RICE UNIVERSITY

Search for the Rare Decay $B_s^0 \to \mu^+\mu^-$ at DØ

by

Michelle Victoria Prewitt

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Approved, Thesis Committee:

Marjorie D. Corcoran, Chair
Professor of Physics and Astronomy

Karl Ecklund
Assistant Professor of Physics and Astronomy

David W. Scott
Noah Harding Professor of Statistics

Houston, Texas

April, 2013

ABSTRACT

Search for the Rare Decay $B_s^0 \to \mu^+\mu^-$ at DØ

by

Michelle Victoria Prewitt

Results of the search for the rare decay $B_s^0 \to \mu^+\mu^-$ using data collected by the D0 detector at the Fermilab Tevatron collider are presented. This analysis covers the full Run II data set, corresponding to approximately 10.4 fb^{-1} of integrated luminosity in $p\bar{p}$ collisions at a center of mass energy of 1.96 TeV. The analysis used new variables and a multivariate technique to improve the background reduction. After seeing fewer events than expected from background, a new Tevatron best observed limit was set on the branching fraction of the decay at $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 15 \times 10^{-9}$ (12×10^{-9}) at the 95% (90%) C.L. [1].

Acknowledgments

It takes a village to raise a scientist. Along the way, many people have contributed to my scientific education by imparting small but important pieces of knowledge or challenging my perspective just enough to help me see the solution to a problem. I cannot begin to acknowledge them all. However, I would like to take a moment to express my gratitude to a select few, without whom, I would not have successfully become a physicist.

To Marj Corcoran,

You are a great advisor. More importantly, you are an amazing mentor and role model. Thank you the support, encouragement, and occasional kick in the pants.

To George Ginther, Bill Lee, and Geoff Savage,

Whether it was on the field/court, in the control room, or in life, you have all had my back. Thanks for the words of wisdom. You helped me handle the stressful times and always help me focus on what was truly important.

To Yuji Enari, Rich Partridge, and Brendan Casey,

Without you guys I would have never been able to keep the luminosity system working so well. Thanks for making me look good!

To ALL my Fermilab friends,

I am not even going to try to list you guys . . . Whether it was hanging out in the village or at Lincoln Karaoke, you guys made life a blast. Even though I might

not see you guys as much as I would like, you have all made a lasting impression on me. I am sure I have made an impression on most of you, because at some point you will stop and notice how quiet it is without me around. You're welcome.

To My family,

I know how proud of me you are. I'm lucky to have such a crazy group of people supporting me. I promise to use my powers for good. Mom, Dad, Charles, Tiffany, Charlie, Aiden, Chris, and Carrie - Thank you for your support!

To Mark,

Especially to Mark . . . Thank You! I cannot begin to express the joy you have brought to my life. Thanks for letting me do things my way. Your love and support has meant the world to me.

Contents

Abstract ii

Acknowledgments iii

List of Illustrations vii

List of Tables xi

1 The Foundation 1

1.1 Standard Model . . . 1

1.2 Particle Interactions . . . 2

1.3 Beyond the Standard Model . . . 5

1.4 $B_s^0 \to \mu^+\mu^-$ Search Motivation . . . 6

1.5 Previous Measurements . . . 7

2 The Machines 11

2.1 Accelerators . . . 12

2.1.1 The Accelerator Chain . . . 12

2.1.2 Tevatron . . . 15

2.2 DØ . . . 16

2.2.1 Silicon Microstrip Tracker . . . 17

2.2.2 Central Fiber Tracker . . . 18

2.2.3 Calorimeter . . . 20

2.2.4 Muon System . . . 21

2.2.5 Luminosity Monitor . . . 23

2.2.6 Triggering . . . 25

2.2.7 DØ Data . . . 27

3 The Ground Work **29**

3.1 Analysis Overview 29

3.1.1 Signal versus Background 31

3.2 Event Selection 34

3.3 Monte Carlo Simulation 36

3.3.1 p_T Reweighting 38

3.3.2 Isolation Reweighting 39

3.3.3 Mass Smearing 40

3.4 Single Event Sensitivity 43

3.4.1 Determining the Number of $B^{\pm}$ Events 45

3.4.2 Efficiency Ratio 46

4 The Result **49**

4.1 Boosted Decision Tree 49

4.2 Optimizing BDT Performance 50

4.2.1 Pre-training Cuts 50

4.2.2 Variables 51

4.3 Estimating the Number of Events 58

4.4 Other Backgrounds 59

4.5 Expected Limit 64

4.6 Opening the Box 64

4.7 Cross Checks 65

4.8 Conclusions 66

A p_T Reweight Validation **70**

B Unblinded Cross Checks **76**

Bibliography **96**

Illustrations

1.1 Particles of the Standard Model 3
1.2 $B_s^0 \to \mu^+\mu^-$ Feynman Diagrams 7
1.3 BSM Feynman Diagrams 8
1.4 MSSM Parameter Space 9
1.5 Previous $B_s^0 \to \mu^+\mu^-$ Search Results 10

2.1 Fermilab Accelerator Complex 13
2.2 Cockcroft Walton 14
2.3 D0 Detector 16
2.4 Silicon Microstrip Tracker 18
2.5 CFT Fiber Layer 19
2.6 Tracking System 20
2.7 Calorimeter 21
2.8 Muon System 22
2.9 Luminosity Monitor 23
2.10 Trigger System 26
2.11 D0 RunII Luminosity 28

3.1 Signal Mass Distribution 30
3.2 Signal and Background Cartoons 31
3.3 Sideband Subtraction 37
3.4 Trigger Efficiency Reweighting 39

3.5 p_T Reweighting 40
3.6 Isolation Reweighting 41
3.7 Individual Muon Isolation 42
3.8 Mass Smearing 43
3.9 Reweighting Checks 44
3.10 $B^{\pm}$ Mass Distribution 45
3.11 Four Track Mass Distribution 47

4.1 Pre-training Cuts 51
4.2 Selected Useful Variables 55
4.3 Correlation Matrix (Signal) 56
4.4 Correlation Matrix (Background) 57
4.5 BDT Response 61
4.6 BDT Response Correlation 62
4.7 Expected Background 63
4.8 Unblinded Distribution 65
4.9 Variable Cross Checks 67
4.10 BDT Cut Cross Check 68
4.11 Current $B_s^0 \to \mu^+\mu^-$ Search Results 69

A.1 p_T Reweighting RunIIa 71
A.2 p_T Reweighting RunIIb1 72
A.3 p_T Reweighting RunIIb2 73
A.4 p_T Reweighting RunIIb3 74
A.5 p_T Reweighting RunIIb4 75

B.1 Unblinded BDT1 Response 76
B.2 Unblinded BDT2 Response 77

B.3 Unblinded Dimuon p_T . . . 77
B.4 Unblinded 2D Pointing Angle . . . 78
B.5 Unblinded 3D Pointing Angle . . . 78
B.6 Unblinded 2D Decay Length . . . 79
B.7 Unblinded 3D Decay Length . . . 80
B.8 Unblinded 2D Decay Length Significance . . . 80
B.9 Unblinded 2D Impact Parameter . . . 81
B.10 Unblinded 2D Impact Parameter Significance . . . 81
B.11 Unblinded Vertex χ^2 . . . 82
B.12 Unblinded 2D p_t^2 . . . 82
B.13 Unblinded 3D p_t^2 . . . 83
B.14 Unblinded Dimuon Isolation . . . 84
B.15 Unblinded $\mu 1$ Isolation . . . 85
B.16 Unblinded $\mu 2$ Isolation . . . 85
B.17 Unblinded Muon Isolation Sum . . . 86
B.18 Unblinded $\mu 1$ p_T . . . 86
B.19 Unblinded $\mu 2$ p_T . . . 87
B.20 Unblinded $\mu 1$ Impact Parameter . . . 87
B.21 Unblinded $\mu 2$ Impact Parameter . . . 88
B.22 Unblinded $\mu 1$ Impact Parameter Significance . . . 89
B.23 Unblinded $\mu 2$ Impact Parameter Significance . . . 90
B.24 Unblinded Lesser Impact Parameter Significance . . . 90
B.25 Unblinded Delta ϕ . . . 91
B.26 Unblinded New Pointing Angle . . . 91
B.27 Unblinded Tertiary Vertex Mass . . . 92
B.28 Unblinded $\mu 1$ Tertiary Vertex Mass . . . 92
B.29 Unblinded $\mu 2$ Tertiary Vertex Mass . . . 93
B.30 Unblinded $\mu 1$ Vertex χ^2 . . . 94

B.31 Unblinded $\mu 2$ Vertex χ^2 95
B.32 Unblinded $\mu 2$ Pointing Angle 95

Tables

1.1 Standard Model Gauge Bosons 4

3.1 Single Event Sensitivity by Epoch 48

Chapter 1

The Foundation

Human beings are inquisitive, ever seeking to understand the mystery of the world around them. This intrinsic curiosity leads to categorization. With commonality and organization comes rules. In this way, order is brought to the chaos. However, it is when the rules are broken that the mystery is solved.

In 1911, an alpha particle did not pass straight through a gold foil. The unexpected behavior led Rutherford to surmise atoms existed in distinct sections: a very small positive central charge with a negative outer region. The experimental outcome did not match prediction, i.e. it broke the rules. In the study of physics, and science in general, a set of rules that well-describes all known experimental results and also makes testable predictions is called a theory. When new evidence is found that doesn't fit with the rules, the theory must be altered to accommodate the new facts. This openness to new information allows science to adapt while keeping a rigorous requirement on the integrity of new information and continually challenging mankind to attempt to break the rules.

1.1 Standard Model

During the 1960s and early 1970s a set of rules was developed to organize the interactions of the basic building blocks of matter. The Standard Model was born. The Standard Model (SM) identifies 12 fundamental particles, the quarks and leptons, and

describes interactions between particles using four gauge bosons. Figure 1.1 shows that quarks and leptons, which are fermions, are divided into three generations. Particles in the second and third generations mimic the properties of the first generation, but with larger mass. The up (u), charm (c), and top (t) quarks have electric charge, which is measured by the value of the characteristic charge of the electron, of $+2/3$. The down (d), strange (s), and bottom (b) quarks have charge of -1/3. The neutrinos (ν) are neutral, while the electron (e), muon (μ), and tau (τ) have charge of -1. The fermions all have spin that is a half integer of $\hbar$, while bosons have intrinsic angular momentum that is of integer value of $\hbar$. Leptons have a property called lepton flavor, that is related to their generation and charge. Lepton flavor conservation in interactions is conserved. Quarks also have flavor, but each type of quark is a distinct flavor. Quarks flavor change is observed. Quarks carry an additional charge called color which is related to the strong interaction. Each of the quarks and leptons also have an anti-particle counterpart. The anti-particles have the same intrinsic properties, like mass, but have opposite electric and color charge.

1.2 Particle Interactions

The quarks, leptons and their anti-particles combine to make up all the visible matter in the universe. Most of the everyday world is made from the up and down quark and the electron. For example, the proton is a combination of two up quarks and a down quark (uud). Neutrons are a ddu. Protons, neutrons and electrons combine to form atoms. Atoms combine to form molecules and compounds, which in turn combine to form the paper or computer screen this is written on.

A group of bound quarks or anti-quarks is referred to as a hadron. A baryon is a bound state of three quarks (qqq) or anti-quarks ($\bar{q}\bar{q}\bar{q}$) into a particle. Quark

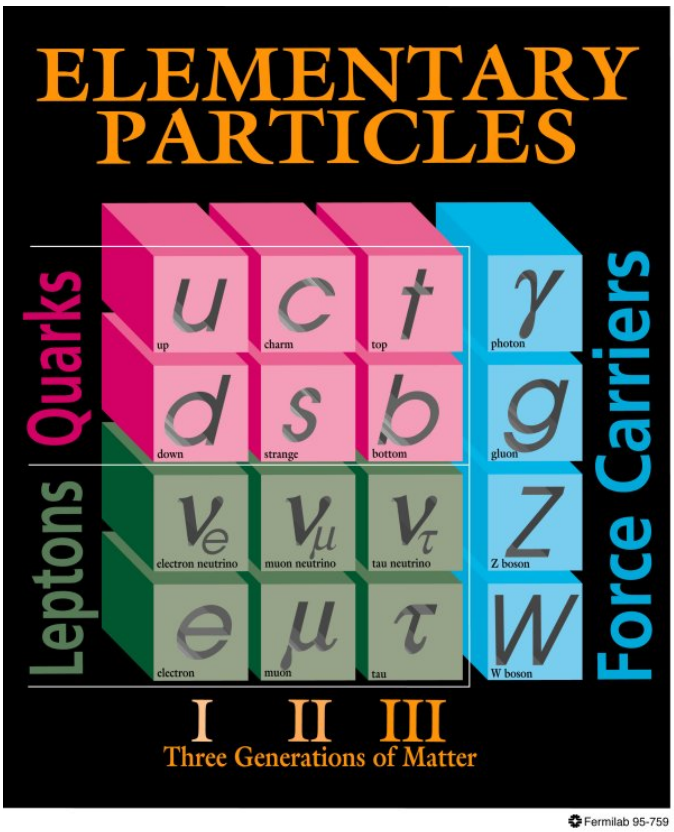


Figure 1.1 : The Standard Model describes particle interactions using 12 fundamental particles, the quarks and leptons, and 4 force carrying bosons. Image courtesy of Fermilab.

anti-quark pairings ($q\bar{q}$) create a meson. Bound states of quarks are held together by the strong force, which is mediated by gluons (g). The SM has 8 varieties of gluons, which carry a color/anti-color combination. Only color neutral particles can bind together. The color charge values are red (R), blue (B), and green (G). Anti-quarks have anti-color values of cyan ($\bar{R}$), yellow ($\bar{B}$), and magenta ($\bar{G}$). A baryon is the color singlet with the wavefunction $\frac{1}{\sqrt{6}}[RBG - RGB - BRG + BGR + GRB - GBR]$. Mesons have a neutral color wavefunction of $\frac{1}{3}[R\bar{R} + B\bar{B} + G\bar{G}]$. Quarks only bind to each other because of the color charge, however, they interact with other particles

Boson	**Charge**	**Mass (GeV)**	**Force**	**Strength**	**Range (m)**
γ	0	0	Electromagnetic	1/137	∞
$g_{i=1-8}$	0	0	Strong	1	10^{-15}
Z	0	91.2	Weak	10^{-6}	10^{-18}
$W^{\pm}$	0	80.4	Weak	10^{-6}	10^{-18}

Table 1.1 : Properties of the Standard Model gauge bosons.

through the other bosons.

The strong interaction is a short range force that only couples to quarks, anti-quarks and gluons since the gluons are self interacting due to color charge. The strong force is the glue that holds together hadrons. The photon (γ) mediates electromagnetic interactions. Since the photon is massless and not self interacting, the electromagnetic force has infinite range. The $W^{\pm}$ and Z bosons, which are massive, mediate weak interactions and couple to all fermions. The weak force is shorter range and weaker than than the strong force due to the exchange of the heavy vector bosons. Properties of the force carriers are shown in Table 1.2.

The weak interaction is responsible for flavor change, changing one type of quark or lepton to another [2, 3]. For example, it is responsible for radioactive decay, in which a neutron changes to a proton emitting an electron and an electron anti-neutrino, ($udd \rightarrow uud + e + \bar{\nu_e}$). The weak force is particular in its coupling to particles of certain helicity. Helicity is a particle property related to spin and momentum. If the spin and direction a particle is traveling is aligned, the particle is said to be right-handed. If these qualities are opposite, the particle is left-handed. The weak force preferentially couples to left-handed fermions and right-handed anti-fermions. The weak force is also needed to have flavor changing neutral current (FCNC) processes. FCNC decays change the flavor of the fermion without changing the charge. These

types of processes are forbidden at first order in the SM but are possible through higher-order processes such as W boson loops.

1.3 Beyond the Standard Model

The Standard Model is a well vetted theory. It predicted the mass and properties of the $W^{\pm}$ and Z bosons as part of the unification of the electromagnetic and weak forces. It also predicted the top quark, which was discovered at Fermilab in 1995 [4]. Despite the remarkable strength of the SM, it is incomplete. The incompleteness of the SM is related to the many free parameters in the theory: masses, coupling constants, flavor mixing matrix elements, etc. For example, the SM assumes neutrinos are massless. However, neutrinos oscillate between flavors so they must have mass, even if it is small. The SM allows for for matter/anti-matter asymmetry through the quark mixing matrix, but the measured values of these parameters are not large enough to account for the asymmetry seen in the universe. The Higgs boson is another area where the SM is not completely confirmed. The SM calls for a scalar which couples to particles to give them mass. This particle is an excitation of the field that is required to have massive bosons for the weak force. A Higgs boson consistent with the SM has been observed at the Large Hadron Collider in CERN, but further testing is needed to determine if the particle is in fact "the" Higgs boson. In an attempt to improve upon the SM to solve these issues, many formulations Beyond the Standard Model (BSM) have been drafted by theorists. Since these models must still explain the experimental results that are seen, they have testable properties. New particles are often predicted in BSM scenarios. This is one way to test the models and search for new physics. Another method to test both the SM and BSM scenarios is to compare predicted and observed properties.

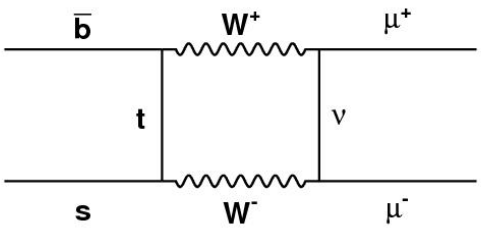
1.4 $B_s^0 \to \mu^+\mu^-$ Search Motivation

The B_s^0 meson is composed of a $\bar{b}s$ quark pair, and the $\bar{B}_s^0$ is a $\bar{s}b$ combination. The B_s^0 can interchange to a $\bar{B}_s^0$ and back, a process called mixing, through double W exchange. The mass eigenstates of the meson are the light ($B_s^0 + \bar{B}_s^0$) and heavy ($B_s^0 - \bar{B}_s^0$) combination, which have different lifetimes. The decay $B_s^0 \to \mu^+\mu^-$ proceeds by a flavor changing neutral current (FCNC) process illustrated by the Feynman diagram in Fig. 1.2. In the Standard Model, FCNC decays are heavily suppressed. $B_s^0 \to \mu^+\mu^-$ is further suppressed by the helicity requirements of the final state. The SM branching fraction is found to be $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 3.5 \pm 0.2 \times 10^{-9}$ [5, 6]. Branching fractions are a measurement of the rate at which a particular event occurs. Until recently, a value of $\mathcal{B} = 3.2 \pm 0.2 \times 10^{-9}$ was used as the SM value [7]. The most recent value for the branching ratio has been corrected to account for the different lifetimes, $\Delta\Gamma_s$, of the heavy and light mass eigenstates.

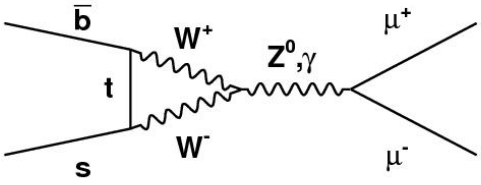
In BSM physics, the decay of the B_s^0 can proceed through new particles in the W loop. In the Minimal Supersymmetric Standard Model (MSSM), for example, the decay can proceed with a light Higgs boson, as shown in Fig. 1.3. Any new particle that can enter in the loops can have a significant effect on the branching fraction.

Many BSM scenarios, including MSSM, predict substantial enhancement to the branching fraction of the $B_s^0 \to \mu^+\mu^-$. Figure 1.4 shows the parameter space of a selection of MSSM.

Measuring the branching fraction of $B_s^0 \to \mu^+\mu^-$ allows limits to be placed on the physics models. Previous D0 results and results from other experiments have already excluded some of the models plotted in Fig. 1.4. However, many models are still allowed or can be tweaked to be viable given the limits on this branching fraction [9, 10, 11]. Perhaps more interesting than the models which would enhance the



(a)



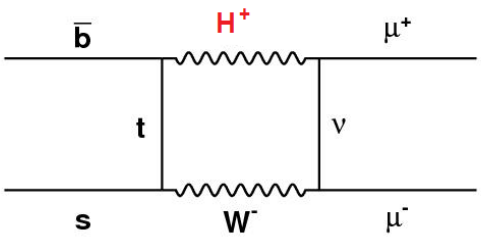
(b)

Figure 1.2 : The (a) box diagram and (b) electroweak penguin diagram are examples of the FCNC processes through which the decay $B_s^0 \to \mu^+\mu^-$ can proceed.

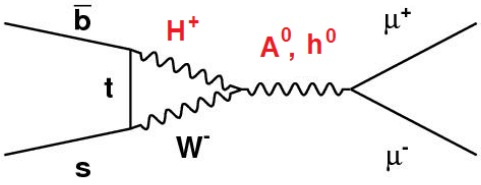
branching fraction of this decay above the SM value are models that predict further suppression of the branching fraction [12]. Thus, accurately measuring $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is an excellent way to look for new physics.

1.5 Previous Measurements

Confirmation of the $B_s^0 \to \mu^+\mu^-$ decay would provide an interesting test of new physics. The search has been done both at the Tevatron experiments D0 and CDF and by the ATLAS, CMS, and LCHb collaborations at the LHC. The results for this search at the Tevatron, shown in Fig. 1.5(a), show how this search has improved with new techniques and more data. Figure 1.5(b) shows the recent results for the $B_s^0 \to \mu^+\mu^-$ search at the time this analysis was made public for both the Tevatron



(a)



(b)

Figure 1.3 : The (a) box diagram and (b) electroweak penguin diagram showing the decay of $B_s^0 \to \mu^+\mu^-$ propagated with a light Higgs as is possible in MSSM.

and LHC experiments.

The last search for $B_s^0 \to \mu^+\mu^-$ at D0 in 2010 yielded a 95% C.L. upper limit on the branching fraction of 5.1×10^{-8} [13]. The CDF collaboration saw an excess over background in their last result, giving a branching fraction of $(18^{+11}_{-9}) \times 10^{-9}$ and a 95% C.L. upper limit of 40×10^{-9} [14, 15]. More recently, they released a final update of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (13^{+9}_{-7}) \times 10^{-9}$ and a 95% C.L. upper limit of 31×10^{-9} [16] that also sees an excess. At the LHC, the CMS and ATLAS experiments set upper limits at 95% C.L. of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 7.7 \times 10^{-9}$ [17] and 22×10^{-9} [18], respectively. The LHCb collaboration recently presented the first evidence for this decay of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.2^{+1.5}_{-1.2}) \times 10^{-9}$ [19], which is consistent with the SM predicted branching fraction; however, this result has not yet been confirmed.

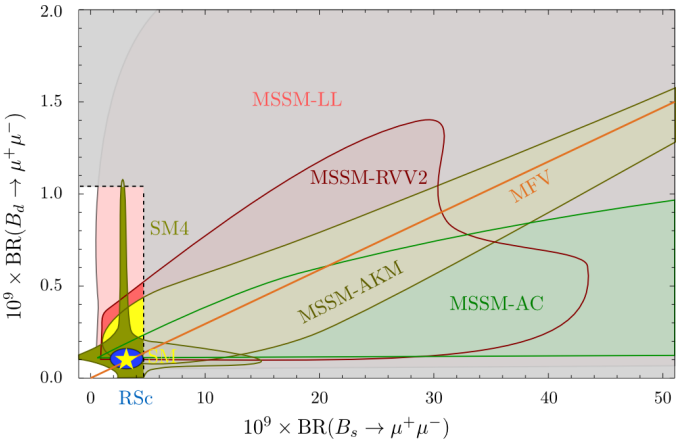
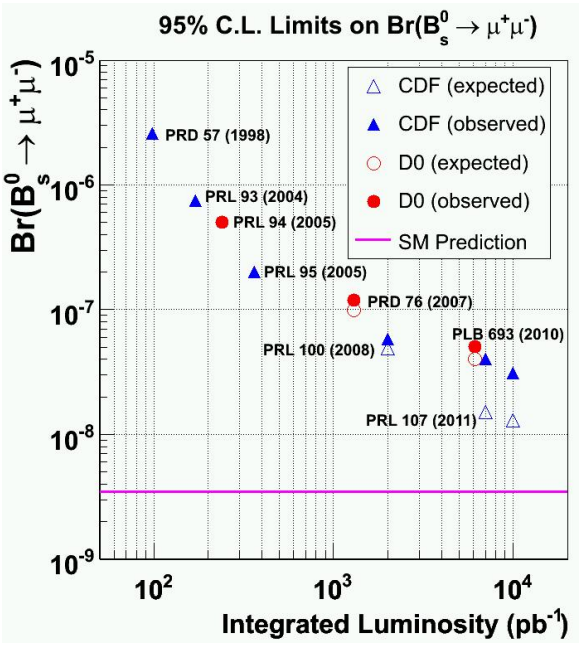
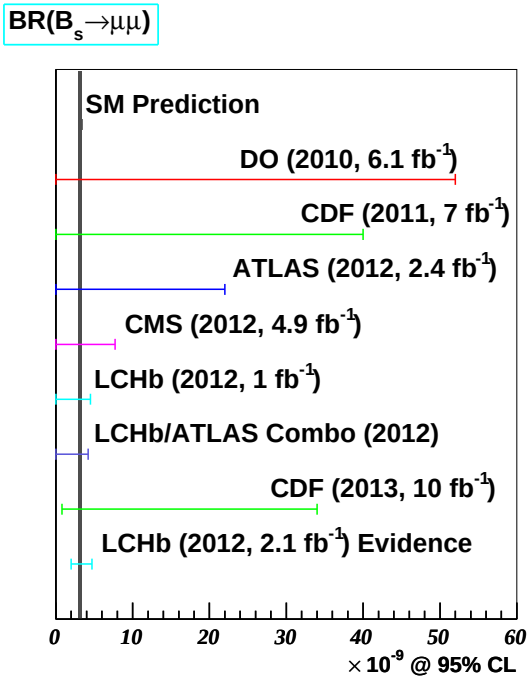


Figure 1.4 : A plot from Straub showing a selection of MSSM scenarios in the parameter space of branching fractions of $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$. The areas that are grayed out had been excluded as of Winter 2012. [8]



(a)



(b)

Figure 1.5 : The previous (a) Tevatron and (b) recent world results for the search for $B_s^0 \to \mu^+\mu^-$. Most values are the 95% C.L. upper limit on the branching fraction. The LHCb 2.1 fb^{-1} value is evidence for the decay at the SM branching fraction.

Chapter 2

The Machines

Fermi National Accelerator Laboratory (FNAL) is known to its employees, users, and visitors as Fermilab. Located in Batavia, IL, it is the premier particle physics laboratory in the United States. Research in particle physics is split into three frontiers: Cosmic, Intensity, and Energy. The Cosmic Frontier focuses on particle astrophysics. Telescopes and other devices are used to study galaxies, black holes, and particles created in the cosmos. The Intensity Frontier uses intense beams to search for rare processes, such as neutrino oscillations or the planned search for the coherent conversion of $\mu \rightarrow e$ and the precise measurement of the anomalous magnetic moment of the muon. The Energy Frontier is used to probe the smallest scales. As with a microscope, higher energy means seeing smaller objects. High energy particle physics is looking at the most fundamental particles in nature. This study requires a great deal of energy and very large machines.

Two types of machines, accelerators and detectors, are most important in the study of high energy physics. Physicists study fundamental particles and interactions by smashing leptons and/or composite particles together at high energies. Accelerators are used to give energy to the particles that are going to be smashed. Detectors are used to determine what is created after the smash. The more energy that is given to the smashing particles, the more variety of particles that can be created. Energy is conserved such that sum of the energy of the particles created from the collision must be less than or equal to the sum of the energy of the particles which collide.

2.1 Accelerators

The accelerators at Fermilab have continued to be remodeled and improved as the focus of the research done at the lab has shifted since the lab's opening in the late 1960s. Currently, research at Fermilab focuses on the Intensity and Cosmic Frontiers. Fermilab is still involved in research on the Energy Frontier and was the world leading producer of high energy physics data with the operation of the Tevatron. The Tevatron is an accelerator at Fermilab where protons and anti-protons were collided at high energy for research. The Tevatron was operational at a lower energy in the 1990s, where data led to the discovery of the top quark. This analysis uses data from the Run II period of the Tevatron from 2002 till the shutdown of the Tevatron on September 30, 2011. The Tevatron is the last accelerator in a series of accelerators shown in Fig. 2.1 used to accelerate particles to almost 1 TeV (1×10^9 electron Volts) per beam.

2.1.1 The Accelerator Chain

Physics at the Tevatron was done by colliding protons (p) and anti-protons ($\bar{p}$), both of which need to be collected and accelerated. Protons are very abundant, but anti-protons which are the anti-particle of the proton, are more rare since they annihilate with matter. Thus, much effort went into collecting and storing anti-protons.

Starting from a bottle of hydrogen, a Cockcroft-Walton generator, shown in Fig. 2.2, is used to produce negatively charged hydrogen ions and accelerate the ions to an energy of 750 keV. The negatively charged hydrogen ions then enter a linear accelerator, called the Linac. The Linac takes the ions from 750 keV up to 400 MeV and transports them to the Booster in pulses. Before entering the booster the ions are stripped of the electrons by means of a foil, so only the protons enter the

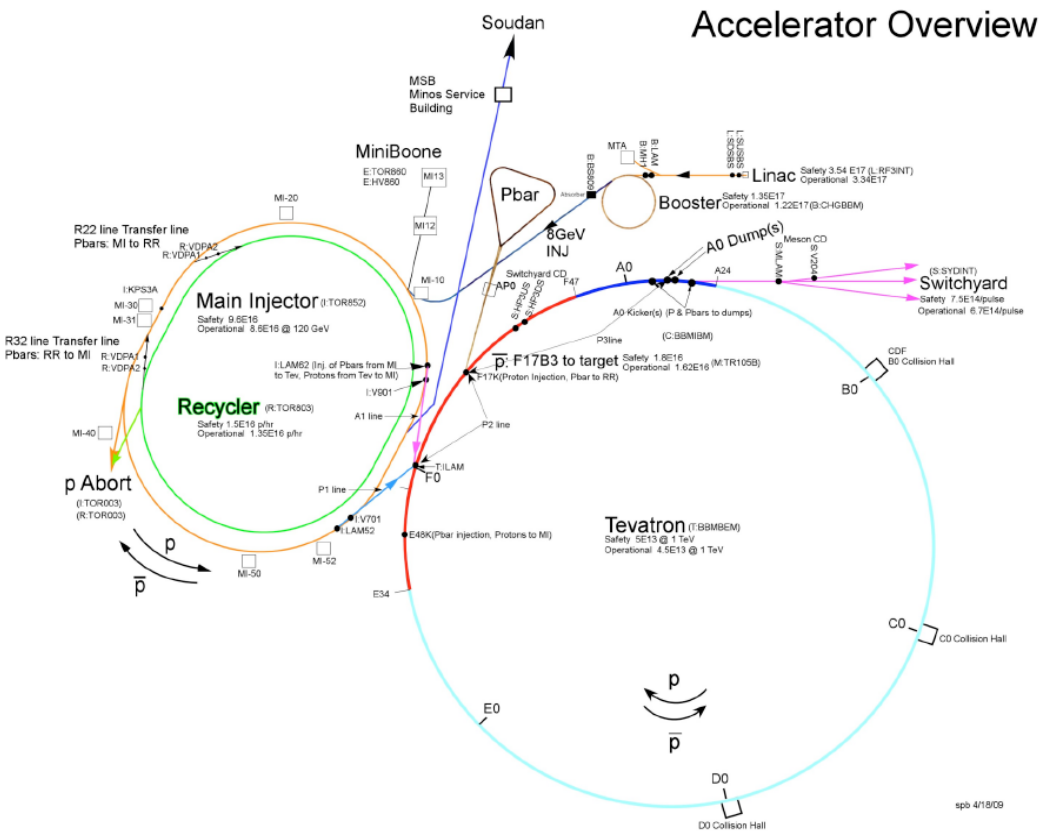


Figure 2.1 : Schematic of the Fermilab Accelerator Complex. A series of accelerators is used to produce colliding beams of protons and anti-protons in the Tevatron at a center of mass energy of $\sqrt{s} = 1.96$ TeV [20].

Booster. The Booster raises the energy of the protons to 8 GeV. Since the protons have entered the Booster in pulses, the protons are now grouped in clusters called bunches. The rest of the accelerators deal with bunches of particles.

Bunches are important because the accelerators work by oscillating electric fields through "on" and "off" cycles. Magnetic fields are used for steering and electric fields are used to give the particles a kick of energy. It is difficult, costly, and sometimes impossible to get magnetic and electric fields to run at constant strength. Having bunches of particles means that with careful timing, the fields can be switched on in

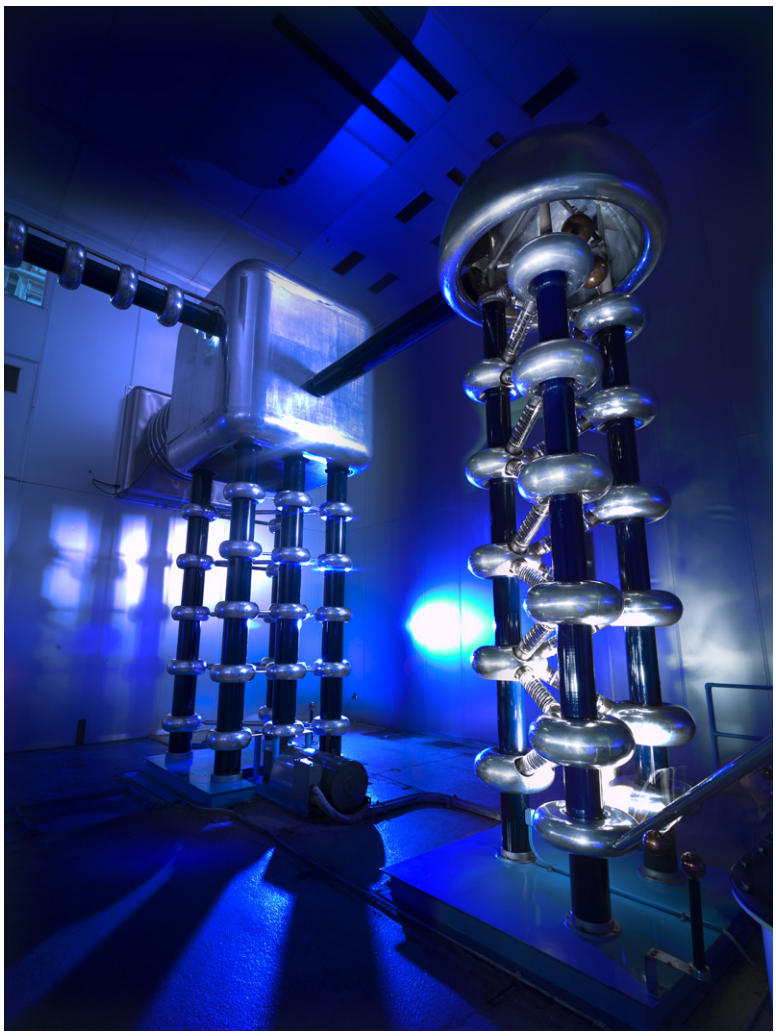


Figure 2.2 : The Fermilab Cockcroft-Walton is the first step in a series of accelerators. The Cockcroft Walton is one of the more interesting looking pieces in the accelerator series used during Tevatron operations. Image courtesy of Fermilab Today [21].

time to move the bunch, but can be off after the bunch passes.

After the Booster, the proton bunches move into the Main Injector. The Main Injector is both a storage area for the protons before going into the Tevatron and the start of anti-proton production. In order to make anti-protons, protons in the Main Injector are accelerated to an energy of 120 GeV and then shot at a nickel tar-

get. The protons interact with the nickel atoms creating anti-protons as a secondary by-product. The anti-protons are separated from the other particles created using magnetic fields and focused to the Debuncher if they have an energy around 8 GeV. Anti-protons are sent from the Debuncher to the Accumulator, both of which are triangular rings. The Accumulator uses stochastic cooling to increase the density of the anti-proton bunches. Groups of dense bunches are then moved to the Recycler where the bunches are electron cooled and stored until an appropriate number of anti-protons are available. With enough anti-protons accumulated, the bunches are transfered into the Main Injector and accelerated from 8 GeV to 150 GeV. The Main Injector is used to put both proton and anti-proton bunches into the Tevatron at an energy of 150 GeV [20].

2.1.2 Tevatron

During Run II operations, 36 bunches of both protons and anti-protons were put into the Tevatron to be collided [22]. The groups of bunches are referred to as a beam. The Tevatron would accelerate these beams from 150 GeV to 980 GeV each. The Tevatron is so named because the energy of each beam is nearly 1 Tera-electron-Volt (TeV). Thus the total final center of mass energy of the Tevatron was 1.96 TeV. This number is determined by adding the energy of the proton beam and the energy of the anti-proton beam. The point of all the work of the Fermilab accelerators is to be able to collide high energy beams of protons and anti-protons in the Tevatron. Collisions are only useful if the resulting wreckage can be studied; therefore, the beams are only allowed to collide at places in the Tevatron where detectors are present. The two experimental groups with detectors in the Tevatron are CDF, the Collider Detector at Fermilab experiment, and D0, the D-Zero experiment.

2.2 DØ

The D0 detector, featured in Fig. 2.3, is a multi-layered machine that combines information from its subdetectors to look at the particles created from the collision of the proton and anti-proton beams. As with a camera, the more sensors that are used to take the picture, the clearer the image. The same holds true for building detectors. The picture is "taken" by the particles that come out of the collision interacting with sensor material in the detector, but unlike a camera, a particle physics detector has to be able to see much more than photons. A detailed description of the D0 detector can be found here [23].

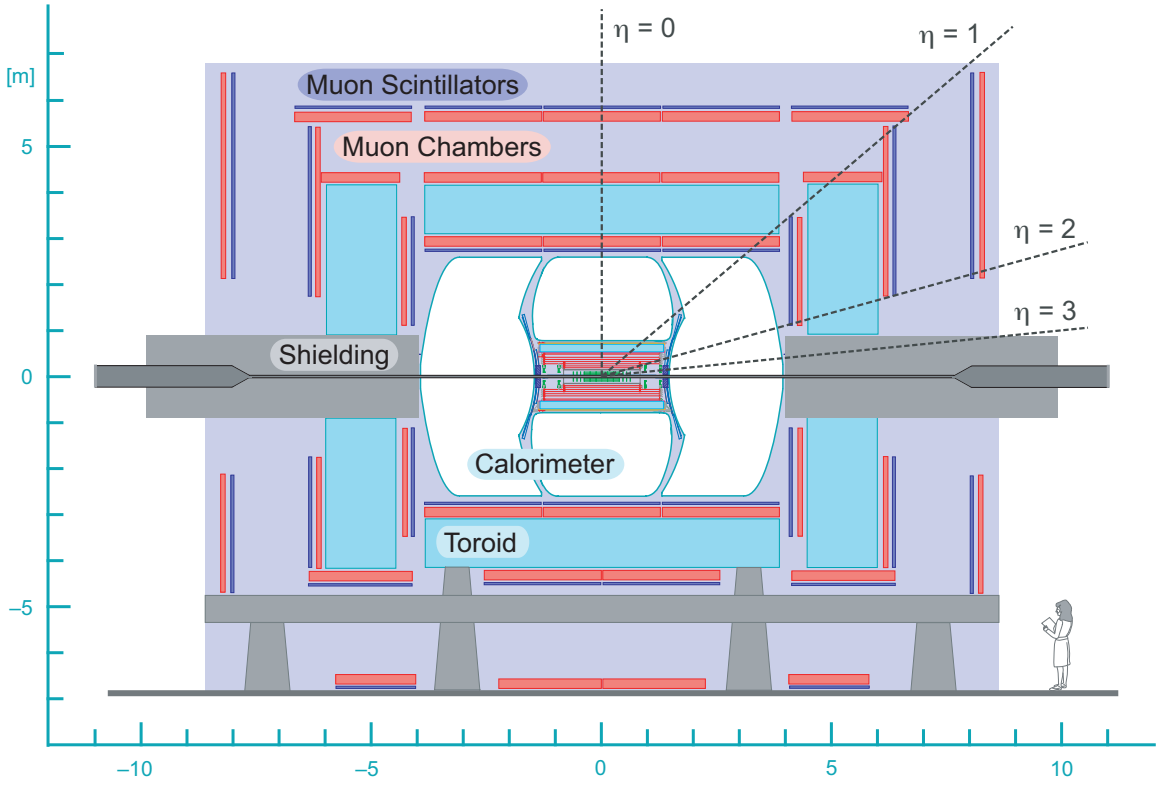


Figure 2.3 : Schematic of the D0 Detector

The subdetectors surround the collision region in concentric cylinders. The main subsystems are the tracking system, the calorimeter, and the muon system. The tracking system is made up of the Silicon Microstrip Tracker and the Central Fiber Tracker, both of which sit inside a solenoid magnet. The solenoid creates a 2 Tesla magnetic field at the center. The subdetectors are discussed in more detail below.

In particle physics, detector coverage is often expressed by a variable known as pseudorapidity. Pseudorapidity is defined as

$$\eta = -ln[tan(\frac{\theta}{2})], \tag{2.1}$$

where η is the pseudorapidity and θ is the angle measured with respect to the beam direction. Pseudorapidity is a useful variable because the number of particles created in beam interactions is roughly constant as a function of pseudorapidity. Cylindrical coordinates (r, ϕ, z) are also useful when describing the D0 detector. The $z = 0$ and $r = 0$ coordinate is centered in the detector in the middle of the beam line where particle interactions are designed to occur.

2.2.1 Silicon Microstrip Tracker

The Silicon Microstrip Tracker (SMT) is the first subdetector a particle will hit after being created in the beam collision. The detector is made up of approximately 793,000 strips of silicon sensors. The SMT has sensors both parallel and perpendicular to the beamline. The parallel sensors are arranged in barrels. Originally in Run II, the SMT barrels had 8 sensor layers, but an additional layer of sensors, called Layer 0, was added in 2006. Layer 0 improved the vertex resolution of the tracking system and compensated for aging and radiation damage in the rest of the barrels. Including

Layer 0, the SMT starts at a radius of 1.6 cm from the center of the beamline and extends to a radial distance of 10.5 cm [24]. The silicon sensors that are perpendicular to the beamline are arranged in disks. Twelve of these disks intersect the barrels in regular spacings, while 2 larger disks sit in the forward region giving the SMT coverage up to $|\eta| < 3$. The arrangement of the disks and barrels after the addition of Layer 0 can be seen in Fig. 2.4. For scaling, the SMT barrel and F-disk region shown covers a 1.2 m distance.

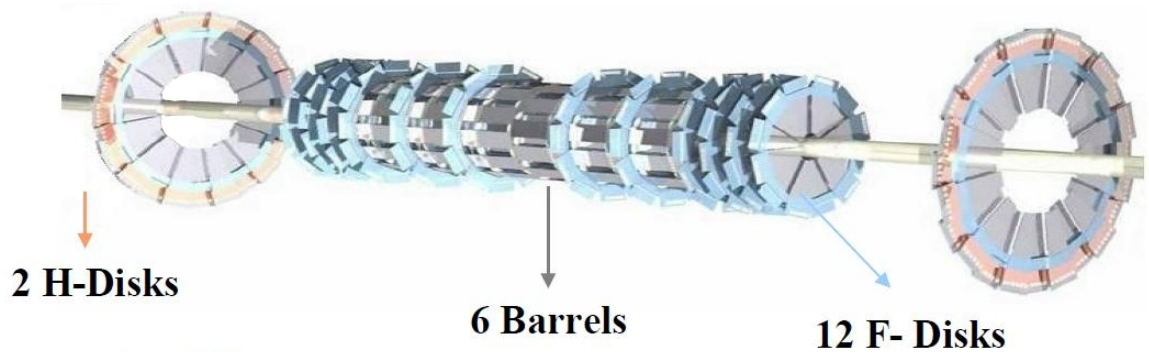


Figure 2.4 : The SMT is the detector closest to the beamline. It determines the track of a particle using strips of silicon sensors.

The SMT is used to track particles just after they leave the beampipe. The fine segmentation in both the perpendicular and parallel direction allows the SMT to precisely determine track vertices. The SMT has a 35 μm spatial resolution.

2.2.2 Central Fiber Tracker

The Central Fiber Tracker (CFT) is located beyond the SMT. The CFT is made of fibers of scintillating plastic arranged on 8 barrels. Each barrel has two layers of fibers, one that is parallel to the beamline and one that is offset by $\pm 3°$. The

offset orientation is alternated between positive and negative from barrel to barrel. The fibers have a diameter of 835 μm and are placed so that each layer is actually a doublet, as shown in Fig. 2.5. One end of each fiber is painted with a reflective coating such that the signal is only collected at the other end. The signal is transported from the uncoated end of the fiber by waveguides to hardware with processing software. Readout systems will be discussed in Sec. 2.2.6.

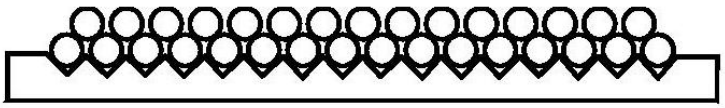


Figure 2.5 : A schematic of the orientation of the fibers in one layer of the CFT detector. Each layer is actually a doublet so that the detector does not have gaps in coverage.

The tracking system, shown in Fig. 2.6, includes the SMT and CFT, which both sit inside a 2 T solenoid magnet. The CFT, which covers a radial range of 20 to 52 cm, is used for determining the momentum of longer lived particles. Charged particles bend in the magnetic field of the solenoid. The CFT tracks the particles and can trace the arc. The radius of curvature is proportional to the momentum. The CFT covers a pseudorapidity range of $|\eta| < 2.5$ with the first layers of the CFT being 1.66 m long, while the last 6 layers are 2.52 m in length.

The CFT also includes preshower detectors in both the central (CPS) and forward (FPS) regions. The preshowers are comprised of triangular prisms of scintillating plastic with a wavelength shifting fiber in the center to transmit signal. The CPS is located after the solenoid. The FPS is in front of the End Cap Calorimeter, discussed

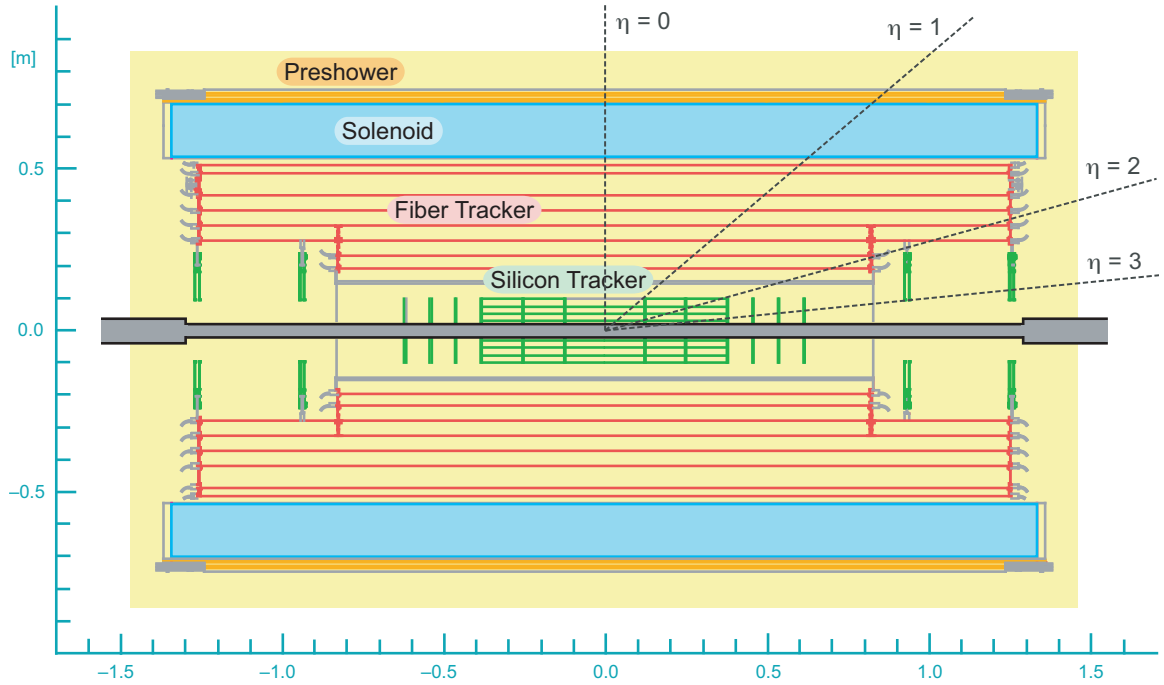


Figure 2.6 : The SMT and CFT both sit inside a 2 T solenoid magnet. This set of subdetectors make up the D0 tracking system. They accurately measure particle vertices and momentum.

in Section 2.2.3.

2.2.3 Calorimeter

Outside of the tracking system sits the calorimeter. The calorimeter is a finely segmented liquid argon and uranium compensating sampling calorimeter. The calorimeter measures the energy of particles by causing them to shower and then measuring the resulting energy deposit. Figure 2.7 shows the calorimeter is separated into three sections known as cryostats. The central cryostat (CC) and two end cryostats (EC) each contain an electromagnetic, fine hadronic, and coarse hadronic region. The regions have slightly different uranium plate composite and thickness to optimize the

region to different particle energy measurements. The electromagnetic region stops electrons, positrons, and photons. Hadrons interact in the two hadronic regions. Muons, which are minimally ionizing, and neutrinos, which are weakly interacting, pass through this detector. For this analysis, information from the calorimeter is not used, since the final products being looked for are muons.

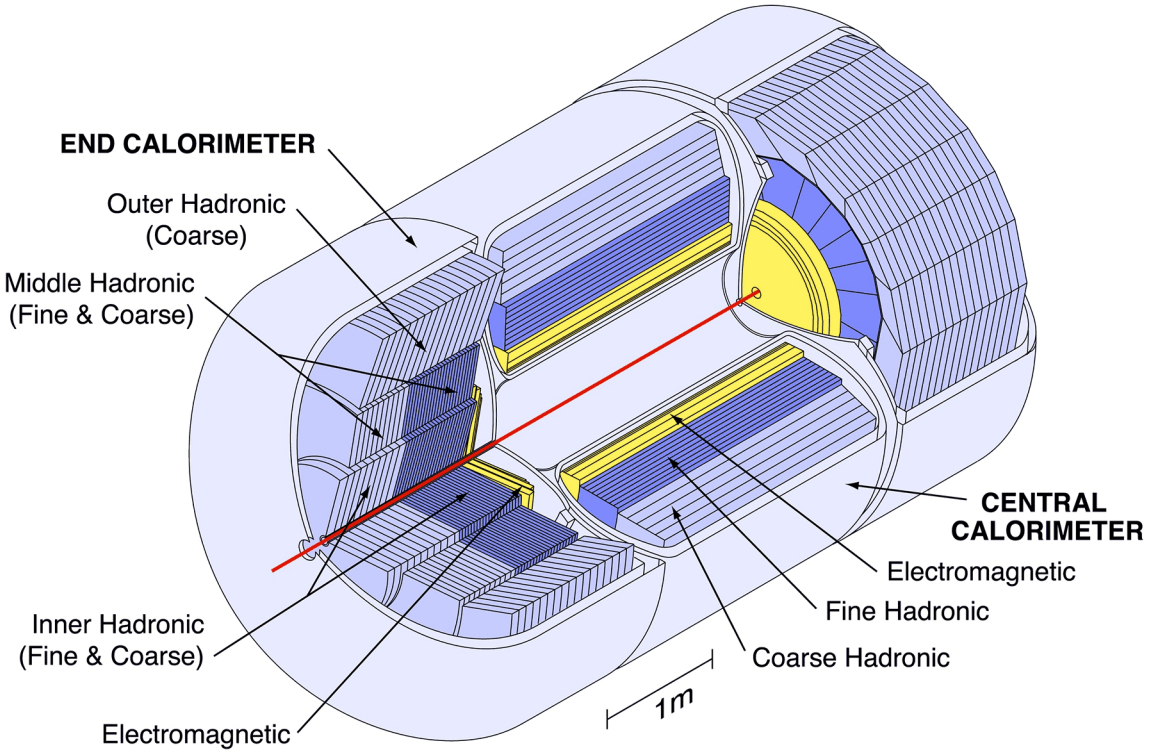


Figure 2.7 : The calorimeter determines the energy of a particle by stopping the particle and measuring the energy deposited. The calorimeter has three sections: electronic, fine hadronic, and coarse hadronic.

2.2.4 Muon System

The muon system is used to identify muons, which are the only interacting particle which will still be traversing the detector. It consists of three layers, A, B, and C, each

with trigger scintillators and tracking detectors. The central region has proportional drift tubes (PDTs) to reconstruct tracks. In the forward region, mini drift tube (MDT) chambers are used for track detection. The arrangement of the drift tubes is shown in Fig. 2.8. Between Layer A and B sits a 1.8 T toroidal magnet. The muon system provided particle identification out to $|\eta| < 2$.

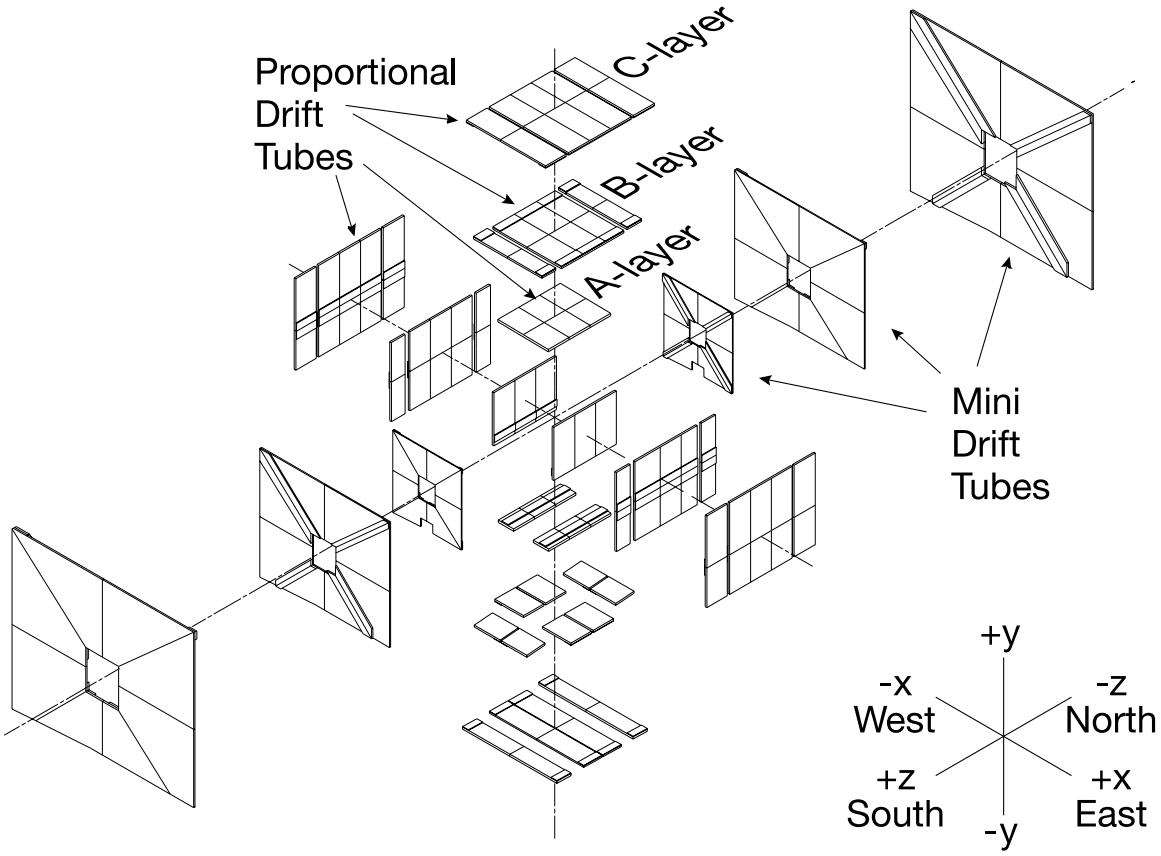
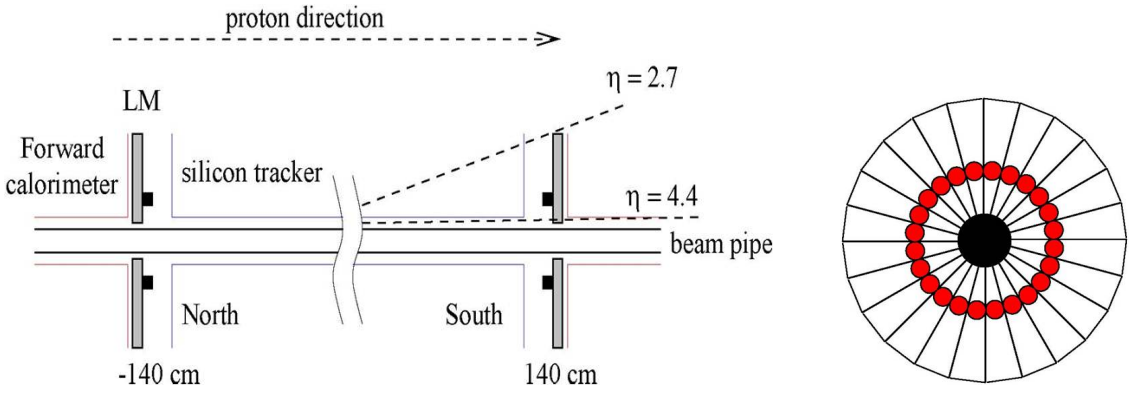


Figure 2.8 : The muon system has many sections making up the three layers in the forward and central region.

2.2.5 Luminosity Monitor

The last major subdetector at D0 is the Luminosity Monitor (LM). The LM sits in the forward region of the detector, along the beam direction. The luminosity monitor has a north and a south section, shown in Fig. 2.9, that sit at $z = \pm 1.4$ m surrounding the beampipe. The detectors cover a pseudorapidity range of $2.7 < |\eta| < 4.4$. The LM uses scintillating wedges with fine-mesh photo-multiplier tube (PMT) readout to detect particles from inelastic beam collisions. Each section has 24 readout channels, for a total of 48, which is quite a small number compared to the other subsystems at D0.



(a) r - z view (b) r - ϕ view

Figure 2.9 : The luminosity monitor consists of two arrays, each with 24 scintillator wedges read out by PMTs, sitting at ± 1.4 m surrounding the beampipe. The placement of the PMTs along the scintillating wedges is shown in red.

The luminosity monitor has a unique role compared to the other subdetectors. The SMT, CFT, calorimeter, and muon system all determine information about the particles created in beam collisions including production vertex, momentum, energy, and type. The luminosity system is in place to determine the luminosity of the

collisions. Luminosity is related to the number of interactions in the beam crossings as defined in Equation 2.2:

$$N = \sigma L, \tag{2.2}$$

where N is the number of interactions, σ is the cross section for the interaction, and L is the luminosity. For measuring the luminosity, inelastic collisions of a known cross section are used. For other processes, once the luminosity is determined, the cross sections of different types of interactions can be determined. Cross sections are one of the ways that theory is tested.

The luminosity is measured using a technique called "counting empties". The probability that no particles hit the LM is related through Poisson statistics to the number of inelastic interactions in the beam and can be used to determine the luminosity, as is shown in Equation 2.3. A "hit" in the luminosity system is required to fire both sides of the LM in coincidence. In the majority of cases, a beam interaction will produce particles that interact in both sides of the LM. An event that only sends particles toward one side of the luminosity monitor is called a single-sided interaction.

$$P(0) = e^{-\sigma_{LM}L/\nu} \times (2e^{(-\sigma_{SS}/2\nu)L} - e^{-\sigma_{SS}L/\nu}) \tag{2.3}$$

Using the counting empties method reduces the uncertainty in the measurement from multiple interactions happening in one beam crossing. The factor in parenthesis is a correction for two single-sided interactions, where σ_{SS} is the cross section of a single-sided interaction, occurring in the same beam crossing and mimicking an event

that would hit both sides of the luminosity monitor. The effective cross section for an interaction producing particles that hit both sides of the luminosity monitor is given by σ_{LM}, which is called the luminosity constant. The luminosity constant was recently determined to be $\sigma_{LM} = 48.3 \pm 2.0$ mb for the RunIIb data period [25].

2.2.6 Triggering

Beam interactions occur at a rate of 17 MHz at the Tevatron. With readout from around a million channels at this rate, it is impossible and unnecessary to save all the information. Triggering is the method by which interesting interactions are selected to be saved for future processing and physics analysis. Triggering is performed in three successive steps at D0 and organized through the Trigger Framework, shown in Fig 2.10. The first level of triggering, Level 1 or L1, happens in each of the subdetectors. Hardware for the subdetector looks at the readout for the channels in the system and attempts to make patterns. For example, in the CFT the hardware that determines if an event is interesting is the Level 1 Central Track Trigger (L1CTT). The L1CTT looks for clusters of charge in the CFT. If tracks can be made to connect clusters in the different layers of the CFT, the event will pass the trigger. If the clusters are randomly dispersed or do not exist, the event would fail the trigger. The Level 1 trigger makes decisions after every beam crossing and reduces the data rate to 2 kHz.

Events have to pass at least one L1 trigger to be considered at Level 2. Level 2 is a combination of hardware and software triggers which creates objects by combining information from related subdetectors. About 1 kHz of data passes Level 2. After passing L1 and L2, events are sent to the Level 3 farm where information from all the subdetectors is combined and a simplified event reconstruction is performed. At L3,

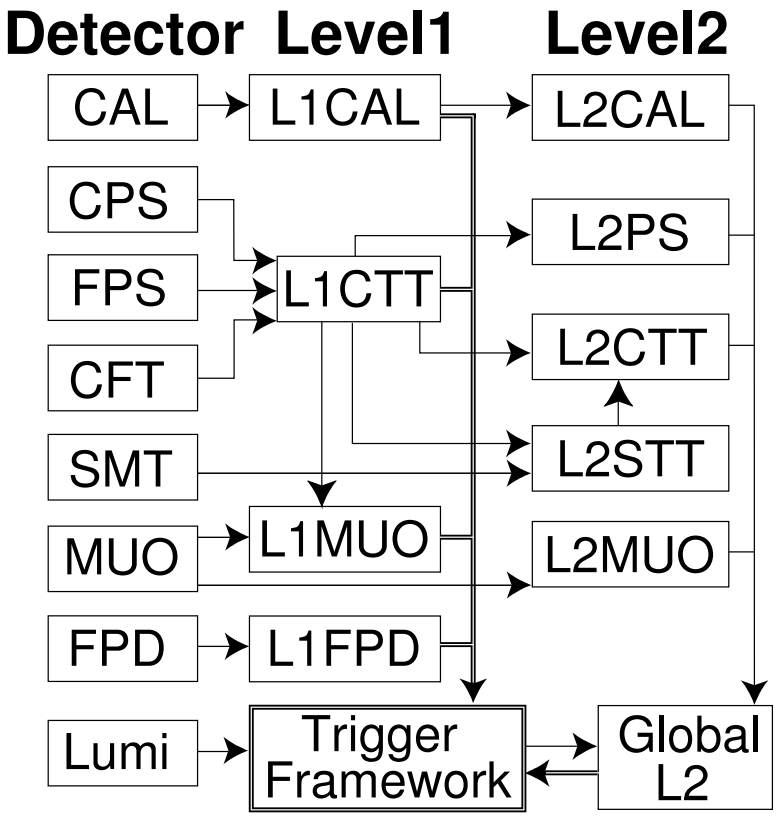


Figure 2.10 : The trigger system at D0 is used to determine which events are interesting and select them to be saved for physics analysis. After passing Level1 and Level 2 triggering an event is sent to be fully reconstructed.

events must pass a trigger requirement based on the physics objects in the event. For example, a trigger requirement could be that an event has two high quality muons. If the event does, then it would be written out to tape. Some triggers are prescaled, meaning only once in every x times, where x is the prescale, will an event that satisfies that trigger requirement be written to tape. The trigger list includes a few special triggers, such as zero-bias and minimum-bias, both of which are highly prescaled. In zero-bias an event needs only to have occurred at a time when the beams were crossing. The requirement for minimum-bias is that a beam interaction occurred.

The L3 trigger brings the rate of data saved down to 50 – 100 Hz.

2.2.7 DØ Data

The data in this analysis is from a data-taking period from 2002 to 2011 known as Run II at D0. The Run II data set is further subdivided into RunIIa and RunIIb. The division between RunIIa and RunIIb is a shutdown in 2006. During this shutdown, changes were made to the detector geometry, including the addition of Layer 0 to the SMT. Additional shorter shutdowns occurred during RunIIb with smaller changes to the D0 detector, such as recovery of dead channels and changes in the trigger requirements. These shutdowns break up the data further. In total, the RunII data-taking period is broken up into five epochs: IIa, IIb1, IIb2, IIb3, and IIb4. The division between the different data epochs are distinguished in Fig. 2.11 by vertical red lines.

The luminosity (L) determined by the LM is also called the instantaneous luminosity. Integrated luminosity ($\mathcal{L}$) is a measure of the total luminosity over a period of time. At D0, almost 12 fb^{-1} was delivered by the Tevatron. Of that, the detector was recording data for 10.7 fb^{-1}. Not all of this data passes quality cuts. Thus for this analysis, the full data set is considered to be 10.4 fb^{-1}.

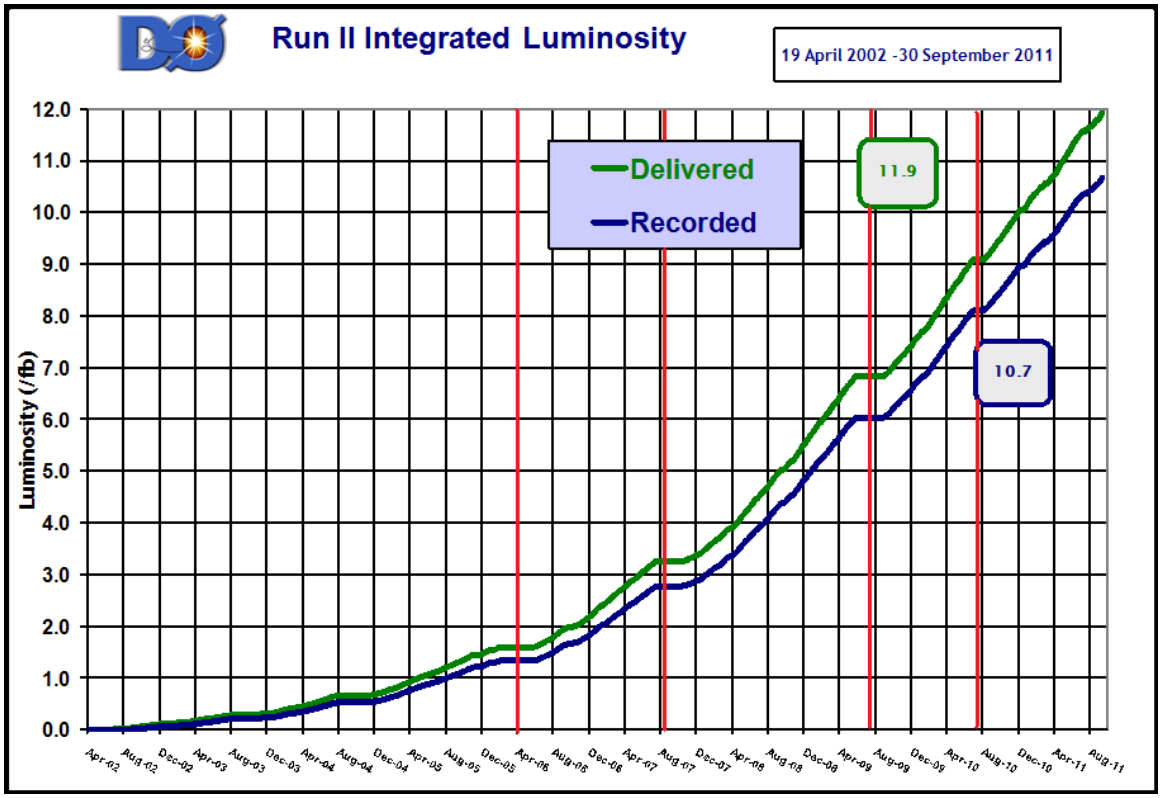


Figure 2.11 : A graph of the integrated luminosity for D0 over the Run II data taking period. The vertical red lines denote the different data taking epochs.

Chapter 3

The Ground Work

3.1 Analysis Overview

The search for $B_s^0 \rightarrow \mu^+\mu^-$ is preformed as a blinded analysis. This means that the dimuon mass range near the B_s^0 mass of 5.35 GeV is not examined until all the procedures for the analysis are finalized. Blinding the signal region helps to prevent bias in the analysis technique. A wide window in dimuon mass of 4.0 – 7.0 GeV is looked at to study the backgrounds. The region from 4.9 – 5.8 GeV is blinded since it includes the B_s^0 mass. As shown in Fig. 3.1, using B_s^0 Monte Carlo simulation, the blinded range covers the $\pm 3\sigma$ mass width for the decay, where the width of the mass peak reflects the resolution of the D0 detector.

A normalization mode is needed for this analysis to determine the number of B_s^0 mesons. The decay $B^\pm \rightarrow J/\psi K^\pm$ with $J/\psi \rightarrow \mu^+\mu^-$ is the normalization channel for this analysis. This channel is chosen because the kinematics are similar to the signal mode and it is an abundant decay. Having similar kinematics between the normalization and signal modes, such as both channels having two muons in the final states, allows for cancellation of most systematic uncertainties. Having a large sample of $B^\pm$ events further reduces the statistical uncertainties.

The normalization mode is used to determine the number of B_s^0 mesons in the data sample, and using the SM branching fraction, the expected signal events, which will be described in Section 3.4. It is also used to validate the Monte Carlo (MC)

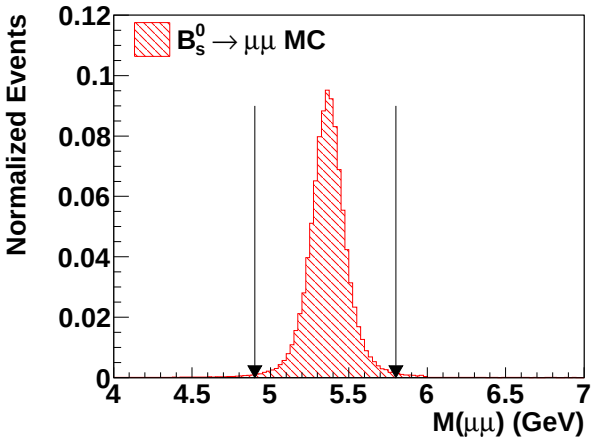


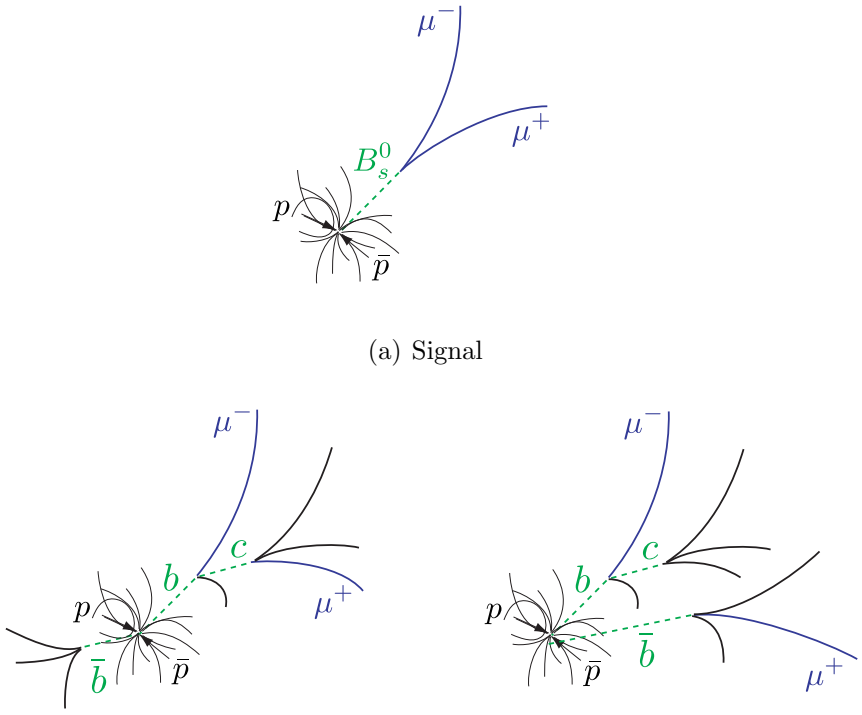
Figure 3.1 : Dimuon mass window with B_s^0 Monte Carlo simulation to show the region that was blinded for the analysis.

simulation. Details of the MC simulation and validation are discussed in Section 3.3. Similar kinematics in the signal and normalization channels allow corrections to MC simulation derived in the normalization mode to be easily applicable to the signal channel.

The heart of the analysis is improving the background rejection so that $B_s^0 \to \mu^+\mu^-$ events might be seen. The strategy is to create new variables to differentiate signal and background. A multivariate technique, discussed in Section 4.1, is also used to distinguish signal from background. For these steps, MC simulation is used to model the signal, and data from the mass ranges of 4.0 – 4.9 and 5.8 – 7.0 GeV, the mass sidebands, are used to model the background. After optimization of the multivariate technique, the expected background and resulting expected limits are determined. Only then is the box opened and the observed limit set.

3.1.1 Signal versus Background

The $B_s^0 \to \mu^+\mu^-$ decay leaves a clean signal, namely two oppositely charged muons that form a good vertex away approximately 1 mm from the beam interaction vertex. Differentiating signal from types of background which produce a dimuon vertex is vital to the analysis. In the dimuon mass region of interest, 4.0 – 7.0 GeV, the background comes from two main decay modes: sequential decay and double semileptonic b decay. Figure 3.2 depicts the signal and the two main backgrounds.



(a) Signal

(b) Background: Sequential Decay

(c) Background: Double Semileptonic Decay

Figure 3.2 : Schematic diagrams showing (a) the signal decay, $B_s^0 \to \mu^+\mu^-$, and the main backgrounds: (b) sequential decay, $b \to c\mu^-\bar{\nu}$ followed by $c \to \mu^+\nu X$ or $\bar{b} \to \bar{c}\mu^+\nu$ with $\bar{c} \to \mu^-\bar{\nu}X$, and (c) double semileptonic decay, $b \to \mu^-\bar{\nu}X$ and $\bar{b} \to \mu^+\nu X$. Images courtesy of Brendan Casey.

The sequential decay background comes from $b \to \mu^- \bar{\nu} c$ with $c \to \mu^+ \nu X$ or $\bar{b} \to \bar{c} \mu^+ \nu$ with $\bar{c} \to \mu^- \bar{\nu} X$. This background dominates the mass sideband from 4.0 – 4.9 GeV. Double semileptonic decay, which dominates the mass sideband from 5.8 – 7.0 GeV, occurs when $b \to \mu^- \bar{\nu} X$ and $\bar{b} \to \mu^+ \nu X$. In both types of background the muons are produced close enough together to fake a good dimuon vertex, though they do not form a real vertex.

In the beam interactions the $b\bar{b}$ pair can be produced in two ways: $q\bar{q}$ annihilation or gluon splitting ($g \to b\bar{b}$). For double semileptonic decay, it is only the $b\bar{b}$ pair produced by gluon splitting that creates a difficult background. This is because particles created in gluon splitting move in roughly the same direction where particles created by annihilation are produced back-to-back.

New variables are introduced in this search to distinguish signal events from background events. In Fig. 3.2(a), the signal is very clean, whereas in Fig. 3.2(b) and Fig. 3.2(c), the background has other particles that will make tracks. To that end, additional vertices are searched for in the events in two ways: tertiary vertices and vertices which include one of the muons plus additional tracks. The beam interaction vertex is referred to as the primary vertex. The dimuon or B vertex is the secondary vertex. An additional vertex found near the dimuon vertex would be a tertiary vertex, for example the $c \to X$ decay. Tracks associated with the same primary vertex that have at least 2 hits in the SMT and 2 hits in the CFT are selected to attempt to find additional vertices. For the tertiary vertex checks, tracks that are not part of the B vertex are combined to see if a good vertex can be made. A search is also done for tertiary vertices including one of the muons. Three tertiary vertex checks are done in total. The check for any vertex with the muon does not require the track to be near the dimuon vertex. The tracks are scanned to see if any make a good vertex with

either of the muons, so there are two vertices searched for in this manner.

The additional vertex checks include five vertex possibilities. For each of the vertices, a number of variables are calculated, including mass, transverse momentum (p_T), and a vertex quality variable called vertex χ^2. Because sometimes additional vertices are not found, these variables are initialized to unphysical values.

In addition to the search for additional vertices, new isolation variables are used in this analysis. Isolation is defined with respect to a momentum vector $\vec{p}$ in a cone of ϕ and η. Isolation is defined as

$$\mathcal{I} = p_T/[p_T + p_T(\text{cone})], \tag{3.1}$$

where p_T(cone) is the scalar sum of all the tracks in the cone of radius $\mathcal{R}$ excluding the track of interest. The radius of the cone is defined by

$$\mathcal{R} = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{3.2}$$

Here η is the pseudorapidity defined in Equation 2.1 and ϕ is the azimuthal angle discussed in Section 2.2. For this analysis the radius was chosen to be $\mathcal{R} = 1$. In addition to the isolation of the B meson, which was not used in the last version of this analysis, isolation is also calculated around each of the muons individually. Adding the individual muon isolation helps identify backgrounds when the muons are produced back-to-back so that the fake dimuon vertex is isolated, but the individual muons are not isolated.

3.2 Event Selection

An inclusive trigger strategy is used to choose $B_s^0 \to \mu^+\mu^-$ candidates. An "OR" of muon triggers are considered such that any event that has a muon in the final state is considered in this search. The code reviews the data searching for any event that has two oppositely charged muons that make a good vertex in the mass range of 4.0 – 7.0 GeV. In the normalization channel, the $J/\psi \to \mu^+\mu^-$ is first reconstructed and a third track, assumed to be the charged kaon (K), is added to the vertex. The dimuon mass range for the J/ψ is 2.7 – 3.45 GeV.

A set of requirements, known as preselection cuts, are imposed on the events to insure good data quality. Preselection cuts are applied to both the normalization and signal modes and to data and MC simulation. The preselection cuts for the signal are listed below, with further explanation following.

- Each muon must have at least 2 hits in the SMT, 2 hits in the CFT, and a match in the muon system.
- Each muon must have $p_T > 1.5$ GeV and $|\eta| < 2$.
- At least one muon must have a hit in the muon system after the toroid.
- The B_s^0 p_T must be less that 20 GeV.
- The χ^2/dof of the B_s^0 vertex must be less than 14.
- The B_s^0 transverse decay length significance (L_T/σ_{L_T}) must be greater than 3.
- The uncertainty on the transverse decay length (σ_{L_T}) of the B system must be less than 1 mm.

The normalization mode is required to pass all the preselection cuts for the signal and has the following additional requirements on the K.

- The kaon must have at least two SMT hits and at least two CFT hits.
- The kaon must have $p_T > 1.0$ GeV and $|\eta| < 2$.
- The χ^2/dof of the B vertex must be less than 6.7.

Tracks are required to have 2 hits in the SMT to ensure good resolution for the secondary vertex. With less than 2 CFT hits, the p_T of the particle can not be accurately measured. Requiring a track match between the tracking system and the muon system ensures the muon was accurately measured through the detector and the particle is a real muon. The η cuts on the tracks ensure that they stay inside the muon system coverage. The p_T cuts are set so that the particles will pass the lower thresholds for the tracking and muon systems. The requirement that one muon has a hit after the toroid reduces fake muon signals.

Requiring the $p_T(B)$ to be less than 20 GeV reduces backgrounds from muon production in $W \to \mu\nu$ decays. The χ^2 variable is a measure of the quality of the vertex. A cut at 14 per degree of freedom (dof) ensures a good dimuon vertex. Transverse decay length (L_T), defined in Equation 3.3, is a measure of the distance from the primary vertex to the B vertex in the transverse plane:

$$L_T = \vec{l}_T \cdot \vec{p}_T / |\vec{p}_T|. \tag{3.3}$$

Here the vector $\vec{l}_T$ is from the primary vertex to the B vertex and $\vec{p}_T$ is the transverse momentum vector of the B system. The uncertainty on the transverse decay

length (σ_{L_T}) is a determined from the uncertainty in the primary vertex position and track parameter uncertainties. Requiring L_T/σ_{L_T} to be greater than 3 with a small uncertainty on σ_{L_T} reduces the background from the primary vertex.

The preselection cuts are quite loose in order to maintain high signal efficiency. Further cuts on the system are made as part of the optimization discussed in Chapter 4.

3.3 Monte Carlo Simulation

Monte Carlo (MC) simulation is generated for both the $B_s^0 \to \mu^+\mu^-$ and the $B^\pm \to J/\psi K^\pm$ channels. Events are generated using the PYTHIA [26] event generator integrated with the EVTGEN [27] decay package. A GEANT [28] package is used to simulate detector response. MC simulation events are overlaid with data from the zero-bias and minimum-bias triggers that were discussed in Section 2.2.6 to model multiple interactions and detector noise. The MC simulation includes $b\bar{b}$ pair production from both annihilation and gluon splitting since both production modes contribute to the main backgrounds. MC simulation is also obtained to study the peaking background $B_s^0 \to K^+K^-$, which will be discussed in Section 4.4, and for $B^0 \to J/\psi K^{0*}$ with $K^{0*} \to K\pi$ to study the track reconstruction efficiency, which will be discussed in Section 3.4.

The $B^\pm \to J/\psi K^\pm$ normalization mode is used to validate the MC simulation by comparing the simulation to data. A technique called sideband subtraction is used to produce the distributions for the $B^\pm$ candidates from the data. Events from sidebands on the low mass and high mass side of the $B^\pm$ mass peak are added together and subtracted from the peak. Additional requirements are made before the subtraction to reduce the background. The J/ψ mass range is restricted to 2.9 – 3.2 GeV. A cut

requiring the pointing angle, defined in Section 4.2.2, to be greater than 0.95 is also made. With these cuts the subtracted amount is smaller, since the backgrounds are reduced, which reduces the uncertainty. Figure 3.3 shows the $B^{\pm} \to J/\psi K^{\pm}$ distribution, with the peak and sideband subtraction regions noted, after the J/ψ mass and pointing angle cuts. For sideband subtraction the mass range for the $B^{\pm}$ peak is 5.2 – 5.35 GeV. The lower mass subtraction sideband range is 5.1 – 5.163 GeV, and the higher mass subtraction sideband covers a range of 5.4 – 5.475 GeV. The spacing around the peak avoids subtracting out true $B^{\pm}$ events. The slightly uneven widths of the lower and higher mass subtraction sidebands compensates for nonlinearity of the backgrounds.

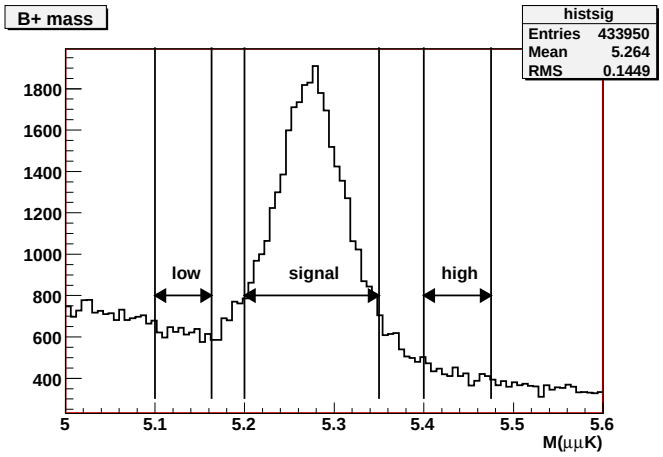


Figure 3.3 : $B^{\pm}$ mass distribution, in GeV, for the normalization channel after a pointing angle and J/ψ mass cuts. The "low" and "high" region events are added together and subtracted from the "signal" for the $B^{\pm} \to J/\psi K^{\pm}$ normalization mode. The sideband subtracted distributions can then be compared to the MC simulation distributions to validate the MC simulation for all variables.

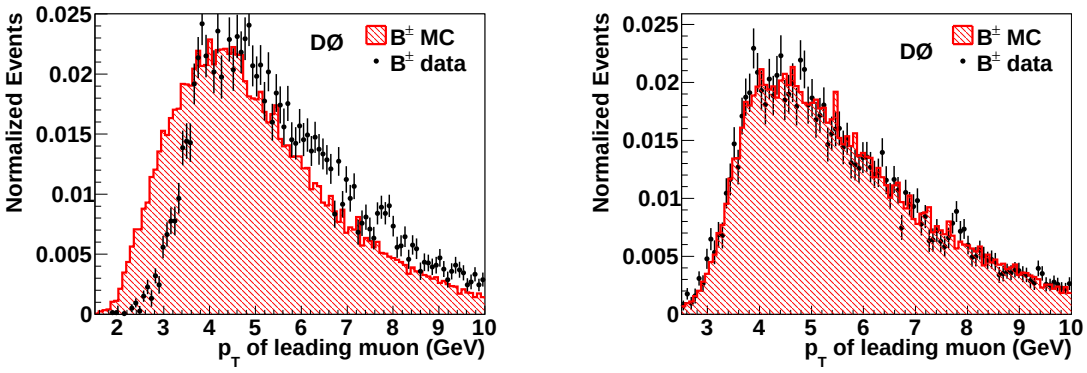
Upon comparing the sideband subtracted data distributions to the MC distributions for $B^{\pm} \to J/\psi K^{\pm}$, it was determined that reweighting would be necessary in some variables to get good agreement between data and MC simulation. The

reweights are determined and validated in the normalization channel and then applied to the $B_s^0 \to \mu^+\mu^-$ signal MC. The detector geometry and trigger acceptance changes between data epochs as discussed in Section 2.2.7, so the MC simulation is reweighted separately for each data epoch.

3.3.1 p_T Reweighting

MC simulation does not have trigger efficiency effects. By reweighting the MC simulation as a function of muon p_T, the trigger efficiencies are taken into account. The corrections are calculated by comparing zero-bias and minimum-bias events with data that passed particular trigger requirements [29]. The result is turn-on curves that are applied to the leading muon, the one with a higher p_T, and the trailing muon, which has the lower of the two muon p_Ts. The trigger efficiency is more noticeable in the leading muon where the efficiency goes from almost zero at $p_T = 2$ GeV, to 50% around $p_T = 3$ GeV, reaching full efficiency around a p_T of 4 or 5 GeV. For the trailing muon, the efficiency is about 50% at $p_T = 1.5$ GeV and reaches full efficiency around $p_T = 3.5$ GeV. Figure 3.4 shows the p_T comparison in the normalization mode for the leading muon before and after the trigger efficiency corrections. This example is for the RunIIb4 data epoch, but all epochs show similar agreement after p_T reweighting. Comparison plots after p_T reweighting for all epochs can be found in Appendix A.

Correcting the leading and trailing muon p_T distributions to account for the trigger efficiencies greatly improves the agreement in the p_T of the other particles. A small additional correction has to be made to the B p_T distribution to correct for the soft p_T spectrum that PYTHIA produces for B mesons. Figure 3.5 shows the agreement in the $B^\pm \to J/\psi K^\pm$ channel for the sideband subtracted data and the MC simulation



(a) (b)

Figure 3.4 : Comparison of $B^{\pm}$ sideband subtracted data and MC simulation for the leading muon (a) before and (b) after applying a trigger efficiency correction.

after the trigger efficiency and B p_T reweighting is applied to the MC simulation for the RunIIb4 data epoch.

3.3.2 Isolation Reweighting

Isolation is an important variable for distinguishing $B_s^0 \to \mu^+\mu^-$ signal from background. It is important to have the isolation variable modeled well in the MC simulation. The MC simulation produces an accurate shape distribution for the isolation, but it overestimates the number of events that are totally isolated. Totally isolated events, those with no other tracks in the isolation cone, have isolation $\mathcal{I} = 1$. To match data, the number of events in the $\mathcal{I} = 1$ bin in the MC distributions is reduced. Figure 3.6 shows the comparison of the B isolation before and after reweighting for the RunIIb4 data epoch.

The fraction of the events removed from the $\mathcal{I} = 1$ bin varies over the data epochs from a low of 4% to a high of 20%. The isolation reweighting was only done for

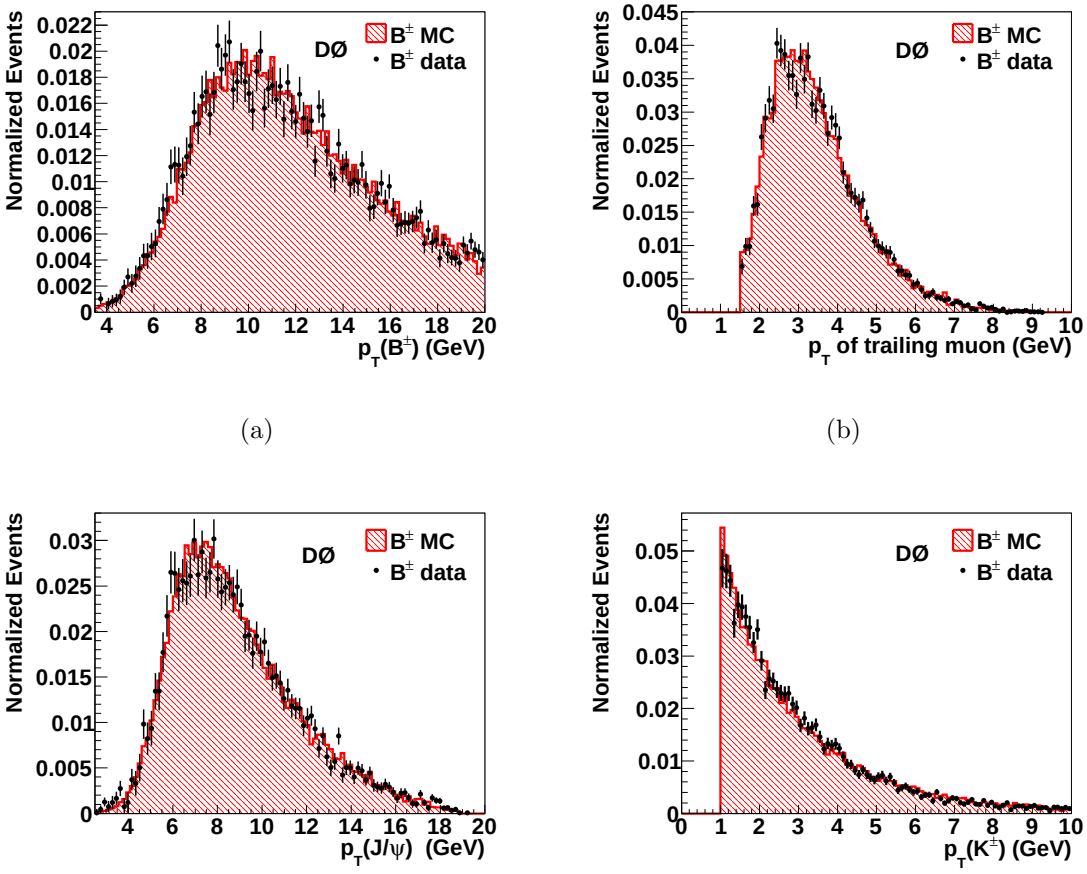
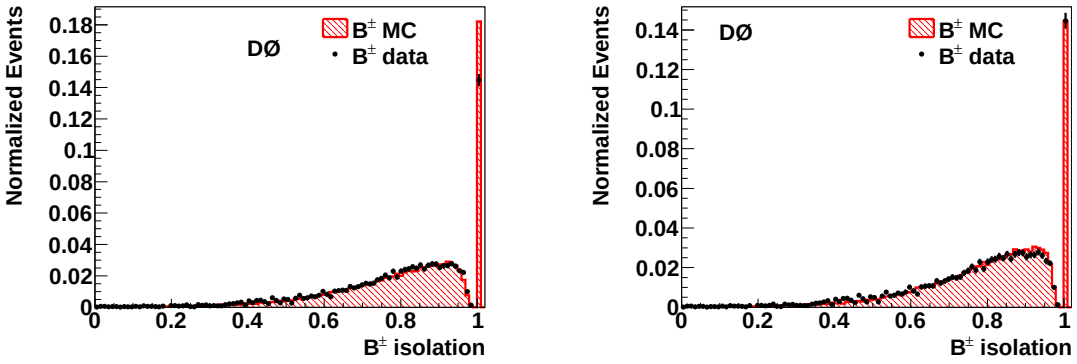


Figure 3.5 : Comparison of $B^{\pm}$ sideband subtracted data and MC simulation after applying all p_T corrections for the (a) $B^{\pm}$, (b) trailing muon, (c) J/ψ, and (d) $K^{\pm}$.

the isolation variable defined with respect to the B meson, but this results in good agreement in the individual muon isolation, which can be seen in Fig. 3.7.

3.3.3 Mass Smearing

The central value and mass distribution of the dimuon system in the MC simulation is also corrected to match the data. The correction is determined using the J/ψ mass



(a) (b)

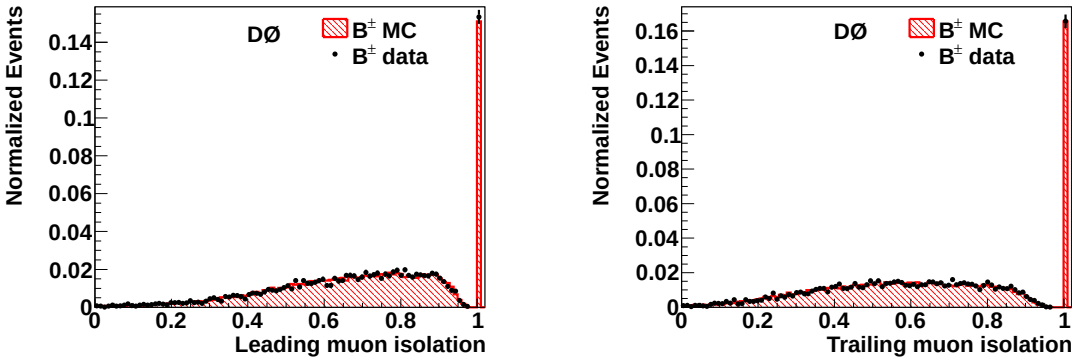
Figure 3.6 : Comparison of $B^{\pm}$ sideband subtracted data and MC simulation for the B meson isolation (a) before and (b) after applying isolation reweighting to the MC simulation.

in the normalization mode. For the J/ψ, the MC simulation overestimates the peak by 10 MeV and underestimates the width by 13%. To correct this, the J/ψ mass peak in the MC simulation is shifted and smeared. The smearing factor is determined by fitting the J/ψ mass peak with a double Gaussian where the width σ_{av} is defined in Equation 3.4:

$$\sigma_{av} = f\sigma_1 + (1-f)\sigma_2, \tag{3.4}$$

where the width of the narrow Gaussian is given by σ_1, which has f fraction of the events. The wider Gaussian has width σ_2.

The resulting shift to the $B_s^0 \to \mu^+\mu^-$ mass peak can be scaled simply, and is determined to be 17 MeV. To appropriately scale the smearing, it is assumed that the MC simulation correctly determines the ratio of the average width to mass. The ratio in Equation 3.5 is determined from MC simulation. The value of the ratio for



(a) (b)

Figure 3.7 : Comparison of $B^{\pm}$ sideband subtracted data and MC simulation after applying isolation reweighting to the B meson system for the individual muon isolation for the (a) leading muon and (b) trailing muon.

the J/ψ can be taken from data. Then the average width of $B_s^0 \to \mu^+\mu^-$ mass peak in data can be determined:

$$\left(\frac{\sigma_{av}}{M}\right)_{J/\psi} / \left(\frac{\sigma_{av}}{M}\right)_{B_s}. \tag{3.5}$$

The corrected σ_{av} for $B_s^0 \to \mu^+\mu^-$ is determined to be 125 ± 3 MeV. The difference between the uncorrected and corrected MC simulation peak can be seen in Fig. 3.8.

The dimuon mass shift and smearing correction is applied to MC simulation for each epoch. After all reweights and corrections were applied to the MC simulation, all the variables used in the analysis were compared in the normalization mode and found to be in good agreement. Figure 3.9 shows a few of the the variables checked after the reweighting.

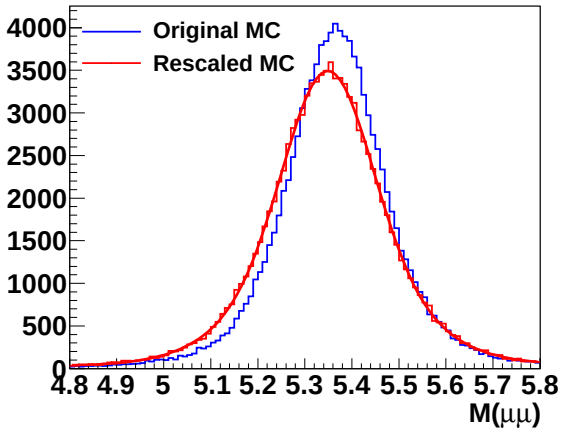


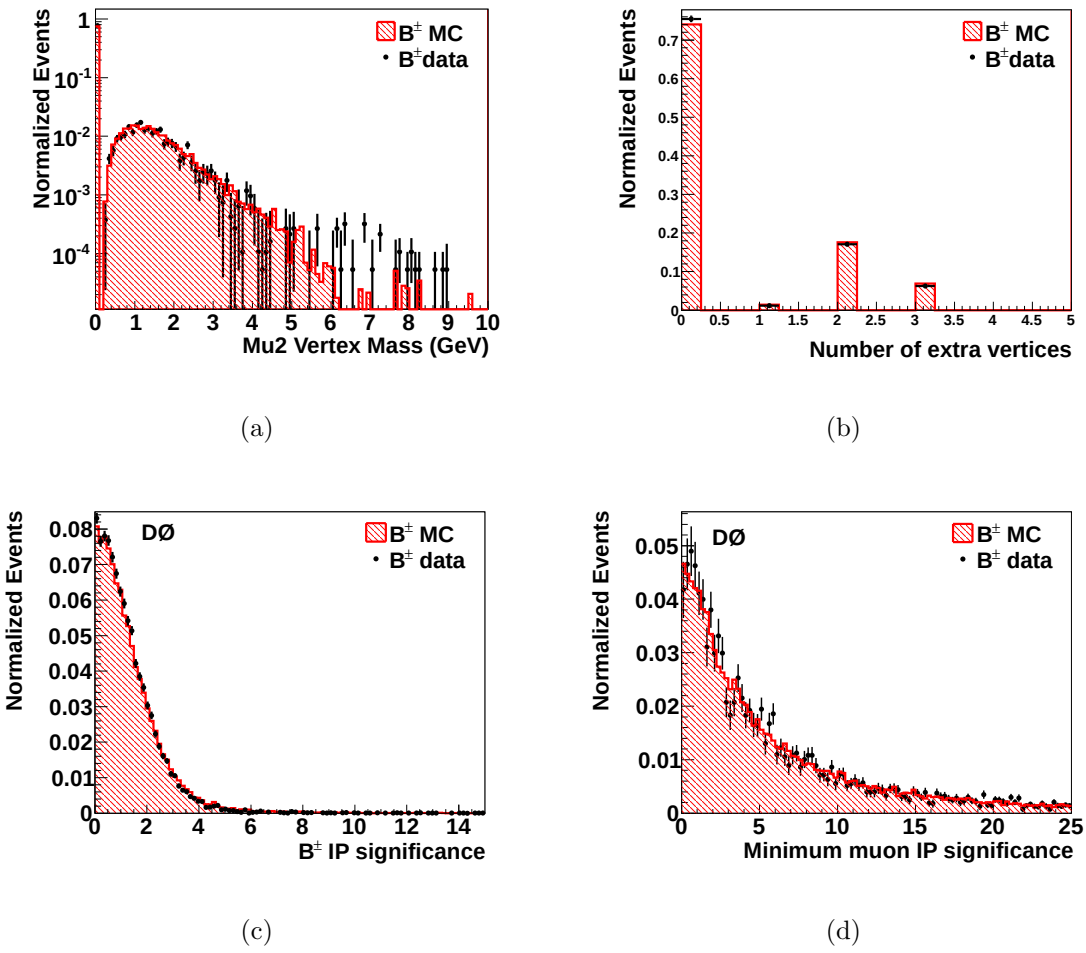
Figure 3.8 : Dimuon mass distribution, in GeV, for $B_s^0 \to \mu^+\mu^-$ MC simulation before and after the dimuon mass smearing correction.

3.4 Single Event Sensitivity

The Single Event Sensitivity (SES) is the branching ratio at which 1 event is expected in the data sample. It can be used to determine the number of B_s^0 decays in the data set. From the SES and the SM branching fraction, the number of expected $B_s^0 \to \mu^+\mu^-$ signal events is determined. The SES is expressed in Equation 3.6.

$$\text{SES} = \frac{1}{N(B^\pm)} \times \frac{\epsilon(B^\pm)}{\epsilon(B_s^0)} \times \frac{f(b \to B^\pm)}{f(b \to B_s^0)} \times \mathcal{B}(B^\pm \to J/\psi K^\pm) \times \mathcal{B}(J/\psi \to \mu^+\mu^-) \quad (3.6)$$

Section 3.4.1 explains how $N(B^\pm)$, the number of $B^\pm$ events in the data, is determined. The ratio of the efficiency of finding the normalization mode events, $\epsilon(B^\pm)$, over the efficiency of finding the signal mode events, $\epsilon(B_s^0)$, is discussed in Section 3.4.2. The ratio of the fragmentation fraction, $\frac{f(b\to B^\pm)}{f(b\to B_s^0)}$, is the relative probability of the b quark fragmenting to a $B^\pm$ compared to a B_s^0. The fragmentation ratio is taken to be the "high-energy" average of $\frac{f(b\to B^\pm)}{f(b\to B_s^0)} = 0.263 \pm 0.017$ from the Heavy



(a) (b)

(c) (d)

Figure 3.9 : Comparison of $B^{\pm}$ sideband subtracted data and MC simulation after applying all reweighting corrections for the (a) the mass of a tertiary vertex containing the trailing muon, (b) the number of additional vertices found in the event, (c) the impact parameter significance for the $B^{\pm}$, and (d) the lesser of the two muons impact parameter significance values.

Flavor Averaging Group [30]. The value was provided to the Particle Data Group for the 2012 compilation [31], which is consistent with other recent measurements [32]. The product of the branching fractions $\mathcal{B}(B^{\pm} \to J/\psi K^{\pm}) \times \mathcal{B}(J/\psi \to \mu^{+}\mu^{-}) = (6.01 \pm 0.21) \times 10^{-5}$ is also taken from the Particle Data Group [31].

3.4.1 Determining the Number of $B^{\pm}$ Events

The number of $B^{\pm}$ events is determined from the data by fitting the $B^{\pm}$ mass distribution. Figure 3.10 shows the fit of the $B^{\pm} \to J/\psi K^{\pm}$ mass distribution for the entire RunII data set. The $B^{\pm}$ peak is fit with a double Gaussian function. The background is fit with an exponential plus a threshold function. The threshold function is necessary to account for partially reconstructed B events, such as $B^0 \to J/\psi K^{0*}$ with $K^{0*} \to K\pi$.

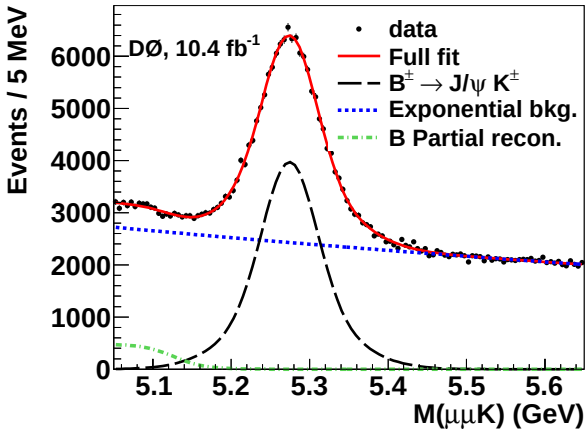


Figure 3.10 : $B^{\pm}$ mass distribution for the entire RunII data set. The $B^{\pm}$ peak is fit with a double Gaussian. The background is fit with an exponential plus a threshold function. The threshold function accounts for partially reconstructed B events.

Variations in the fit binning, mass range, and background model are tested. The difference in the number of $B^{\pm}$ events from these variations is taken into account as part of the systematic uncertainties. The variations to the background model includes allowing the inflection point of the threshold function to vary, fitting without the threshold function, and including contributions from $B^{\pm} \to J/\psi\pi^{\pm}$. $B^{\pm}$ candidate selection is also taken into account in the uncertainty on $N(B^{\pm})$. In the event that

more than one $B^{\pm} \to J/\psi K^{\pm}$ candidate is found, the "best" $B^{\pm} \to J/\psi K^{\pm}$ candidate is selected in the standard reconstruction. The vertex χ^2 is used to determine the best candidate, with the lower χ^2 candidate being retained. An uncertainty for this choice is determined by comparison to data samples in two data epochs where all candidates are kept. The final number of $B^{\pm}$events is determined to be $N(B^{\pm}) = (87.4 \pm 3.0) \times 10^3$ for the full data set, including statistical and systematic uncertainties.

3.4.2 Efficiency Ratio

The ratio of the reconstruction efficiencies, $\frac{\epsilon(B^{\pm})}{\epsilon(B_s^0)}$, is determined from MC simulation at the preselection level. The efficiency of additional cuts made as part of the optimization procedure, discussed in Section 4.2, will be taken into account later. The efficiency is defined as the number of events that pass preselection over the total number of B events generated. The number of events that pass preselection is easily determined and the number of B events recorded in the MC is known. However, before a generated B is recorded, it must pass very loose cuts at the generator level. The efficiency ratio of the generator level cuts is found to be 0.695 ± 0.015, where the uncertainties are statistical. The final efficiency ratio also includes the trigger turn-on effects, which favors the harder p_T of the muons in the $B_s^0 \to \mu^+\mu^-$ signal mode.

An advantage of using $B^{\pm} \to J/\psi K^{\pm}$ with $J/\psi \to \mu^+\mu^-$ as the normalization mode is that many of the systematics will cancel in the ratio since the kinematics are very similar. One systematic that does not cancel is that of finding the K track in the $B^{\pm} \to J/\psi K^{\pm}$. To determine this efficiency, it is assumed that the data over MC simulation efficiency ratio for adding a fourth track is the same as the ratio for adding a third track. Using the four track decay of $B^0 \to J/\psi K^{0*}$ with $K^{0*} \to K\pi$, the data/MC ratio is found to be 0.88 ± 0.06 on average, though it varies over the

data periods. To determine the ratio, the number of B^0 events is found by fitting the data shown in Fig. 3.11 for the RunIIb2 data set.

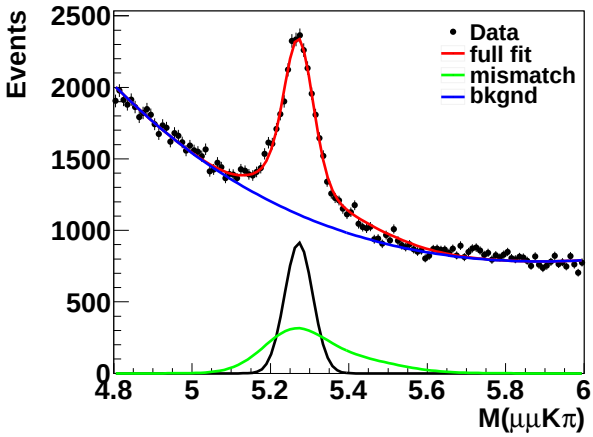


Figure 3.11 : Mass distribution, in GeV, for $B^0 \to J/\psi K^{0*}$ with $K^{0*} \to K\pi$ candidates. The data is fit by including a "signal" with the correct mass assignment and a "mismatch" with the wrong mass assignment each with the same normalization. An exponential function is fit to the background.

When reconstructing the four track vertex, the K and π can not be uniquely identified. In the mass distribution two entries are made for each event. In one entry the mass is assigned correctly, and in the other entry the wrong mass assignment is made. The shape of the wrong sign combination is determined from MC simulation, where the particle truth can be determined. The fit of the mass distribution then requires the wrong sign and correct sign combination to have the same normalization.

The efficiency ratio for the reconstruction varies by data epoch, but is on average $(13 \pm 0.5)\%$. The value for the number of $B^{\pm}$ events, efficiency ratio, tracking efficiency, and SES for each epoch can be found in Table 3.4.2. The SES for the full data set is determined to be SES $= (3.36 \pm 0.29) \times 10^{-10}$. This give a SM expected number of 10.4 ± 1.1 $B_s^0 \to \mu^+\mu^-$ events.

Run period	$N(B^{\pm})$	$\epsilon(B^{\pm})/\epsilon(B_s)$	SES $\times 10^{-9}$
RunIIa	13330 ± 809	0.137 ± 0.01	2.30 ± 0.28
RunIIb1	10866 ± 662	0.239 ± 0.01	2.93 ± 0.36
RunIIb2	24699 ± 1091	0.126 ± 0.007	1.17 ± 0.12
RunIIb3	17944 ± 1161	0.121 ± 0.009	1.53 ± 0.19
RunIIb4	20525 ± 1283	0.130 ± 0.01	1.45 ± 0.18
IIa +IIb	87364 ± 2297		0.336 ± 0.029

Table 3.1 : Column 2 gives the number of $B^{\pm} \to J/\psi K^{\pm}$ events observed in each data epoch, including statistical and systematic errors. Column 3 gives the ratio of efficiencies for normalization mode and signal mode, including statistical and systematic errors. Column 4 gives the SES value for each data epoch.

Chapter 4

The Result

4.1 Boosted Decision Tree

A Boosted Decision Tree (BDT) multivariate technique is used to separate the $B_s^0 \to \mu^+\mu^-$ signal from the main dimuon backgrounds. A BDT is a sophisticated cut-based technique. Given a signal sample, background sample, and variable list, the BDT cycles through the list of variables making a decision as to whether that variable is more signal-like or background-like. After the BDT is trained, this decision process can be applied to a test sample in the form of weights. Based on the BDT weights, a "likelihood" value, called BDT response is calculated for each event. The BDT response is used to separate signal from background. Events closer to -1 in BDT response are more background-like, while events closer to 1 are more signal-like.

The data sidebands are used as background to train the BDT. The corrected $B_s^0 \to \mu^+\mu^-$ MC simulation is used as signal in the BDT training. The MC simulations for the different data epochs are added together with each epoch having the correct number of events by luminosity. The data sideband and MC simulation samples are each divided into three independent samples. Sample A, containing about 25% of the events, is used to train the BDT. Sample B, which also contains approximately 25% of the full sample, is used to optimize the cut in the BDT response. Sample C includes the remaining 50% of the events. Sample C is used to estimate the number of expected signal and background events after the BDT response cut. In splitting

the data samples, care was taken to ensure that each data sample had an appropriate mix of events from the five data epochs.

The BDT is implemented through the TMVA package of ROOT [33].

4.2 Optimizing BDT Performance

A great deal of work goes into the optimization of the BDT. Different types of BDT are tested as part of the optimization study. Over 100 different variables are studied to determine which gave good discriminating power. During the BDT performance testing the significance ($S/\sqrt{S+B}$), where S is the number of signal events and B is the number of background events, was used as the metric for judging improvement.

The best discrimination between signal and background comes from using two different BDTs to discriminate against the two types of backgrounds. BDT1 is trained against the sequential decay background shown in Fig. 3.2(b). BDT2 is trained to fight the double semileptonic decay background shown in Fig. 3.2(c). Before training each BDT, additional cuts are made to remove a large portion of the background, see Section 4.2.1. Each BDT uses the same 30 variables, given in Section 4.2.2.

4.2.1 Pre-training Cuts

The following cuts were made before the BDT training to reduce the "easy" backgrounds:

- Cosine of the dimuon pointing angle in the transverse plane > 0.95.
- Cosine of the dimuon pointing angle using 3D information > 0.9.
- Dimuon $p_T > 5$ GeV.

- Dimuon $p_t^2 < 20$ GeV^2 in the transverse plane.
- Dimuon $p_t^2 < 30$ GeV^2 using 3D information.

The pointing angle is the angle between the line that connects the primary vertex to the dimuon vertex and the vector corresponding to the dimuon momentum. Dimuon p_t^2 is the square of the dimuon momentum component perpendicular to the line from the primary vertex to the dimuon vertex.

These cuts remove approximately 96% of the background in the blinded region while leaving 78% of the signal. Figure 4.1 shows the dimuon mass distribution in data after these cuts are made.

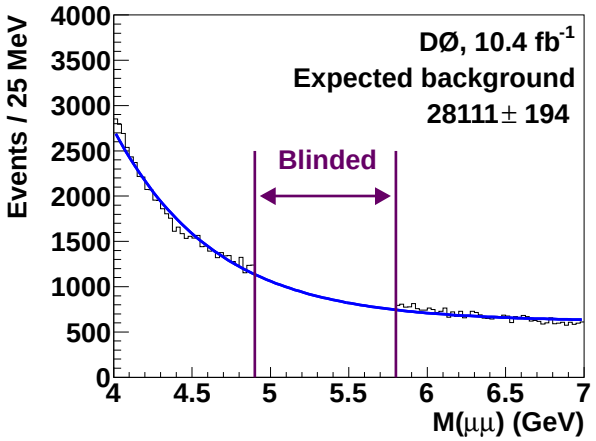


Figure 4.1 : Dimuon mass distribution in data after the pre-training cuts are applied. The plot only shows data from Sample C, but the value given for the background is for the full data set.

4.2.2 Variables

In the final BDTs, 30 variables are used to separate signal from the dimuon backgrounds. These variables are listed below, with explanation to follow. The variable

names given in parenthesis correspond to the names in Fig. 4.3 and Fig. 4.4.

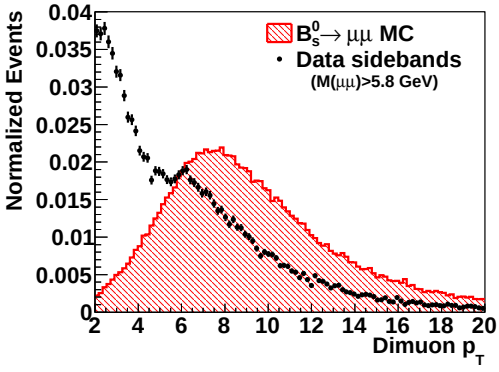
- Dimuon system p_T. (ptmumu)
- Cosine of the dimuon pointing angle calculated using only information in the transverse plane. (cxymumu)
- Cosine of the dimuon pointing angle using 3D information. (c3dmumu)
- Dimuon decay length calculated using only information in the transverse plane. (lxymumu)
- Dimuon decay length using 3D information. (l3dmumu)
- Dimuon decay length significance calculated using only information in the transverse plane. (sigxymumu)
- Dimuon impact parameter calculated using only information in the transverse plane. (ip1mumu)
- Dimuon impact parameter significance calculated using only information in the transverse plane. (ipsig1mumu)
- Dimuon vertex χ^2. (chi2)
- Dimuon p_t^2 calculated using only information in the transverse plane. (ptxysq)
- Dimuon p_t^2 using 3D information. (pt3dsq)
- Dimuon system isolation. (isolation)
- Leading muon isolation. (isomu1)
- Trailing muon isolation. (isomu2)

- Sum of the two individual muon isolations. (isosum)
- Leading muon p_T. (ptmu1)
- Trailing muon p_T. (ptmu2)
- Leading muon impact parameter. (ip1mu1)
- Trailing muon impact parameter. (ip1mu2)
- Leading muon impact parameter significance. (ipsig1mu1)
- Trailing muon impact parameter significance. (ipsig1mu2)
- Smaller of the two impact parameters of the two muons. (ipsigless)
- Difference in azimuthal angles between the two muons. (deltaphi)
- Cosine of the pointing angle for a vertex with both muons and the closest track to the dimuon vertex. (cxyNew)
- Invariant mass of the tracks associated with a tertiary vertex that does not include either muon. (mTer)
- Invariant mass of the tracks associated with a tertiary vertex that includes the leading muon. (mTermu1)
- Invariant mass of the tracks associated with a tertiary vertex that includes the trailing muon. (mTermu2)
- χ^2 of the vertex of tracks with the leading muon. (chi2mu1iso)
- χ^2 of the vertex of tracks with the trailing muon. (chi2mu2iso)

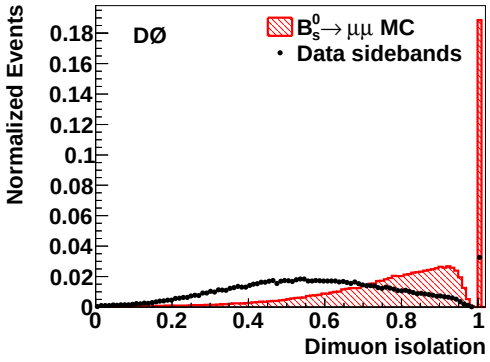
- Cosine of the pointing angle for the vertex of tracks with the trailing muon. (csymu2iso)

The dimuon decay length is the length between the primary vertex and the dimuon vertex. Decay length significance is the decay length over the uncertainty in the decay length measurement. The impact parameter is the distance of closest approach between the particle in question and the primary vertex. The impact parameter significance is the impact parameter over the uncertainty in the impact parameter measurement. Variables are often calculated in the transverse plane rather than in all three dimensions because of the uncertainty of the measurement along the beam direction, the z component. Isolation around an object, defined in Section 3.1.1, is $\mathcal{I} = p_T(\text{object})/[p_T(\text{object}) + p_T(\text{cone})]$ where the cone size is set to $\mathcal{R} = 1$. Since the calorimeter information is not used in this analysis, the isolation is only tracking based. Figure 4.2 shows difference between the $B_s^0 \to \mu^+\mu^-$ MC simulation and the background distributions for a selection of the listed variables.

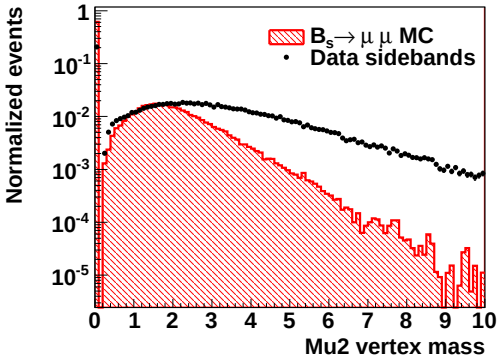
A great deal of care goes into the variable selection. Each variable is checked in the normalization mode to make sure the MC simulation produces an accurate distribution. The TMVA architecture produces a matrix of variable correlation in the signal sample and the background sample. While it is acceptable for variables to be correlated to each other, it is important to check that a variable is not correlated to mass. In some cases when variables are correlated with other variables used by the BDT, the amount of correlation in the signal sample and background sample is different, which can be useful information for the separating signal and background. Figure 4.3 and 4.4 show the correlation matrix for the signal and background samples for the BDT trained against the sequential decay backgrounds.



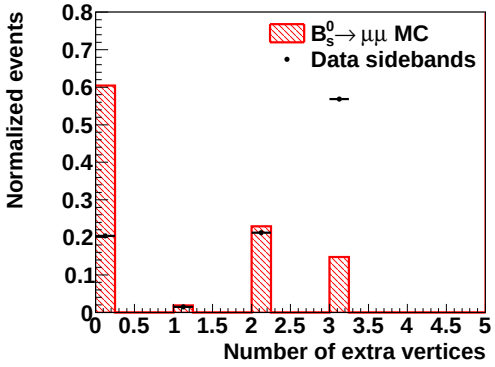
(a)



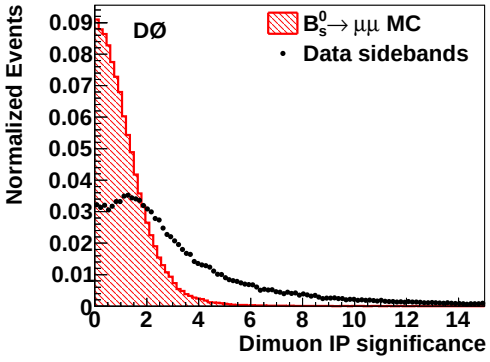
(b)



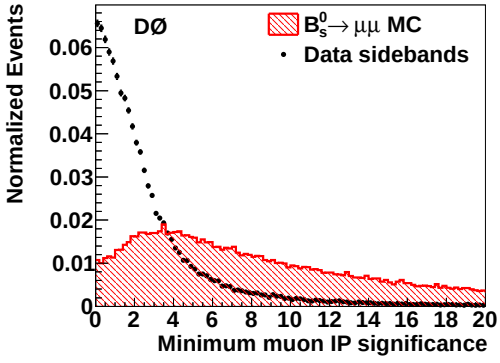
(c)



(d)



(e)



(f)

Figure 4.2 : Comparison of $B_s^0 \to \mu^+\mu^-$ MC simulation and data sidebands to highlight the difference in the distribution for some of the useful variables. Variables plotted are the (a) dimuon p_T (GeV), (b) dimuon isolation, (c) mass of a tertiary vertex containing the trailing muon (GeV), (d) number of additional vertices found in the event, (e) dimuon impact parameter significance, and (d) lesser of the two muons impact parameter significance.

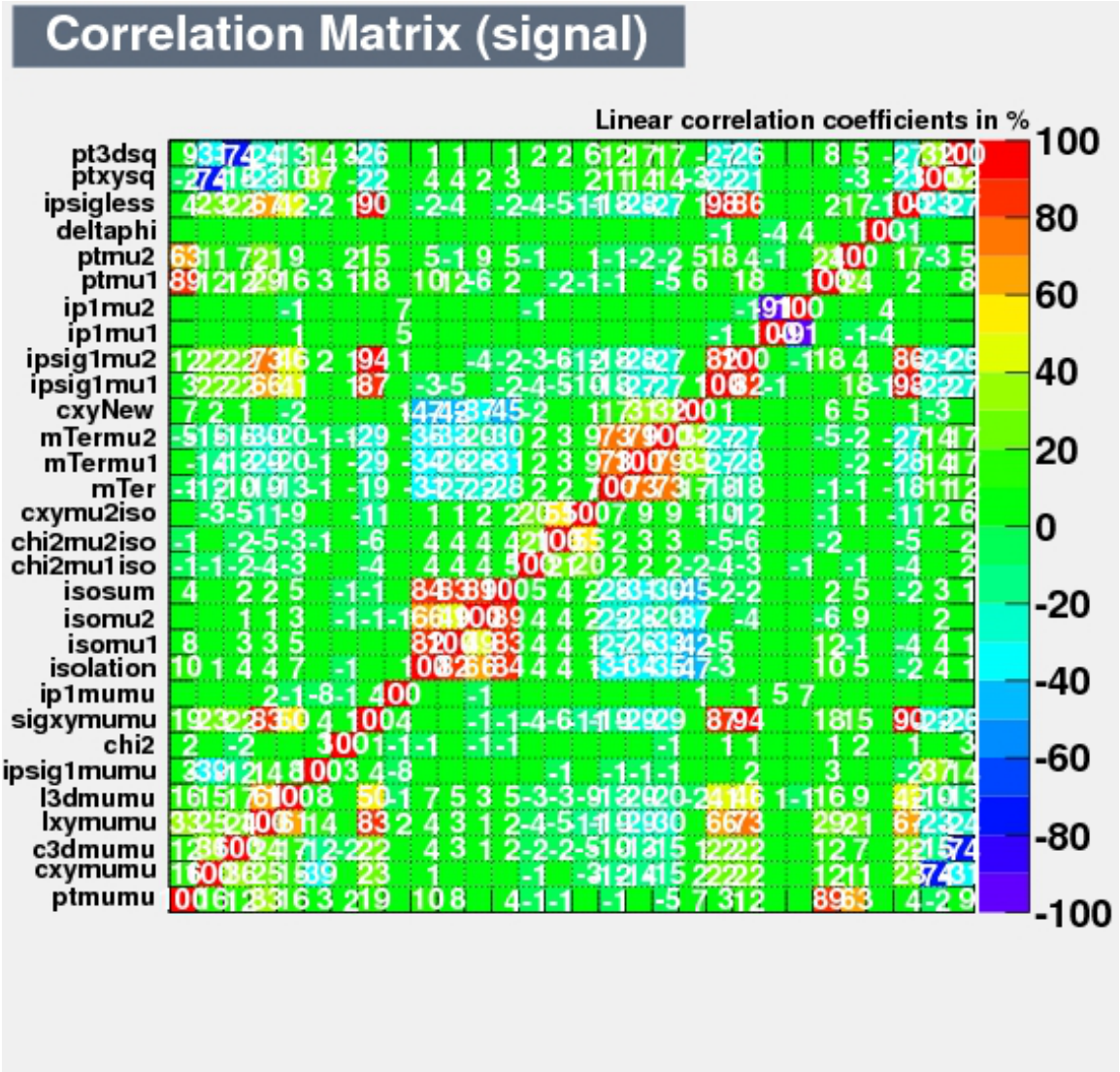


Figure 4.3 : Matrix of variable correlation for the MC simulation signal sample for the BDT trained against the sequential decay backgrounds. The variables along the bottom axis are the same left to right as going from bottom to top on the side axis.

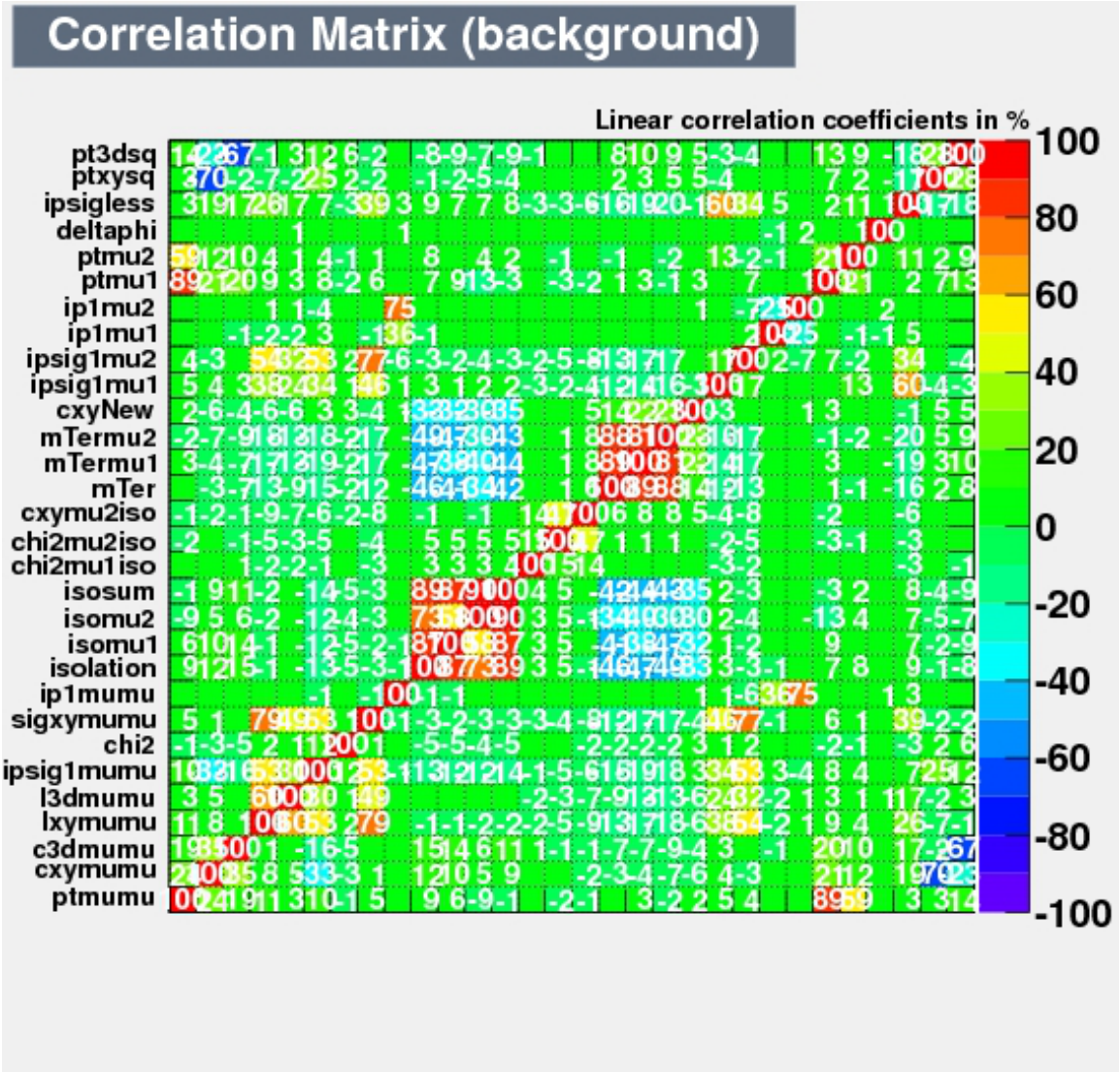


Figure 4.4 : Matrix of variable correlation for the background sample for the BDT trained against the sequential decay backgrounds. The variables along the bottom axis are the same left to right as going from bottom to top on the side axis.

4.3 Estimating the Number of Events

The data sidebands in Fig. 4.1 are fitted to estimate the number of background events in the signal region. This fit is done using a loglikelihood fit to an exponential plus constant function. The data are fitted over the sidebands and the fit is interpolated across the blinded region. The number of background events in the signal region can then be determined from integration over the appropriate boundaries. The fitting to estimate the number of events is done on Sample C. The values given for the number of expected signal and background take into account the full data set.

As part of the optimization, the signal region is taken to be smaller than the full blinded region. Because the signal is a Gaussian with a peak centered about 5.35 GeV, but the background is an exponential plus a constant, taking only the central region of the Gaussian improves the signal significance. The blinded region is from 4.9 – 5.8 GeV. The signal mass region is from 5.15 – 5.55 GeV, which covers a $\pm 1.6\sigma$ width of the B_s^0 peak.

After the BDT is trained with Sample A, the BDT weights are applied to Sample B and C. Sample B is used to determine where in BDT response to cut. Since there are two BDTs, an event is required to pass both BDT response cuts to be considered as part of the final sample. Both a coarse and fine optimization are used to determine where to make the BDT response cuts. The coarse optimization uses the significance ($S/\sqrt{S+B}$) to determine the best pair of cuts. The coarse procedure took the BDT response range for BDT1 and BDT2 and stepped over each using 100 steps, testing a total of 10,000 pairs of BDT response cuts. Once the best cuts are determined from the coarse optimization, a fine tuning of the optimization is done. In the fine optimization, the limit rather than the significance is optimized. Smaller steps around the coarse optimized cuts are studied tested in the fine optimization.

The BDT response is shown in Fig. 4.5 for both BDT1 and BDT2. As indicated by the arrows in the figure, events with BDT1 response > 0.19 and BDT2 response > 0.26 are kept. Straight cuts are chosen for this analysis. The correlation between BDT1 and BDT2 response is shown in Fig. 4.6.

After applying the BDT response cuts, the data sidebands are fit to determine the expected number of background events in the signal region. Figure 4.7 shows the remaining events in Sample C after applying the BDT response cuts. The expected dimuon background for the full data set in the signal region is 4.0 ± 1.5 events.

To determine the expected number of signal events, the BDT response cuts are applied to the signal MC. The number of expected SM signal events from the SES calculation is multiplied by the ratio of the number of MC events that pass all cuts over the total number of events in the sample after preselection. The ratio is done carefully to account for all events lost after preselection, since the SES takes into account all systematics up to preselection. So events that are removed because of the additional pre-training requirements before BDT training, the signal window restriction, and the BDT response cuts are taken into account in calculating this ratio. The expected number of SM signal events is 1.23 ± 0.13 events.

4.4 Other Backgrounds

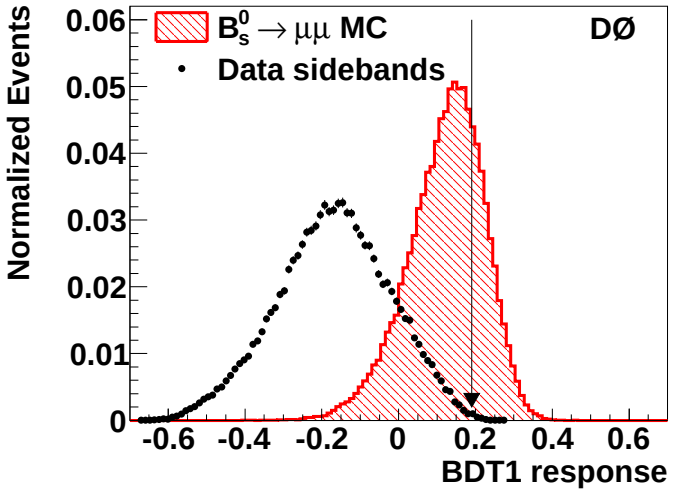
The BDTs are used to separate the signal from the two main dimuon backgrounds. Backgrounds that peak in the B_s^0 mass range also need to be studied. These peaking backgrounds include $B_d^0 \to \mu^+\mu^-$, $B_d^0 \to hh$, and $B_s^0 \to hh$, where h is a hadron. Due to the mass resolution at D0, the B_d^0 decay falls inside the mass window of the B_s^0, even though the B_d^0 peaks at a lower mass. For this analysis, the contribution from B_d^0 decay is assumed to be negligible. The $B_d^0 \to \mu^+\mu^-$ process is suppressed

by the Cabibbo-Kobayashi-Maskawa matrix element ratio $|V_{td}/V_{ts}|^2 \approx 0.04$ [31] and has a SM branching ratio of $\mathcal{B}(B_d^0 \to \mu^+\mu^-) = (1.07 \pm 0.10) \times 10^{-10}$ [7]. This suppression is supported by experimental evidence, with the LHCb collaboration placing the most stringent limit on the branching fraction on the $B_d^0 \to \mu^+\mu^-$ decay at $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 9.4 \times 10^{-10}$ [19] at the 95% C.L.

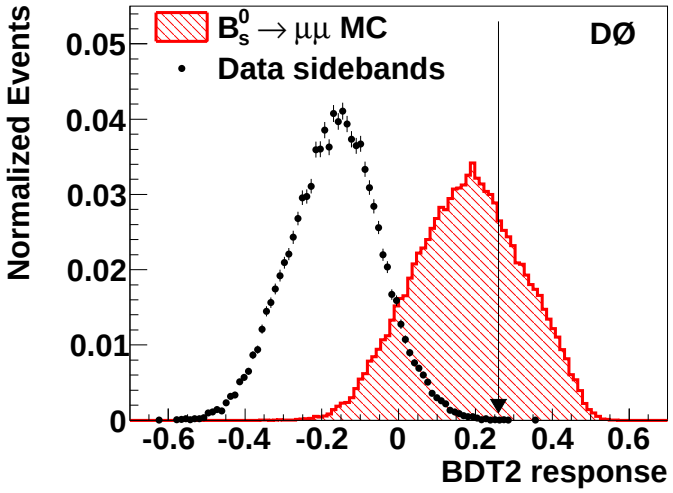
Of the B_s^0 to hadron decays, $B_s^0 \to KK$ is the largest contribution. The branching fraction for this decay is $\mathcal{B}(B_s^0 \to KK) = (2.64 \pm 0.28) \times 10^{-5}$ [31]. The $B_s^0 \to KK$ fakes a $B_s^0 \to \mu^+\mu^-$ signal when both of the kaons fake a muon. A kaon fakes a muon by decaying in flight to a muon after leaving the tracker. Thus, there are good hits in the tracking system and a match in the muon system. The kaon can also fake a muon by penetrating far enough into the detector to leave a signal in the muon system. In order to establish the contribution of this peaking background, the $K \to \mu$ fake rate must be determined.

The $K \to \mu$ fake rate is determined from data using $B \to \mu D^0 X$ with $D^0 \to K\pi$. The muon identification efficiency is determined from $J/\psi \to \mu^+\mu^-$decays in the data. One muon is required to penetrate to through the toriod, which helps reduce the fake rate. The fake rate is found to be $\epsilon(KK \to \mu\mu)/\epsilon(\mu\mu \to \mu\mu) = (3.0 \pm 1.1) \times 10^{-5}$ [29].

Using the expected number of signal events and correcting for the branching fraction difference and muon fake rate, the expected background in the signal region from $B_s^0 \to KK$ is 0.3 ± 0.1 events. The other hadron peaking backgrounds, namely $B_s^0 \to K\pi$, $B_d^0 \to K\pi$ and $B_d^0 \to \pi\pi$, are taken to be negligible due to the lower muon fake rate for π and smaller branching fractions.



(a)



(b)

Figure 4.5 : BDT response for (a) BDT1 trained to remove sequential decays and (b) BDT2 trained to remove double semileptonic decay. The MC simulation is used to model the signal and the data sidebands are used to model the backgrounds. The arrows denote where the cut in BDT response is made, with more signal like events on the right side of the cut being kept.

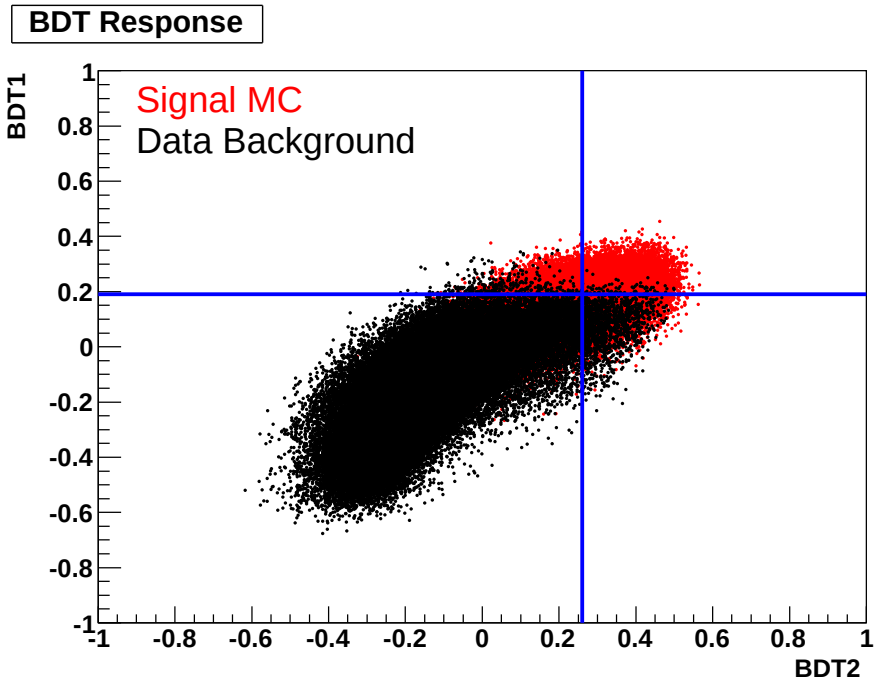


Figure 4.6 : Plot of BDT1 response versus BDT2 response showing the cuts in blue. The upper right-hand quadrant is kept. This looks like a lot of signal events, but the signal MC represents the ≈ 10 events that are in the blinded region.

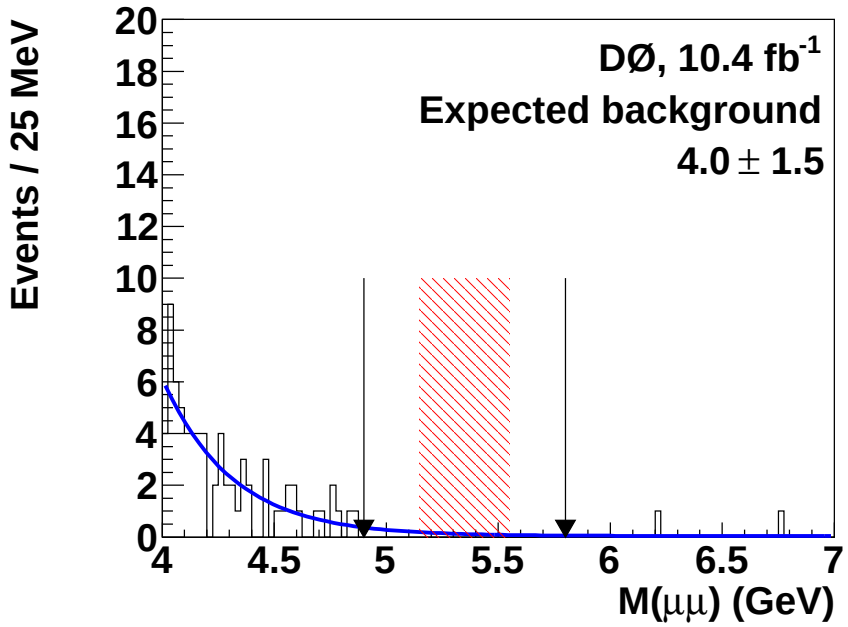


Figure 4.7 : Dimuon background after the BDT response cuts have been applied. The blue line is the exponential plus constant that is fit to the data sidebands and extrapolated into the blinded region. The arrows here denote the edges of the blinded region. The red shaded area denotes the signal region.

4.5 Expected Limit

The expected SM signal is 1.23±0.13 events in the signal window. The expected background is the combination of the dimuon background and the peaking background. The expected dimuon background is 4.0 ± 1.5 events, and the expected peaking background is 0.3 ± 0.1 events. This gives a total expected background of 4.3 ± 1.6 events. The expected limit is calculated using a modified frequentist method [34]. The calculation includes a convolution over probability distributions that represent the uncertainty in the background and signal estimates. For the dimuon background the likelihood function from the fit in Fig. 4.7 is used in the convolution. For the peaking background and signal the uncertainty is taken to be Gaussian. Weighting the possible outcomes with their Poisson probabilities gives an expected upper limit on the branching fraction of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 23 \times 10^{-9}$ (18×10^{-9}) at the 95% (90%) confidence level (C.L.).

4.6 Opening the Box

Before opening the box, a cautionary step of looking into the control region is taken. The control region is the part of the blinded region that is not in the signal region, a mass range of 4.9 – 5.15 and 5.55 – 5.8 GeV. From the dimuon fit, the expected background in the control region was 6.7 ± 2.6 events. Of these, 5.3 ± 1.9 events were expected in the lower mass side of the control region and 1.4±1.4 events were expected on the high mass side of the control region. Limits are set before the control region is unblinded for the maximum number of events that could be observed to proceed to unblind the signal region. The limit is 11 events, and 9 events are observed in the control region. With that, unblinding of the signal region can proceed.

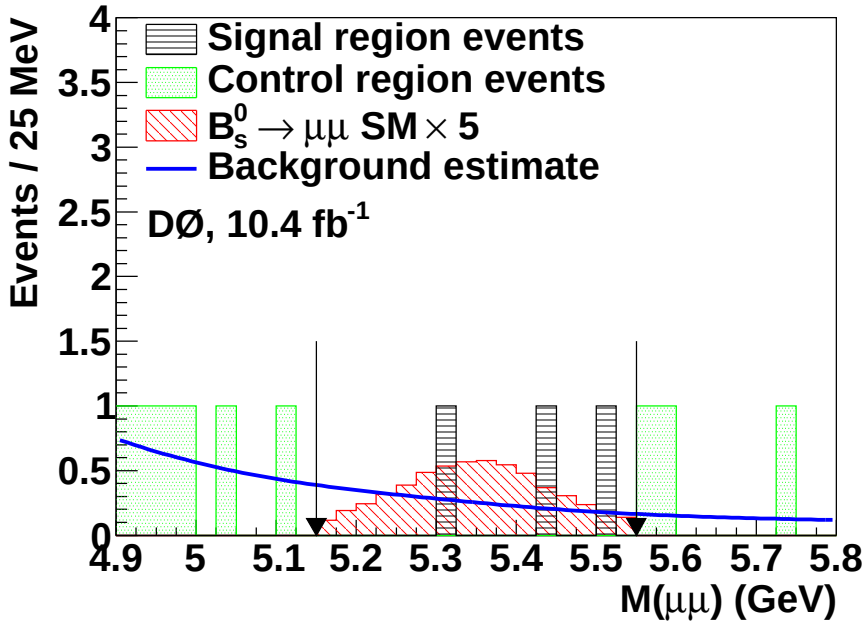


Figure 4.8 : Distribution of events after unblinding the control and signal region. The blue line represents the fit to the background. The arrows signify the edges of the signal region. The MC distribution is multiplied by a factor of 5 to be visible on the plot.

Figure 4.8 shows the events in the blinded region of 4.9 – 5.8 GeV after unblinding. A total of 3 events are observed in the signal region from 5.15 – 5.55 GeV.

4.7 Cross Checks

Several cross checks are performed on the events found in the blinded region before calculating the final limit. The first cross check looks at the data taking quality for the 3 signal region events by checking the bad run list and reading the data taking log book to look for any signs of suspicious detector behavior. This did turn up interesting information about the event at 5.5 GeV. I was on shift when this event

was recorded. No issues are discovered during this set of cross checks.

Along with checking for any data quality issues, the value of the events for all variables used in the analysis are checked. Figure 4.9 shows the distribution for a selection of variables with the value for all the events in the blinded region shown. Events in the blinded region are shown with a value of 0.01 to be visible on the plots. Appendix B has the results of this cross check for all variables.

Additionally, a check is made to ensure that the number of observed events agreed with expectations for a range of BDT cuts. For this check, shown in Fig. 4.10, the number of expected and observed events is compared as a function of the BDT response cuts.

4.8 Conclusions

The expected limit in this analysis represents a factor of 3.4 improvement over the previous D0 result. This is a factor of 1.7 better than the improvement due to increased luminosity alone. This improvement is achieved by creating new variables to differentiate signal and background, training separate BDTs for the two main backgrounds, and cutting the "easy" backgrounds before training the BDTs. Before opening the box, 4.3 ± 1.6 background and 1.23 ± 0.13 signal events are expected in the signal region. This gave an expected upper limit on the branching fraction of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 23 \times 10^{-9}$ at the 95% C.L. Upon opening the box, 3 events were observed in the signal region. Figure 4.11 compares the observed limit with recent results. The result is the best observed Tevatron limit of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 15 \times 10^{-9}$ (12×10^{-9}) at the 95% (90%) C.L. [1].

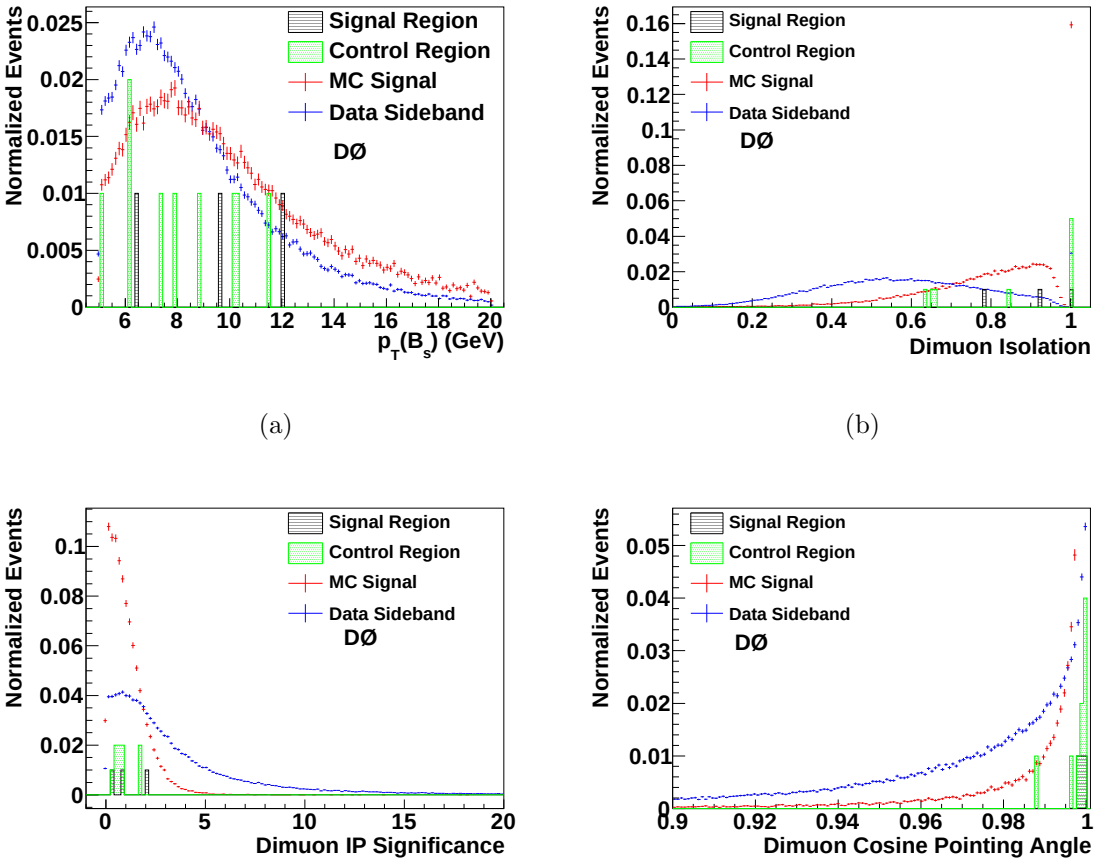


Figure 4.9 : Comparison of the distributions for the dimuon (a) p_T, (b) isolation, (c) impact parameter, and (d) cosine of the pointing angle for signal MC simulation, background, and the events found in the blinded region. Events from the signal region are denoted with black shading. Blinded region events are plotted with as 0.01 normalized events.

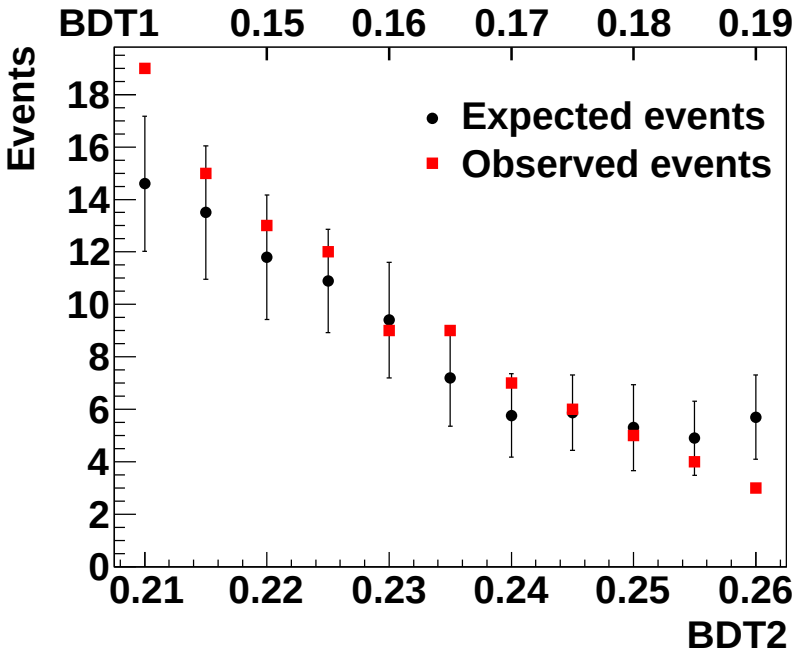


Figure 4.10 : Plot comparing the expected and observed number of events in the signal region as a function of BDT response cut. The BDT response cut test is done with even steps of BDT response.

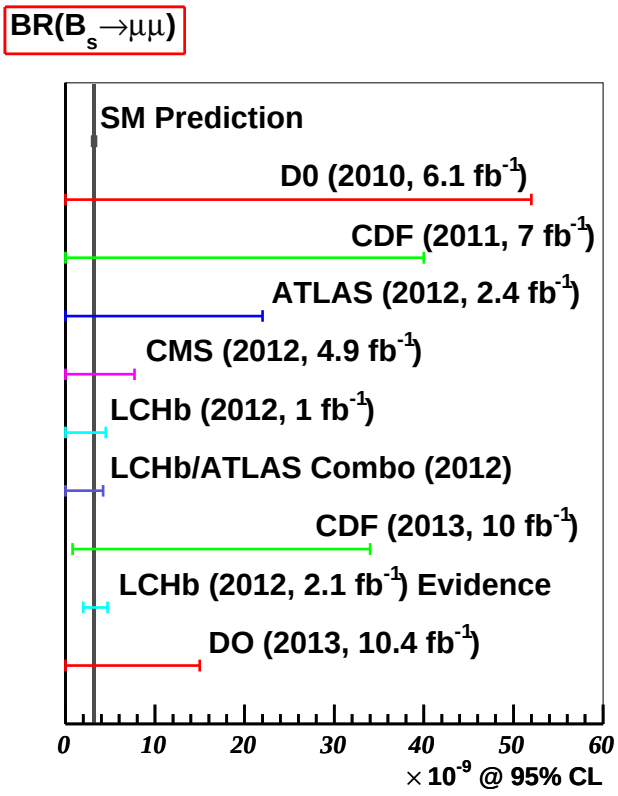
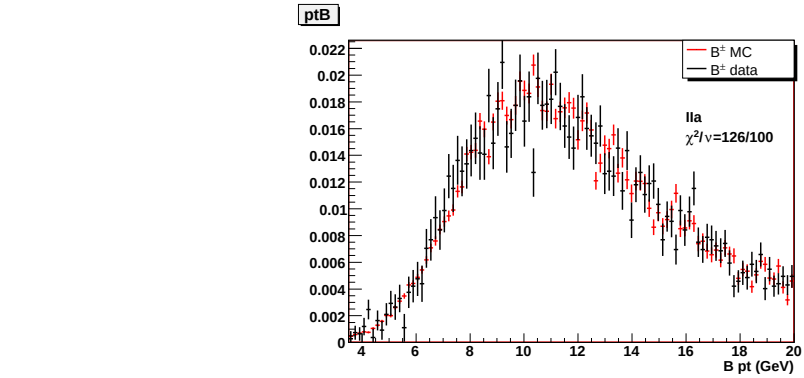


Figure 4.11 : Comparision of this observed limit with recent results.

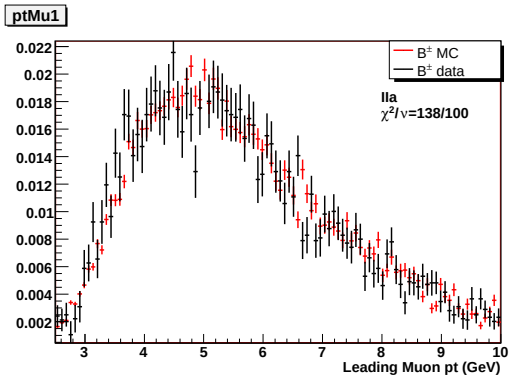
Appendix A

p_T Reweight Validation

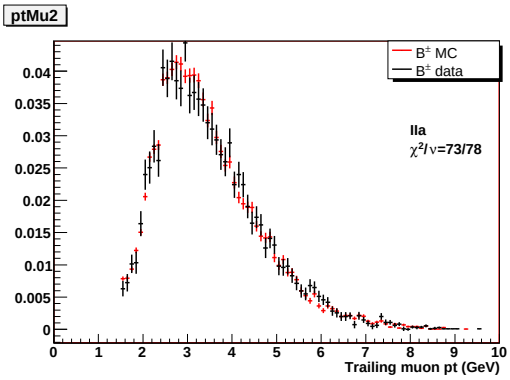
Comparisions between data and MC simulation in the normalization mode of the various p_T distributions in all 5 data epochs.



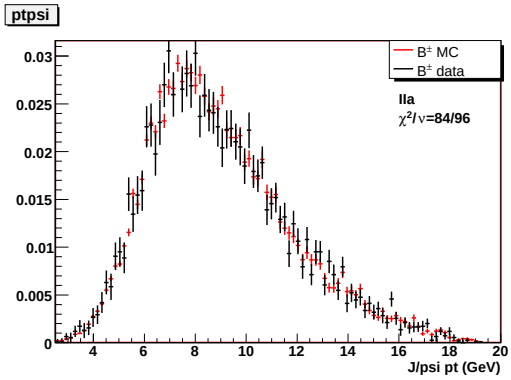
(a)



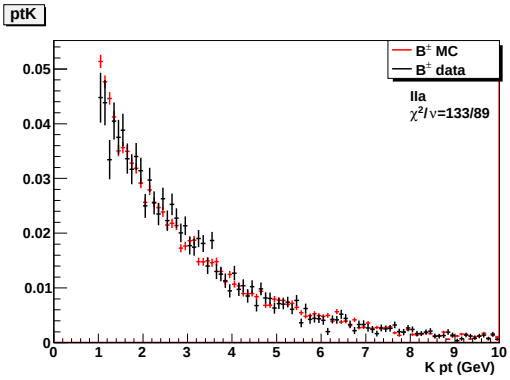
(b)



(c)

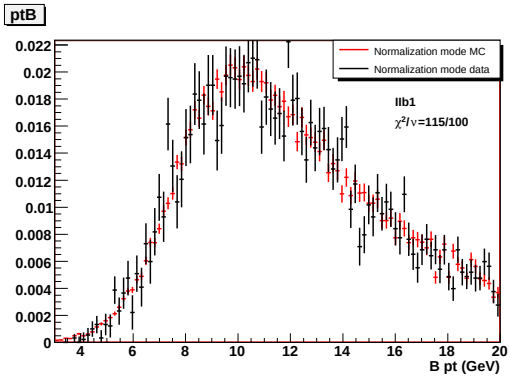


(d)

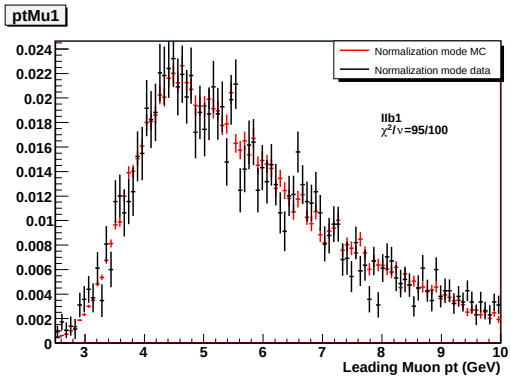


(e)

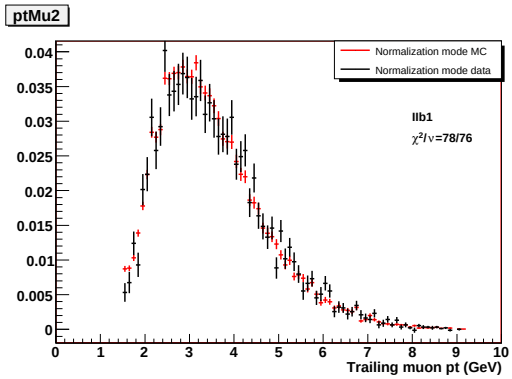
Figure A.1 : Comparison of $B^{\pm}$ sideband subtracted data and MC simulation after applying all p_T correction for the RunIIa data epoch for the (a) $B^{\pm}$, (b) leading muon, (c) trailing muon, (d) J/ψ, and (e) $K^{\pm}$.



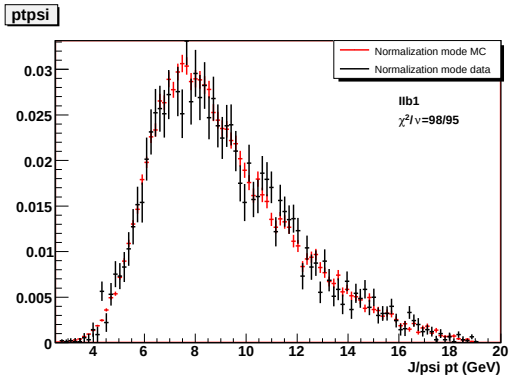
(a)



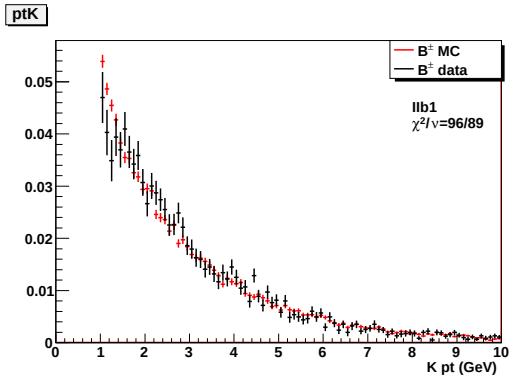
(b)



(c)

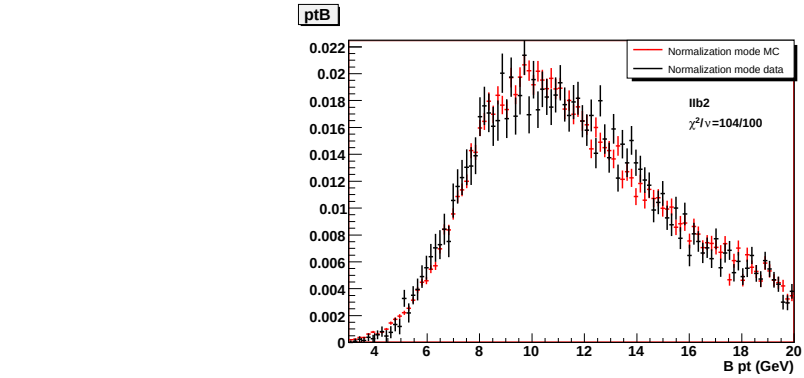


(d)

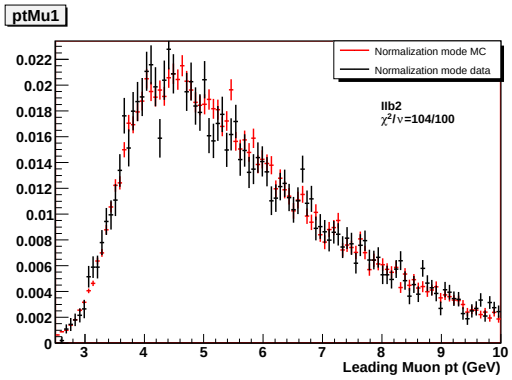


(e)

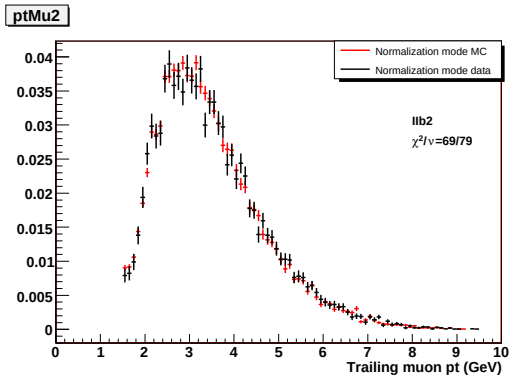
Figure A.2 : Comparison of $B^{\pm}$ sideband subtracted data and MC simulation after applying all p_T correction for the RunIIb1 data epoch for the (a) $B^{\pm}$, (b) leading muon, (c) trailing muon, (d) J/ψ, and (e) $K^{\pm}$.



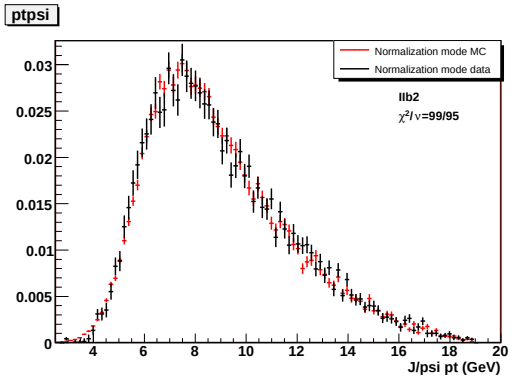
(a)



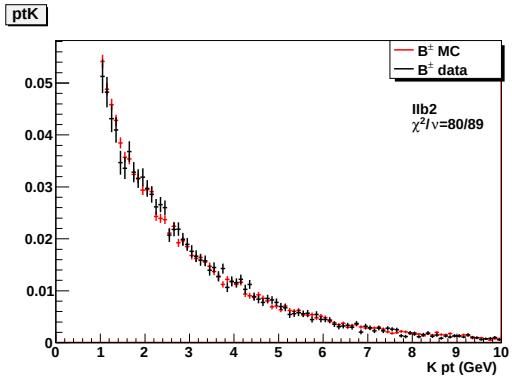
(b)



(c)

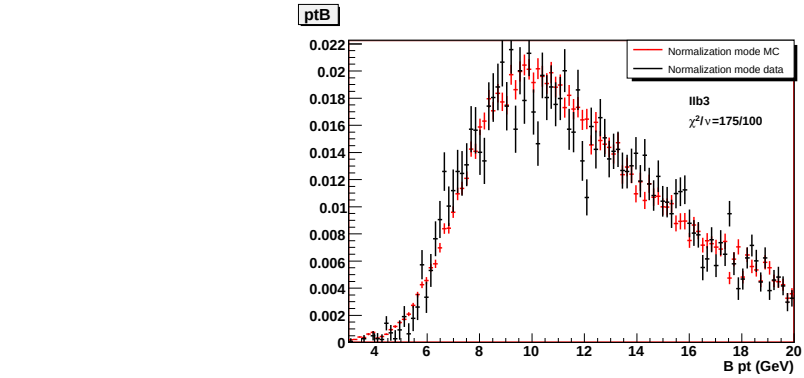


(d)

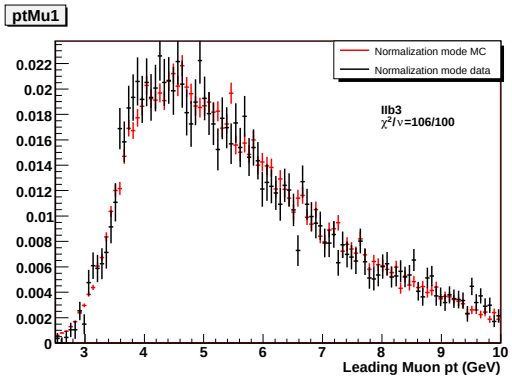


(e)

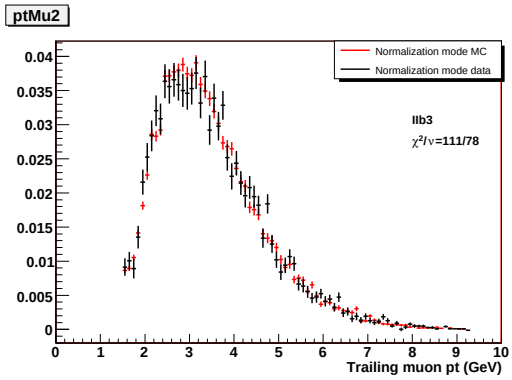
Figure A.3 : Comparison of $B^{\pm}$ sideband subtracted data and MC simulation after applying all p_T correction for the RunIIb2 data epoch for the (a) $B^{\pm}$, (b) leading muon, (c) trailing muon, (d) J/ψ, and (e) $K^{\pm}$.



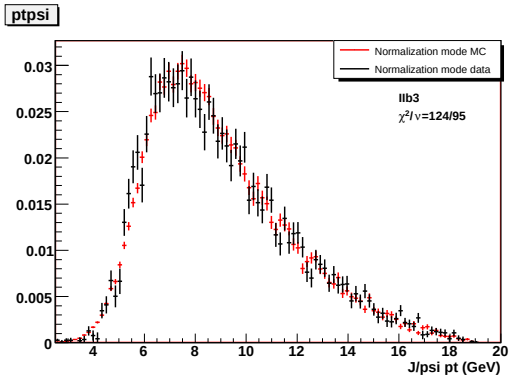
(a)



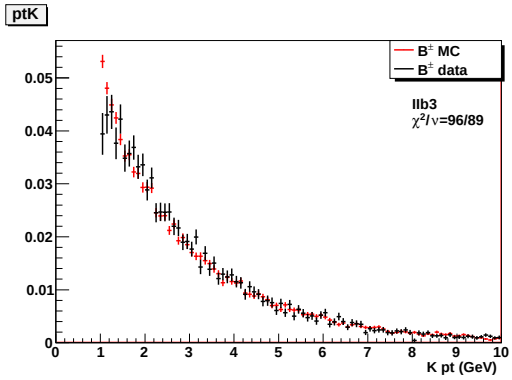
(b)



(c)

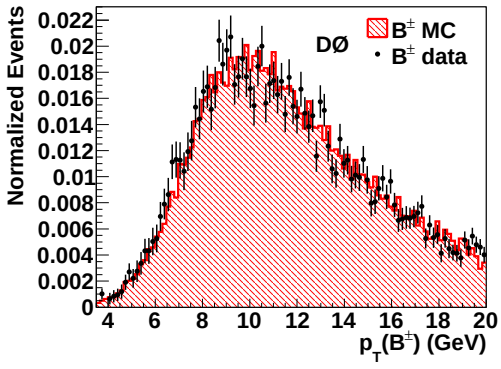


(d)

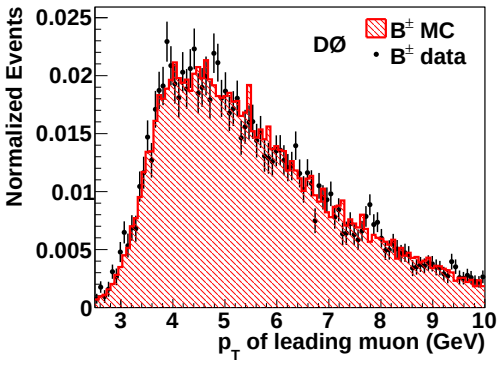


(e)

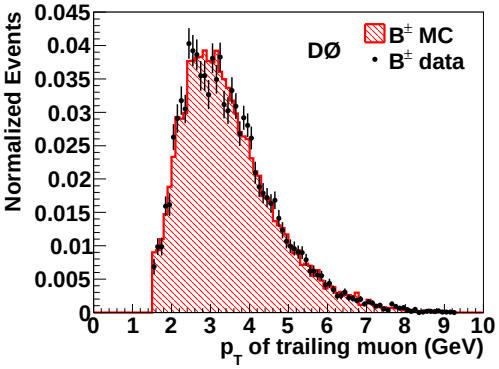
Figure A.4 : Comparison of $B^{\pm}$ sideband subtracted data and MC simulation after applying all p_T correction for the RunIIb3 data epoch for the (a) $B^{\pm}$, (b) leading muon, (c) trailing muon, (d) J/ψ, and (e) $K^{\pm}$.



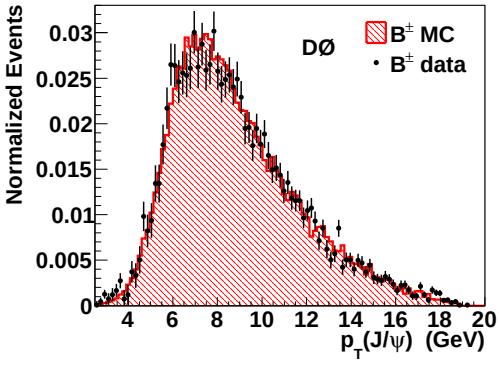
(a)



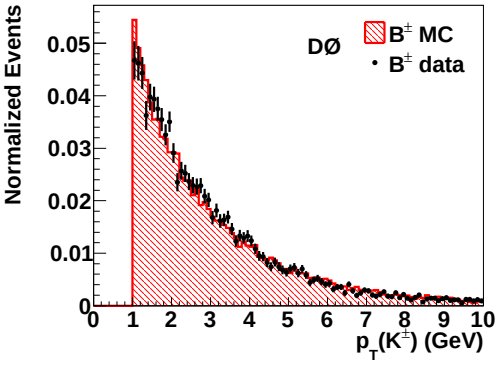
(b)



(c)



(d)



(e)

Figure A.5 : Comparison of $B^{\pm}$ sideband subtracted data and MC simulation after applying all p_T correction for the RunIIb4 data epoch for the (a) $B^{\pm}$, (b) leading muon, (c) trailing muon, (d) J/ψ, and (e) $K^{\pm}$.

Appendix B

Unblinded Cross Checks

Plots of all the variables used in the analysis comparing the signal MC distribution, the data sidebands, and the events found in the unblinded region. Unblinded events are ploted as 0.01 normalized events. Signal region events are in black. Control region events are in green.

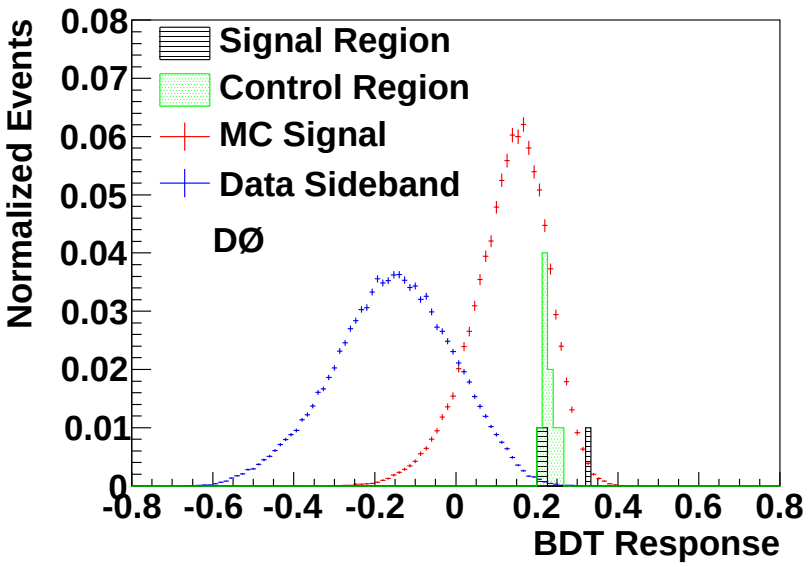


Figure B.1 : BDT1 Response (Sequential Decay)

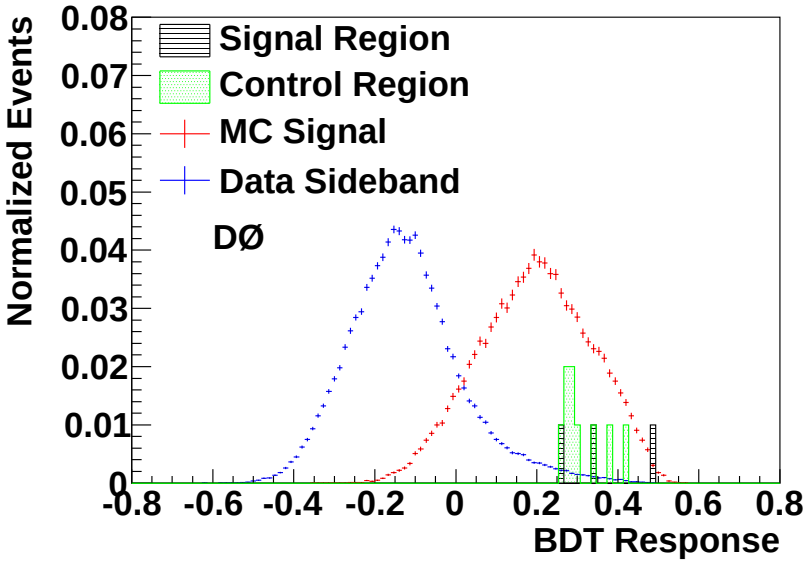


Figure B.2 : BDT2 Response (Double b Decay)

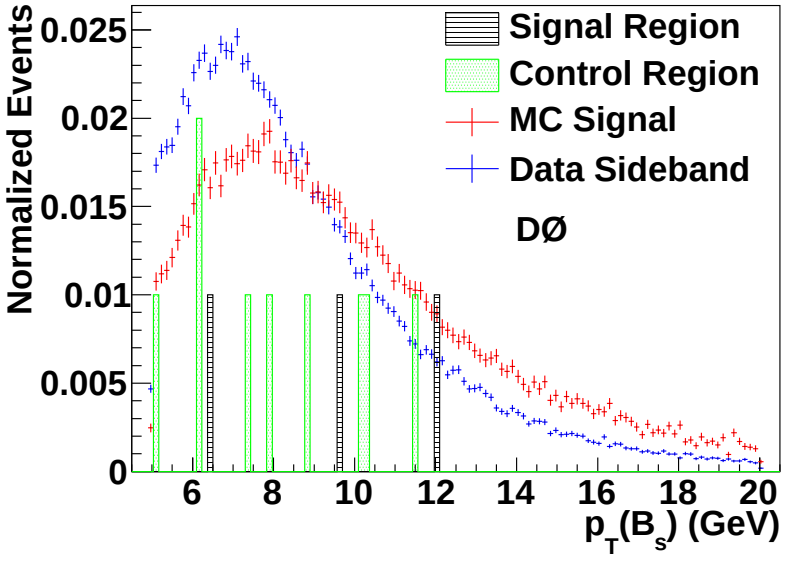


Figure B.3 : Dimuon p_T

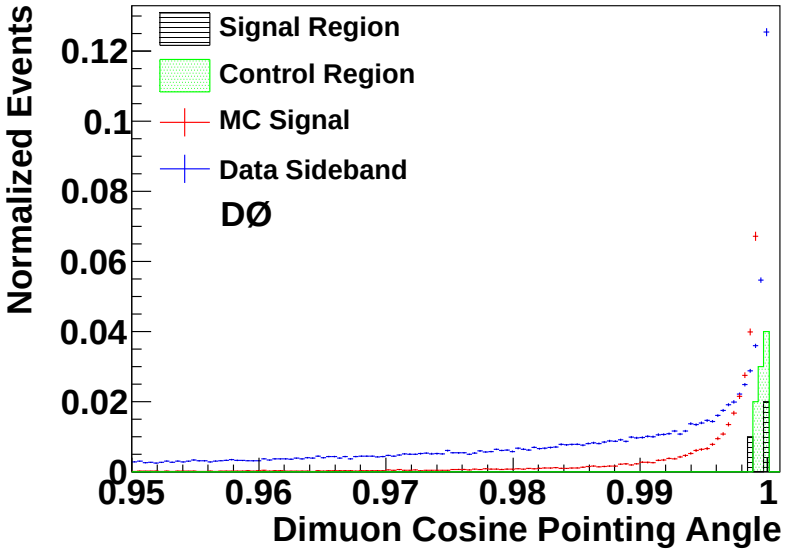


Figure B.4 : Cosine of the dimuon pointing angle (2D)

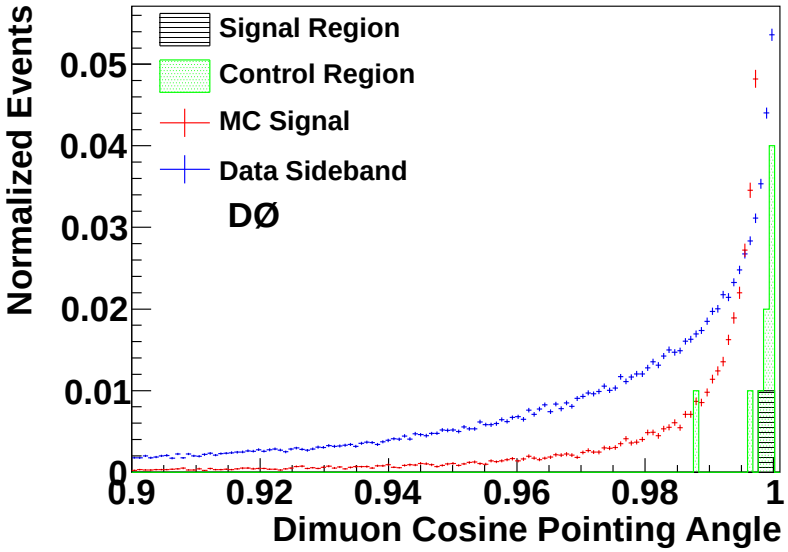


Figure B.5 : Cosine of the dimuon pointing angle (3D)

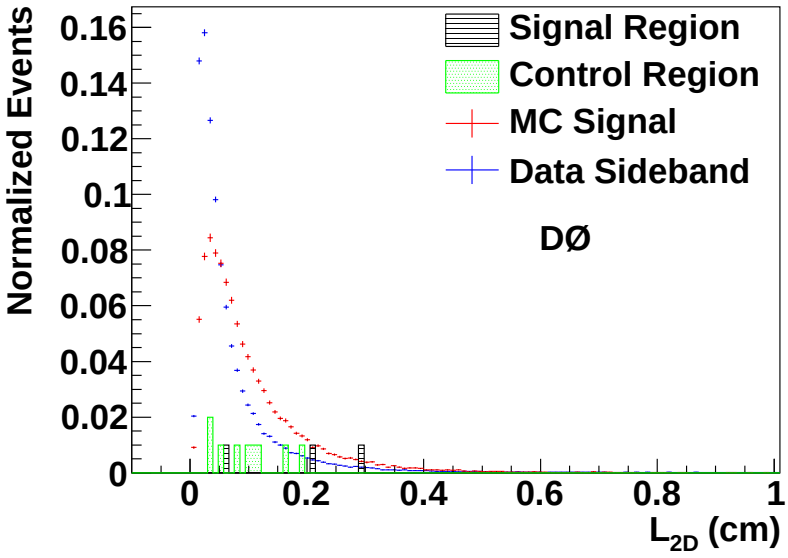


Figure B.6 : Dimuon decay length (2D)

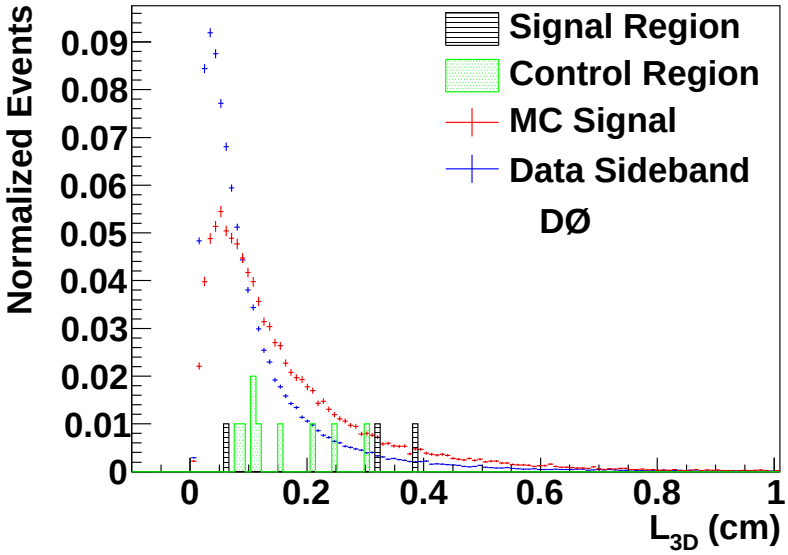


Figure B.7 : Dimuon decay length (3D)

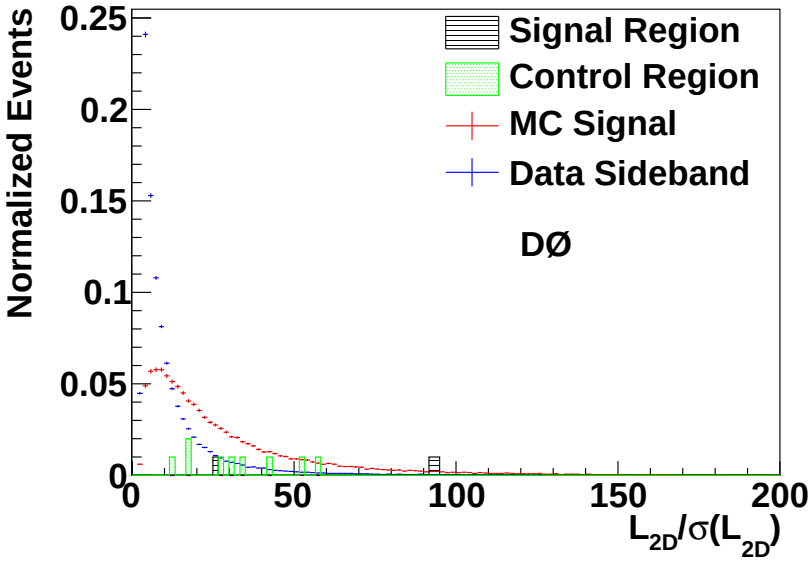


Figure B.8 : Dimuon decay length significance (2D)

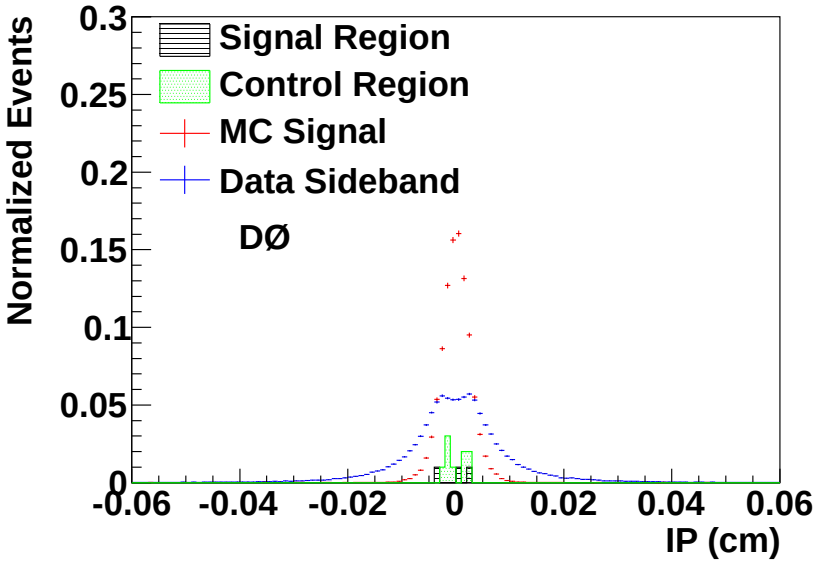


Figure B.9 : Dimuon impact parameter (2D)

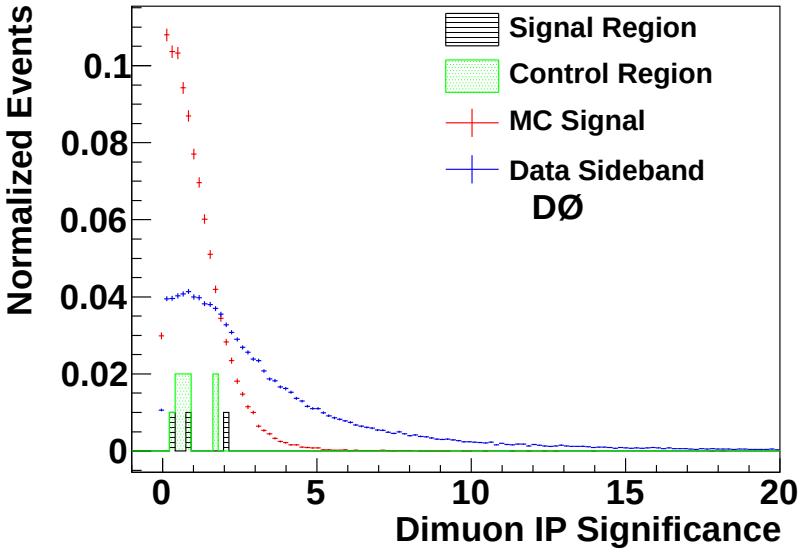


Figure B.10 : Dimuon impact parameter significance (2D)

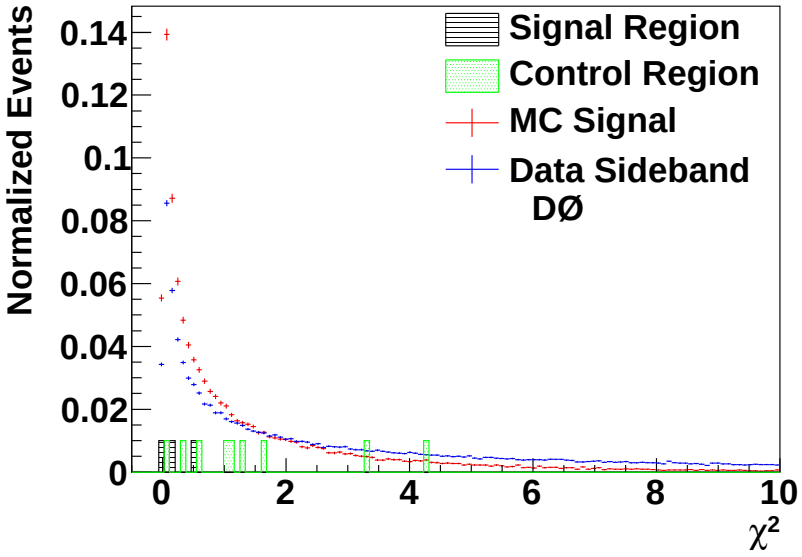


Figure B.11 : Dimuon vertex χ^2

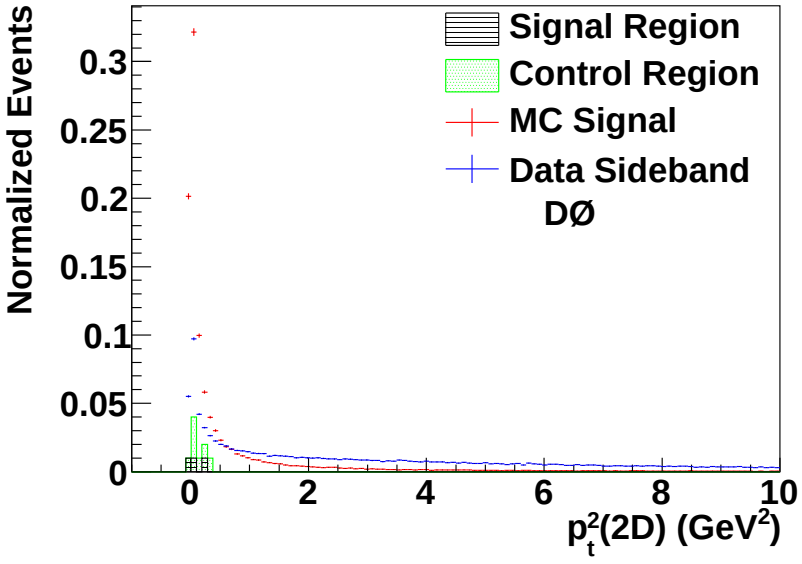


Figure B.12 : Dimuon p_t^2 (2D)

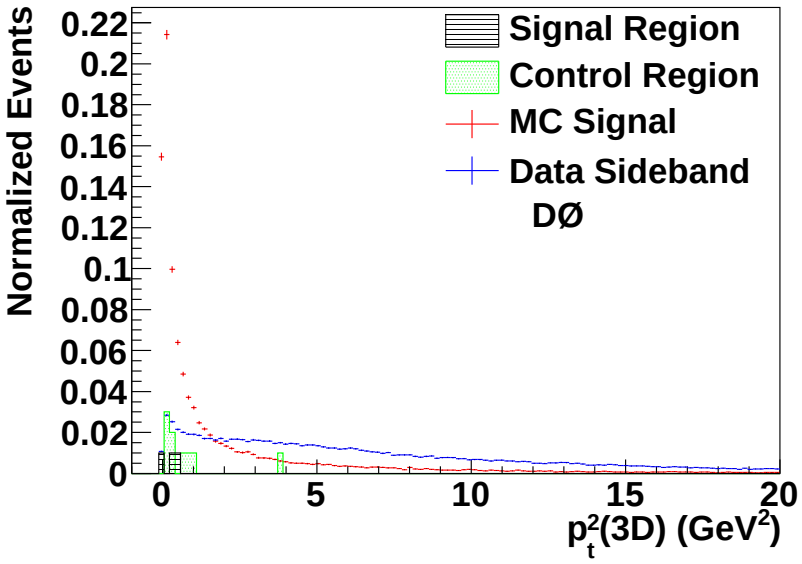


Figure B.13 : Dimuon p_t^2 (3D)

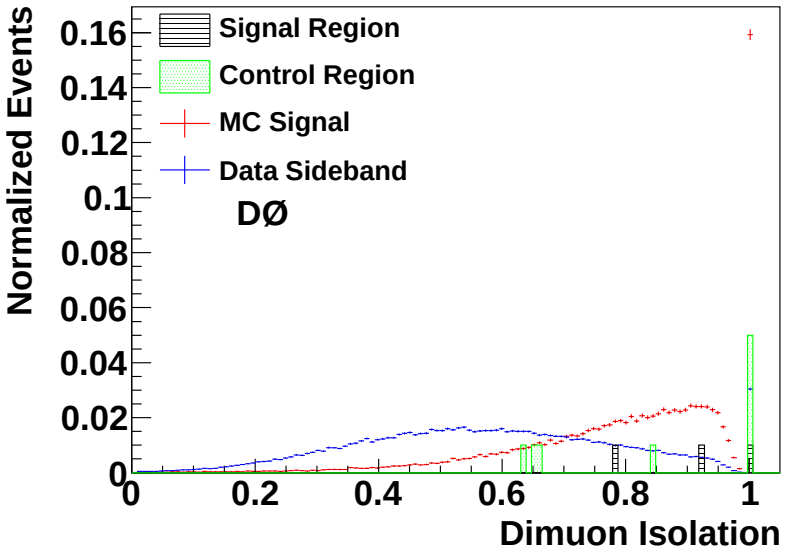


Figure B.14 : Dimuon isolation

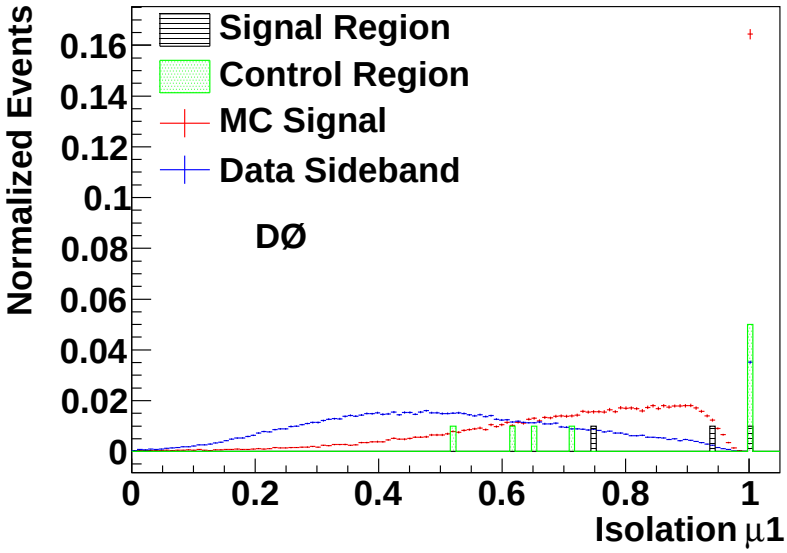


Figure B.15 : Leading muon isolation

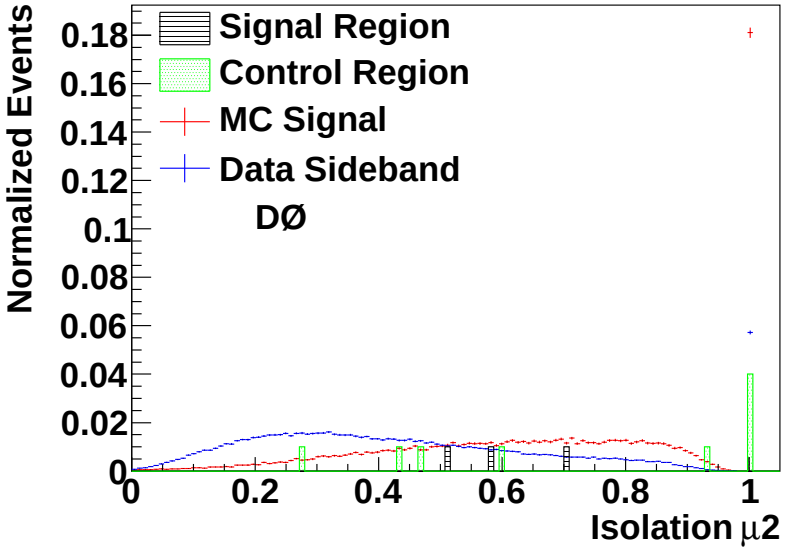


Figure B.16 : Trailing muon isolation

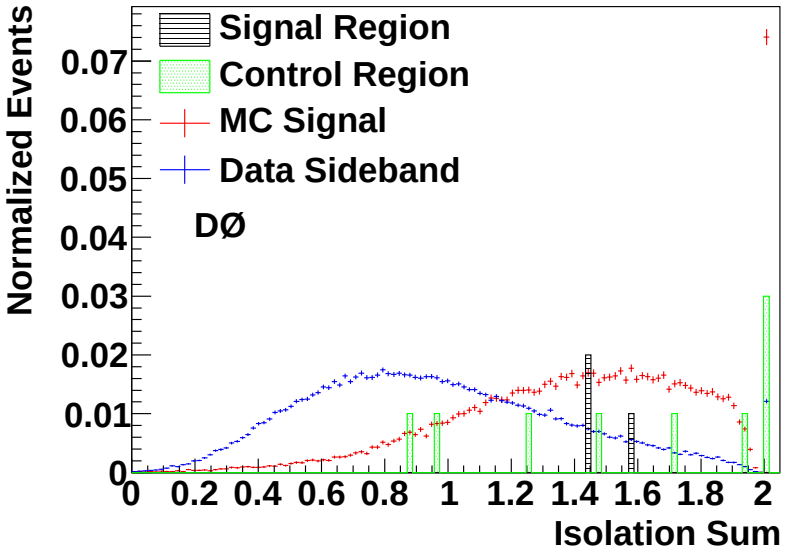


Figure B.17 : Sum of the two individual muon isolation

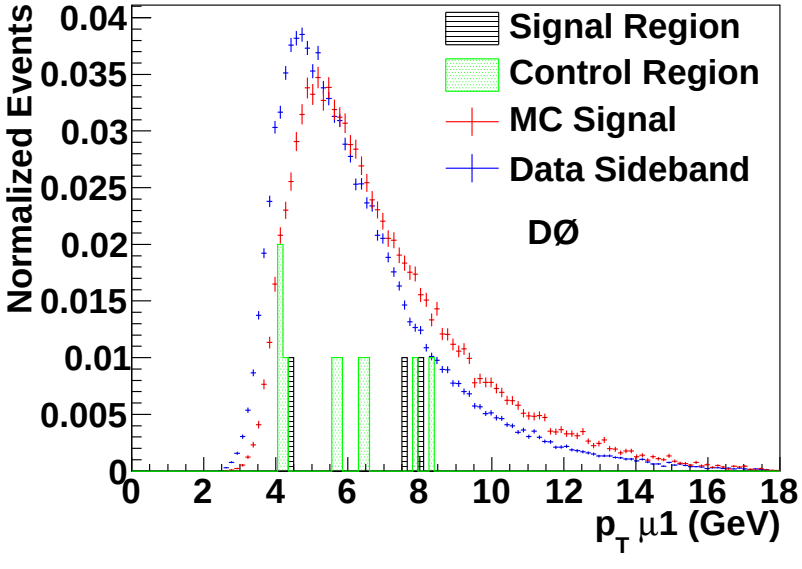


Figure B.18 : Leading muon p_T

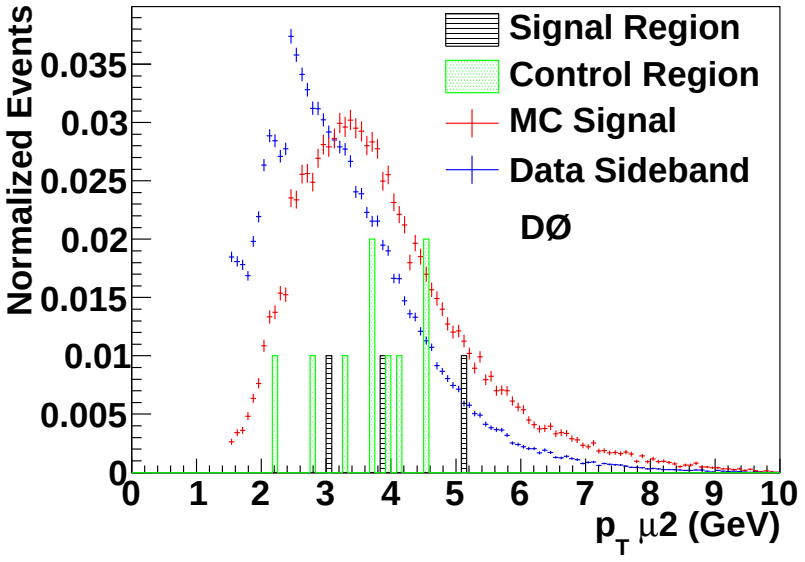


Figure B.19 : Trailing muon p_T

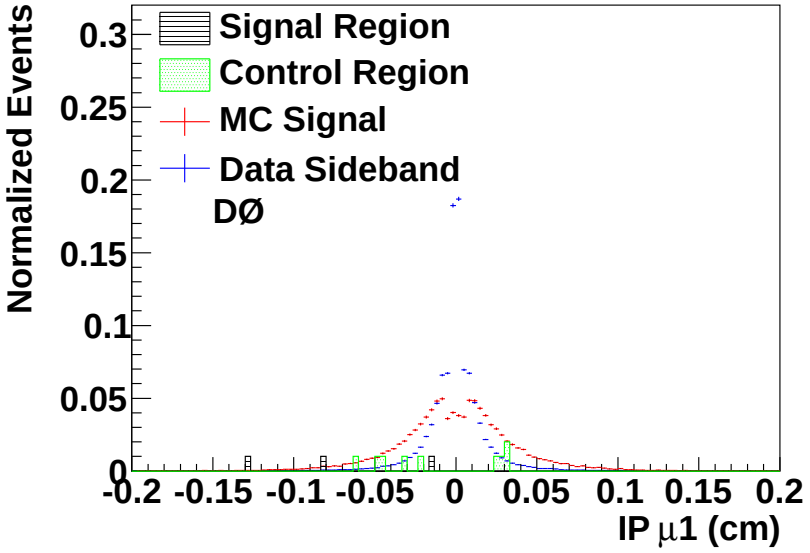


Figure B.20 : Leading muon impact parameter

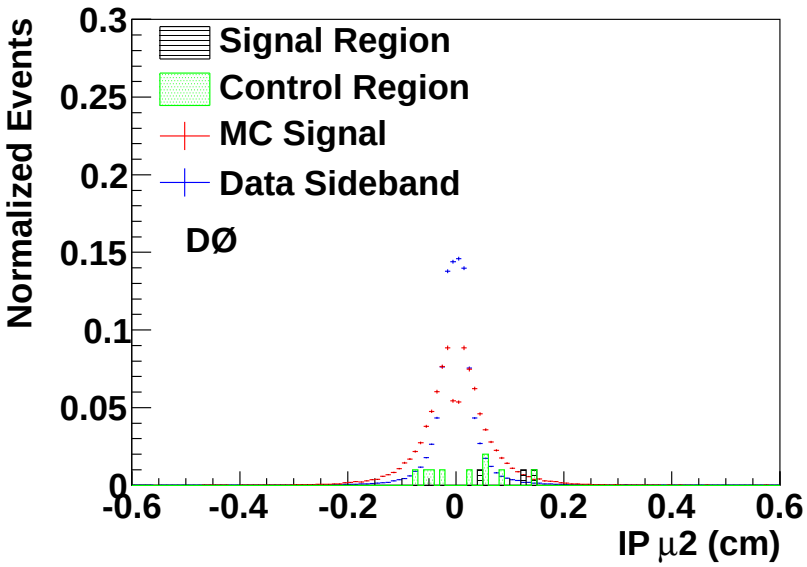


Figure B.21 : Trailing muon impact paramter

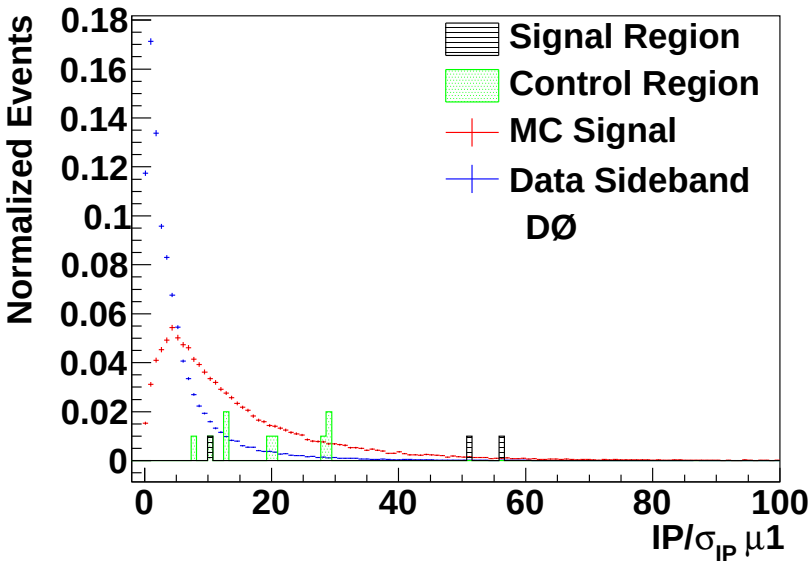


Figure B.22 : Leading muon impact parameter significance

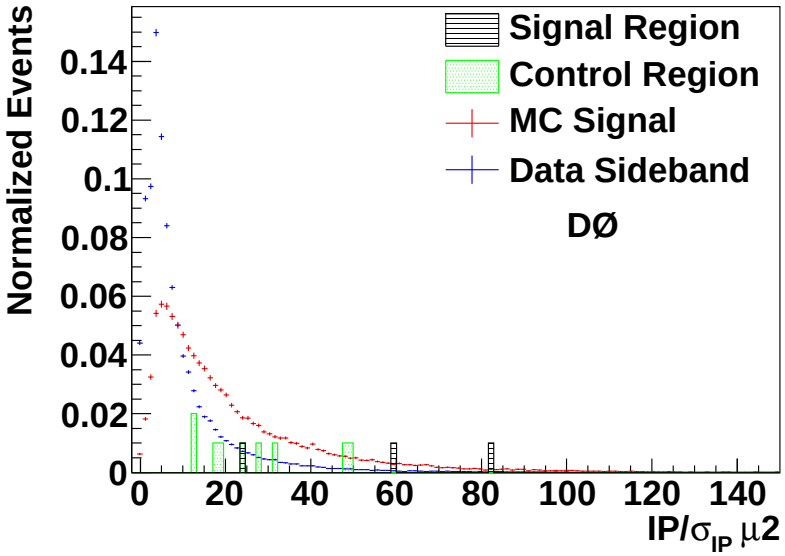


Figure B.23 : Trailing muon impact parameter significance

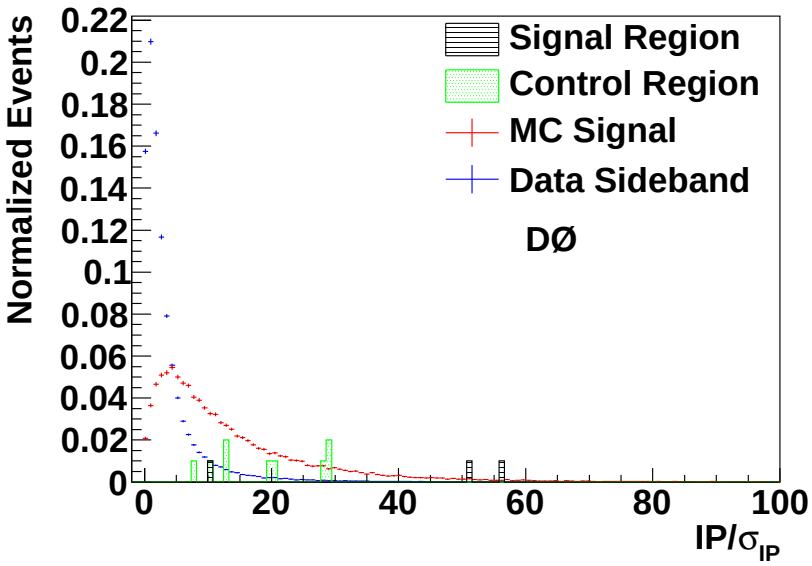


Figure B.24 : Smaller of the two muons impact parameter significance

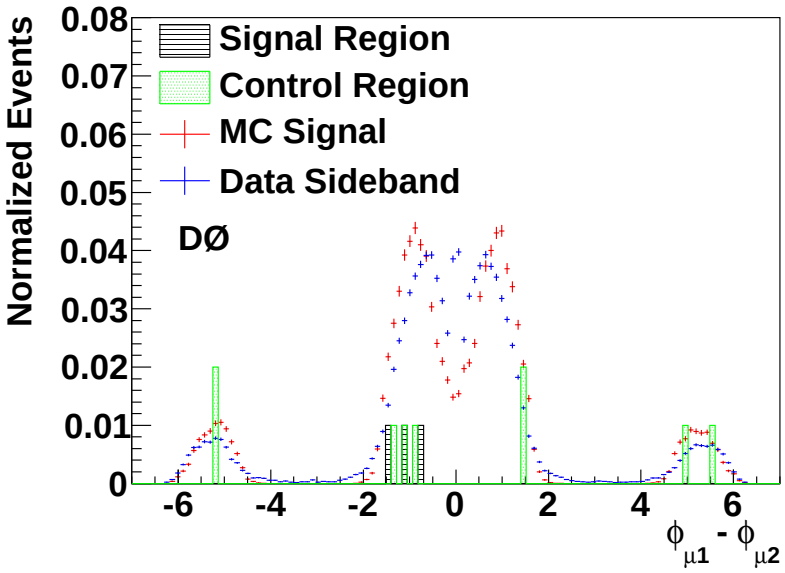


Figure B.25 : Difference in azmuthal angle between the two muons

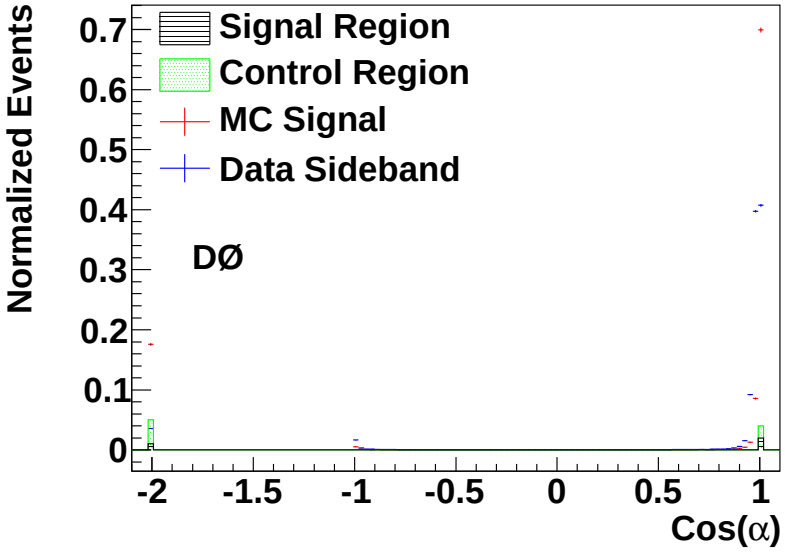


Figure B.26 : Cosine of the pointing angle of the dimuon plus closest track vertex

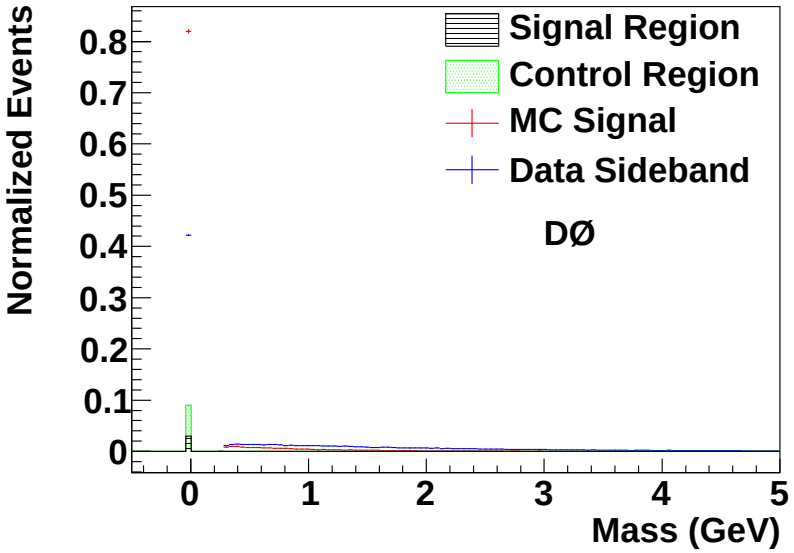


Figure B.27 : Invariant mass of the tertiary vertex

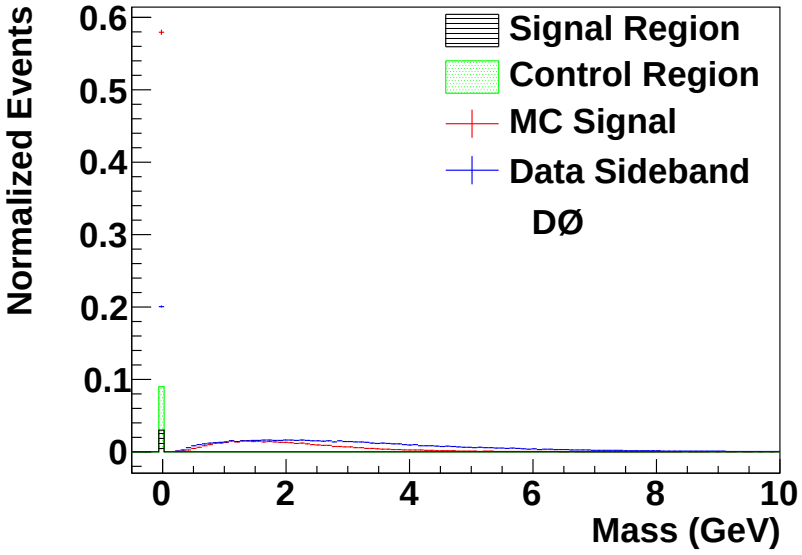


Figure B.28 : Invariant mass of the tertiary vertex with the leading muon

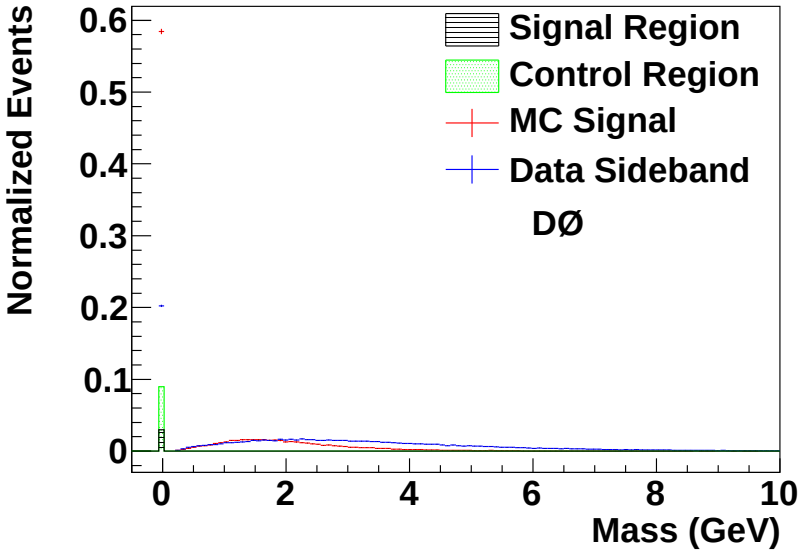


Figure B.29 : Invariant mass of the tertiary vertex with the trailing muon

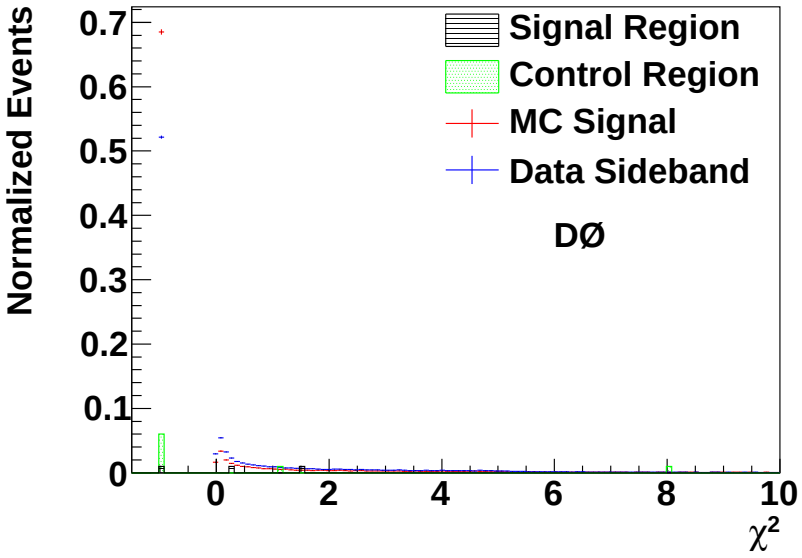


Figure B.30 : Vertex χ^2 of tracks with the leading muon

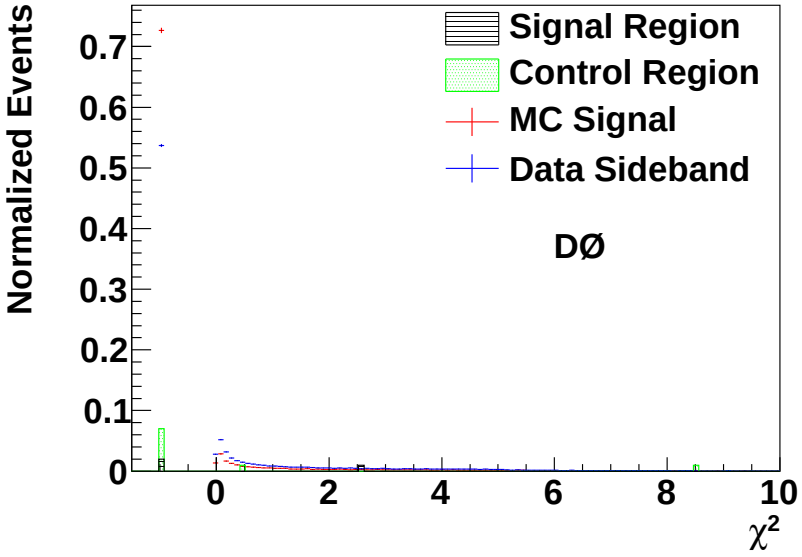


Figure B.31 : Vertex χ^2 of tracks with the trailing muon

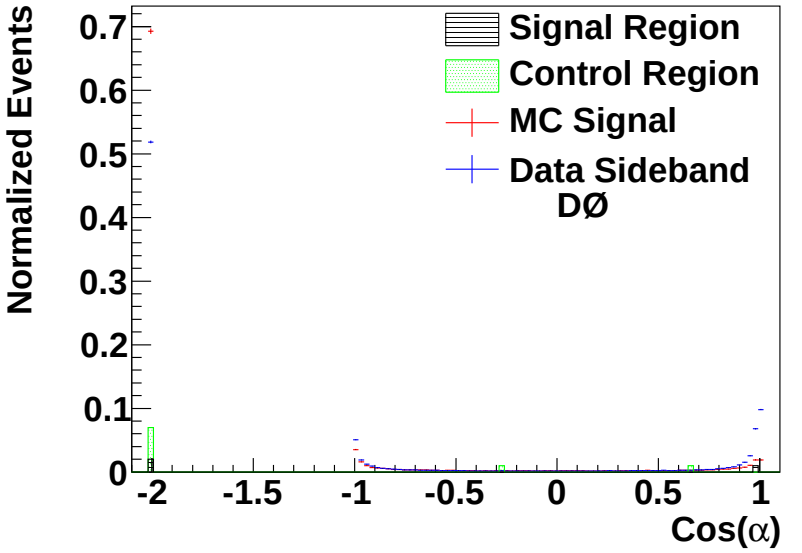


Figure B.32 : Cosine of the pointing angle for the vertex of tracks with the trailing muon

Bibliography

[1] V. Abazov *et al.*, "Search for the rare decay $B_s^0 \to \mu^+\mu^-$," *Phys. Rev. D*, vol. 87, p. 072006, 2013.

[2] S. Glashow, "Partial Symmetries of Weak Interactions," *Nucl.Phys.*, vol. 22, pp. 579–588, 1961.

[3] S. Weinberg, "A Model of Leptons," *Phys.Rev.Lett.*, vol. 19, pp. 1264–1266, 1967.

[4] S. Abachi *et al.*, "Observation of the Top Quark," *Phys. Rev. Lett.*, vol. 74, pp. 2632–2637, 1995.

[5] K. D. Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, A. Pellegrino, and N. Tuning, "Probing New Physics via the $B_s^0 \to \mu^+\mu^-$ Effective Lifetime," *Phys. Rev. Lett.*, vol. 109, p. 041801, 2012.

[6] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, "On the Standard Model prediction for $\mathcal{B}(B_{s,d} \to \mu^+\mu^-)$," *Eur. Phys. J.*, vol. C72, p. 2172, 2012.

[7] A. J. Buras, "Minimal flavour violation and beyond: Towards a flavour code for short distance dynamics," *Acta Phys.Polon.*, vol. B41, pp. 2487–2561, 2010.

[8] D. M. Straub, "Overview of constraints on new physics in rare B decays," *Proceedings: Rencontres de Moriond on Electroweak Interactions and Unified Theories*, vol. 47, 2012.

[9] S. R. Choudhury and N. Gaur, "Dileptonic decay of B_s meson in SUSY models with large tanβ," *Phys. Lett. B*, vol. 451, pp. 86–92, 1999.

[10] J. K. Parry, "Lepton flavour violating Higgs boson decays, $\tau \to \mu\gamma$ and $B_s \to \mu^+\mu^-$ in the constrained MSSM+NR with large tanβ," *Nucl. Phys.*, vol. B760, pp. 38–63, 2007.

[11] R. L. Arnowitt, B. Dutta, T. Kamon, and M. Tanaka, "Detection of $B_s \to \mu^+\mu^-$ at the Tevatron run II and constraints on the SUSY parameter space," *Phys. Lett. B*, vol. 538, pp. 121–129, 2002.

[12] J. Ellis, J. S. Lee, and A. Pilaftsis, "B-meson observables in the maximally CP-violating MSSM with minimal flavor violation," *Phys. Rev. D*, vol. 76, p. 115011, 2007.

[13] V. Abazov *et al.*, "Search for the rare decay $B_s^0 \to \mu^+\mu^-$," *Phys. Lett. B*, vol. 693, pp. 539–544, 2010.

[14] T. Aaltonen *et al.*, "Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ Decays with CDF II," *Phys. Rev. Lett.*, vol. 107, p. 191801, 2011.

[15] T. Aaltonen *et al.*, "Publisher's Note: Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ Decays with CDF II," *Phys. Rev. Lett.*, vol. 107, p. 239903(E), 2011.

[16] T. Aaltonen *et al.*, "Search for $B_s \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ decays with the full CDF Run II data set," *Phys. Rev. D*, 2013.

[17] S. Chatrchyan *et al.*, "Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays," *J. High Energy Phys.*, vol. 04, p. 033, 2012.

[18] G. Aad *et al.*, "Search for the decay $B_s^0 \to \mu^+\mu^-$ with the ATLAS detector," *Phys. Lett. B*, vol. 713, pp. 387–407, 2012.

[19] R. Aaij *et al.*, "First evidence for the decay $B_s^0 \to \mu^+\mu^-$," *Phys.Rev.Lett.*, vol. 110, p. 021801, 2013.

[20] Fermilab Accelerator Division, "Concepts Rookie Book," 2010. `http://www-bdnew.fnal.gov/operations/rookie\_books/rbooks.html`.

[21] R. Hahn. Fermilab Today, 2012. `http://www.fnal.gov/pub/today/archive/archive_2012/today12-08-21.html`.

[22] Fermilab Accelerator Division, "Tevatron Rookie Book," 2009. `http://www-bdnew.fnal.gov/operations/rookie\_books/rbooks.html`.

[23] V. Abazov *et al.*, "The Upgraded DØ Detector," *Nucl. Instrum. Meth.*, vol. A565, pp. 463–537, 2006.

[24] R. Angstadt *et al.*, "The Layer 0 Inner Silicon Detector of the D0 Experiment," *Nucl. Instrum. Meth.*, vol. A622, pp. 298–310, 2010.

[25] B. Casey, M. Corcoran, K. DeVaughan, Y. Enari, E. Gallas, I. Katsanos, J. Linnemann, J. Orduna, R. Partridge, M. Prewitt, H. Schellman, G. Snow, and M. Verzocchi, "The D0 Run IIb Luminosity Measurement," *Nucl.Instrum.Meth.*, vol. A698, pp. 208–223, 2013.

[26] T. Sjöstrand, P. Edén, C. Friberg, L. Lönnblad, G. Miu, S. Mrenna, and E. Norrbin, "High-energy-physics event generation with Pythia 6.1," *Comput. Phys. Commun.*, vol. 135, pp. 238–259, 2001.

[27] D. J. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Meth.*, vol. A462, p. 152, 2001.

[28] R. Brun and F. Carminati. CERN Program Library Writeup W5013, 1993.

[29] M. Aoki, "Search for the rare decay $B_s^0 \to \mu^+\mu^-$ using 6 fb^{-1} of D0 Run II data," Internal Note 6048, DØ Collaboration, 2010.

[30] Heavy Flavor Averaging Group. `http://www.slac.stanford.edu/xorg/hfag/osc`.

[31] J. Beringer *et al.*, "Review of Particle Physics," *Phys. Rev. D*, vol. 86, p. 010001, 2012.

[32] R. Aaij *et al.*, "Determination of f_s/f_d for 7 TeV pp Collisions and Measurement of the $B^0 \to D^-K^+$ Branching Fraction," *Phys. Rev. Lett.*, vol. 107, p. 211801, 2011.

[33] A. Hoecker *et al.*, "Toolkit for Multivariate Data Analysis." arXiv:physics/0703039v5, 2007.

[34] T. Junk, "Confidence level computation for combining searches with small statistics," *Nucl. Instrum. Meth.*, vol. A434, p. 435, 1999.