg.	6.40±0.56	6.32±0.55	6.24±0.55	6.17±0.54	6.09±0.53	31.22
0.850–0.900	Obs.	3	4	3	2	1	13
CF NN bin	Exp. Bkg.	4.88±0.48	4.82±0.48	4.76±0.47	4.70±0.47	4.64±0.46	23.80
0.900–0.940	Obs.	3	8	7	8	5	31
CF NN bin	Exp. Bkg.	4.14±0.44	4.09±0.44	4.04±0.43	3.99±0.43	3.94±0.42	20.20
0.940–0.970	Obs.	5	7	2	1	2	17
CF NN bin	Exp. Bkg.	2.89±0.46	2.85±0.46	2.82±0.45	2.78±0.45	2.74±0.44	14.07
0.970–0.987	Obs.	2	1	3	1	4	11
CF NN bin	Exp. Bkg.	0.88±0.26	0.87±0.25	0.86±0.25	0.85±0.25	0.83±0.24	4.30
0.987–0.995	Obs.	4	0	1	0	1	6
CF NN bin	Exp. Bkg.	0.82±0.37	0.81±0.36	0.79±0.36	0.75±0.35	0.72±0.35	3.89
0.995–1.000	Obs.	1	0	0	0	1	2

Table 5.2: Expected total backgrounds and number of observed events in the B_s^0 signal window for CC (top) and CF (bottom).

	Mass bin (GeV/c^2)	5.310–5.334	5.334–5.358	5.358–5.382	5.382–5.406	5.406–5.430	Total
CC NN bin	Exp. Bkg.	10.43±0.72	10.33±0.71	10.23±0.71	10.14±0.70	10.04±0.69	51.17
0.700–0.760	Obs.	13	8	7	6	7	41
CC NN bin	Exp. Bkg.	11.03±0.74	10.93±0.74	10.82±0.73	10.72±0.72	10.62±0.72	54.13
0.760–0.850	Obs.	9	8	12	15	8	52
CC NN bin	Exp. Bkg.	4.70±0.46	4.65±0.45	4.61±0.45	4.56±0.44	4.52±0.44	23.03
0.850–0.900	Obs.	6	8	5	6	5	30
CC NN bin	Exp. Bkg.	4.50±0.45	4.45±0.44	4.41±0.44	4.37±0.43	4.33±0.43	22.05
0.900–0.940	Obs.	5	5	5	6	8	29
CC NN bin	Exp. Bkg.	3.86±0.41	3.82±0.41	3.78±0.40	3.74±0.40	3.71±0.39	18.91
0.940–0.970	Obs.	5	7	2	3	4	21
CC NN bin	Exp. Bkg.	2.22±0.38	2.19±0.37	2.17±0.37	2.15±0.37	2.12±0.36	10.84
0.970–0.987	Obs.	1	4	8	2	3	18
CC NN bin	Exp. Bkg.	0.94±0.26	0.92±0.26	0.91±0.26	0.90±0.25	0.89±0.25	4.56
0.987–0.995	Obs.	1	1	3	1	0	6
CC NN bin	Exp. Bkg.	0.31±0.14	0.26±0.14	0.25±0.14	0.24±0.14	0.23±0.14	1.29
0.995–1.000	Obs.	0	1	2	0	1	4
CF NN bin	Exp. Bkg.	10.18±0.72	10.05±0.71	9.93±0.7	9.80±0.69	9.68±0.68	49.64
0.700–0.760	Obs.	10	16	12	11	10	59
CF NN bin	Exp. Bkg.	11.22±0.76	11.08±0.75	10.94±0.74	10.8±0.73	10.67±0.72	54.71
0.760–0.850	Obs.	8	13	9	13	4	47
CF NN bin	Exp. Bkg.	6.11±0.54	6.03±0.53	5.96±0.52	5.88±0.52	5.81±0.51	29.79
0.850–0.90	Obs.	1	5	9	3	6	24
CF NN bin	Exp. Bkg.	4.65±0.46	4.60±0.46	4.54±0.45	4.48±0.44	4.42±0.44	22.69
0.900–0.940	Obs.	6	2	8	5	4	25
CF NN bin	Exp. Bkg.	3.95±0.42	3.90±0.42	3.85±0.41	3.80±0.41	3.75±0.40	19.25
0.940–0.970	Obs.	1	6	3	4	5	19
CF NN bin	Exp. Bkg.	2.75±0.44	2.71±0.44	2.68±0.43	2.64±0.43	2.61±0.42	13.38
0.970–0.987	Obs.	1	6	3	1	3	14
CF NN bin	Exp. Bkg.	0.83±0.25	0.82±0.24	0.81±0.24	0.80±0.24	0.79±0.23	4.06
0.987–0.995	Obs.	1	1	1	1	0	4
CF NN bin	Exp. Bkg.	0.73±0.35	0.71±0.34	0.69±0.34	0.68±0.34	0.67±0.33	3.48
0.995–1.000	Obs.	1	1	0	1	1	4

BIBLIOGRAPHY

[1] Ch. Weinheimer, *et al.*, Phys. Lett. B **460**, 219 (1999).

[2] PYTHIA 6.2 Physics and Manual, arXiv:hep-ph/0108264v1.

[3] K. Nakamura *et al.* (Particle Data Group), J. Phys. G **37**, 075021, (2010).

[4] A. J. Buras, arXiv:hep-ph/9806471v1.

[5] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, arXiv:1208.0934 [hep-ph].

[6] E. Gamiz, *et al.* (HPQCD Collaboration), Phys. Rev. D **80**, 014503 (2009); A. J. Buras, M. V. Carlucci, S. Gori, and G. Isidori, J. High Energy Phys. **10**, 009, (2010).

[7] A. J. Buras, Acta Phys. Polon. B **41**, 2487-2561, (2010), arXiv:1012.1447v2 [hep-ph].

[8] M. Perelstein, Pramana **67**, (2006), 813-820, arXiv:hep-ph/0703138v1.

[9] V. Abazov *et al.* (D0 Collaboration), Phys. Lett. B **693**, 539 (2010).

[10] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett **107**, 191801 (2011).

[11] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **107**, 191802, (2011).

[12] M. Bettler, arXiv:1110.2411v2.

[13] CMS and LHCb Collaborations, CERN-LHCb-CONF-2011-047.

[14] S. W. Herb *et al.*, Phys. Rev. Lett. **39**, 252 (1977).

[15] F. Abe *et al.* (CDF Collaboration), Phys. Rev. Lett. 74, 2626 (1995); S. Abachi *et al.* (D0 Collaboration), Phys. Rev. Lett. 74, 2632 (1995).

[16] Kodama, K. *et al.* (DONUT Collaboration), Phys. Lett. B **504** 218, (2001), arXiv:hep-ex/0012035.

[17] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **99**, 202001, (2007).

[18] C.W. Schmidt and C.D. Curtis. IEEE Transactions on Nuclear Science **NS-26**, 4120, (1979).

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

[20] F. Abe *et al.*, Nucl. Instrum. Methods A **271**, 387 (1988).

[21] F. Abe *et al.*, Phys. Rev. D **50**, 2966 (1994b).

[22] C. S. Hill *et al.* (CDF Collaboration), Nucl. Instrum. Methods A **530**, 1 (2004).

[23] A. Sill *et al.* (CDF Collaboration), Nucl. Instrum. Methods A **447**, 1 (2000).

[24] A. A. Affolder *et al.* (CDF Collaboration), Nucl. Instrum. Methods A **453**, 84 (2000).

[25] T. Affolder *et al.*, Nucl. Instrum. Methods A **526**, 249 (2004).

[26] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. **96**, 202001 (2006).

[27] L. Balka *et al*, Nucl. Instrum. Methods A **267**, 272 (1988).

[28] P. de Barbaro *et al*, IEEE Trans. Nucl. Sci. **42**, 510 (1995).

[29] S. Bertolucci *et al*, Nucl. Instrum. Methods A **267**, 301 (1988).

[30] G. Ascoli *et al*, Nucl. Instrum. Methods A **268**, 33 (1988).

[31] A. Artikov *et al*, Nucl. Instrum. Methods A **538**, 358 (2005).

[32] D. Acosta *et al*, Int. J. Mod. Phys. **A16S1C**, 1119 (2001).

[33] D. Acosta *et al*, Nucl. Instrum. Methods A **494**, 57 (2002).

[34] E. J. Thomson *et al*, IEEE Trans. Nucl. Sci. **49**, 1063 (2002).

[35] P. S. Maybeck, *Stochastic Models, Estimation, and Control*.

[36] J. Marriner, CDF NOTE 1996 (1993).

[37] http://www-cdf.fnal.gov/internal/dqm/goodrun/good.html.

[38] D. Lange, Nucl. Instrum. and Meth. A **462**, 152 (2001).

[39] R. Field (CDF Collaboration), Acta Phys. Polon. B **36**, 167 (2005).

[40] D. Smirnov and K. Anikeev, ROOTSNNS v2.0, CDF internal note 7999.

[41] A Neural Bayesian Estimator for Conditional Probability Densities, arXiv:physics/0402093v1.

[42] M. Feindta and U. Kerzelb, Nucl. Instrum. and Meth. A **559**, 190 (2006).

[43] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **106**, 181802 (2011).

[44] T. Junk, Nucl. Instrum. Methods A **434, 435** (1999).

[45] G. Aad *et al.* (ATLAS Collaboration), Phys. Rev. B **713**, 387 (2012).

[46] S. Chatrchyan *et al.* (CMS Collaboration), J. High Energy Phys. 1204 (2012) 033.

[47] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. Lett. **108**, 231801 (2012).

[48] CMS Collaboration, *CMS prospects for heavy flavour physics*, CMS-NOTE-2011-008.

APPENDIX A

ANALYSIS WITH NEW DATA ONLY

The analysis is performed separately with only the new data for comparison with the 7 fb^{-1} analysis. All aspects of the analysis remained the same and new combinatorial and peaking backgrounds are estimated. The same signal sensitivities are used with the exception of the normalization mode yields. The new combinatorial background estimates for B_s^0 and B_d^0 are given in Table A.1 and A.2, respectively. The $B \to h^+h'^-$ background estimates have are shown in Table A.3 and Table A.4. Finally, the unblinded results are given in Table A.5 and Table A.6.

We repeat all the statistical analysis done with the full data set. The expected limit using the CLs methodology are $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.4 \times 10^{-8}$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 10.0\times10^{-9}$. The observed limits are $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 2.9\times10^{-8}$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 9.0\times10^{-9}$ and no significant excess over the background-only hypothesis is found. We calculated p-values of 47% and 82% for B_d^0 and B_s^0 respectively in the background-only hypothesis. The p-value for the SM+background hypothesis is also evaluated at 48% for the B_s^0 window.

To compare to the full analysis we calculate the p-value for the 3rd highest NN bin in the CC channel. There was no excess in the new data corresponding to a p-value of 44%.

A.1 NN mass bias check

The NN output distributions as a function of dimuon mass are fitted to both a constant (Fig. 4.6) and a first order polynomial for OS+ and OS– samples.

Table A.1: Combinatorial background estimates and statistical uncertainty for B_s^0 signal window for all NN bins using only the new data.

NN bin \ Mass bin (GeV/c^2)	5.310–5.334	5.334–5.358	5.358–5.382	5.382–5.406	5.406–5.430
CC					
0.700< ν_N <0.760	2.56±0.34	2.52±0.33	2.49±0.33	2.46±0.32	2.42±0.32
0.760< ν_N <0.850	2.77±0.35	2.73±0.35	2.69±0.34	2.66±0.34	2.62±0.33
0.850< ν_N <0.900	1.22±0.23	1.20±0.23	1.18±0.22	1.17±0.22	1.15±0.22
0.900< ν_N <0.940	1.05±0.21	1.03±0.21	1.02±0.21	1.01±0.20	0.99±0.20
0.940< ν_N <0.970	1.05±0.21	1.03±0.21	1.02±0.21	1.01±0.20	0.99±0.20
0.970< ν_N <0.987	0.63±0.16	0.62±0.16	0.61±0.16	0.60±0.16	0.60±0.15
0.987< ν_N <0.995	0.13±0.07	0.12±0.07	0.12±0.07	0.12±0.07	0.12±0.07
0.995< ν_N <1.000	0.08±0.06	0.08±0.06	0.08±0.06	0.08±0.06	0.08±0.06
CF					
0.700< ν_N <0.760	1.74±0.28	1.72±0.27	1.69±0.27	1.67±0.27	1.64±0.26
0.760< ν_N <0.850	1.83±0.28	1.80±0.28	1.77±0.28	1.75±0.27	1.72±0.27
0.850< ν_N <0.900	1.23±0.23	1.21±0.23	1.20±0.23	1.18±0.22	1.16±0.22
0.900< ν_N <0.940	0.81±0.19	0.80±0.18	0.78±0.18	0.77±0.18	0.76±0.18
0.940< ν_N <0.970	0.68±0.17	0.67±0.17	0.66±0.17	0.65±0.16	0.64±0.16
0.970< ν_N <0.987	0.38±0.13	0.38±0.13	0.37±0.12	0.37±0.12	0.36±0.12
0.987< ν_N <0.995	0.17±0.09	0.17±0.08	0.17±0.08	0.16±0.08	0.16±0.08
0.995< ν_N <1.000	0.17±0.09	0.17±0.08	0.17±0.08	0.16±0.08	0.16±0.08

Figure A.1 show the result for the first order polynomial fits. The χ^2's of the fits to constants are similar to the χ^2's of first order polynomial fits. We also calculate the linear correlation coefficient (r) for all three of these samples. The resulting correlation coefficients are 0.022 and 0.004 for OS+ and OS–, respectively. We estimate the 68% C.L. for the correlation coefficients at $0.004 < r < 0.039$ and $-0.005 < r < 0.013$ for OS+ and OS–, respectively. For the OS+ sample the estimated r value is not within the 68% C.L. The estimated r is 1.24σ from the

Table A.2: Combinatorial background estimates and statistical uncertainty for B_d^0 signal window for all NN bins using only the new data.

NN bin \ Mass bin (GeV/c^2)	5.219–5.243	5.243–5.267	5.267–5.291	5.291–5.315	5.315–5.339
CC					
0.700< ν_N <0.760	2.68±0.35	2.65±0.35	2.62±0.34	2.58±0.34	2.55±0.34
0.760< ν_N <0.850	2.90±0.37	2.87±0.36	2.83±0.36	2.79±0.35	2.76±0.35
0.850< ν_N <0.900	1.27±0.24	1.26±0.24	1.24±0.23	1.23±0.23	1.21±0.23
0.900< ν_N <0.940	1.10±0.22	1.08±0.22	1.07±0.22	1.06±0.21	1.04±0.21
0.940< ν_N <0.970	1.10±0.22	1.08±0.22	1.07±0.22	1.06±0.21	1.04±0.21
0.970< ν_N <0.987	0.66±0.17	0.65±0.17	0.64±0.17	0.64±0.17	0.63±0.16
0.987< ν_N <0.995	0.13±0.08	0.13±0.07	0.13±0.07	0.13±0.07	0.12±0.07
0.995< ν_N <1.000	0.09±0.06	0.09±0.06	0.09±0.06	0.09±0.06	0.08±0.06
CF					
0.700< ν_N <0.760	1.83±0.29	1.81±0.29	1.79±0.28	1.76±0.28	1.74±0.28
0.760< ν_N <0.850	1.92±0.30	1.90±0.29	1.87±0.29	1.85±0.29	1.82±0.28
0.850< ν_N <0.900	1.30±0.24	1.28±0.24	1.26±0.24	1.25±0.23	1.23±0.23
0.900< ν_N <0.940	0.85±0.20	0.84±0.19	0.83±0.19	0.82±0.19	0.81±0.19
0.940< ν_N <0.970	0.72±0.18	0.71±0.18	0.70±0.17	0.69±0.17	0.68±0.17
0.970< ν_N <0.987	0.40±0.14	0.40±0.13	0.39±0.13	0.39±0.13	0.38±0.13
0.987< ν_N <0.995	0.18±0.09	0.18±0.09	0.17±0.09	0.17±0.09	0.17±0.09
0.995< ν_N <1.000	0.18±0.09	0.18±0.09	0.17±0.09	0.17±0.09	0.17±0.09

r = 0 (no correlation) hypothesis for the OS+ sample.

Table A.3: $B \to h^+h'^-$ background estimates for B_s^0 signal window for all NN bins using only the new data.

Mass bin (GeV/c^2) / NN bin	5.310–5.334	5.334–5.358	5.358–5.382	5.382–5.406	5.406–5.430
CC					
0.700< ν_N <0.760	0.001± < 0.001	–	–	–	–
0.760< ν_N <0.850	0.002± < 0.001	0.001± < 0.001	–	–	–
0.850< ν_N <0.900	0.001± < 0.001	–	–	–	–
0.900< ν_N <0.940	0.002± < 0.001	0.001± < 0.001	–	–	–
0.940< ν_N <0.970	0.003±0.001	0.001± < 0.001	–	–	–
0.970< ν_N <0.987	0.004±0.001	0.001± < 0.001	–	–	–
0.987< ν_N <0.995	0.006±0.001	0.002± < 0.001	0.001± < 0.001	–	–
0.995< ν_N <1.000	0.021±0.004	0.008±0.001	0.003± < 0.001	0.001± < 0.001	–
CF					
0.700< ν_N <0.760	–	–	–	–	–
0.760< ν_N <0.850	–	–	–	–	–
0.850< ν_N <0.900	–	–	–	–	–
0.900< ν_N <0.940	0.001± < 0.001	–	–	–	–
0.940< ν_N <0.970	0.001± < 0.001	–	–	–	–
0.970< ν_N <0.987	0.001± < 0.001	0.001± < 0.001	–	–	–
0.987< ν_N <0.995	0.001± < 0.001	–	–	–	–
0.995< ν_N <1.000	0.005±0.001	0.002± < 0.001	0.001± < 0.001	–	–

Table A.4: $B \to h^+h'^-$ background estimates for B_d^0 signal window for all NN bins using only the new data.

NN bin \ Mass bin (GeV/c^2)	5.219–5.243	5.243–5.267	5.267–5.291	5.291–5.315	5.315–5.339
CC					
0.700< ν_N <0.760	0.004± < 0.001	0.004± < 0.001	0.003±0.001	0.002± < 0.001	0.001± < 0.001
0.760< ν_N <0.850	0.008±0.001	0.008±0.001	0.006±0.001	0.003±0.001	0.001± < 0.001
0.850< ν_N <0.900	0.006±0.001	0.005±0.001	0.004±0.001	0.002± < 0.001	0.001± < 0.001
0.900< ν_N <0.940	0.009±0.001	0.008±0.001	0.006±0.001	0.004±0.001	0.001± < 0.001
0.940< ν_N <0.970	0.014±0.001	0.011±0.001	0.009±0.001	0.005±0.001	0.001± < 0.001
0.970< ν_N <0.987	0.017±0.002	0.015±0.002	0.012±0.002	0.007±0.001	0.003± < 0.001
0.987< ν_N <0.995	0.024±0.002	0.024±0.003	0.018±0.003	0.010±0.002	0.003±0.001
0.995< ν_N <1.000	0.094±0.009	0.086±0.010	0.064±0.010	0.037±0.006	0.014±0.002
CF					
0.700< ν_N <0.760	0.001± < 0.001	0.001± < 0.001	0.001± < 0.001	0.001± < 0.001	–
0.760< ν_N <0.850	0.002± < 0.001	0.002± < 0.001	0.002± < 0.001	0.001± < 0.001	–
0.850< ν_N <0.900	0.002± < 0.001	0.002± < 0.001	0.001± < 0.001	0.001± < 0.001	–
0.900< ν_N <0.940	0.002± < 0.001	0.002± < 0.001	0.002± < 0.001	0.001± < 0.001	–
0.940< ν_N <0.970	0.003± < 0.001	0.003± < 0.001	0.002± < 0.001	0.001± < 0.001	–
0.970< ν_N <0.987	0.004± < 0.001	0.005±0.001	0.004±0.001	0.002± < 0.001	0.001± < 0.001
0.987< ν_N <0.995	0.004± < 0.001	0.004± < 0.001	0.003±0.001	0.002± < 0.001	0.001± < 0.001
0.995< ν_N <1.000	0.020±0.002	0.022±0.003	0.017±0.003	0.010±0.002	0.004±0.001

Table A.5: B_s^0 signal window for CC(top) and CF(bottom): Expected backgrounds, including $B \to h^+h'^-$, and number of observed events using new data only.

	Mass bin (GeV/c^2)	5.31–5.334	5.334–5.358	5.358–5.382	5.382–5.406	5.406–5.43	Total
CC NN bin	Exp. Bkg.	2.56±0.34	2.52±0.33	2.49±0.33	2.46±0.32	2.42±0.32	12.45
0.7–0.76	Obs.	4	2	1	4	2	13
CC NN bin	Exp. Bkg.	2.77±0.35	2.73±0.35	2.69±0.34	2.66±0.34	2.62±0.33	13.47
0.76–0.85	Obs.	1	2	1	4	1	9
CC NN bin	Exp. Bkg.	1.22±0.23	1.2±0.23	1.18±0.22	1.17±0.22	1.15±0.22	5.92
0.85–0.9	Obs.	1	2	3	1	1	8
CC NN bin	Exp. Bkg.	1.05±0.21	1.03±0.21	1.02±0.21	1.01±0.2	0.99±0.2	5.11
0.9–0.94	Obs.	1	0	1	1	1	4
CC NN bin	Exp. Bkg.	1.05±0.21	1.04±0.21	1.02±0.21	1.01±0.2	0.99±0.2	5.11
0.94–0.97	Obs.	1	2	0	0	0	3
CC NN bin	Exp. Bkg.	0.63±0.18	0.62±0.17	0.61±0.17	0.6±0.17	0.6±0.17	3.07
0.97–0.987	Obs.	0	0	1	1	0	2
CC NN bin	Exp. Bkg.	0.13±0.08	0.13±0.08	0.12±0.07	0.12±0.07	0.12±0.07	0.62
0.987–0.995	Obs.	0	0	0	1	0	1
CC NN bin	Exp. Bkg.	0.11±0.07	0.09±0.07	0.08±0.07	0.08±0.07	0.08±0.07	0.44
0.995–1	Obs.	0	0	0	0	0	0
CF NN bin	Exp. Bkg.	1.74±0.28	1.72±0.27	1.69±0.27	1.67±0.27	1.64±0.26	8.46
0.7–0.76	Obs.	2	3	3	2	1	11
CF NN bin	Exp. Bkg.	1.83±0.28	1.8±0.28	1.78±0.28	1.75±0.27	1.72±0.27	8.88
0.76–0.85	Obs.	1	5	2	2	0	10
CF NN bin	Exp. Bkg.	1.23±0.23	1.21±0.23	1.2±0.23	1.18±0.22	1.16±0.22	5.99
0.85–0.9	Obs.	0	0	3	0	1	4
CF NN bin	Exp. Bkg.	0.81±0.19	0.8±0.18	0.78±0.18	0.77±0.18	0.76±0.18	3.92
0.9–0.94	Obs.	2	1	2	2	1	8
CF NN bin	Exp. Bkg.	0.68±0.17	0.67±0.17	0.66±0.17	0.65±0.16	0.64±0.16	3.3
0.94–0.97	Obs.	1	1	0	0	0	2
CF NN bin	Exp. Bkg.	0.38±0.13	0.38±0.13	0.37±0.13	0.37±0.13	0.36±0.13	1.86
0.97–0.987	Obs.	0	2	0	0	1	3
CF NN bin	Exp. Bkg.	0.17±0.09	0.17±0.09	0.17±0.09	0.16±0.09	0.16±0.09	0.83
0.987–0.995	Obs.	0	0	1	0	0	1
CF NN bin	Exp. Bkg.	0.18±0.11	0.17±0.11	0.17±0.11	0.16±0.11	0.16±0.11	0.83
0.995–1	Obs.	0	0	0	0	0	0

Table A.6: B_d^0 signal window for CC(top) and CF(bottom): Expected backgrounds, including $B \to h^+h'^-$, and number of observed events using new data only.

	Mass bin (GeV/c^2)	5.219–5.243	5.243–5.267	5.267–5.291	5.291–5.315	5.315–5.339	Total
CC NN bin	Exp. Bkg.	2.69±0.35	2.65±0.35	2.62±0.34	2.58±0.34	2.55±0.34	13.09
0.7–0.76	Obs.	4	4	4	2	4	18
CC NN bin	Exp. Bkg.	2.91±0.37	2.87±0.36	2.84±0.36	2.8±0.35	2.76±0.35	14.18
0.76–0.85	Obs.	4	0	2	2	0	8
CC NN bin	Exp. Bkg.	1.28±0.24	1.26±0.24	1.25±0.23	1.23±0.23	1.21±0.23	6.24
0.85–0.9	Obs.	3	1	0	1	1	6
CC NN bin	Exp. Bkg.	1.11±0.22	1.09±0.22	1.08±0.22	1.06±0.21	1.05±0.21	5.39
0.9–0.94	Obs.	1	2	1	3	1	8
CC NN bin	Exp. Bkg.	1.11±0.22	1.1±0.22	1.08±0.22	1.06±0.21	1.05±0.21	5.4
0.94–0.97	Obs.	0	0	0	1	1	2
CC NN bin	Exp. Bkg.	0.68±0.18	0.67±0.18	0.65±0.18	0.64±0.18	0.63±0.17	3.27
0.97–0.987	Obs.	1	0	0	0	0	1
CC NN bin	Exp. Bkg.	0.16±0.08	0.15±0.08	0.15±0.08	0.14±0.08	0.13±0.08	0.72
0.987–0.995	Obs.	1	0	1	1	0	3
CC NN bin	Exp. Bkg.	0.18±0.09	0.17±0.08	0.15±0.08	0.12±0.07	0.1±0.07	0.72
0.995–1	Obs.	1	0	0	0	0	1
CF NN bin	Exp. Bkg.	1.84±0.29	1.81±0.29	1.79±0.28	1.76±0.28	1.74±0.28	8.93
0.7–0.76	Obs.	1	3	2	4	1	11
CF NN bin	Exp. Bkg.	1.93±0.3	1.9±0.3	1.87±0.29	1.85±0.29	1.82±0.28	9.37
0.76–0.85	Obs.	2	3	2	0	2	9
CF NN bin	Exp. Bkg.	1.3±0.24	1.28±0.24	1.26±0.24	1.25±0.23	1.23±0.23	6.32
0.85–0.9	Obs.	0	0	2	0	0	2
CF NN bin	Exp. Bkg.	0.85±0.2	0.84±0.19	0.83±0.19	0.82±0.19	0.81±0.19	4.15
0.9–0.94	Obs.	0	3	2	2	1	8
CF NN bin	Exp. Bkg.	0.72±0.18	0.71±0.18	0.7±0.18	0.69±0.17	0.68±0.17	3.49
0.94–0.97	Obs.	0	1	0	0	1	2
CF NN bin	Exp. Bkg.	0.41±0.14	0.4±0.14	0.4±0.14	0.39±0.14	0.38±0.13	1.98
0.97–0.987	Obs.	1	0	0	0	1	2
CF NN bin	Exp. Bkg.	0.18±0.1	0.18±0.09	0.18±0.09	0.17±0.09	0.17±0.09	0.88
0.987–0.995	Obs.	0	0	0	0	0	0
CF NN bin	Exp. Bkg.	0.2±0.12	0.2±0.12	0.19±0.12	0.18±0.11	0.17±0.11	0.94
0.995–1	Obs.	0	0	0	0	0	0

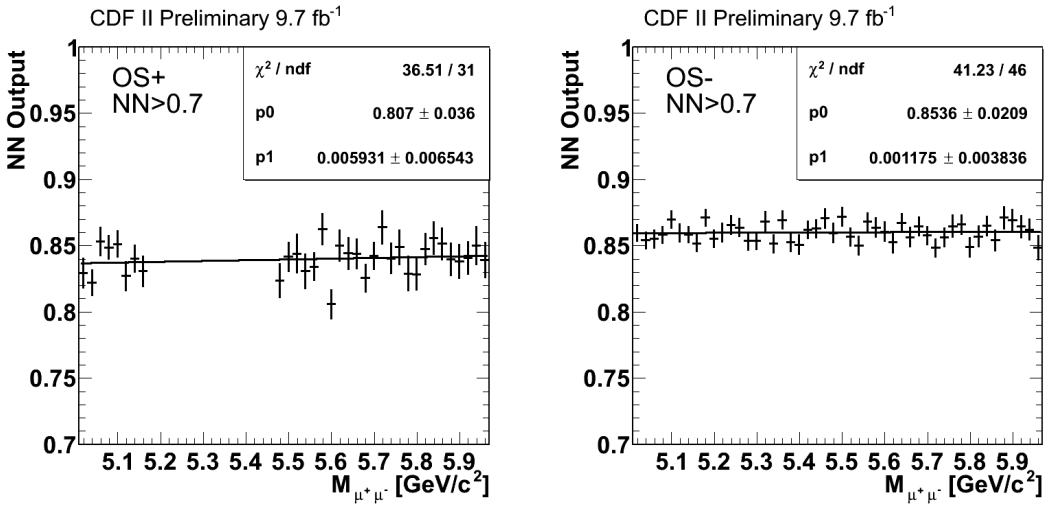

Figure A.1: The NN output as a function of di-muon invariant mass fitted to a first order polynomial for the OS+ and OS– sample.