SEARCH FOR THE PRODUCTION OF THE STANDARD MODEL Z BOSON IN ASSOCIATION WITH $W^{\pm}$ BOSON IN PROTON ANTI-PROTON COLLISIONS AT 1.96 TeV CENTER OF MASS ENERGY

Justin Keung

A DISSERTATION

in

Physics and Astronomy

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2010

Evelyn Thomson, Assistant Professor, Physics and Astronomy
Supervisor of Dissertation

Ravi K. Sheth, Professor, Physics and Astronomy
Graduate Group Chairperson

Dissertation Committee

I. Joseph Kroll, Professor, Physics and Astronomy

Nigel Lockyer, Adjunct Professor, Physics and Astronomy

Ravi K. Sheth, Professor, Physics and Astronomy

Evelyn Thomson, Assistant Professor, Physics and Astronomy

Mark Trodden, Professor, Physics and Astronomy

Acknowledgements

I am humbled to have this opportunity to acknowledge all the people who had helped me during my Ph.D. study in the Physics and Astronomy Department at the University of Pennsylvania.

This thesis would not have been possible without my irreplaceable thesis advisor Professor Evelyn Thomson, who provided me with a truly rigorous education through her great attention to detail. I owe my deepest gratitude to Professor Nigel Lockyer for having absolutely floored me with his dedication. I am grateful to Professor Chris Neu for having been a part of every breakthrough that I have had as a graduate student. I feel genuinely blessed to be able to learn from them.

I must also acknowledge Dr. Peter Dong, Dr. Tom Junk, Dr. Enrique Palencia, Dr. Paolo Mastrandrea, Dr. Homer Wolfe and Dr. Tom Schwarz for keeping me on my toes, and for helping me master the tools needed to contribute to the CDF community.

The always timely and constructive suggestions and encouragements from Dr. Joel Heinrich, Dr. Anadi Canepa, and Dr. Pavel Murat are much appreciated.

And the leadership of Dr. Craig Group and Dr. Wei-Ming Yao in the WH physics subgroup, as well as Dr. Eric James and Dr. Ben Kilminster in the Higgs discovery group shall be well remembered.

Departing from CDF, I must note for the record my appreciations toward Dr. Sergei Nagaitsev and Dr. Jinhao Ruan for teaching me to properly comprehend the elegance of particle accelerator physics.

I am indebted also to Prof. Vijay Balasubraman, Prof. Joe Kroll, Prof. Elliot Lipeles, Prof. Ravi Sheth, Prof. Mark Trodden, Prof. H. H. Williams for their additional supervision in my graduate education.

I will always hold dear the friendship of Anna, Elisabetta, Tatiana, and Yanjun who particularly understand the joys and pains of working with particles. And I will always have an appetite for the repeatedly insightful discussions with Bill, Godwin, Mitch, Nandor, Rick and Walt. And I will always remember the behind the scene assistance of Jean and Pauline.

Most importantly, I recognize my wife and my folks, that only with their full and strong support throughout these years was this possible.

ABSTRACT

SEARCH FOR THE PRODUCTION OF THE STANDARD MODEL Z BOSON IN ASSOCIATION WITH $W^{\pm}$ BOSON IN PROTON ANTI-PROTON COLLISIONS AT 1.96 TeV CENTER OF MASS ENERGY

Justin Keung

Supervisor: Evelyn Thomson

The search for the production of the Standard Model Z boson in association with a W boson is motivated and discussed. This is performed using 4.3 fb^{-1} of Tevatron Run II data collected with the CDF detector in $\sqrt{s} = 1.96$ TeV proton anti-proton collisions. This is a signature-based analysis where the W boson decays semileptonically into a high-P_T electron or muon plus a neutrino, and where the Z boson decays into two b quark jets (b-jets). We increase the signal-to-background ratio by identifying the b-quarks in the jets with a new neural network-based algorithm. Another neural network then uses kinematic information to distinguish WZ to further increase the signal-to-background ratio. Since our sensitivity is still not enough to achieve an observation, we set a 95% Confidence Level upper limit on the product of the WZ production cross section and its branching fraction to the decay products specified above, and express it as a ratio to the theoretical Standard Model prediction. The resulting limit is 3.9xSM (3.9xSM expected).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

We have been asking for a long time what things are and how things work. The ancient Greeks reasoned that everything must be composed of the classical elements, and that everything can be seen as manifestation of their different combinations. Then they reasoned that everything can be chopped down to smaller pieces until the ultimate uncuttable piece, the atom, and that everything is composed of different arrangements of atoms.

At this moment in human history, we have a working theory that describes quite well what the world is and what holds it together. It says that everything that exists can be boiled down to 12 different particles, 6 quarks and 6 leptons. It also says that every type of material is a composite of quarks and leptons, and is held together by force-carrying particles. This theory is called The Standard Model (SM).

The Standard Model is quite a good theory. Numerous experiments have been performed to verify its predictions to incredible precision. But it is still an incomplete theory. One shortcoming of this theory is that it cannot explain the different masses of particles. One proposal to ameliorate this suggests that mass is not an intrinsic property of particles, but that mass arises from the interaction of a particle with a yet unseen field, the Higgs field, via the particle called the Higgs boson.

At present, the existence of a neutral Higgs boson is certainly the largest unresolved problem in the Standard Model. Standard electroweak theory uses a single fundamental scalar particle, the Higgs boson, to motivate the spontaneous electroweak symmetry breaking [1], which is needed to explain how the masses of the W and Z bosons arise. But this Higgs boson has not been directly observed experimentally. Its mass is a priori unknown, and the present constraint on the Higgs boson mass, $m_H > 114.4$ GeV/c^2 at 95% confidence level (C.L.), comes from direct searches at LEP2 experiments [2].

Since the Higgs boson is constrained to be very massive, it will take a large concentration of energy to produce. The Tevatron Run II program at Fermilab, located in Batavia, Illinois, USA, has had the capability of producing it since 2001. But unfortunately, its production is predicted to be so rare, that only one Higgs boson is produced for approximately every 10^{11} interactions carried out. This poses a difficult challenge to those of us who want to identify a set of Higgs bosons for studies.

Just like we can use the properties of a needle (e.g. magnetic) to help us identify it within a large haystack, we can use the predicted features of how the Higgs boson decays and how it is produced to help us identify it within a large background sample. But unfortunately, since the Higgs boson has not been found, its mass is not known. If its mass is higher than 135 GeV/c^2, it will decay mainly into a pair of W bosons, and if it is lighter than 135 GeV/c^2 then it will decay mainly into a pair of bottom quarks. I participate in the search in the lighter mass range ($m_H < 135$ GeV/c^2), and my work involves advancing the technology used to identify the bottom quarks to increase our sensitivity to the Higgs boson.

In searching for the Higgs boson decaying into a pair of bottom quarks, we notice that pairs of bottom quarks are unfortunately produced much more frequently by other processes (10^6x). However, we notice that the Higgs boson is sometimes produced in association with a W boson (Fig. 1.1). Even though it is rarer still than

the production of a Higgs boson alone, there is a far greater reduction in the rate of other processes producing a pair of bottom quarks in association with a W boson. As a result, in searching for the Higgs boson in the lighter mass range we also look for the associated production of a W boson, this search is called the WH search.

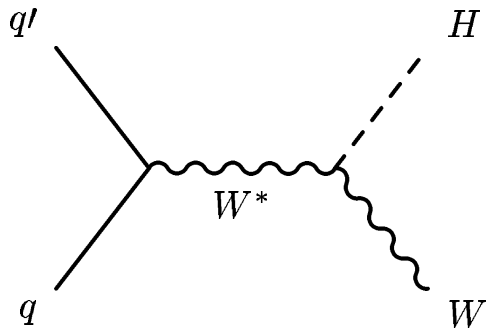


Figure 1.1: Higgs boson produced in association with a W boson, where q and q' indicate quarks coming respectively from the proton and anti-proton, W* indicates a virtual W boson which then decays into a real W boson by radiating a Higgs boson.

For my dissertation I search for the Z boson produced in association with a W boson (Fig. 1.2), which is called the WZ search. This is similar to the WH search, in that both the Z boson and Higgs boson can decay into a pair of bottom quarks. Overall the Z boson is predicted to be five times more abundant in the same production and decay modes than a Higgs boson at $m_H = 120$ GeV/c^2, but the different kinematics arising from the Z boson having a different mass reduces our sensitivity gain to 2.3x instead of the five from production and decay branching ratios alone (see section 2.4).

Since there is a a factor of 2.3 times higher sensitivity in this WZ search, we should expect to observe WZ before observing WH. This work is important because it is a proving ground for the Higgs boson search.

In chapter 2 we introduce the physics motivation behind this work and its current state in the field. In chapter 3 we introduce the Fermilab Tevatron and CDF, the particle accelerator used to produce exotic particles and the detector to observe

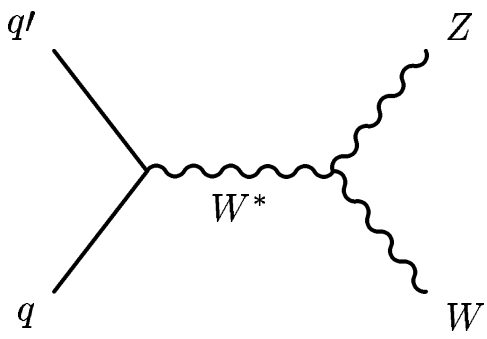


Figure 1.2: Z boson produced in association with a W boson, where q and q' indicate quarks coming respectively from the proton and anti-proton, W* indicates a virtual W boson which then decays into a real W boson by radiating a Z boson.

them. In chapter 4 we discuss how the basic physics objects are identified at CDF. In chapter 5 we discuss in more detail the identification of jets that come from b quarks. In chapter 6 we discuss the validation of the b-jet identification algorithms using the pair production cross section of top quarks. In chapter 7 we discuss data collection, and the signal and background modeling. And finally in chapters 8 and 9 we reveal and discuss our results.

Chapter 2

Theory: The Standard Model

2.1 The Particles

The Standard Model [8] is a working theory that describes quite well what the world is and what holds it together. It says that everything that exists can be boiled down to 12 different particles, 6 quarks and 6 leptons, and for each particle there is a corresponding antimatter particle. Antimatter particles have the same masses but opposite charges as their corresponding matter particles, and when they meet they annihilate each other, which then gives rise to other particles.

Quarks are one type of particle, and there are 6 of them (up, down, charm, strange, top, bottom). Unique to the quarks is that they have "color" charges, and this binds them together such that they are seen only in colorless combinations of pairs (called mesons), or triplets (called baryons). The most familiar quark type compositions are the stable triplets: proton (up + up + down) and neutron (up + down + down). The charm and top quarks are heavier versions of the up quark, and the strange and bottom quarks are heavier versions of the down quark.

Leptons are a different type of particle, and there are also 6 of them, three which have electrical charge (electron, muon, tau) and three that do not (electron neutrino,

muon neutrino, tau neutrino). The muon and tau are heavier versions of the electron, but the masses of the neutrinos are very small and have not been conclusively measured yet. Unlike quarks which are always found in pairs and triplets, leptons can be found by themselves.

There are 4 forces that together can describe every kind of interaction between quarks and leptons, and they are the strong, weak, electromagnetic, and gravitational forces. The strong force is an attractive force that acts on color charges, which only the quarks and gluons have, and binds them together to form protons and neutrons. Weak interactions are responsible for the decay of massive quarks and leptons into their lighter versions. The familiar forces are the electromagnetic force, which acts on electrically charged particles, and the gravitational force, which acts on all particles.

The forces are mediated by particles as well. The carrier for the strong force is the gluon particle, the carriers for the weak force are the W and Z particles, and the carrier for the electromagnetic force is the photon; these have all been observed. The carrier for the gravitational force is postulated to exist and is called the graviton.

The way matter interacts can be described with field equations, which are just multi-dimensional complex versions of the classic Newton equations. Now let's step back and discuss briefly the kinematics (position/momentum/energy relations) of particles starting from the classical picture.

2.2 The Kinematics

The kinematics of a free electron particle will be discussed in several formulations, starting from the Newtonian picture and taking it to the relativistic regime, leading to the quantum treatment, then finally motivating the gauge transformation formulation.

2.2.1 Relativistic Quantum Mechanics

The kinematics of a free particle in the Newtonian picture are dictated by $E = \frac{p^2}{2m}$. Note that for a free particle the kinematics are the same regardless of its position, or frame of reference. Note also that the energy is independent of time, such that a free particle having momentum p at some time will continue to have that same momentum at all other times.

If we take the equation to relativistic energies, then we must use instead Einstein's relation, $E^2 = p^2c^2 + m^2c^4$, or on taking square root

$$E = \sqrt{p^2c^2 + m^2c^4} \tag{2.1}$$

which when Taylor expanded to the second term (assuming $p = mv \ll mc$) is $E = mc^2 + \frac{p^2}{2m}$, which is the same as $E_{kinetic} = \frac{p^2}{2m}$ plus $E_{mass} = mc^2$.

Taking this equation to the quantum regime, we replace the observables with their respective operators, $E \rightarrow i\hbar\frac{\partial}{\partial t}$ and $p_j \rightarrow \frac{\hbar}{i}\frac{\partial}{\partial x_j}(j = 1, 2, 3)$, and get

$$i\hbar\frac{\partial}{\partial t}\psi = \left(\sqrt{-\hbar^2c^2\sum_{j=1}^{3}\left(\frac{\partial^2}{\partial x_j^2}\right) + m^2c^4}\right)\psi \tag{2.2}$$

which again when Taylor expanded to the second term (assuming $p = mv \ll mc$) gives us the famous Schrodinger equation.

But this equation is unable to describe particles at high energies, because it treats time and space differently in the relativistic regime. This is evident when the square root is expanded in a Taylor series without truncation, the order of the partial derivative with respect to space will grow indefinitely and the partial derivative with respect to time is only first order.

One way to solve this square root is to "complete the square" due to Dirac, and it involves solving for γ^μ in $p^2c^2 + m^2c^4 = (c\sum_{j=1}^{3}(\gamma^j p_j) + \gamma^0 mc^2)^2$. It turns out that each γ^μ needs to be a 4x4 matrix, called the "gamma matrices", having the following property

$$(\gamma^0)^2 = 1, (\gamma^1)^2 = (\gamma^2)^2 = (\gamma^3)^2 = -1, \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 0 \; for \; \mu \neq \nu. \tag{2.3}$$

Now we can rewrite equation 2.2 without the square root

$$i\hbar\frac{\partial}{\partial t}\psi = \sum_{j=1}^{3}\left(c\gamma^j\frac{\hbar}{i}\frac{\partial}{\partial x_j} + \gamma^0 mc^2\right)\psi. \tag{2.4}$$

Using the properties of the gamma matrices, and substituting
$\partial_0 = \frac{1}{c}\frac{\partial}{\partial t}$ and $\partial_j = \frac{\partial}{\partial x_j}$, we get

$$\left(\sum_{\mu=0}^{3}(i\hbar\gamma^\mu\partial_\mu) - mc\right)\psi = 0. \tag{2.5}$$

Equation 2.5 is the Dirac Equation. It is the relativistic quantum description of the equation of motion of the free electron. The fact that the world we live in has 3 space dimensions and 1 time dimension requires the gamma matrices to be 4x4 matrices, giving rise to four independent solutions. These we now interpret as two distinct particles each with two possible spin states: the electron with up/down spin, and the anti-electron(positron) with up/down spin.

2.2.2 Interactions: Gauge Transformations

Let us write down the Dirac Lagrangian, where $L = i\hbar c\overline{\psi}\gamma^\mu\partial_\mu\psi - mc^2\overline{\psi}\psi$, where $\overline{\psi}$ is the complex conjugate transpose of ψ, and where the repeated μ is implicitly being summed over the three components of space and one component of time. Note that when we attempt to minimize the action ($\int Ldt$) we get back the Dirac Equation.

We notice that the Lagrangian is invariant under the phase transformation $\psi \rightarrow e^{i\theta}\psi$, for $\overline{\psi} \rightarrow e^{-i\theta}\overline{\psi}$ and the phase cancels out for both $\overline{\psi}\gamma^\mu\partial_\mu\psi$ and $\overline{\psi}\psi$. This transformation is called a global gauge transformation, which is akin to rotating the coordinate axes by a certain angle, and doesn't measurably change anything. Nevertheless, if this phase factor was not a constant but instead depends continuously on its

space-time locations, then $\psi \to e^{i\theta(x)}\psi$ is called the local gauge transformation, and $L \to L - \hbar c(\partial_\mu \theta)\overline{\psi}\gamma^\mu\psi \neq L$, thus the Lagrangian is not invariant under local gauge transformation.

To make the Lagrangian invariant under local gauge transformation we must use the gauge covariant derivative, $D_\mu := \partial_\mu - i\frac{e}{\hbar c}A_\mu$, which is independent of the coordinate system and the local gauge. This simplifies the derivative operation, but it makes a connection with some field A_μ which remains to be dealt with. This field A_μ turns out to be physical and observable. Substituting this D_μ for ∂_μ, we get $L = i\hbar c\overline{\psi}\gamma^\mu\partial_\mu\psi - e\overline{\psi}\gamma^\mu A_\mu\psi - mc^2\overline{\psi}\psi$.

We notice that in order for the Euler-Lagrange equation to be non-trivial in A_μ the Lagrangian needs to have derivatives for the A_μ. We find that under this local gauge transformation, $A_\mu \to A_\mu - \frac{1}{e}(\partial_\mu\theta)$. The simplest term that is invariant under this transformation is $F_{\mu\nu}F^{\mu\nu}$ where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. We recognize that this $F_{\mu\nu}$ has the same form as the Electromagnetic Tensor, and it follows that A_μ is the electromagnetic potential. We multiply this by $\frac{1}{4}$ as it is the convention. And finally, we get a Lagrangian invariant under local gauge transformation:

$$L = i\hbar c\overline{\psi}\gamma^\mu\partial_\mu\psi - e\overline{\psi}\gamma^\mu A_\mu\psi - mc^2\overline{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{2.6}$$

In summary, by asking the Lagrangian that describes the kinematics of the electron to be invariant under local gauge transformation, the electromagnetic interaction shows up. This demonstrates the importance of the symmetry of the local gauge transformation.

2.2.3 Weak Interactions and The Mass Term

The Lagrangian in equation 2.6 would completely describe the kinematics of the electron, if it was only affected by electromagnetism, interacting with the massless photon. But experiments have shown that it is also affected by the weak force, interacting with the three vector bosons. Therefore we need to add additional terms

to the Lagrangian to describe the weak interactions. They are added in a manner similar to the electromagnetic interactions, contributing a potential and a kinetic term in the same manner as the electromagnetic interaction. However, there is one drawback. While the electromagnetic interaction is carried by a massless particle, the photon, the weak interaction is different in that it is carried by very massive particles, the W and Z bosons. If we attempt to add a boson mass then, under local gauge transformation,

$$\frac{1}{2}m^2 A_\mu A^\mu \rightarrow \frac{1}{2}m^2 A_\mu A^\mu - \frac{1}{e}A_\mu(\partial^\mu\theta) - \frac{1}{e}(\partial_\mu\theta)A^\mu + \frac{1}{e^2}(\partial_\mu\theta)(\partial^\mu\theta) \neq \frac{1}{2}m^2 A_\mu A^\mu \quad (2.7)$$

we see that the mass term is not invariant, spoiling the invariance of the Lagrangian.

2.2.4 The Higgs Mechanism

The Higgs Mechanism is a method to add mass terms while keeping the Lagrangian invariant under local gauge transformation, thus allowing for the unification of electromagnetic and weak interactions.

We illustrate the Higgs Mechanism using the simplest case, and show only the part of the Lagrangian that gives mass to the force carrying bosons. Introducing a new complex scalar field called the Higgs field: $\Phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$, the kinematics are described by the Lagrangian

$$L = (D_\mu\Phi) * (D^\mu\Phi) - \mu^2(\Phi^*\Phi) - \lambda(\Phi^*\Phi)^2 \quad (2.8)$$

where $D_\mu := \partial_\mu - igA_\mu$ (absorbing the $e, \hbar, c$ into the g), and μ^2, λ are free parameters describing the Higgs potential $V(\Phi) = \mu^2\Phi^*\Phi + \lambda(\Phi^*\Phi)^2$ (with $\lambda > 0$). This is the simplest Higgs potential capable of producing a vacuum expectation value that is bounded from below.

Now if $\mu^2 > 0$, then the particle has mass μ, but if $\mu^2 < 0$ then there is a circle of minima at $v = \phi_1^2 + \phi_2^2 = \sqrt{\frac{-\mu^2}{2\lambda}}$. This v is called the vacuum expectation value.

Now we can expand the original Φ field about somewhere on the circle of minima, choosing $(\phi_1, \phi_2) = (v, 0)$ (this is choosing the unitary gauge, the convenient case where the tangential component disappears), then $\Phi(x) = \frac{1}{\sqrt{2}}(v + h(x))$ and substituting this into equation 2.8 we get

$$L = \frac{1}{2}(\partial_\mu - igA_\mu(v + h(x)))(\partial^\mu - igA^\mu(v + h(x))) - \frac{1}{2}\mu^2(v + h(x))^2 - \frac{1}{4}\lambda(v + h(x))^4 \tag{2.9}$$

and simplifying we get:

$$L = \frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \mu^2 h^2 + \frac{(gv)^2}{2}A_\mu A^\mu + g^2 v h A_\mu A^\mu + O(h^3) + const \tag{2.10}$$

Here we can look for massive objects of the form $m^2\phi^2$. And there are two: $\frac{(gv)^2}{2}A_\mu A^\mu$ which gives $m_A = gv$, and $\mu^2 h^2$ which gives $m_H = \sqrt{2}\mu$.

This is the Higgs mechanism as described using the simplest case. In this manner, the masses of all the electroweak force carrying boson particles have been predicted, and they are

$$M_{W^\pm} = \frac{1}{2}gv \tag{2.11}$$

$$M_Z = \frac{1}{2}gv\sqrt{1 + tan^2\theta_W} \tag{2.12}$$

$$M_\gamma = 0 \tag{2.13}$$

with θ_W being the weak mixing angle [4].

2.3 Search for The Higgs Boson

The Standard Model with the Higgs mechanism has predicted the results of many measurements, including the masses of the $W^{\pm}$ and Z bosons with incredible accuracy. But it has still not been completely verified, precisely because the Higgs field has not yet been observed. There are two main ways of seeing evidence of the Higgs field, and both involve the Higgs boson associated with the Higgs field. One is by directly producing and observing the Higgs boson, the other is to perform precision measurements to indirectly constraint the Higgs boson mass.

2.3.1 Direct Searches at LEP

The Large Electron Positron Collider (LEP) at CERN operated at center of mass energies up to 209 GeV until November 2000. The experimentalists searched for evidence of a Higgs boson directly produced from electrons and positrons collisions. Their main search channel was the Higgs boson produced in association with a Z boson (Fig. 2.1). Their search resulted in an constraint of $m_H > 114.4$ GeV/c^2 at 95% C.L. [2].

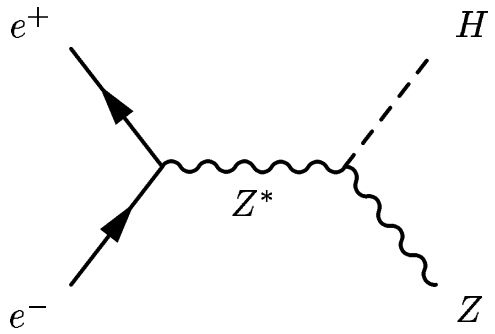


Figure 2.1: Higgs boson produced in association with a Z boson, where Z^* indicates a virtual Z boson which then decays into a real Z boson by radiating a Higgs boson.

2.3.2 Indirect Searches with Precision Measurements

The Higgs boson enters in Standard Model interactions via radiative corrections, and by measuring those interactions with great precision we will be able to place bounds on its mass. The "screening theorem" [5] tells us that the radiative corrections have a logarithmic dependence on M_H, and that by measuring very accurately the masses of the top quark (M_t) and the W boson (M_W), we can constrain the Higgs boson mass as shown in Figure 2.2. The blue ellipse indicates the current best measurements of M_W and M_t, and the green bands are theoretical predictions for their values with a variety of input Higgs boson masses. The fact that the blue ellipse is nearer the top left edge of the green bands indicates that the combination of indirect constraints and direct measurements point towards a lower Higgs mass. Figure 2.3 shows the constraints on the Higgs boson mass using data from all precision electroweak measurements, displaying the quality of the Standard Model constraining fit as a function of input Higgs boson mass. The preferred fit value is $M_H = 84^{+34}_{-26}$ GeV/c^2, and the 95% C.L. upper limit is $M_H < 154$ GeV/c^2. But if the exclusions from the direct searches are included, the 95% C.L. upper limit becomes $M_H < 185$ GeV/c^2.

2.3.3 Direct Searches at Tevatron

The Fermilab Tevatron has been operating at a center of mass energy of 1.96 TeV since 2001. It is expected to deliver 12 fb^{-1} by the end of Run II in 2011.

Figure 2.4 shows the Tevatron production cross section for several processes as a function of input Higgs boson mass. The processes with the largest production cross section are gluon fusion ($gg \to H$) and associated production with W/Z boson ($qq \to WH/ZH$).

Figure 2.5 shows the branching fractions for several decay modes as a function of input Higgs boson mass. The dominant decay mode changes at $M_H = 135$ GeV/c^2 from a pair of bottom quarks ($H \to b\bar{b}$) to a pair of W bosons ($H \to W^+W^-$).

The methods for the direct searches for the Higgs boson at the Tevatron are

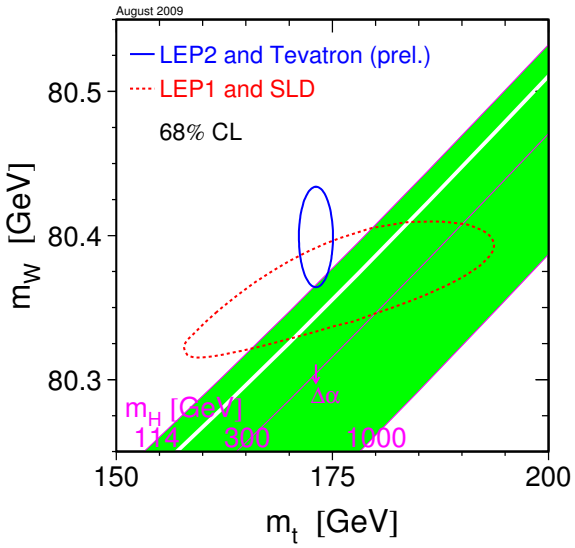


Figure 2.2: Higgs Mass constraints using M_W and M_t [6].

driven by the production cross section and the decay modes at a given Higgs boson mass, and can be grouped into two categories: high-mass ($M_H > 135$ GeV/c^2) and low-mass ($M_H < 135$ GeV/c^2). The high mass category uses the W boson pair decay mode, while the low-mass category uses the bottom quark pair decay mode instead.

Figure 2.6 shows the expected and observed limits obtained using the two detectors at the Fermilab Tevatron, CDF and D0, as of November 2009. In this figure, the dashed line represent expected 95% Confidence Level limit obtainable with present sensitivity, and the green and yellow bands represent one and two standard deviation fluctuations about the expected value. The solid black line, representing the observed value, has crossed below the horizontal SM=1 line in the mass range indicated by the purple band, indicating that the Tevatron has already started to rule out the existence of the Higgs boson in that band at a 95% Confidence Level. In the lower (and higher) mass ranges however, additional sensitivity is needed before we can rule out or provide evidence for the existence of the Higgs boson in those

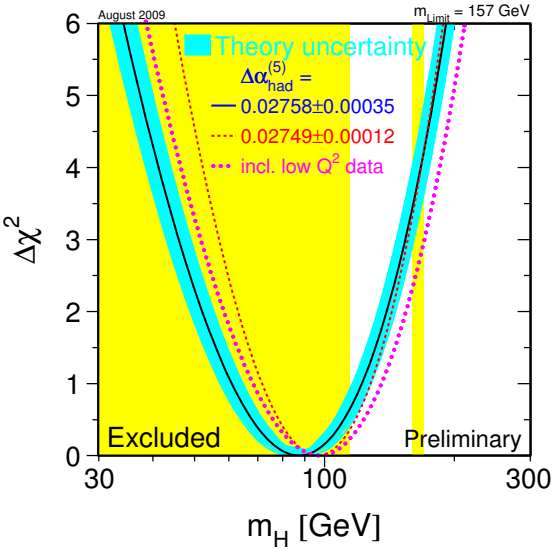


Figure 2.3: Higgs Mass constraints using all precision electroweak measurements [6].

ranges at a 95% Confidence Level. Increases in sensitivity can come from larger data samples as well as improvements in analysis techniques.

2.3.4 Direct Searches In The Low-Mass Category at Tevatron

In the low-mass category the dominant mode of the Higgs boson decay is the bottom quark pair. Pairs of bottom quarks are unfortunately produced much more frequently by other processes (10^6x) at the Tevatron. However, we notice in Figure 2.4 that the associated production with W/Z boson ($qq \rightarrow WH/ZH$) has the second largest production cross section. Even though the associated production of the Higgs boson is rarer than the production of a Higgs boson alone, the rate of other processes involved in the associated production of bottom quarks pairs is reduced much more. This gives the associated production channels a higher signal-to-background ratio.

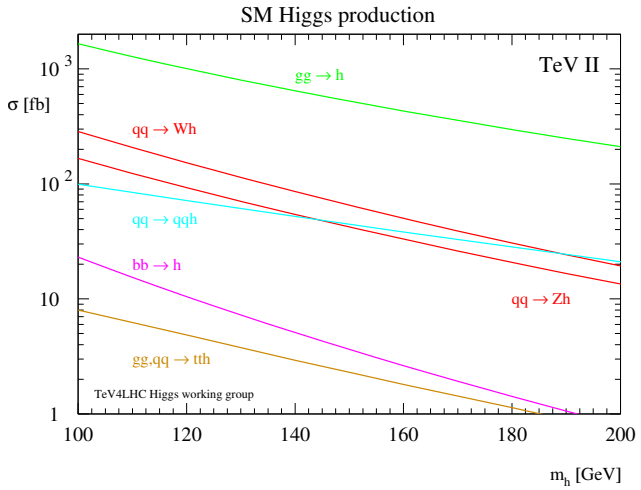


Figure 2.4: Standard Model Higgs boson production cross sections at the Tevatron [7].

The searches for the Higgs boson in the low-mass category can be divided into three channels depending on the mode of decay of the associated W or Z boson (Table 2.1).

	Branching Fraction
$W \rightarrow e\nu$	10.8±0.1%
$W \rightarrow \mu\nu$	10.6±0.2%
$W \rightarrow \tau\nu$	11.3±0.2%
$W \rightarrow hadrons$	67.6±0.3%
$Z \rightarrow ee$	3.363±0.004%
$Z \rightarrow \mu\mu$	3.366±0.007%
$Z \rightarrow \tau\tau$	3.370±0.008%
$Z \rightarrow hadrons$	69.91±0.06%
$Z \rightarrow b\bar{b}$	15.12±0.05%
$H \rightarrow b\bar{b}$ ($m_H = 120$ GeV/c^2)	67.9%

Table 2.1: Listing of key branching fractions [8],[3].

If it is the W boson with which the Higgs boson is produced in association, then the final state searched for is $WH \rightarrow l\nu b\bar{b}$, where the W decays into an electron or

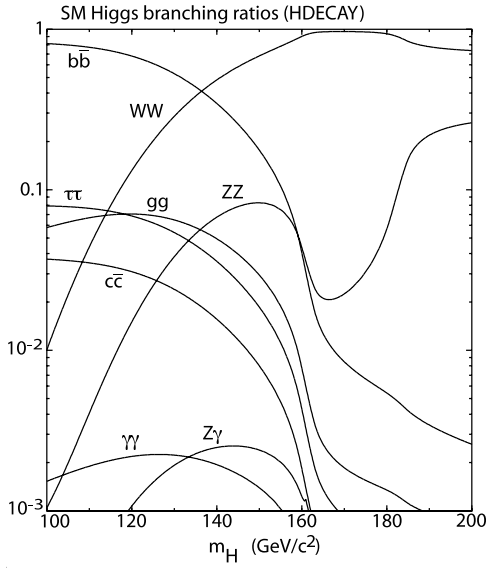


Figure 2.5: Standard Model Higgs boson branching fractions [3].

muon and a neutrino. This channel is the most sensitive channel for Higgs boson search due to the higher production cross section compared to Z associated production. The decay of W into a tau which then decays into an electron or muon is also included, but the hadronic decay modes are not. The inclusion of using the hadronic decay modes of tau is an area of active development. Including only the electron or muon and the leptonic decays of tau, approximately 30 events ($m_H = 120$ GeV/c^2) are expected to be produced in this channel for each fb^{-1} of integrated luminosity [8],[3]. Note that the trigger, detector acceptance, and selection efficiencies are not included.

If it is the Z boson with which the Higgs boson is produced in association, then the final state searched for is $ZH \rightarrow llb\bar{b}$, where the Z decays into a pair of electrons or muons or taus (including the leptonic decays of the tau only). This provides the cleanest signature, since the background processes rarely produce lepton pairs. The decays of the Z boson into a pair of taus are not considered if either tau then decays hadronically, since their identification is more difficult. Approximately 5

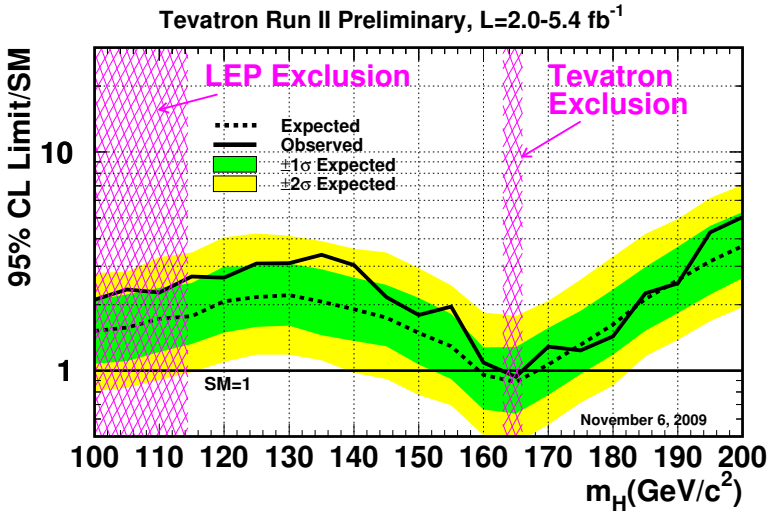


Figure 2.6: The combined Tevatron Standard Model Higgs boson limit as a function of its mass.

events ($m_H = 120$ GeV/c^2) are expected to be produced in this channel for each fb^{-1} of integrated luminosity [8],[3]. Note again that the trigger, detector acceptance, and selection efficiencies are not included.

But be it either the W or Z boson with which the Higgs boson is produced in association, sometimes the leptonic decay products of the W or Z eludes direct identification. Sometimes they fail identification criteria or escapes detection by passing through inactive regions of the detector, or for the Z boson decaying via $Z \rightarrow \nu\nu$ and escaping detection in the face of active regions as well. In these cases we can detect their presence indirectly, from a large momentum imbalance in the transverse plane (see chapter 4). Approximately 30 events ($m_H = 120$ GeV/c^2) are expected to be produced in this channel for each fb^{-1} of integrated luminosity [8],[3]. Note again that the trigger, detector acceptance, and selection efficiencies are not included.

The search with the most sensitivity $WH \rightarrow l\nu b\bar{b}$, has a related search: instead of the Higgs boson decaying into a pair of bottom quarks it is the Z boson decaying into a pair of bottom quarks ($WZ \rightarrow l\nu b\bar{b}$). This related WZ search is a proving ground for the Higgs boson search because unlike the Higgs, which has an unknown mass and may not even exist, WZ is an established SM process with a known production cross-section. In addition, the Z boson is predicted to be more abundant than the Higgs boson in the same production and decay modes.

2.4 Search for the Production of the Standard Model Z Boson in Association with $W^{\pm}$ Boson in Semi-leptonic Final States

The search for the Z boson produced in association with a W boson (Fig. 1.2) is similar to the WH search, in that both the Z boson and Higgs boson can decay into a pair of bottom quarks. The Z boson has a 15.1% decay branching fraction into a pair of bottom quarks [8], and the Higgs boson has, in the mass range of interest, an approximately 40-80% decay branching fraction into a pair of bottom quarks [3]. Overall the Z boson is predicted to be five times more abundant (Table 2.2) in the same production and decay modes than a Higgs boson at $m_H = 120$ GeV/c^2, but the different kinematics arising from the Z boson having a different mass reduces our sensitivity gain.

The decay products of the Z boson will have lower energies since it has a lower mass than the Higgs boson, which causes a reduction in signal acceptance and identification efficiency. Its lower mass also places its reconstructed mass in a kinematic region where there are more backgrounds, further reducing our sensitivity. However, since the CDF detector is able to reconstruct the mass of the parent particle with a

resolution error proportional to the reconstructed mass itself [9], the sensitivity reduces a little less than it would otherwise. The total impact of these various effects is enumerated in Table 2.3: the expected gain in sensitivity for WZ compared to WH ($m_H = 120$ GeV/c^2) is a factor of 2.3 instead of the naive factor of five from production and decay branching ratios alone.

	Cross Section
Theoretical WZ	3.96 pb
Theoretical WH ($m_H = 120$ GeV/c^2)	0.16 pb
Predicted $WZ \to e\nu b\bar{b}$ and $WZ \to \mu\nu b\bar{b}$	0.125 pb
Predicted $WH \to e\nu b\bar{b}$ and $WH \to \mu\nu b\bar{b}$	0.024 pb

Table 2.2: Listing of key cross sections [8],[44],[3].

	WZ Limit Improvement over WH
XSection x BR(Z→bb vs H→bb)	5x
Acceptance (All ID cuts except b-tagging)	0.6x
b-tagging	0.9x
Dijet mass Background (twice at 91 GeV/c^2)	sqrt(0.5)=0.7x
Dijet mass Resolution ($\frac{\sigma(m_{bb})}{m_{bb}}$ is constant [9])	sqrt(120/91)=1.1x
Include WZ→lvcc contribution	1.1x
Total Improvement Expected	**2.3x**

Table 2.3: Sensitivity gain of WZ over WH ($m_H = 120$ GeV/c^2), showing that even as the cross section times branching ratio for WZ is a factor of 5 times larger, the different kinematics reduces the sensitivity gain to a factor of 2.3 times.

Although the branching fraction for $WZ \to l\nu ll$ is smaller than $WZ \to l\nu b\bar{b}$ (6.8% for $Z \to ll$ vs 15.1% for $Z \to b\bar{b}$), the much smaller background in the fully leptonic final state makes it easier to achieve observation. The production of WZ has been observed in the fully leptonic final states ($WZ \to l\nu ll$) already, and the cross section (not including branching fractions) has been measured to be $4.3^{+1.3}_{-1.1}$ pb [10], [11]. Similarly, the production of WW has also been observed in the

fully leptonic final states ($WW \to l\nu l\nu$) already, and the cross section (not including branching fractions) has been measured to be $12.1^{+1.8}_{-1.7}$ pb [12].

It is of interest to note that the production of WW + WZ has been observed and the cross section measured to be 16.0 ± 3.3 pb [13] in the semileptonic final state $WW + WZ \to l\nu jj$. One could imagine that we can distinguish between the W and the Z using their different masses, but due to the limited jet energy resolution (see section 4.3), the two processes are virtually indistinguishable in this way.

Therefore, to observe the production of WZ alone, we must use another distinguishing feature in the hadronic decays of W and Z bosons other than their mass. One feature is that the Z boson can decay into bottom quarks while the W boson has no decay modes involving any bottom quarks.

The search for WZ with the Z boson decaying hadronically into two bottom jets ($WZ \to l\nu b\bar{b}$) is even more challenging than the $WW + WZ \to l\nu jj$ search, since it does not benefit from either the larger production cross section for WW or the larger branching fraction for $Z \to jj$ (69.9%) compared to $Z \to b\bar{b}$ (15.1%), and in addition suffers from having to identify the jets from bottom quarks. This analysis represents the first time this search has been performed at CDF. A central part of this analysis is the calibration of an improved algorithm for identification of jets from bottom quarks in order to increase the selected number of $WZ \to l\nu b\bar{b}$ events.

Chapter 3

Experimental Apparatus

Rare particles such as the top quark and the Higgs boson are produced with proton and anti-proton collisions at a center-of-mass energy of 1.96 TeV, at two designated locations (CDF and DZERO) within the Tevatron [14], located at the Fermi National Accelerator Laboratory (FNAL, or Fermilab) in Batavia, Illinois, USA. The CDF detector, located at one of these collision points, is used to record the interactions resulting from the proton and anti-proton collisions at a center-of-mass energy of 1.96 TeV for studies.

This chapter discusses the accelerator complex that produces the 0.98 TeV protons and anti-protons, and the CDF detector that records their interactions.

3.1 The Accelerator

At designated locations within the Tevatron the proton and anti-proton beams collide at a center-of-mass energy of 1.96 TeV. These beams are generated in the accelerator complex (Fig. 3.1), by a series of particle accelerators leading to the Tevatron, which is a synchrotron accelerating proton and anti-proton beams in opposite directions to an energy of 0.98 TeV.

The protons in the proton beam come from a bottle of hydrogen gas, which is

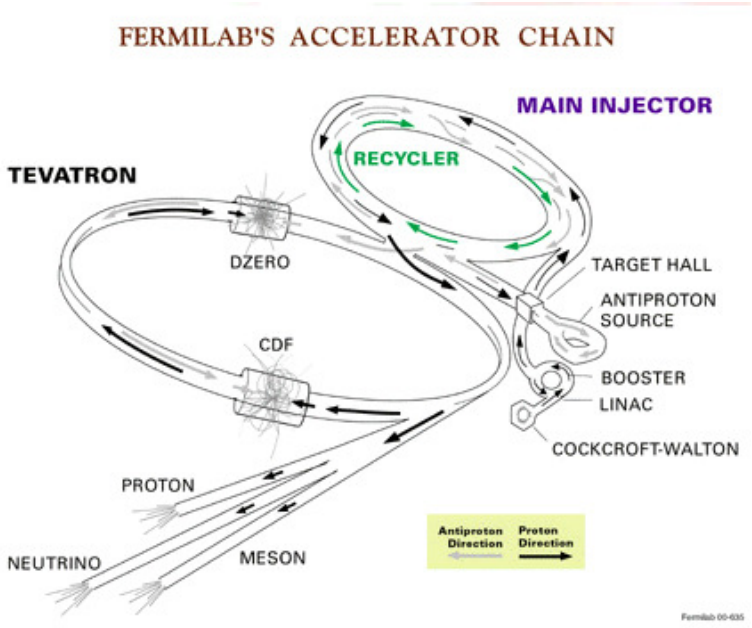


Figure 3.1: The Fermilab accelerator chain.

replaced about every six months. The gas is fed into an ionization chamber lined with molybdenym, where a magnetron is used to generate a plasma near the metal surface to form H^- ions which continuously pass through an electrostatic potential of 25 keV to become a beam. This beam is accelerated to 750 keV in the Cockroft-Walton accelerator, which is an electrostatic accelerating column. The continuous H^- beam is then directed through an Electrostatic Chopper and Buncher to clip off 42 μs long bunches to be fed into the Linac.

The bunched H^- beam is injected into the 150 m long linear accelerator (Linac), consisting of a drift-tube linac to accelerate the H^- beam to 116 MeV, and a side-coupled cavity linac to accelerate the H^- beam to 401 MeV. The beam is then directed to a thin carbon foil in the Booster, where the electrons are stripped off the passing ions, transforming the H^- beam into a proton beam. The Booster is a 150 m diameter circular accelerator that uses magnets to bend the proton beam such that it travels in an approximately circular path. The beam travels around the Booster,

each time getting accelerated slightly, finally increasing the beam energy to 8 GeV.

The proton beam from the Booster is then injected into the Main Injector. The Main Injector has a 3320 m circumference, and it accelerates the proton beam to 120 GeV for the anti-proton production or to 150 GeV for injection into the Tevatron. The Main Injector can also accelerate the anti-proton beam to 150 GeV for injection into the Tevatron.

To produce the anti-protons, 120 GeV protons from the Main Injector strike a nickel target producing numerous types of secondary particles. Then they are passed through a solid rod of lithium, called a Lithium Lens, of about 1 cm radius and carrying along an electric current to produce a large radial magnetic gradient of 1000 T/m, which focuses these particles into a beam. This beam travels through a dipole magnetic field, and the trajectories of the particles will curve by different amounts according to their charge to mass ratio. Only the anti-protons will have a trajectory directed downstream, the rest will head to the beam dump.

The anti-proton beam is then sent to the Debuncher where it is conditioned, meaning that the beam momentum spread and beam width are reduced. The conditioned beam is then sent to the Accumulator, which is a ring used to store the anti-protons. There is a second ring, the Recycler ring, which is used to store the anti-protons when the Accumulator fills up. The amount of anti-protons needed is decided by the run coordinators daily. A collider physics run usually require more than 10^{12} anti-protons, and presently it takes less than one day to accumulate enough for a run.

The 150 GeV proton and anti-proton beams are sent from the Main Injector in 3 trains of protons and anti-protons, each train separated by 2.6 μs. There are 12 bunches per train, each bunch separated by 396 ns. These beams are sent to the Tevatron, a synchrotron with a ring radius of 1 km, which accelerates them to 980 GeV. The proton beam has typically 10^{13} particles, and the anti-proton beam 2×10^{12} particles. They are made to collide at designated locations on the ring,

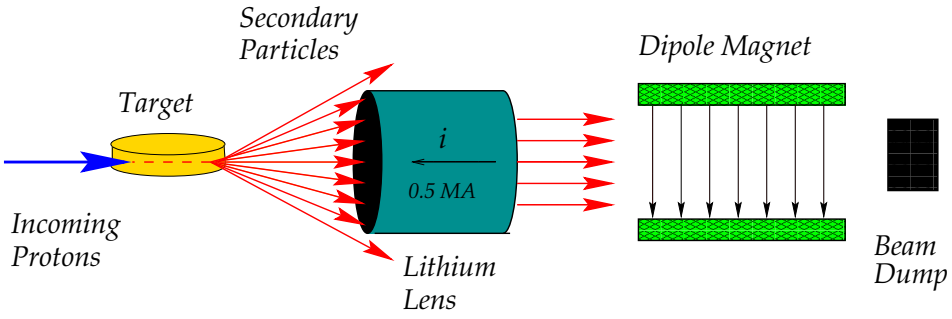


Figure 3.2: Anti-proton production at the target station. The dipole magnets directs the anti-protons to the Debuncher, and sends the rest to the beam dump.

called the B0 and D0 points where the CDF and D0 (sometimes DZERO) detectors are located, to produce rare particles such as the top quark and the Higgs boson at a center-of-mass energy of 1.96 TeV.

These rare particles are produced at the rate of integrated luminosity ($\mathcal{L}_{int}$) $\times$ cross section (σ). To study a rare process with a very small cross section, a high luminosity is needed. The luminosity is dependent on the collision frequency (f), the number of the protons and anti-protons in each beam (N_P, $N_{\overline{p}}$) and the average transverse beam size (A) as:

$$\mathcal{L} = f\frac{N_p N_{\overline{p}}}{4\pi A}. \tag{3.1}$$

To date, the highest instantaneous luminosity achieved is 3.5×10^{32} cm^{-2}s^{-1}, and the luminosity integrated during the past 9 years of Tevatron running since 2001 is 8 fb^{-1}, delivered at a center of mass energy of 1.96 TeV.

3.2 The CDF Detector

The CDF detector [15], located at one of the two collision points of the Tevatron, is used to record the interactions resulting from the proton and anti-proton collisions at

a center-of-mass energy of 1.96 TeV for studies. It is a detector designed to measure the energy, momentum and the identity of particles produced in Tevatron collisions, combining charged particle tracking with projective calorimetry and muon detection. A cross-sectional view of half the detector is shown in Fig. 3.3.

Particles produced in the collisions first pass through the tracking detectors, where the momentum of charged particles is measured from their curvature, then the calorimeters, where the energy of electrons and hadrons are deposited, then the muon detectors, which note the passage of any charged particle that escapes from the calorimeters. The combined responses of the various detectors allow for the identification of the different particles (Fig. 3.4).

The tracking systems are contained in a superconducting solenoid 1.5 m in radius and 4.8 m in length, which generates a 1.4 T magnetic field parallel to the beam axis. Calorimetry and muon systems are outside the solenoid.

3.2.1 The CDF Coordinate System

The CDF detector uses a coordinate system where the $+z$ is in the direction of the proton beam, $+x$ is towards the outside of the Tevatron ring and $+y$ is the vertical direction pointing upwards. The x-y plane is called the transverse plane. A cylindrical coordinate system is frequently used: r is the radial distance from the beam line, θ is the polar angle from the beam line ($\theta = 0$ in the direction of $+z$ and $\theta = 90^\circ$ perpendicular to the beam), and ϕ is the azimuthal angle ($\phi = 0$ in the direction of $+x$ and $\phi = 180^\circ$ in the direction of $-x$). In addition, the quantity pseudorapidity η is frequently used, where $\eta = -\ln(\tan\frac{\theta}{2})$. The polar and azimuthal angular separations are often expressed as $\Delta\eta$ and $\Delta\phi$. A related quantity $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ is useful in comparing the angular separation within different parts of the detector. The parts of the detector with $|\eta|$ near zero is often called the central region, and $|\eta|$ greater than two or three is often called the forward region. The detector is approximately symmetric in η and in the azimuthal angle ϕ.

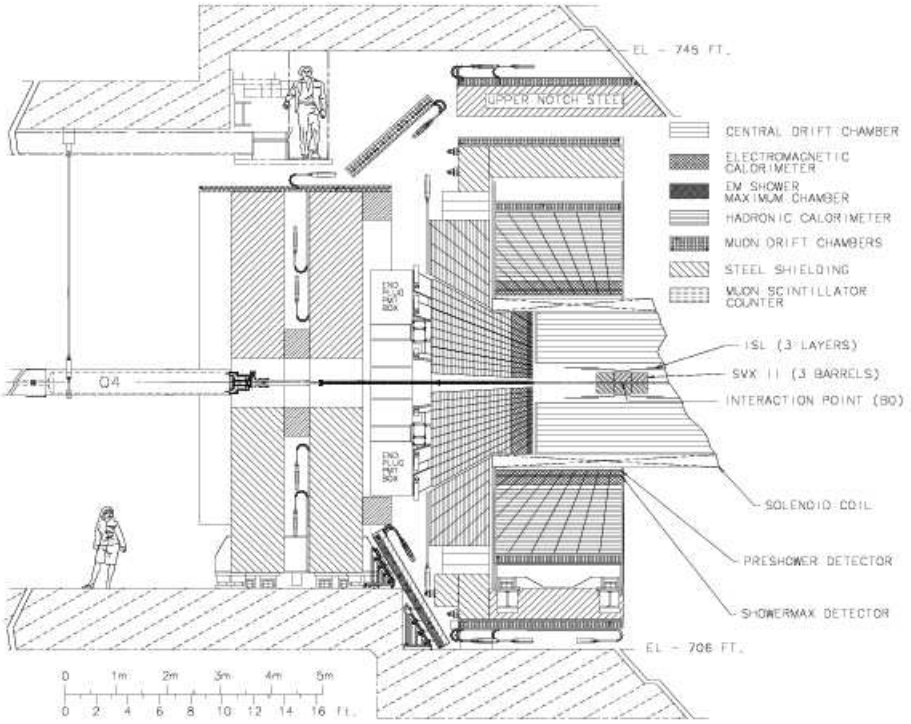


Figure 3.3: A cross-sectional view of the CDF detector.

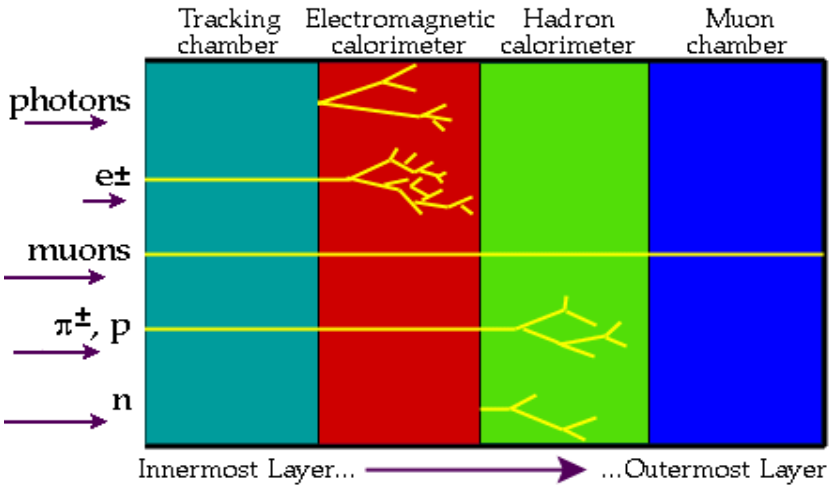


Figure 3.4: A schematic showing the behaviors of various particles passing through different detector subsystems.

3.2.2 Tracking System

The CDF Tracking System is designed to reconstruct the trajectories of charged particles. The system, consisting of the Silicon Detectors and the Central Outer Tracker (COT), is contained in a superconducting solenoid which generates a 1.4 T magnetic field parallel to the beam axis, uniform to within 0.1% throughout the entire tracking volume.

Silicon Detectors

The CDF silicon detectors are designed to perform high-precision tracking, which is very important for the identification of long-lived particles like B hadrons. These B hadrons can travel several millimeters before decaying into several particles, and the precise reconstruction of the charged particles allows the extrapolation of their trajectories to find a common decay origin (secondary vertex) that is well displaced from the location of proton anti-proton collision (primary vertex).

The CDF silicon detectors is composed of silicon microstrip sensors that can be divided into three sub-detectors, the Layer 00 (L00), Silicon Vertex Detector (SVX II), and Intermediate Silicon Layers (ISL). L00, located on the beam pipe at radius $r = 1.35$ cm, covers $|\eta| < 4.0$. SVX II, located outside of the L00 at the radius from $r = 2.1$ cm to 17.3 cm, consists of 5 layers and covers $|\eta| < 2.0$. Each layer has two types of strips, one along the z-axis ("axial" strips), the other tilted ("stereo" strips), allowing for a 3-dimensional track reconstruction. The ISL, located after the SVX II consists of one layer at $r = 22$ cm and $|\eta| < 1.0$, and two layers from $r = 20$ cm to $r = 28$ cm and $1.0 < |\eta| < 2.0$. The r-z view of the silicon detectors is show in Fig. 3.5.

When charged particles pass through a silicon sensor, they produce electron-hole pairs. We apply an electric field across the body of the silicon, causing electrons to drift towards the anode and holes towards the cathode. The charge generated by the incident particle is thus collected by the electrodes to give information about

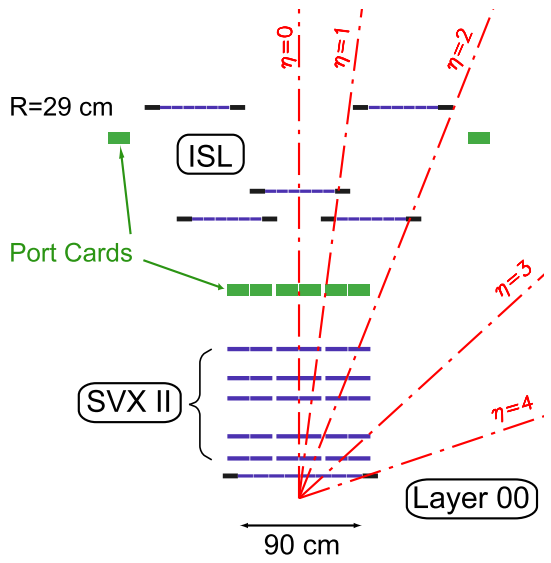


Figure 3.5: The r-z view of the silicon detectors. The Port Cards regulate power on, read out from, and control the silicon sensors.

the amount of energy deposited, but more importantly it tells us that the particle has traversed that space within the detector. The track of the particle can then be reconstructed using the information from many sensors. The CDF silicon detectors has a track resolution of 10 microns.

Central Outer Tracker

The Central Outer Tracker (COT) is a cylindrical drift chamber located outside of the silicon detectors at a radius from $r = 40$ cm to 137 cm and covers $|\eta| < 2.0$. It consists of 8 superlayers, 4 parallel to the z-axis ("axial" superlayers) and 4 with an angle $\pm 2°$ with respect to the z-axis ("stereo" superlayers). Superlayers are made of a varying number of cells (superlayer 1 has 168 cells and superlayer 8 has 480 cells). Each cell consists of a field sheet and a wire plane with alternating sense wires and field wires. Each cell has a total of 12 sense wires. The chamber is filled with a (50 : 50) mixture of Argon and Ethane gas.

A charged particle passing through the COT interacts and ionizes the gas mixture. Positively charged ions and free electrons are created. If an electric field is applied in the gas volume, electrons will drift toward the anode. In the high-field region near the anode, the electron ionizes other atoms and produces an avalanche, which creates a large signal on the wire. The electrons are collected on the anode wire giving an indication of the passage of a particle near that volume. Electrons drift faster than ions due to their lower mass. The electron drift velocity depends on the electric field gradient and the properties of the gas molecules, and for the COT it is ~ 50 μm/s. The signals deposited by a particle are collected in less than 200 ns.

The COT is useful in measuring the momentum of the charged particles. Since the COT is placed in the 1.4 T magnetic field, the charged particles travel in a helix with the radius $r = \frac{p_T}{|q|B}$, where $p_T = \sqrt{p_x^2 + p_y^2}$ (transverse momentum), q is the particle charge and B is the magnetic field. By reconstructing the track's curvature in the r-ϕ plane, p_T can be determined.

3.2.3 Time of Flight System

The Time of Flight system (TOF) is a scintillator detector positioned outside the COT at $r = 1.4$ m that measures the time of arrival of a particle with respect to the collision time. It is useful for particle identification, especially for distinguishing $K^{\pm}$ from $\pi^{\pm}$, which is very important in studying the bottom hadrons. The mass of particles can be determined by $m = \frac{p}{c}\sqrt{\frac{c^2t^2}{L^2} - 1}$, where p is the momentum, L is the path length and t is the time of flight.

3.2.4 Calorimeter System

The Calorimeter System, located outside the Tracking System, measures the energy and direction of the particles going out of the tracking system. Covering a pseudorapidity range $|\eta| < 3.6$, it has electromagnetic and hadronic compartments for measuring the energy of electrons/photons and hadrons, respectively.

The Calorimeter System consists of the Central Electromagnetic Calorimeter(CEM) ($|\eta| < 1.1$) composed of alternating lead-scintillator layers, followed by the Central Hadronic Calorimeter (CHA) ($|\eta| < 0.9$) and the Endwall Hadronic Calorimeter (WHA) ($0.8 < |\eta| < 1.2$) both composed of alternating iron-scintillator layers. To extend the coverage of the calorimeter to a larger pseudorapidity, the Plug Electromagnetic Calorimeter (PEM) is placed outside of the Central Calorimeter, covering $1.1 < |\eta| < 3.6$, and the Plug Hadronic Calorimeter (PHA) follows the PEM, covering $1.2 < |\eta| < 3.6$.

The Calorimeter System is segmented into towers. The Central Calorimeter is segmented into towers each 15° in azimuthal angle and 0.11 in pseudorapidity. The 4 most forward towers slices are segmented in 15° similar to the central calorimeters, while the remaining 8 towers each have a finer segmentation of 7.5°.

When the particles (e.g. photon, electron, hadron) pass through the calorimeter, they interact with the material in the calorimeter (lead in the CEM or iron in the CHA) by electromagnetic interactions (in the CEM) or hadronic interactions (in the CHA). These interactions produce sprays of particles whose energy profile and total energy provide information about the kinematics and identity of the primary particle. These sprays of particles are called showers.

The depth of the showers in the calorimeter depends on the the radiation (interaction) length of the materials used in the electromagnetic (hadronic) calorimeter. The shorter these lengths the more compact the calorimeter can be. At CDF the showers are developed in the lead (iron), chosen for their short radiation (nuclear interaction) lengths.

The energy profile and total energy of the showers are measured by the scintillators sandwiched between the lead or iron. When the charged particles from the showers pass through the scintillator, the scintillator emits photons. The produced photons are subsequently coupled via light guides to be collected by photomultiplier tubes (PMTs). By adding up the signal collected in the surrounding photomultiplier

tubes, we can measure the energy of the charged particles in the showers, and infer the kinematics and identity of the primary particle.

3.2.5 Muon System

The Muon System is the outermost subdetector, located outside the Calorimeter System. Because muons have a long life time (2.2 μs) and interact much less with materials than electrons/photons/hadrons, they can often travel a long distance unimpeded, passing through the tracking and calorimetery and reach the muon detectors. Muon detection has low background because most other particles would have been absorbed before reaching the muon detectors. The Muon System at CDF consists of four sub-detectors: the Central Muon Detector (CMU) ($|\eta| < 0.6$), the Central Muon Upgrade (CMP) ($|\eta| < 0.6$), the Central Muon Extension (CMX) ($0.6 < |\eta| < 1.0$), and the Barrel Muon Detector (BMU) ($1.0 < |\eta| < 1.5$).

The muon system uses a drift chamber design similar to the COT, but has only a single wire in each drift chamber. When a muon interacts with the gas in the drift chamber, it produces electrons that will drift toward the anode wire, and be collected by it. The muon system provides very strong indicators that the particles reaching it are muons.

3.2.6 Luminosity System

The Cherenkov Luminosity Counter (CLC) detector [16] is responsible for measuring the luminosity at CDF. It is a set of gaseous Cherenkov detectors filled with isobutane. They are located at a very forward region, at $3.7 < |\eta| < 4.7$. The luminosity $\mathcal{L}$ is inferred by counting the number of interactions per bunch crossing (μ), and using the known rate of bunch crossings in the Tevatron ($f_{BC} = 1.7$ MHz) and the measured cross section for inelastic proton anti-proton scattering ($\sigma_{in} = 60\ mb$), given by:

$$\mathcal{L} = \frac{\mu \cdot f_{BC}}{\sigma_{in}}. \tag{3.2}$$

3.2.7 Trigger

The Tevatron produces 1.7 million collisions between bunches of proton and anti-proton every second. The data recorded by the CDF detector for each bunch crossing is called an event. Each collision event as recorded by CDF requires 250 kB to record. The CDF trigger system is designed to have near zero deadtime, to quickly decide which events to record, in order to ignore most of the collisions and record only the ones most likely to yield interesting results, since we are able to record at most at a rate of $\sim$ 100 MB per second.

The three Level trigger system at CDF is able to relieve the strain on the data storage system by reducing the 1.7 million events per second to a manageable rate of several hundreds of potentially interesting events per second. Decisions are made at each level, reconstructing an increasingly number of physics objects.

Level 1 (L1) is a buffered synchronous system with an event read in every beam crossing. It is a set of integrated circuit boards that uses the information from a subset of the readout electronics of the sub-detectors and makes a decision by simply counting the physics objects, such as electron, muon, photon, or missing transverse energy. This trigger can decide whether to record the event within 5 μs. This is able to reduce the rate of potentially interesting events to 30000 per second.

Level 2 (L2) is a buffered asynchronous system which processes events that have passed L1 in a time ordered fashion. It is a set of integrated circuit boards that further identifies the features of the events passing L1. The events passing L1 will have their tracks processed for displacement from the proton anti-proton collision point, and their calorimeter towers processed for a jet candidate. Then a decision on whether to record the event will be made. Since the time for further reconstructing each event takes up to 35 μs, the events accepted by L1 are stored in one out of four

L2 buffers. This is able to reduce the rate of potentially interesting events to several thousand per second.

Level 3 (L3) is a dedicated computing farm. If an event is accepted by L2, the information from the entire detector is read out. L3 uses additional algorithms such as tracking, calorimeter, muon stub finding algorithms to further reconstruct the physics objects, and select the events with certain characteristics. The reconstruction and selection takes on average one second per event. This reduces the rate of events to be recorded to a manageable amount of several hundred per second.

3.2.8 Production Algorithms

The events passing the L3 trigger are then recorded to permanent storage media. To determine the interactions occurring at the collisions accurately, we need to reconstruct the event in a more precise way compared to the trigger level. At CDF, there are two stages for reconstructing the events off-line. First, the event information from all the subdetectors is delivered to the production algorithms, which reconstruct the basic objects such as electromagnetic clusters, muon candidates, and jet candidates. Then the basic objects reconstructed by the production algorithms need to pass some identification quality criteria, to be separated into jets and lepton candidates.

Chapter 4

Physics Object Identification

This search for a Standard Model Z boson production in association with a W boson is a signature-based analysis, where the W boson decays into a lepton (l) and a neutrino (ν) and the Z boson decays into two b quarks ($b\bar{b}$) that hadronize to form two b-jets. This chapter discusses the identification of the lepton, neutrino, and the jets. The identification that the jets came from b quarks is discussed in the next chapter.

4.1 Lepton Identification

For this analysis we seek to use the leptons identified in the detector with the highest signal-to-noise ratio. For CDF, the most well-identified leptons are the electrons and muons identified with the central region of the detector.

4.1.1 Electron

Electron reconstruction starts by matching a cluster of energy in the CEM calorimeter to a COT track [17]. To obtain a purer sample of electrons, the reconstructed electrons are required to pass the standard CDF electron identification cuts [18].

4.1.2 Muon

Muons are identified using the muon detectors. In this analysis, we require muons from $|\eta| \leq 0.6$ to have hits in both the CMU and CMP, and muons from $0.6 < |\eta| < 1.0$ to have CMX hits forming a track segment consistent with an extrapolated COT track. To obtain a purer sample of muons, the reconstructed muons are required to pass the standard CDF muon identification cuts [19].

4.2 Neutrinos and Missing Transverse Energy

Neutrinos interact weakly with every type of material, and as a result they often depart from the detector unseen.

We can infer the component of the neutrino's momentum that is transverse to the proton beam direction from the imbalance in the measurements of the energy and momenta of all the other particles. Since the incoming proton and anti-proton have negligible transverse momenta, conservation of momentum implies that the sum of the transverse momenta of the outgoing particles should be zero as well. The neutrino's transverse momentum then points in the opposite direction to the summed transverse momentum of all the other particles.

Convention defines missing transverse energy as the magnitude of the missing transverse momentum. Note that the neutrino's longitudinal momentum cannot be inferred in this way, since the summed longitudinal momentum of the incoming constituents in the proton and anti-proton is unknown.

4.3 Jet Identification

When quarks are produced from high-energy interactions, they will hadronize into sprays of particles called jets. A jet manifests itself as clusters of energy in the electromagnetic and hadronic calorimeters. Jets are identified using the JetClu

algorithm [20], and defined in this analysis as calorimeter clusters within a cone of radius $\Delta R = 0.4$ (section 3.2.1). The energy of the jet is corrected for losses in the gaps, multiple interactions, and detector non-linearity using the JetCorr algorithm [21] at level-5. This analysis uses jets from $|\eta| < 2.0$ that have $E_T > 20$ GeV, which are referred to in CDF as tight jets. In this region, the jet energy resolution in the transverse plane is approximately $\frac{\sigma_{E_T}}{E_T} = \frac{50\%}{\sqrt{E_T}} \oplus 3\%$.

The next chapter discusses how to identify jets from bottom hadrons apart from jets from charm hadrons and light flavor hadrons (from gluons and up, down, and strange quarks).

Chapter 5

Bottom Jet Identification

At CDF it is often necessary to identify jets that come from b quarks. This is called b-tagging. The identification of b-jets is an essential component for measurements involving the top quark and searches for a low-mass Higgs boson. The signatures of these interesting signal processes all contain b-jets; the ability to discriminate b-jets from the overwhelming inclusive jet background helps to increase the purity of the selected event sample. For a signature like WH or $WZ \to l\nu b\bar{b}$, the most useful sample has two b-tagged jets, the number of such signal events is proportional to the square of the b-tag efficiency.

Two properties that help in identifying b-jets apart from the other flavored jets are that bottom quarks have a lifetime of ~ 1.5 ps, and that bottom hadrons have mass of ~ 5 GeV/c^2. For example, a b-jet with 50 GeV will have a γ boost of ~ 10, then the lifetime of ~ 1 ps allows it to travel on average 4.5 mm before decaying. And the fact that it has a mass of ~ 5 GeV/c^2 means that it will have numerous charged decay products, which will leave numerous tracks. These tracks will be displaced from the primary vertex, and can be used to reconstruct a secondary decay vertex.

At CDF there is an algorithm that calculates the probability of a jet containing a b quark using the displacement of its tracks from the primary vertex called JetProb (short for Jet Probability), and an algorithm that identifies the jet as containing a b

quark by reconstructing its secondary decay vertex called SecVtx. These algorithms judges whether the jet contains a b quark by requiring that the probability passes a minimum threshold (cut).

Another property that we use is that 40% of b-jets will have an electron or muon within it, resulting from the weak decay of the b quark. And by identifying the electron or muon within the jet we can infer that it was a b-jet. At CDF there is an algorithm that looks for a muon within the b-jet to differentiate it from other flavors called SLTμ (short for Soft Lepton Tagger, muon).

Each of the existing CDF b-tag algorithms uses only one property of the b quark for identification. This chapter discusses the work done to improve upon these existing algorithms, using all the above mentioned properties of the b quark together, instead of one at a time.

5.1 Roma Neural Network

The Roma neural network (RomaNN) is a new algorithm [22] designed by the INFN-Roma group to separate b-jets from the other flavors. In its original form, it does not judge whether the jet contains a b quark but provides only a indicator of how consistent the jet is with coming from a b-quark. Because it did well in initial testing, I adapted it to perform b-tagging (binary mode), and commissioned it in order for it to become another standard CDF b-tagging algorithm.

The RomaNN consists of three Neural Networks, for vertex identification (VerticesNN), displaced track identification (TracksNN), and flavor identification. The Neural Networks are based on the commercial NeuroBayes package [23]. Fig. 5.1 gives an overview of the information flow within RomaNN.

The RomaNN first identifies the vertices produced by the decay of heavy flavor hadrons using VerticesNN, which is based on the CTVMFT package [24]. The tracks not associated to a vertex are then fed into a second Neural Network, TracksNN,

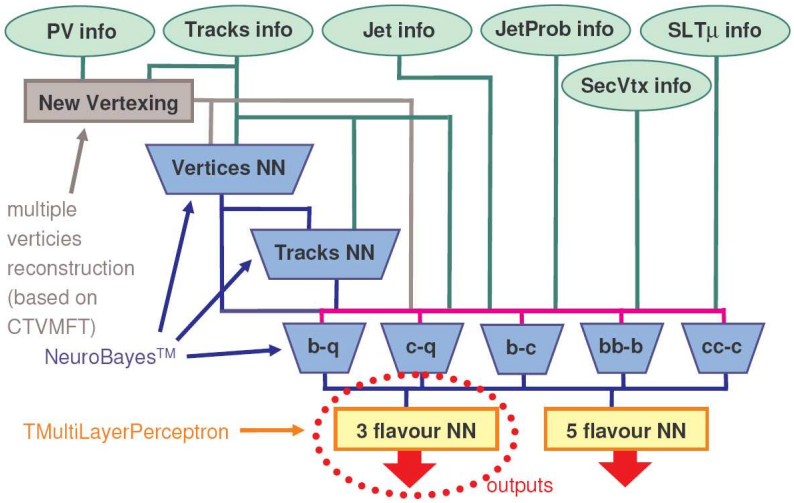


Figure 5.1: Flow chart of the RomaNN. The output from the "3 flavour NN" is used to separate b-jets from the other flavors.

to distinguish between tracks produced by heavy flavor hadron decay and prompt tracks. The information from these two Neural Networks is combined with information from existing CDF b-tag tools to form a third Neural Network. Fig. 5.2 is the output of RomaNN for simulated jets passing W+2jets selection (described in section 7.2), showing that b-jets have output values higher than charm or light-jets. The input variables used in vertex, track, and flavor identification are listed for reference in Tables 5.1, 5.2, and 5.3 respectively.

RomaNN is designed to provide a per-jet output value in the range from -1 to +1; a value near +1 indicates that the jet is consistent with coming from a b-quark, and values away from +1 indicate that the jet is more consistent with originating from some other flavor. The fact that there is a negligible amount of jets near -1 means that very few jets are definitely not coming from a b-quark. This is because even though b-quarks have a longer lifetime compared to light quarks, there is a significant number of b-quarks decaying quickly enough such that their tracks are not displaced enough to be distinguishable from those coming from light quark jets.

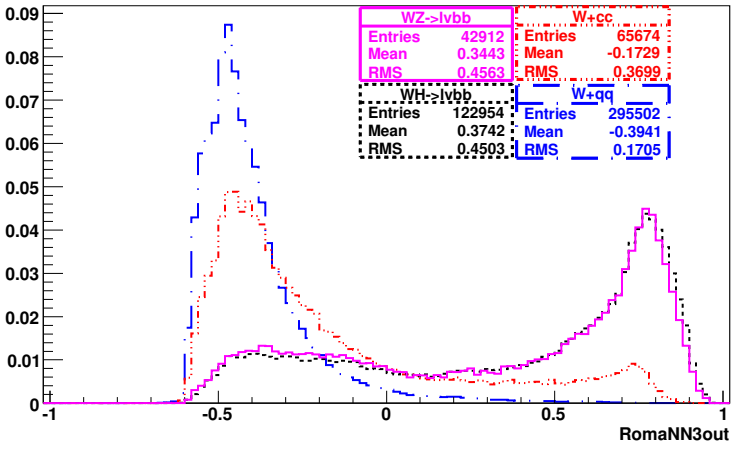


Figure 5.2: Output of RomaNN for simulated jets, normalized to unit area: Bottom jets from WH(120GeV) and WZ, charm jets from W+cc and light flavor jets from W+qq, showing that b-jets have output values higher than charm or light-jets.

Each analysis can choose its own minimum threshold (cut) on the RomaNN output value (binary mode RomaNN), giving an opportunity to customize the level of b purity. Three operating points are used, RomaNN3out>0.40 (TightRomaNN), RomaNN3out>0.20 (LooseRomaNN), and RomaNN3out>0.00 (UltraLooseRomaNN). These first two operating points were chosen because they have approximately the same misidentification rates of the existing operating points of the SecVtx (see Fig.5.37). The UltraLoose operating point is the cut with double the misidentification rate of the Loose operating point.

The b purity is dependent on both the efficiency of identification and the probability of misidentification. Therefore to accurately optimize for the cut on the RomaNN, these two quantities must be properly calibrated. The following sections describe their measurements and how our simulations are calibrated.

Rank	*Variable*
1	Transverse separation between secondary and primary vertex
2	Significance of the d_0 of the 2nd Most Displaced Track in Vertex
3	Angle Between Vertex Momentum Vector and Vertex Displacement Vector
4	Significance of the 3D separation between secondary and primary vertex
5	Invariant Mass of Vertex

Table 5.1: Input variables to the vertex NN, ranked in order of significance by NeuroBayes [23].

Rank	*Variable*
1	track d_0 significance
2	D/L
3	vertex NN
4	α
5	D

Table 5.2: Input variables to the track NN, ranked in order of significance by NeuroBayes [23]. These variables are defined in Fig. 5.3.

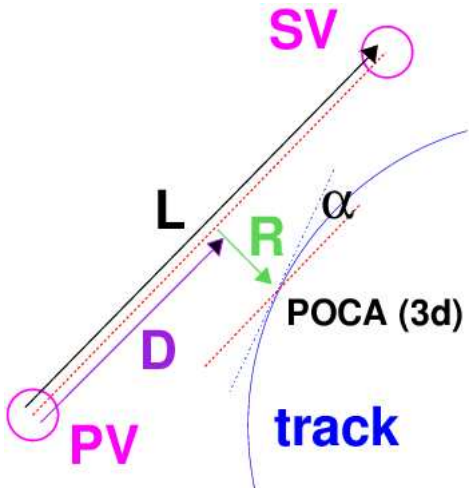


Figure 5.3: Definition of track observables. PV is primary vertex. SV is secondary vertex. L is the 3D separation between PV and SV. POCA is point of closest approach of the track to the line segment connecting PV to SV. R is the perpendicular to L that ends at POCA. D is the separation from PV to the base of R. α is the angle between the tangent at POCA and the line on POCA parallel to L.

Rank	*Variable*
1	JetProbability (when calculated with at least 2 tracks)
2	Transverse separation between best secondary vertex and primary vertex
3	LooseSecVtx Tag
4	Number of Muons Identified By SLTμ
5	$\Sigma_{selected\ tracks} p_T / \Sigma_{all\ jet\ tracks} p_T$ (scalar sums)
6	Significance of the 3D separation between best secondary and primary vertex
7	Invariant Mass of Best Vertex
8	Invariant Mass of Selected Tracks (both Vertexed and Unvertexed)
9	Number Of Selected Tracks
10	3D separation between best secondary and primary vertex
11	Number Of Good Tracks In Jet
12	$\Sigma_{all\ jet\ tracks} p_T$ (scalar sums)
13	3D separation between second best secondary and primary vertex
14	p_T of (Highest p_T) Muon With Respect To Jet Axis
15	Mass of SecVtx Vertex
16	Jet E_T

Table 5.3: Input variables to the flavor NNs, ranked in order of significance by NeuroBayes [23].

5.2 Efficiency Measurement

One of the most important performance parameters of the identification of b-jets is the efficiency at which they are correctly identified. This, along with the misidentification probability, characterizes the purity of the identified b-jets.

The b-tag efficiencies are needed to estimate the yields of signal and background, which are typically estimated using Monte Carlo (MC) simulations. The simulated b-tag efficiencies have to be calibrated in data to accurately extrapolate the MC yields to those expected in the data.

The approach widely used at CDF is to measure the efficiency for tagging b-jets (Eqn. 5.4) in the data and in the simulation, and encode any mismatch in a data/MC correction factor. This correction factor is in CDF jargon called a "scale factor", and it is helpful in understanding the precision of our estimates made using simulation.

We follow this approach, and describe in this section a measurement of the data-to-MC scale factor for the binary mode RomaNN. In MC, the efficiency for tagging b-jets can be calculated simply by counting the b content of a sample from the simulation truth bank, before and after applying the b-tag. In data, we rely on the lepton p_T^{rel} technique described below to measure the efficiency for tagging b-jets.

5.2.1 Lepton p_T^{rel} Technique

In data, the technique we use to measure the efficiency for tagging b-jets is called the lepton p_T^{rel} technique. This technique relies on the fact that jets resulting from B hadron decay will on average impart its daughter particles with a higher transverse momentum relative to the jet vector, compared to the decay of a generic hadron, since B hadrons have masses higher than generic hadrons.

Since CDF can measure the momentum of leptons with higher precision than hadrons, we use the subset of jets where there is a lepton (μ or e) within the jet. These are called lepton jets. The lepton transverse momentum relative to the jet

vector (lepton p_T^{rel}) is defined in Eqn. 5.1,

$$lepton\ p_T^{rel} = \left|\overrightarrow{P}_{lepton}\right| \sqrt{1 - \left(\frac{\overrightarrow{P}_{lepton} \cdot \overrightarrow{P}_{jet}}{\left|\overrightarrow{P}_{lepton}\right|\left|\overrightarrow{P}_{jet}\right|}\right)^2} \tag{5.1}$$

where $\overrightarrow{P}_{lepton}$ is the momentum of the lepton and $\overrightarrow{P}_{jet}$ is the momentum of the lepton-jet.

The momentum of the electron within the jet is obtained from a combination of tracking and calorimetry, whereas the momentum of the muon within the jet is obtained from tracking alone.

The energy and momentum of the jet containing the electron or muon, is obtained from a combination of tracking and calorimetry, processed through the JetCorr algorithm [21]. A jet containing a muon in particular, must be corrected for the muon escaping the calorimeter with most of its momentum according to Eqn. 5.2 and 5.3. In these equations, 2.0 GeV is taken as the most likely energy deposition in the calorimeter by the muon [25]. This was not necessary for the electron, since electrons deposit all their energy in the calorimeter.

$$\overrightarrow{P}_{corr} = \overrightarrow{P}_{jet} + \left(1 - \frac{2\ \frac{GeV}{c}}{|p_\mu|}\right)\overrightarrow{P}_{\mu} \tag{5.2}$$

$$E_{T,corr} = E_{T,jet}\left(\frac{E_{jet} - 2\ GeV}{E_{jet}}\right) + p_{T,\mu} \tag{5.3}$$

The amount of b-jets in a collection of jets can be determined by using the discriminating shape of the p_T^{rel} between b and non-b-jets. Splitting the data sample into two subsamples, "tagged" and "not-tagged", we can fit their p_T^{rel} distribution to find the b content of each. This allows us to calculate the b-tagging efficiency in data from the yields of b-jets in each subsample (Eqn. 5.4).

$$\epsilon_b \ = \ \frac{N_{Tag}^b}{N_{total}^b} \ = \ \frac{N_{Tag}^b}{N_{Tag}^b + N_{NoTag}^b} \tag{5.4}$$

where N^b_{Tag} is the number of b-jets that are tagged (RomaNN3out>RomaNNcut); N^b_{NoTag} is the number of b-jets that are not tagged (RomaNN3out≤RomaNNcut); N^b_{total} is the total number of b-jets. Then we can obtain the data-to-MC scale factor by dividing the efficiency for tagging b-jets in data by the efficiency for tagging b-jets in MC (Eqn. 5.5).

$$SF = \frac{\epsilon_{data}}{\epsilon_{MC}} \tag{5.5}$$

To be sure, this is the efficiency and scale factor for identifying a b-jet with the RomaNN when the b quark has decayed semileptonically, and these are not necessarily the same as for generic jets. We make in the next section a brief justification of why semileptonically decaying jets can represent generic jets, but detailed explanations can be found in [26],[27],[33].

5.2.2 Electron vs Muon

The p^{rel}_T technique has been used in CDF since 2005 [27] to measure the efficiency for tagging b-jets, with the jets having a muon. The muon jets can extrapolate (with some suitable systematic errors applied) to the generic candidate jets for the SecVtx b-tagger because the tagger does not explicitly use the muon information [27]. However, because the RomaNN uses explicitly the number of muons as an input, jets with a muon are no longer good handles to measure the efficiency for tagging b-jets in general.

For this reason, we have developed the technique further, to use jets having an electron to measure the efficiency for tagging b-jets. The RomaNN3out distributions from b-hadronic (no muon and no electron) jets, b-jets containing an electron, and b-jets containing a muon are shown in Fig. 5.4. It is clear that generic jets and jets with electrons have similar distributions, while jets with a muon have a distribution shifted toward positive (b-like) values. Because of this bias seen in the jets with a muon, we measure the scale factor of two sub-samples of jets: muon-jets, and generic

jets as represented by electron-jets.

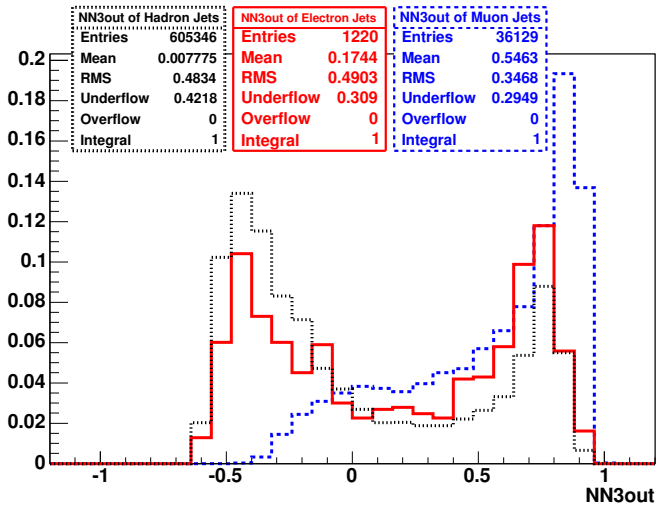


Figure 5.4: RomaNN3out distribution comparison between generic, electron, and muon, from WZ Monte Carlo with jet $|\eta| < 1.2$ and $E_T > 9$ GeV.

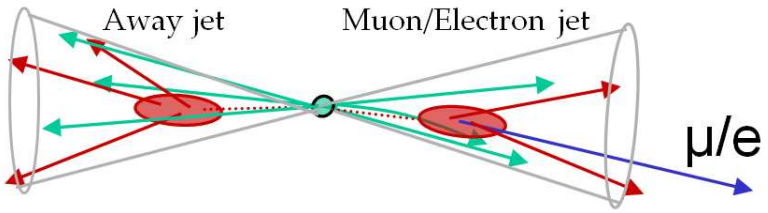


Figure 5.5: Probe-jet (muon/electron) and away-jet.

5.2.3 Event Selection

To measure efficiency for tagging b-jets, it is necessary to get a sample of jets with enriched b content. We obtain a relatively pure sample of b-jets in data by taking advantage of the fact that b-jets are typically produced in pairs. While there are many more pairs of jets from light flavor $q\overline{q}$ production, we can reduce this background by requiring one jet to pass b-jet identification (away-jet) and the other jet to contain a muon or electron (referred to as the "probe jet" in the discussion below).

To enrich the sample with b-jets, we select events with two jets, called di-jet events, having "back to back" ($|\phi_{probe} - \phi_{away}| > 2.0$ radians) jets (Fig. 5.5). We have in the di-jet events a probe-jet and an away-jet, and require that the away-jet be tagged by the Loose operating point of SecVtx. We require that the probe-jet contain a lepton within $\Delta R < 0.4$ of the center of the jet. Should there be more than one lepton inside the candidate probe-jet, the lepton with the highest E_T is selected. The selection criteria for electron-jets, muon-jets, and away-jet are listed in Tables 5.4, 5.5 and 5.6.

Samples

The MC samples used the generating process of PYTHIA [36] di-jet $P_T > 20$ GeV, $|\eta| < 2.0$, filtered for the electron-jet sample to have an electron with $P_T > 8$ GeV and $|\eta| < 1.2$, and filtered for the muon-jet sample to have a muon with $P_T > 9$ GeV and $|\eta| < 0.6$. The data samples are taken from the 8 GeV Electron trigger path and the 8 GeV Muon trigger path, with an integrated luminosity of 4.3 fb^{-1}.

5.2.4 Lepton p_T^{rel} Procedures

The lepton p_T^{rel} procedures consist of constructing signal and background templates, then fitting for the signal fraction in the data distribution. We describe each of these steps below.

	Requirements
-	raw Jet $E_T > 15$ GeV
-	Contains An Electron (ΔR < 0.4)
	- Electron CollType 1 (defEmObject)
	- CEM Electron ($\|\eta\| < 1.2$)
	- $E_T > 9$ GeV
	- $P_T > 8$ GeV
	- $0.5 <$ E/p < 2
	- Had/em < 0.05
	- Lshr < 0.2
	- |Signed CES Δx| < 3 cm
	- |CES Δz| < 5 cm
	- Strip $\chi^2 < 10$
	- z_0 Within 5 cm of Primary Vertex
	- Isolation > 0.1
	- Fiducial to SVX
	- Conversion Veto

Table 5.4: Electron-jet requirements [34].

	Requirements
-	raw Jet $E_T > 15$ GeV
-	Contains an SLT Muon
-	Muon track $\chi^2 < 2$

Table 5.5: Muon-jet requirements [35].

Electron p_T^{rel} Templates Construction

Sources of electrons embedded in the jets include semileptonic B and D hadron decays, fakes from light flavor jets, and electrons from photon conversions in the detector material.

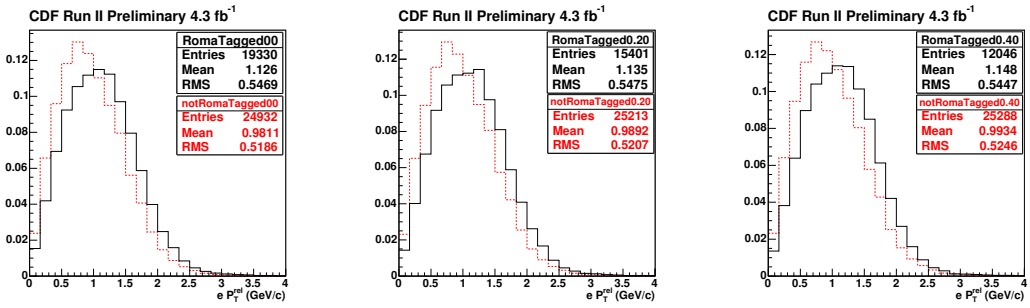
The bottom-jet templates (b-templates) are constructed from the electron-jets matched with a simulation truth b-jet, and the away-jet to be positively tagged by the LooseSecVtx as a b-jet; having the same b-enrichment event selection as in data.

We construct separate templates for RomaNN tagged and not-tagged b-jets. This

	Requirements
-	Jet Energy Level 5 Corrected > 15 GeV
-	$\|\eta\| < 1.5$
-	$\|\phi_{Away} - \phi_{Probe}\| > 2.0$ radians
-	LooseSecVtx Tagged

Table 5.6: Away-jet requirements.

is because the b-tag efficiency increases rapidly with jet E_T and since the electron energy is included in the jet, the lepton p_T^{rel} is proportional to jet E_T. Figure 5.6 shows a significant difference of 0.15 GeV between the tagged and not-tagged b-jets.



(a) UltraLooseRomaNN. (b) LooseRomaNN. (c) TightRomaNN.

Figure 5.6: Comparison of the b electron p_T^{rel} templates, normalized to unit area.

The charm-jet template (c-template) is constructed from the electron-jets matched with a simulation truth c-jets, without any away-jet requirement. The light-flavor-jet template is constructed from the electron-jets vetoed upon any simulation truth level match to b or c-jets, without any away-jet requirement.

The conversion electron template is constructed from data electron-jets, requiring the away-jet to be negatively tagged by LooseSecVtx (identified as unlikely to be a b-jet), and the electron inside the electron-jet to have 0 silicon hits registered where there should have been hits. This sample is dominated by photon conversions

where the electron originates in the detector and is not from the primary vertex. The electron p_T^{rel} templates for charm, light-flavor, and conversions are shown in Fig. 5.7. No significant difference is seen between tagged and not-tagged for charm, light flavor, and conversion-jets.

The light-flavor template is very similar to the conversion template. Because of this, we can choose either of the two to represent non-b and non-c: we chose the conversion electron template since it has much more statistics than the MC light-flavor-jet template. We will call the conversion electron template our l-template.

These b,c,l-templates are referred to as the default p_T^{rel} templates.

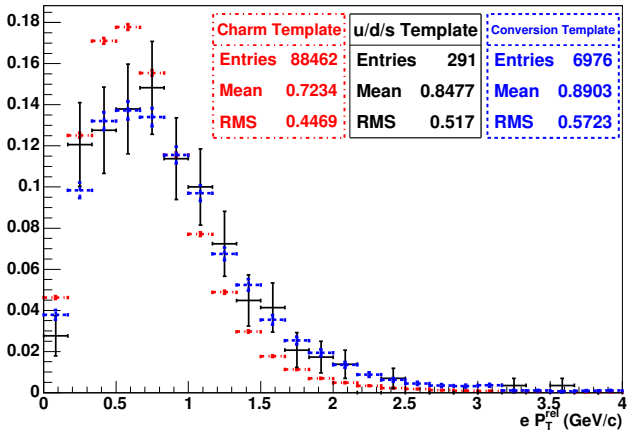


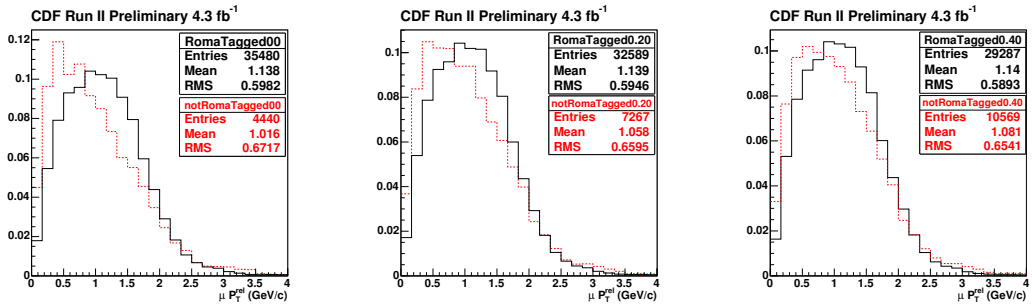
Figure 5.7: Comparison of the non-b electron p_T^{rel} templates, normalized to unit area.

Muon p_T^{rel} Templates Construction

Sources of muons embedded in the jets include semileptonic B and D hadron decays, and fakes from light flavor jets.

The bottom-jet templates (b-templates) have the same b-enrichment event selection as in data. They are constructed from the muon-jets matched with a simulation truth b-jet, and with the away-jet positively tagged by LooseSecVtx. The RomaNN tag information is used to separate the b-jets, which are then used to construct the

tagged and not-tagged muon p_T^{rel} b-templates as shown in Fig. 5.8.



(a) UltraLooseRomaNN. (b) LooseRomaNN. (c) TightRomaNN.

Figure 5.8: Comparison of the b muon p_T^{rel} templates, normalized to unit area.

The charm-jet template (c-template) is constructed from the muon-jets matched with a simulation truth c-jets, without any away-jet requirement. The light-flavor-jet template is constructed from the muon-jets vetoed upon any simulation truth level match to b or c-jets, without any away-jet requirement. The muon p_T^{rel} templates for charm and light-flavor are shown in Fig. 5.9. No significant difference is seen between tagged and not-tagged for the charm and light flavor jets.

Fitting The Data p_T^{rel} Distribution

We need to estimate the number of jets in our data sample which are b-jets. So we use the p_T^{rel} templates constructed above to get the b fraction of the data sample, and we multiply the number of jets by the b fraction. The b fraction in the sample of jets is f_b such that it maximizes the likelihood fit in Eqn. 6.2,

$$\mathcal{L} = \prod_{i=0}^{Nbins} \mathbf{P}(n_i, \mu_i) \tag{5.6}$$

where $\mathbf{P}(a, b)$ is the Poisson probability of observing a events when expecting b; n_i is the number of data jets in bin i; μ_i is the expected number of jets in bin i according

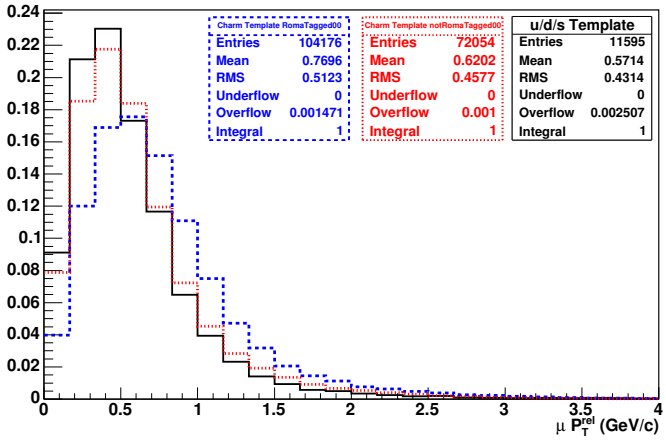


Figure 5.9: Comparison of the non-b muon p_T^{rel} templates, normalized to unit area.

to (Eqn. 5.7)

$$\mu_i = N_{data}\left(f_b T_b^i + f_c T_c^i + (1 - f_b - f_c)T_l^i\right) \tag{5.7}$$

where N_{data} is the total number of data jets in all bins; f_b is the b-fraction and f_c is the c-fraction; T_x^i is the size of the i^{th} bin of template x normalized to unit area.

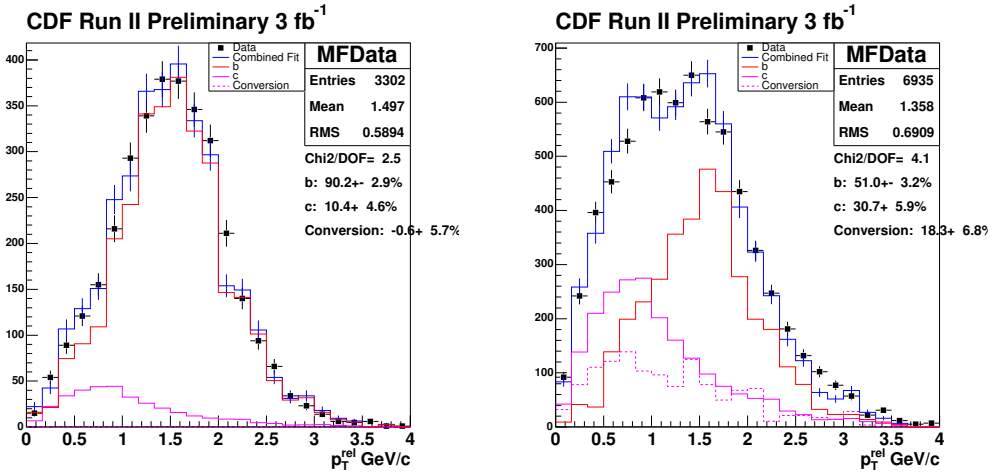
Fig. 5.10 shows a pair of sample fits using electron p_T^{rel}, and Fig. 5.11 shows a pair of sample fits using muon p_T^{rel}.

5.2.5 Results

The data efficiency of the RomaNN for b-jets has been measured using the lepton p_T^{rel} technique, using both the muon and the electron as the lepton. The MC and data efficiency for tagging b-jets are calculated according to Eqn. 5.4 and the results then produce a scale factor according to Eqn. 5.5.

Generic-Jets (as represented by Electron-Jets)

We show in Fig. 5.12 the data efficiency, simulation efficiency and their scale factor (Eqn. 5.5) for several regions of jet E_T. Though the b-tag efficiency varies with jet E_T, the scale factor is well-described by a constant. Note that for each region in

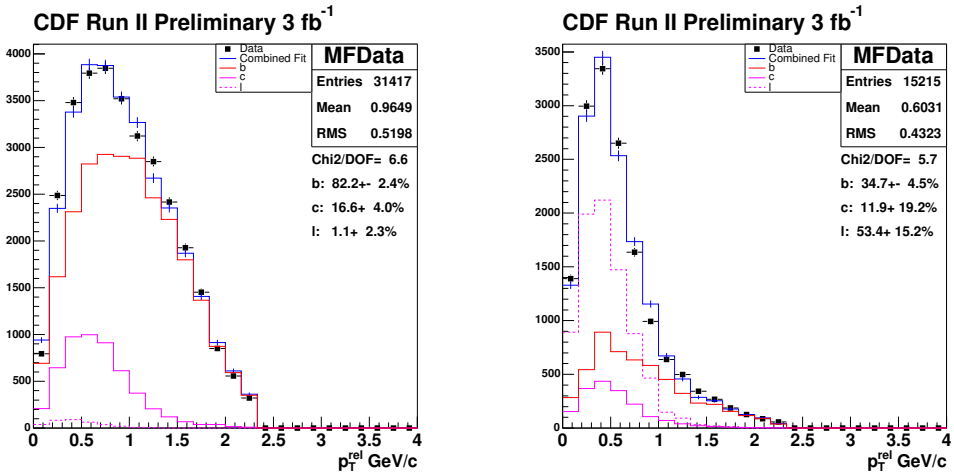


(a) Tagged.

(b) Not tagged.

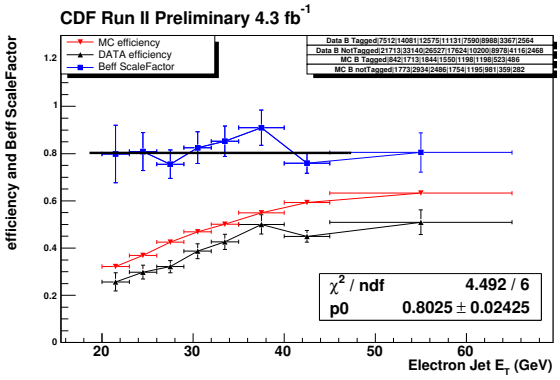
Figure 5.10: A pair of sample fits using electron p_T^{rel}. In this sample there are 3302 (jets tagged) $\times$ 90.2% b = 2978 b-jets tagged, and 6935 (jets not tagged) $\times$ 51.0% $b = 3537$ b-jets not tagged, giving an efficiency of $\frac{2978}{2978+3537} = 46\%$.

jet E_T, the templates are constructed only from jets in that region of jet E_T. We find scale factors of 0.803 ± 0.024(stat), 0.796 ± 0.024(stat), 0.785 ± 0.026(stat), for UltraLooseRomaNN, LooseRomaNN, and TightRomaNN respectively.

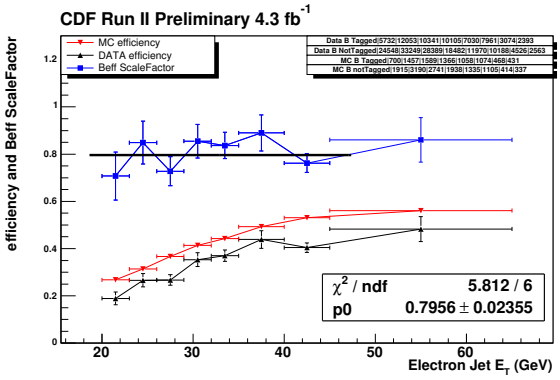


(a) UltraLooseRomaNN. (b) LooseRomaNN.

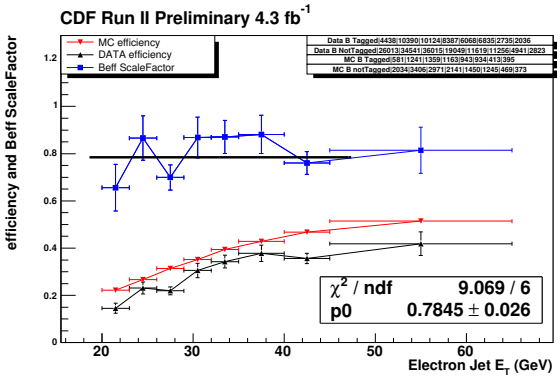
Figure 5.11: A pair of sample fits using muon p_T^{rel}. In this sample there are 31417 (jets tagged) $\times$ 82.2% $b = 25825$ b-jets tagged, and 15215 (jets not tagged) $\times$ 34.7% $b = 5280$ b-jets not tagged, giving an efficiency of $\frac{25825}{25825+5280} = 83\%$.



(a) UltraLooseRomaNN.



(b) LooseRomaNN.



(c) TightRomaNN.

Figure 5.12: Scale Factor vs. Jet E_T for electron jets, fitted with flat line.

The number of z vertices is an important variable that measures how many simultaneous particle interactions are taking place within a crossing of the proton and anti-proton bunches. Fig. 5.13 compares the distribution of the number of z vertices in events from di-jet data, di-jet MC, and $t\bar{t}$ MC. We find that the scale factor varies significantly with the number of z vertices in the event, see Fig. 5.14. This is an indication that the simulation is optimistic with regards to tagging jets within an environment where there are a lot of interactions within the detector. We provide correction factor of $\frac{0.8518-0.1298\times(nZvertex-1.8)}{0.8518}$, $\frac{0.8617-0.1278\times(nZvertex-1.8)}{0.8617}$, $\frac{0.8483-0.1545\times(nZvertex-1.8)}{0.8483}$, for UltraLooseRomaNN, LooseRomaNN, and TightRomaNN respectively. The correction factor is centered on 1.8 because it is the mean number of z vertices for our di-jet data sample and it minimizes the parameterization error.

To get an estimate of the uncertainty of using the parameterization to extrapolate to another sample, we take the combination of the highest/lowest constant value and slopes in Fig. 5.14, weighted by the z vertices distribution of $t\bar{t}$ production, and take the standard deviation. The standard deviation is 0.030, 0.032, 0.036, for UltraLooseRomaNN, LooseRomaNN, and TightRomaNN respectively, which we take as the uncertainty of parameterization. In summary, we have scale factors for generic-jets of

$0.803 \times \frac{0.8518-0.1298\times(nZvertex-1.8)}{0.8518} \pm 0.024(stat) \pm 0.030(nZvertex)$ for UltraLooseRomaNN,

$0.796 \times \frac{0.8617-0.1278\times(nZvertex-1.8)}{0.8617} \pm 0.024(stat) \pm 0.032(nZvertex)$ for LooseRomaNN,

$0.785 \times \frac{0.8483-0.1545\times(nZvertex-1.8)}{0.8483} \pm 0.026(stat) \pm 0.036(nZvertex)$ for TightRomaNN.

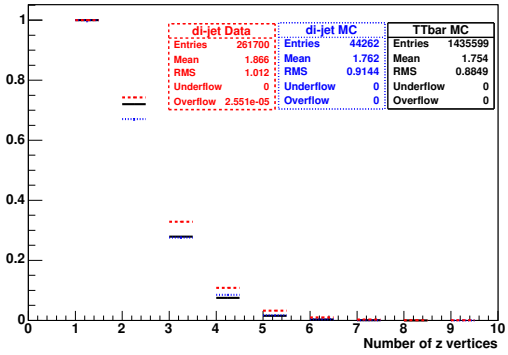
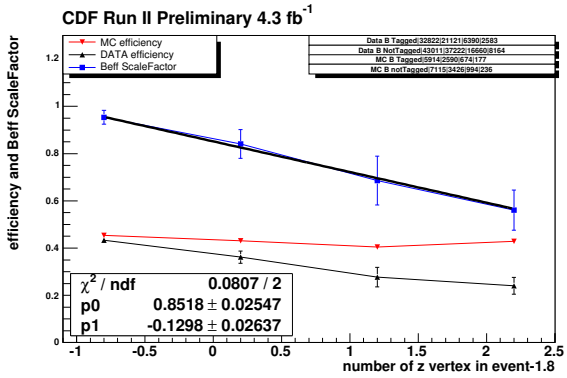
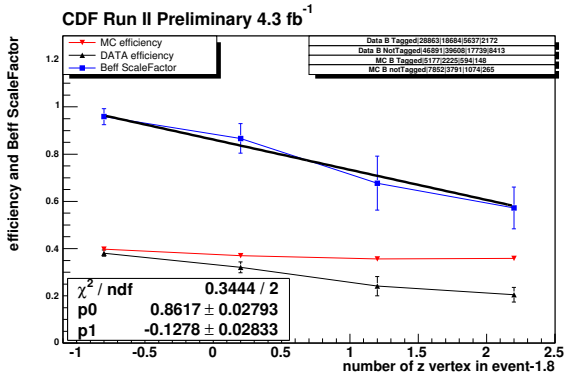


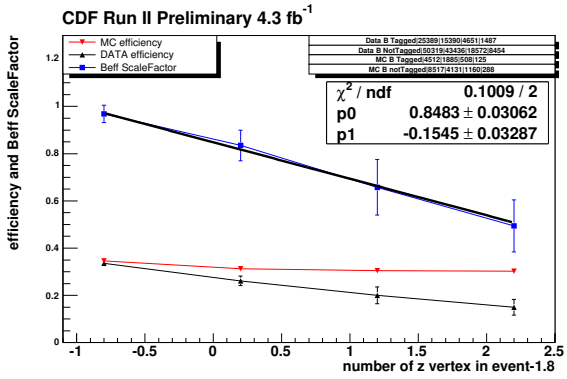
Figure 5.13: Comparison of the distribution of the number of z vertices in events from di-jet data, di-jet MC, and $t\bar{t}$ MC.



(a) UltraLooseRomaNN.



(b) LooseRomaNN.



(c) TightRomaNN.

Figure 5.14: Scale Factor vs. the number of z vertices in the event for electron jets, x-axis shifted to minimize the parameterization error.

Muon-Jets

We show in Figure 5.15 the data efficiency, simulation efficiency and their scale factor (Eqn. 5.5) for several regions of jet E_T. For these muon jets, the scale factor is not well-described by a constant but decreases as jet E_T increases. We provide parameterizations of
$(0.8881 - 0.004603 \times (JetE_T(inGeV) - 45))$ for UltraLooseRomaNN,
$(0.8776 - 0.004243 \times (JetE_T(inGeV) - 45))$ for LooseRomaNN,
$(0.8716 - 0.001106 \times (JetE_T(inGeV) - 45))$ for TightRomaNN, (see Fig. 5.15).
The parameterization is centered on 45 GeV because it minimizes the parameterization error.

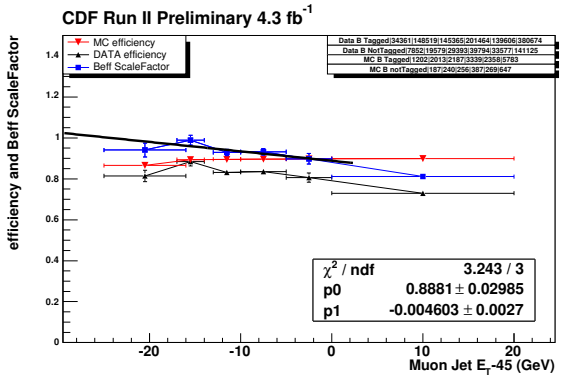
To get an estimate of the uncertainty of using the parameterization to extrapolate to another sample, we take the combination of the highest/lowest constant value and slopes in Fig. 5.15, weighted by the E_T spectrum of $t\bar{t}$ production, and take the standard deviation. This takes into account the effect of the parameterization uncertainty, which is a statistical error, as well as the systematic error in extrapolating to another sample. The standard deviation is 0.0085, 0.0119, 0.0123, for UltraLooseRomaNN, LooseRomaNN, and TightRomaNN respectively, which we take as the uncertainty of parameterization. In summary, we have a scale factor for muon-jets of
$(0.8881 - 0.004603 \times (JetE_T(inGeV) - 45)) \pm 0.0085(stat + JetE_T\ param)$ for UltraLooseRomaNN,
$(0.8776 - 0.004243 \times (JetE_T(inGeV) - 45)) \pm 0.0119(stat + JetE_T\ param)$ for LooseRomaNN,
$(0.8716 - 0.001106 \times (JetE_T(inGeV) - 45)) \pm 0.0123(stat + JetE_T\ param)$ for TightRomaNN.

5.2.6 Systematic Errors and Cross Checks

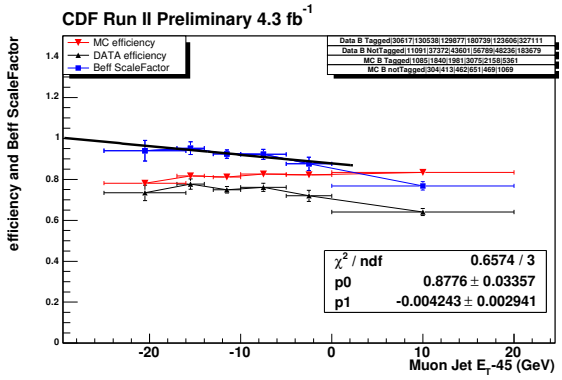
We investigate several sources of systematic uncertainty on the scale factor. The systematic uncertainties come from track multiplicity modeling, the modeling of b-jets, as well as for electrons the dependence on jet E_T and description of conversion electrons. These errors are assumed to be independent, and added in quadrature.

The systematic errors in the scale factor measurement of the binary mode RomaNN for electron-jets are listed in Tables 5.7, 5.8, 5.9, and muon-jets in Tables 5.10, 5.11, 5.12.

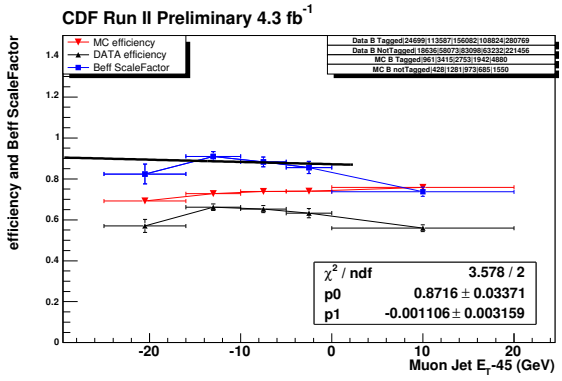
We will compare the distribution of various quantities from the di-jet data, di-jet MC, and $t\bar{t}$ MC. For di-jet data, we show only jets with electrons. For di-jet MC, we show only b-jets with electrons. For $t\bar{t}$ MC we show only b-jets.



(a) UltraLooseRomaNN.



(b) LooseRomaNN.



(c) TightRomaNN.

Figure 5.15: Scale Factor vs. Jet E_T for muon jets, x-axis shifted to minimize the parameterization error.

	Source	*Systematic Error*
-	Jet Transverse Energy	0.011
-	Track Multiplicity	0.027
-	Conversion Model	0.018
-	MC b-jet Model	0.041
	Total	0.053

Table 5.7: Systematic errors in UltraLooseRomaNN scale factor measurement for electron p_T^{rel}.

	Source	*Systematic Error*
-	Jet Transverse Energy	0.011
-	Track Multiplicity	0.028
-	Conversion Model	0.018
-	MC b-jet Model	0.048
	Total	0.059

Table 5.8: Systematic errors in LooseRomaNN scale factor measurement for electron p_T^{rel}.

Jet Transverse Energy

For electrons, we assume a constant scale factor with jet E_T. However, the jet E_T spectrum in our signal sample (W+2 jets) is peaked at a higher jet E_T value than for the calibration di-jet sample. To estimate a systematic uncertainty we use a linear fit to the scale factor as a function of E_T shown in Fig. 5.17, weighted by the E_T spectrum of $t\bar{t}$ events shown in Fig. 5.16, to compute a weighted scale factor. The shift is taken as the systematic uncertainty.

For muons, this systematic uncertainty was built into the description of the scale factor as a function of jet E_T, so it is already accounted for.

	Source	*Systematic Error*
-	Jet Transverse Energy	0.011
-	Track Multiplicity	0.030
-	Conversion Model	0.019
-	MC b-jet Model	0.048
	Total	0.061

Table 5.9: Systematic errors in TightRomaNN scale factor measurement for electron p_T^{rel}.

	Source	*Systematic Error*
-	Track Multiplicity	0.003
-	MC b-jet Model	0.012
	Total	0.012

Table 5.10: Systematic errors in UltraLooseRomaNN scale factor measurement for muon p_T^{rel}.

Track Multiplicity

The track multiplicity spectrum in our signal sample (W+2jets) is different from the calibration di-jet sample. However, We had assumed a constant scale factor with track multiplicity ($nTrack$). To estimate a systematic uncertainty of this assumption, we use the track multiplicity spectrum of $t\bar{t}$ events, shown in Fig. 5.18, 5.19 and weight the $nTrack$ spectrum (Fig. 5.20, 5.21) bin by bin with the di-jet MC and the $t\bar{t}$ MC, and compute weighted scale factors. The shift is taken as the systematic uncertainty.

Conversion Template Shape

The systematic error taken for the shape of the conversion template is needed for electron jets only. We estimate how sensitive the scale factor is to the choice of conversion template. The default light-flavor templates are constructed from data by requiring the away-jet to be negatively tagged by LooseSecVtx, and the electron

	Source	*Systematic Error*
-	Track Multiplicity	0.006
-	MC b-jet Model	0.011
	Total	0.013

Table 5.11: Systematic errors in LooseRomaNN scale factor measurement for muon p_T^{rel}.

	Source	*Systematic Error*
-	Track Multiplicity	0.012
-	MC b-jet Model	0.012
	Total	0.017

Table 5.12: Systematic errors in TightRomaNN scale factor measurement for muon p_T^{rel}.

inside the electron-jet to have 0 silicon hits registered where there should have been hits. Figure 5.22 shows the distribution of number of the silicon hits for the electron in the data sample.

We build 3 other light-flavor templates with 3/4/5 silicon hits registered instead of zero silicon hits registered. The four l-templates (Fig. 5.23) along with the default b and c-templates were used to generate 4 sets of pseudodata. These 4 sets of pseudodata are then fitted with the default templates. The shift in scale factor would then be a good estimate of the error caused by relying on a particular method of l-template construction.

MC b-jet Model

We need to estimate how sensitive the scale factor is to our model of lepton b-jets. We have two methods of studying this: one based on comparison between data and MC, another one based on MC-tuning.

For the data based study, we obtain a data sample very enriched in b-jets by

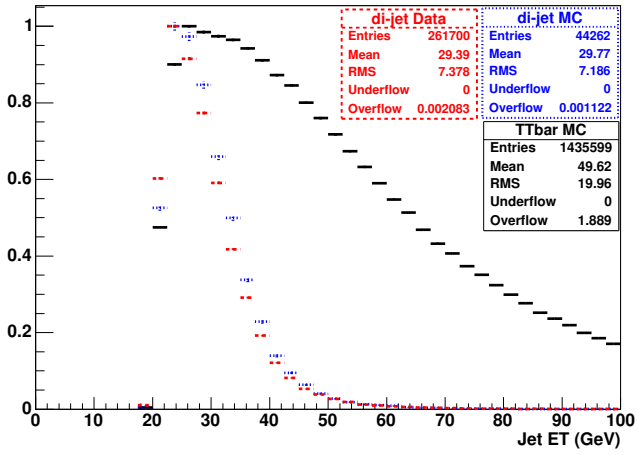


Figure 5.16: Comparison of the E_T distribution of electron-jets from di-jet data, di-jet MC, and $t\bar{t}$ MC.

requiring the away-jet to be tagged by the UltraTightSecVtx tagger, and that the lepton-jet be tagged with a really tight cut (RomaNNcut=0.8). Both of these cuts are made to ensure the sample is very pure in b-jets. From Fig. 5.2, we can assume that the RomaNN tagged lepton-jets in this sample are all from b-jets, with no contamination from charm or light. We use the p_T^{rel} of these tagged lepton-jets as an alternate model of tagged b-jets as shown in Fig. 5.24, 5.25. We take this shift to be the systematic uncertainty estimation using this data based method.

For the MC based study we estimate the scale factor sensitivity to the fragmentation tuning, which affects the fraction of the b-quarks momentum that is eventually carried off by the B hadron. We use three different models of the fragmentation, and they are Bowler-Lund, Peterson(PFP=0.0025), and Peterson(PFP=0.0041) [28]. As these models were readily available in existing W+bb MC, we use b-jets from that MC to populate the pseudo-experiments. Because lepton p_T^{rel} is dependent on the jet E_T, and that the W+bb MC has a jet E_T spectrum that is peaked at higher values of jet E_T than our default di-jet sample, we use the W+bb Bowler-Lund as the default b-jet template, as this allows a more accurate determination of the systematic

shift from the two alternative Peterson models. This allows us to not mix in the separately evaluated systematic uncertainty due to the jet E_T.

In order to estimate the systematic uncertainty due to the fragmentation model used in the MC simulations, we would ideally use MCs with the default Bowler-Lund within a band of its parameters' uncertainty, and accordingly perform pseudoexperiments. But we only have one MC generated by the Bowler-Lund model.

So the strategy used is to estimate the scale factor shift within a band of parameters' uncertainty of an alternative fragmentation model, the Peterson model, of which we fortunately have 2 existing $W+b\bar{b}$ MC. These two MCs used PFP=(0.0025,0.0041), the lower bound of ALEPH (0.0031$\pm$0.0006) and the central value of OPAL (0.00412$\pm$ 0.00037) respectively [28]. Then we would have the scale factor shift as a function of PFP shift, with which we can get the scale factor uncertainty as a result of PFP uncertainty. We would use the PFP uncertainty from OPAL since the Bowler-Lund model uses parameters as fitted from OPAL. Fig. 5.26 shows the 3 different b-templates.

5.2.7 Efficiency Measurement Conclusion

The efficiency and scale factor measurement is performed for the binary mode RomaNN using the electron p_T^{rel} technique. This is measured for three operating points, RomaNN3out>0.40 (TightRomaNN), RomaNN3out>0.20 (LooseRomaNN), and RomaNN3out>0.00 (UltraLooseRomaNN).

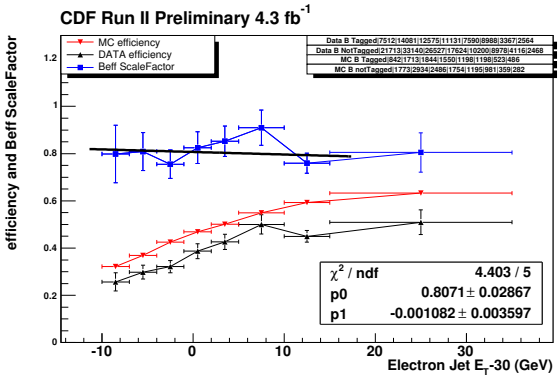
For all muon jets, we measured a scale factor of
$(0.8881 - 0.004603 \times (JetE_T(inGeV) - 45))\pm 0.0085(stat+JetE_T\,param)\pm 0.012(syst)$
for UltraLooseRomaNN,
$(0.8776 - 0.004243 \times (JetE_T(inGeV) - 45))\pm 0.0119(stat+JetE_T\,param)\pm 0.013(syst)$
for LooseRomaNN,
$(0.8716 - 0.001106 \times (JetE_T(inGeV) - 45))\pm 0.0123(stat+JetE_T\,param)\pm 0.017(syst)$
for TightRomaNN. And for all other jets a scale factor of

$0.803 \times \frac{0.8518-0.1298\times(nZvertex-1.8)}{0.8518} \pm 0.024(stat) \pm 0.030(nZvertex) \pm 0.053(syst)$ for UltraLooseRomaNN,

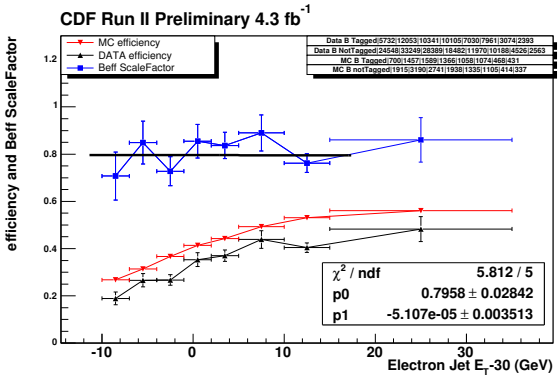
$0.796 \times \frac{0.8617-0.1278\times(nZvertex-1.8)}{0.8617} \pm 0.024(stat) \pm 0.032(nZvertex) \pm 0.059(syst)$ for LooseRomaNN,

$0.785 \times \frac{0.8483-0.1545\times(nZvertex-1.8)}{0.8483} \pm 0.026(stat) \pm 0.036(nZvertex) \pm 0.061(syst)$ for TightRomaNN.

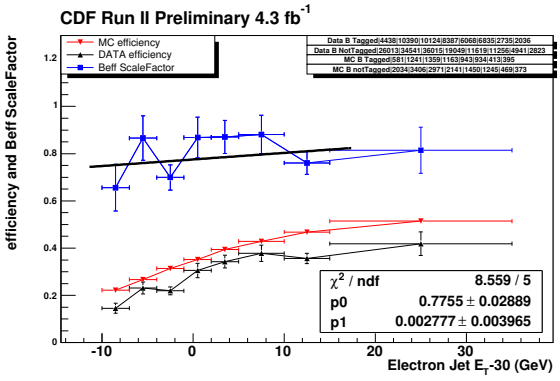
Then we can obtain the data efficiency by multiplying the efficiency obtained from simulations with the scale factor (Eqn. 5.5).



(a) UltraLooseRomaNN.



(b) LooseRomaNN.



(c) TightRomaNN.

Figure 5.17: Scale Factor vs. Jet E_T for electron jets, fitted with line.

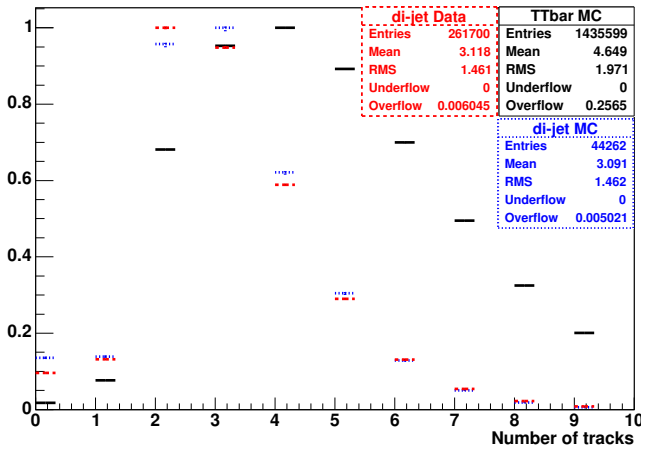


Figure 5.18: Comparison of the distribution of the number of tracks inside the electron-jets from di-jet data, di-jet MC, and $t\bar{t}$ MC.

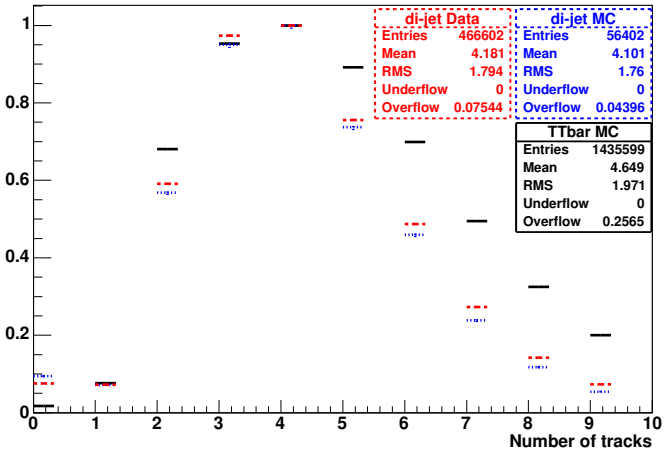
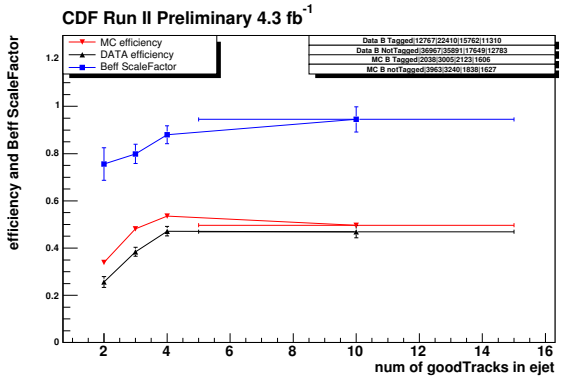
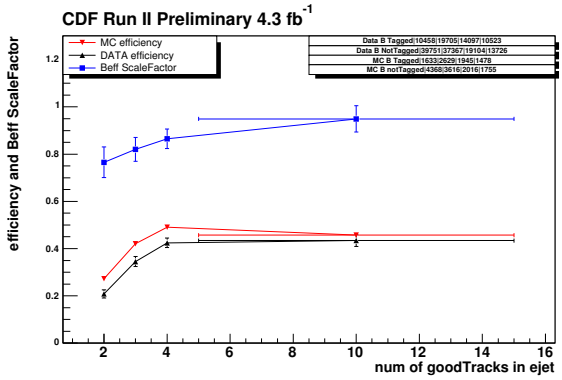


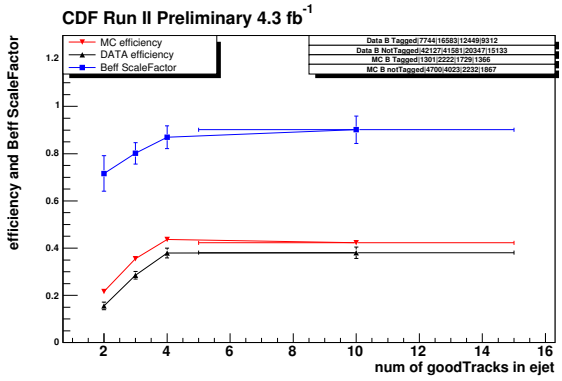
Figure 5.19: Comparison of the distribution of the number of tracks inside the muon-jets from di-jet data, di-jet MC, and $t\bar{t}$ MC.



(a) UltraLooseRomaNN.

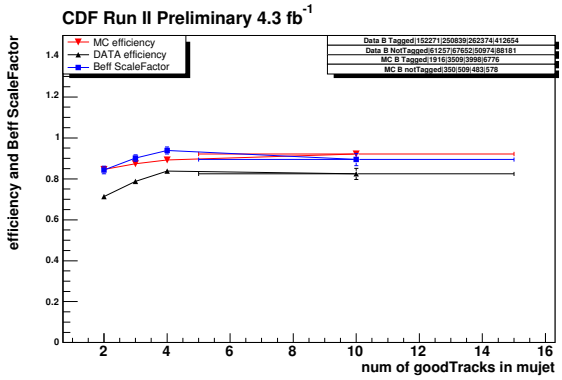


(b) LooseRomaNN.

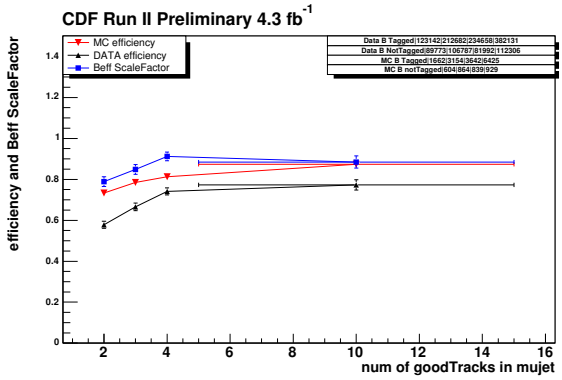


(c) TightRomaNN.

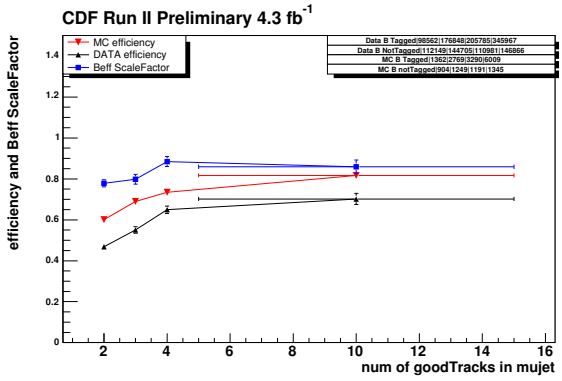
Figure 5.20: Scale Factor vs. Number of RomaNN good tracks for electron jets.



(a) UltraLooseRomaNN.



(b) LooseRomaNN.



(c) TightRomaNN.

Figure 5.21: Scale Factor vs. Number of RomaNN good tracks for muon jets.

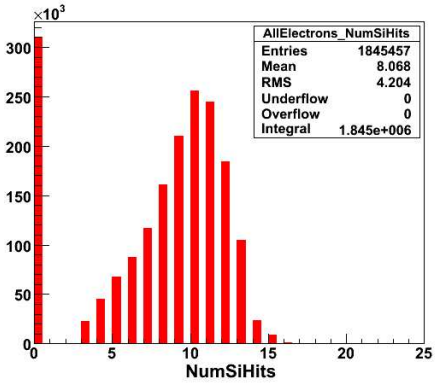


Figure 5.22: The distribution of number of Si-Hits in the electron-jet in the di-jet data sample.

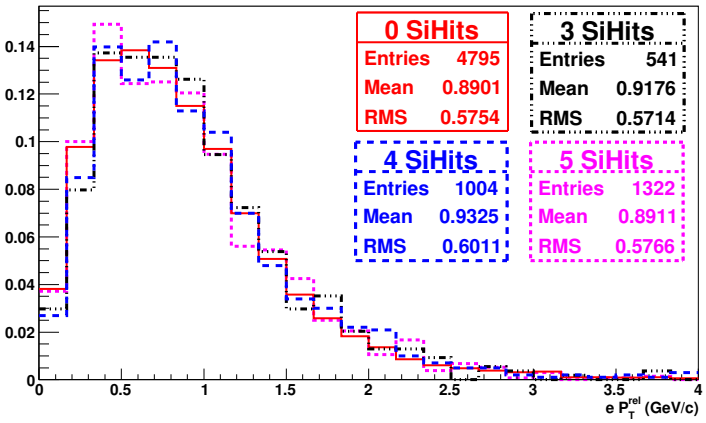


Figure 5.23: 4 different l-templates, built from data. The default template has 0 SiHits.

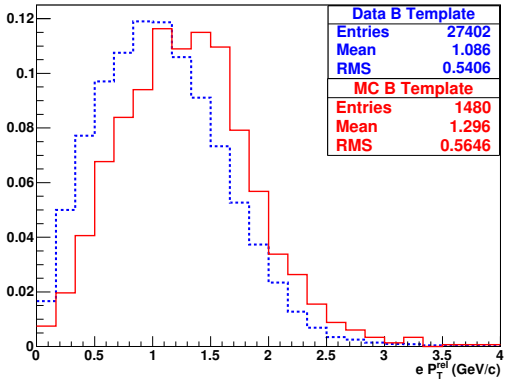


Figure 5.24: Comparison of the tagged templates for electron jets, red = MC based(default) b-template, blue = data based b-template.

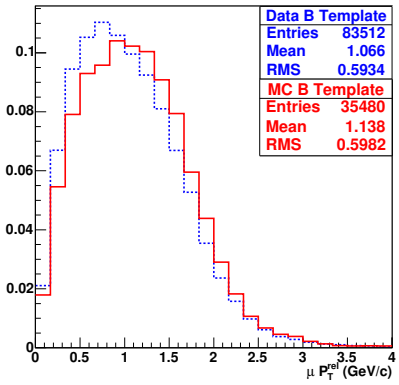
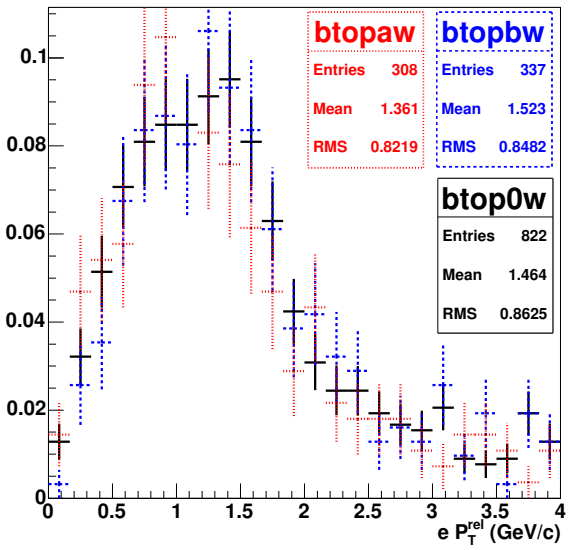
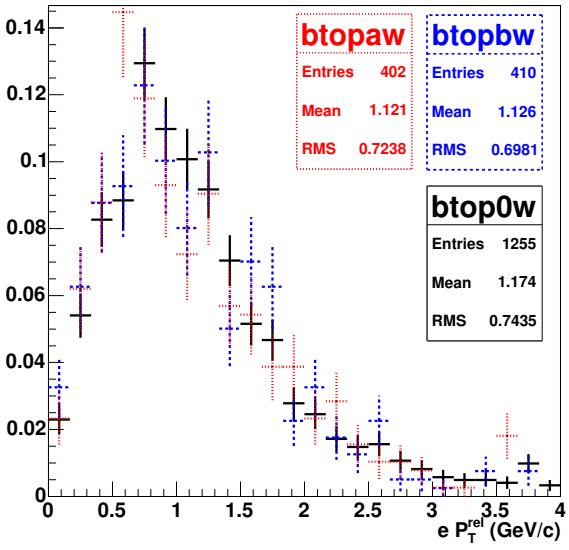


Figure 5.25: Comparison of the tagged templates for muon jets, red = MC based(default) b-template, blue = data based b-template.

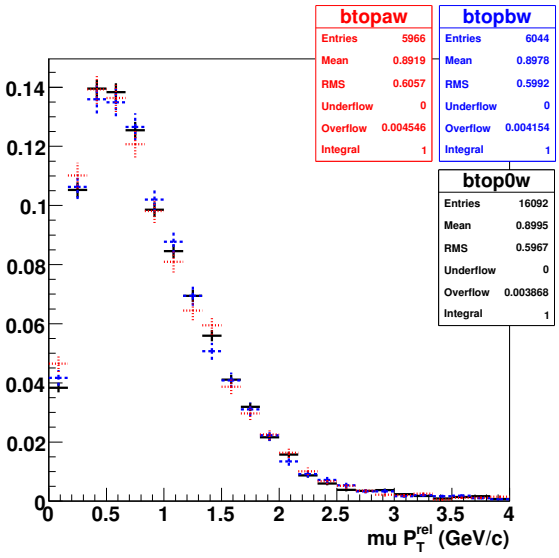


(a) Tagged templates.

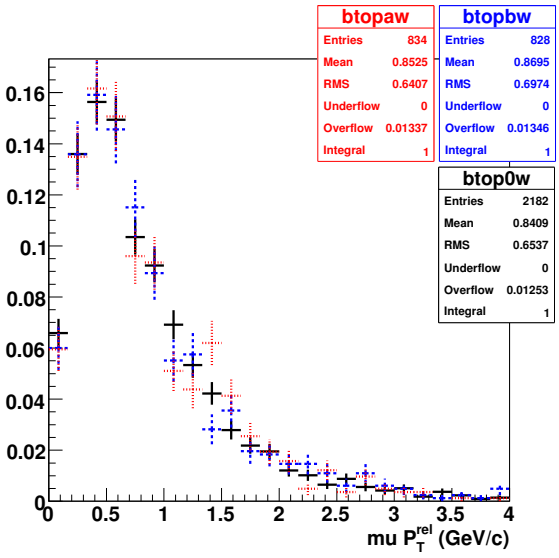


(b) Not Tagged templates.

Figure 5.26: Three different electron p_T^{rel} b-templates, built from $W + b\bar{b}$ MCs. The Bowler-Lund model is labeled btop0w, Peterson(PFP=0.0025) model labeled btopaw, and Peterson(PFP=0.0041) model labeled btopbw.



(a) Tagged templates.



(b) Not Tagged templates.

Figure 5.27: Three different muon p_T^{rel} b-templates, built from $W + b\bar{b}$ MCs. The Bowler-Lund model is labeled btop0w, Peterson(PFP=0.0025) model labeled btopaw, and Peterson(PFP=0.0041) model labeled btopbw.

5.3 Misidentification Rate Measurement

One of the most important performance parameters of the identification of b-jets is the rate at which they are misidentified. This, along with the identification efficiency, characterizes the purity of the identified b-jets.

There will always be a trade-off between high b-jet identification efficiency and misidentification rate (mistag rate). In this section we describe the estimation of the mistag rate. In a physics analysis, the number of mistags is obtained by applying the mistag rate to the number of observed jets. As the mistag rate grows with jet E_T and number of tracks, the mistag rate is parameterized in terms of these variables as well as several others in a matrix lookup format, as shown in Table 5.13.

Mistags are due to spurious large impact parameter tracks that arise from the limited detector resolution, long-lived light particle decays, and material interactions.

For taggers that look only for a secondary decay vertex with the jet, mistags due to the limited detector resolution can be expected to be symmetric in their signed 2D displacement of the vector separating the primary and secondary vertices. The effects due to long-lived light particle decays, and material interactions can then be estimated by measuring the b-fraction in the oppositely signed b-tags.

For a tagger like the Roma Neural Network, mistags can no longer be estimated in the same way because the mistags due to the limited detector resolution cannot be expected to be symmetric in any single variable. Thus we have to measure the mistags directly. The strategy used is to measure the overall tag rate, then subtracting from it the tag rate due to real b-jets.

5.3.1 Methods

The RomaNN mistag rate is estimated from the tag rate of RomaNN on jets dominated by light flavor with a correction for the contribution from heavy flavor:

$$rate^{mistag}_{RomaNN} = rate^{TotalTag}_{RomaNN} - rate^{heavy} \times \epsilon^{heavy}_{RomaNN;data}, \tag{5.8}$$

where $rate^{mistag}_{RomaNN}$ is the RomaNN mistag rate, $rate^{TotalTag}_{RomaNN}$ is the raw tag rate of RomaNN on jets, $rate^{heavy}$ is the rate of heavy flavor jets in the sample independent of the tagger, and $\epsilon^{heavy}_{RomaNN;data} = (\epsilon^{heavy}_{RomaNN;MC} \times ScaleFactor_{RomaNN})$ is the heavy flavor tagging efficiency of the RomaNN in data.

The correction for heavy flavor is based on the difference between the total and mistag rates of the SecVtx tagger:

$$rate^{heavy} = \frac{(rate^{+}_{SecVtx}) - \alpha\beta(rate^{-}_{SecVtx})}{\epsilon^{heavy}_{SecVtx;data}}, \tag{5.9}$$

where $rate^{+}_{SecVtx}$ is the TightSecVtx positive tag rate, $\alpha\beta(rate^{-}_{SecVtx})$ is the TightSecVtx mistag rate, and $\epsilon^{heavy}_{SecVtx;data} = (\epsilon^{heavy}_{SecVtx;MC} \times ScaleFactor_{SecVtx})$ is the heavy flavor tagging efficiency of TightSecVtx in data.

Derivations

For any b-tagger, the total tag rate ($rate^{TotalTag}$) is the sum of the tag rate of heavy flavor jets ($rate^{heavy}$) and the mistag rate $rate^{mistag}$,

$$rate^{TotalTag} = rate^{heavy} + rate^{mistag}. \tag{5.10}$$

Let's state specifically the tag rates for SecVtx and RomaNN,

$$rate^{TotalTag}_{SecVtx} = rate^{heavytag}_{SecVtx} + rate^{mistag}_{SecVtx}, \tag{5.11}$$

$$rate^{TotalTag}_{RomaNN} = rate^{heavytag}_{RomaNN} + rate^{mistag}_{RomaNN}. \tag{5.12}$$

For any sample, the total rate of jet collection ($rate^{total}$) is the sum of the rate of heavy flavor jet collection ($rate^{heavy}$) and the rate of light jet collection ($rate^{light}$),

$$rate^{total} = rate^{heavy} + rate^{light}. \tag{5.13}$$

The tag rate of heavy flavor jets ($rate^{heavytag}$) is related to the rate of heavy flavor jet collection by the data heavy flavor tag efficiency (ϵ^{heavy}_{data}),

$$rate^{heavytag} = \epsilon^{heavy}_{data} \times rate^{heavy}. \tag{5.14}$$

Again let's state specifically the relations for SecVtx and RomaNN,

$$rate^{heavytag}_{SecVtx} = \epsilon^{heavy}_{SecVtx;data} \times rate^{heavy}, \tag{5.15}$$

$$rate^{heavytag}_{RomaNN} = \epsilon^{heavy}_{RomaNN;data} \times rate^{heavy}. \tag{5.16}$$

We seek the RomaNN mistag rate ($rate^{mistag}_{RomaNN}$), obtained from inverting Eqn. 5.12,

$$rate^{mistag}_{RomaNN} = rate^{TotalTag}_{RomaNN} - rate^{heavytag}_{RomaNN}, \tag{5.17}$$

of which $rate^{TotalTag}_{RomaNN}$ can be measured directly, and $rate^{heavytag}_{RomaNN}$ is from Eqn. 5.16, of which $\epsilon^{heavy}_{RomaNN;data}$ can be estimated from MC simulations and corrected by the MC-to-data scalefactor, and $rate^{heavy}$ can be estimated by inverting Eqn. 5.15 to get

$$rate^{heavy} = \frac{rate^{heavytag}_{SecVtx}}{\epsilon^{heavy}_{SecVtx;data}}, \tag{5.18}$$

where $\epsilon^{heavy}_{SecVtx;data}$ can be estimated from MC simulations and corrected by the MC-to-data scalefactor, and $rate^{heavytag}_{SecVtx}$ can be measured by inverting Eqn. 5.11 to get

$$rate^{heavytag}_{SecVtx} = rate^{TotalTag}_{SecVtx} - rate^{mistag}_{SecVtx} = rate^{+}_{SecVtx} - \alpha\beta(rate^{-}_{SecVtx}). \tag{5.19}$$

5.3.2 Introduction to Tag Rates

We describe a new set of mistag matrices filled using a total of 4.3 fb^{-1} of data. The same matrix parameterization as used previously for the SecVtx family [30] is used

here for the RomaNN family, and is summarized in Table 5.13. Matrices are constructed for the SecVtx (tight operating point) and RomaNN (UltraLooseRomaNN, LooseRomaNN and TightRomaNN operating point) taggers, separately for jets with muons and jets without muons. They are constructed from the data events where there is at least one jet, called the inclusive jet samples. The Jet20/25/70/100 samples correspond to events containing jets, where the least energetic jet has transverse energy of 20/25/70/100 respectively. The SumEt sample also contains jets, with the jets having a broad spectrum of energy, which is similar to a random subsample of the four Jet20/25/70/100 samples. The tag rates here are measured from fiducial jets. Fiducial jets have raw $E_T > 10$ GeV, and $|\eta| < 2.4$.

Variable	Bin Edges
Jet E_T (GeV)	0, 15, 22, 30, 40, 60, 90, 130, 500
Num Tracks/jet	0, 1, 2, 3, 4, 5, 6, 7, 8, 10, 13, 50
$\lvert\eta_{\text{jet}}\rvert$	0.0, 0.3, 0.6, 0.8, 1.1 (, 1.5, 2.0, 2.4)*
$n_{\text{PrimaryVertex}}$	1, 2, 3, 4-5, 6-10
$\sum E_T$ (GeV)	0, 80, 140, 220, 500
z_{prim} (cm)	-100, -60, -25, -10, 10, 25, 60, 100

Table 5.13: Variables and binning used in the mistag matrices.
(* Jets without any muons only. Since muon coverage only goes up to $|\eta| < 0.6$, and with our definition that jets span at most $\eta = 0.4$, those bins are unused.)

5.3.3 Validation

We validate the tag rate matrices by creating a sub-matrix using even numbered events and apply them to odd numbered events. The predicted and observed tag rates are compared for the variables used to parameterize the matrix. Tables 5.14, 5.15, 5.16, 5.17, 5.18, 5.19 (RomaNN), 5.20, 5.21 (TightSecVtx) show the tag rates predicted using even numbered events and observed with odd numbered events. The observed/predicted ratios are very close to unity as expected. We see one clear trend, and it is that the tag rates are higher for jets with larger E_T.

Sample	Obs (odd) +Rate	Pred (even) +Rate	Ratio
All	0.40840 ± 0.00004	0.40861 ± 0.00004	0.99950 ± 0.00013
Jet20	0.33863 ± 0.00131	0.34988 ± 0.00132	0.96787 ± 0.00522
Jet50	0.40247 ± 0.00121	0.40270 ± 0.00121	0.99942 ± 0.00423
Jet70	0.43661 ± 0.00113	0.43261 ± 0.00113	1.00924 ± 0.00370
Jet100	0.45753 ± 0.00093	0.45031 ± 0.00093	1.01603 ± 0.00294

Table 5.14: Tag-rates for UltraLooseRomaNN, muon jets: tag-rate matrix predicted with even events observed with odd events.

Sample	Obs (odd) +Rate	Pred (even) +Rate	Ratio
All	0.27407 ± 0.00003	0.27689 ± 0.00003	0.98982 ± 0.00017
Jet20	0.19065 ± 0.00064	0.20237 ± 0.00065	0.94210 ± 0.00436
Jet50	0.27033 ± 0.00067	0.27722 ± 0.00067	0.97514 ± 0.00339
Jet70	0.30393 ± 0.00048	0.30534 ± 0.00048	0.99538 ± 0.00222
Jet100	0.33393 ± 0.00041	0.32661 ± 0.00041	1.02243 ± 0.00179

Table 5.15: Tag-rates for LooseRomaNN, muon jets: tag-rate matrix predicted with even events observed with odd events.

Sample	Obs (odd) +Rate	Pred (even) +Rate	Ratio
All	0.17184 ± 0.00003	0.17453 ± 0.00003	0.98458 ± 0.00023
Jet20	0.10809 ± 0.00050	0.11787 ± 0.00052	0.91707 ± 0.00588
Jet50	0.17045 ± 0.00057	0.17764 ± 0.00058	0.95950 ± 0.00446
Jet70	0.19433 ± 0.00041	0.19634 ± 0.00041	0.98975 ± 0.00296
Jet100	0.21693 ± 0.00036	0.21039 ± 0.00035	1.03109 ± 0.00243

Table 5.16: Tag-rates for TightRomaNN, muon jets: tag-rate matrix predicted with even events observed with odd events.

Sample	Obs (odd) +Rate	Pred (even) +Rate	Ratio
All	0.04327 ± 0.00002	0.04322 ± 0.00002	1.00126 ± 0.00050
Jet20	0.01786 ± 0.00002	0.01820 ± 0.00003	0.98164 ± 0.00192
Jet50	0.03854 ± 0.00005	0.03878 ± 0.00005	0.99381 ± 0.00173
Jet70	0.05078 ± 0.00006	0.05090 ± 0.00006	0.99759 ± 0.00162
Jet100	0.06506 ± 0.00006	0.06429 ± 0.00006	1.01195 ± 0.00136

Table 5.17: Tag-rates for UltraLooseRomaNN, nomuon jets: tag-rate matrix predicted with even events observed with odd events.

Sample	Obs (odd) +Rate	Pred (even) +Rate	Ratio
All	0.02513 ± 0.00001	0.02512 ± 0.00001	1.00037 ± 0.00066
Jet20	0.00828 ± 0.00001	0.00851 ± 0.00001	0.97325 ± 0.00155
Jet50	0.02127 ± 0.00002	0.02166 ± 0.00002	0.98204 ± 0.00138
Jet70	0.02980 ± 0.00002	0.02985 ± 0.00002	0.99844 ± 0.00097
Jet100	0.04027 ± 0.00002	0.03973 ± 0.00002	1.01353 ± 0.00083

Table 5.18: Tag-rates for LooseRomaNN, nomuon jets: tag-rate matrix predicted with even events observed with odd events.

Sample	Obs (odd) +Rate	Pred (even) +Rate	Ratio
All	0.01436 ± 0.00001	0.01435 ± 0.00001	1.00068 ± 0.00087
Jet20	0.00491 ± 0.00001	0.00502 ± 0.00001	0.97792 ± 0.00202
Jet50	0.01249 ± 0.00002	0.01268 ± 0.00002	0.98480 ± 0.00181
Jet70	0.01703 ± 0.00002	0.01708 ± 0.00002	0.99704 ± 0.00129
Jet100	0.02263 ± 0.00002	0.02232 ± 0.00002	1.01367 ± 0.00111

Table 5.19: Tag-rates for TightRomaNN, nomuon jets: tag-rate matrix predicted with even events observed with odd events.

Sample	Obs (odd) +Rate	Pred (even) +Rate	Ratio
All	0.04632 ± 0.00002	0.04632 ± 0.00002	0.99989 ± 0.00048
Jet20	0.02448 ± 0.00031	0.02652 ± 0.00032	0.92288 ± 0.01605
Jet50	0.04854 ± 0.00043	0.05010 ± 0.00043	0.96877 ± 0.01193
Jet70	0.05381 ± 0.00042	0.05345 ± 0.00042	1.00675 ± 0.01112
Jet100	0.06027 ± 0.00038	0.05759 ± 0.00037	1.04648 ± 0.00938
Sample	Obs (odd) -Rate	Pred (even) -Rate	Ratio
All	0.00869 ± 0.00001	0.00881 ± 0.00001	0.98624 ± 0.00111
Jet20	0.00218 ± 0.00009	0.00229 ± 0.00009	0.95396 ± 0.05669
Jet50	0.00728 ± 0.00017	0.00751 ± 0.00017	0.97026 ± 0.03153
Jet70	0.01040 ± 0.00019	0.01043 ± 0.00019	0.99694 ± 0.02556
Jet100	0.01456 ± 0.00019	0.01471 ± 0.00019	0.98997 ± 0.01822

Table 5.20: Tag-rates for TightSecVtx, muon jets: tag-rate matrix predicted with even events observed with odd events.

Sample	Obs (odd) +Rate	Pred (even) +Rate	Ratio
All	0.01667 ± 0.00001	0.01660 ± 0.00001	1.00375 ± 0.00081
Jet20	0.00651 ± 0.00002	0.00663 ± 0.00002	0.98314 ± 0.00337
Jet50	0.01510 ± 0.00003	0.01519 ± 0.00003	0.99397 ± 0.00302
Jet70	0.01983 ± 0.00004	0.01980 ± 0.00004	1.00141 ± 0.00274
Jet100	0.02504 ± 0.00004	0.02468 ± 0.00004	1.01453 ± 0.00237
Sample	Obs (odd) -Rate	Pred (even) -Rate	Ratio
All	0.00545 ± 0.00001	0.00542 ± 0.00001	1.00498 ± 0.00143
Jet20	0.00130 ± 0.00001	0.00134 ± 0.00001	0.96461 ± 0.00739
Jet50	0.00437 ± 0.00002	0.00439 ± 0.00002	0.99591 ± 0.00566
Jet70	0.00652 ± 0.00002	0.00654 ± 0.00002	0.99676 ± 0.00478
Jet100	0.00932 ± 0.00003	0.00915 ± 0.00003	1.01842 ± 0.00394

Table 5.21: Tag-rates for TightSecVtx, nomuon jets: tag-rate matrix predicted with even events observed with odd events.

5.3.4 Tag Rate Uncertainties

We consider three sources of systematic error for the RomaNN tag rates: jet sample dependence, $\sum E_T$ dependence, and trigger bias. These errors are assumed to be independent, and added in quadrature. This is the same treatment as was done for the SecVtx family previously [29]. The systematic uncertainties are evaluated using the first 1 fb^{-1} of CDF data.

Systematic Uncertainty due to Jet Sample Dependence

The systematic uncertainty due to jet sample dependence is evaluated by attempting to predict the mistag rate in a different sample from the inclusive jet samples used to build the mistag matrices. We apply the mistag matrices to the jets in the ($\sum E_T$) samples, and compare the predicted rates to the observed rates in Tables 5.22, 5.23 (RomaNN), 5.24, 5.25 (TightSecVtx).

Operating Point	Obs +Rate	Pred +Rate	Ratio	Syst Error
UltraLoose	0.42022 ± 0.00111	0.42978 ± 0.00112	0.97777 ± 0.00363	0.022
Loose	0.29034 ± 0.00003	0.30124 ± 0.00003	0.96381 ± 0.00016	0.036
Tight	0.19194 ± 0.00003	0.20159 ± 0.00003	0.95210 ± 0.00020	0.048

Table 5.22: Tag-rates for RomaNN, muon jets: regular tag-rate matrix observed with a different sample ($\sum E_T$).

Operating Point	Obs +Rate	Pred +Rate	Ratio	Syst Error
UltraLoose	0.03749 ± 0.00004	0.03702 ± 0.00004	1.01259 ± 0.00157	0.013
Loose	0.02502 ± 0.00001	0.02430 ± 0.00001	1.02962 ± 0.00068	0.030
Tight	0.01638 ± 0.00001	0.01559 ± 0.00001	1.05055 ± 0.00087	0.051

Table 5.23: Tag-rates for RomaNN, nomuon jets: regular tag-rate matrix observed with a different sample ($\sum E_T$).

Sample	Obs +Rate	Pred +Rate	Ratio	Systematic Error
All	0.05226 ± 0.00040	0.05337 ± 0.00041	0.97920 ± 0.01060	0.021
Sample	Obs -Rate	Pred -Rate	Ratio	Systematic Error
All	0.00781 ± 0.00016	0.00839 ± 0.00016	0.93077 ± 0.02635	0.069

Table 5.24: Tag-rates for TightSecVtx, muon jets: regular tag-rate matrix observed with a different sample ($\sum E_T$).

Sample	Obs +Rate	Pred +Rate	Ratio	Systematic Error
All	0.01511 ± 0.00003	0.01440 ± 0.00003	1.04933 ± 0.00263	0.049
Sample	Obs -Rate	Pred -Rate	Ratio	Systematic Error
All	0.00428 ± 0.00001	0.00434 ± 0.00001	0.98539 ± 0.00459	0.015

Table 5.25: Tag-rates for TightSecVtx, nomuon jets: regular tag-rate matrix observed with a different sample ($\sum E_T$).

Systematic Uncertainty due to $\sum E_T$ Dependence

Simulated events are known to not have perfect agreement with data regarding the activity of low E_T jet in the event, and this manifests in disagreement in the event variable $\sum E_T$, the scalar sum of jet E_T for all jets. We study systematic uncertainty due to $\sum E_T$ dependence because we want to use these mistag matrices on jets from simulated events as well as data events.

For the $\sum E_T$ dependence, we varied the event $\sum E_T$ in the inclusive jet samples. We apply the mistag matrices to the jets in the inclusive jet samples with the $\sum E_T$ scaled up by the mean jet energy in each event. The mean jet energy is $\frac{\sum E_T}{N_{jets}}$, the $\sum E_T$ divided by the number of jets in the event. We compare the predicted rates to the observed rates in Tables 5.26, 5.27 (RomaNN), 5.28, 5.29 (Tight SecVtx).

Operating Point	Obs +Rate	Pred +Rate	Ratio	Syst Error
UltraLoose	0.40828 ± 0.00004	0.42401 ± 0.00004	0.96292 ± 0.00012	0.019
Loose	0.27391 ± 0.00003	0.29087 ± 0.00003	0.94168 ± 0.00016	0.029
Tight	0.17673 ± 0.00003	0.19111 ± 0.00003	0.92472 ± 0.00021	0.038

Table 5.26: Tag-rates for RomaNN, muon jets: regular tag-rate matrix observed with events with $\sum E_T$ scaled up.

Operating Point	Obs +Rate	Pred +Rate	Ratio	Syst Error
UltraLoose	0.04327 ± 0.00002	0.04526 ± 0.00002	0.95604 ± 0.00047	0.022
Loose	0.02801 ± 0.00001	0.02934 ± 0.00001	0.95444 ± 0.00059	0.023
Tight	0.01658 ± 0.00001	0.01735 ± 0.00001	0.95553 ± 0.00077	0.022

Table 5.27: Tag-rates for RomaNN, nomuon jets: regular tag-rate matrix observed with events with $\sum E_T$ scaled up.

Sample	Obs +Rate	Pred +Rate	Ratio	Systematic Error
All	0.04498 ± 0.00002	0.04837 ± 0.00002	0.92986 ± 0.00044	0.035
Sample	Obs -Rate	Pred -Rate	Ratio	Systematic Error
All	0.00858 ± 0.00001	0.00919 ± 0.00001	0.93340 ± 0.00104	0.033

Table 5.28: Tag-rates for TightSecVtx, muon jets: regular tag-rate matrix observed with events with $\sum E_T$ scaled up.

Sample	Obs +Rate	Pred +Rate	Ratio	Systematic Error
All	0.01667 ± 0.00001	0.01734 ± 0.00001	0.96132 ± 0.00077	0.019
Sample	Obs -Rate	Pred -Rate	Ratio	Systematic Error
All	0.00544 ± 0.00001	0.00578 ± 0.00001	0.94111 ± 0.00132	0.029

Table 5.29: Tag-rates for TightSecVtx, nomuon jets: regular tag-rate matrix observed with events with $\sum E_T$ scaled up.

Systematic Uncertainty due to Trigger Bias

There is potentially a bias on the trigger jets, because trigger jets have passed a certain trigger selection cuts, whereas the typical jet does not necessarily meet those cuts. The trigger bias systematic uncertainty is calculated by examining the trigger jets, which are the jets closest to the Level 2 calorimeter cluster that fired the jet trigger. We apply the mistag matrices to only the trigger jets in the inclusive jet samples and compare the predicted rates to the observed rates for the same trigger jets in Tables 5.30, 5.31 (RomaNN), 5.32, 5.33 (Tight SecVtx).

Operating Point	Obs +Rate	Pred +Rate	Ratio	Syst Error
UltraLoose	0.38882 ± 0.00004	0.40348 ± 0.00004	0.96366 ± 0.00013	0.036
Loose	0.25477 ± 0.00003	0.27105 ± 0.00003	0.93994 ± 0.00017	0.060
Tight	0.15720 ± 0.00003	0.17204 ± 0.00003	0.91372 ± 0.00022	0.086

Table 5.30: Tag-rates for RomaNN, muon jets: regular tag-rate matrix observed with trigger jets.

Operating Point	Obs +Rate	Pred +Rate	Ratio	Syst Error
UltraLoose	0.04704 ± 0.00002	0.04791 ± 0.00002	0.98192 ± 0.00046	0.018
Loose	0.03026 ± 0.00001	0.03089 ± 0.00001	0.97971 ± 0.00058	0.020
Tight	0.01743 ± 0.00001	0.01785 ± 0.00001	0.97641 ± 0.00077	0.024

Table 5.31: Tag-rates for RomaNN, nomuon jets: regular tag-rate matrix observed with trigger jets.

Sample	Obs +Rate	Pred +Rate	Ratio	Systematic Error
All	0.03597 ± 0.00001	0.03964 ± 0.00001	0.90746 ± 0.00048	0.093
Sample	Obs -Rate	Pred -Rate	Ratio	Systematic Error
All	0.00936 ± 0.00001	0.00960 ± 0.00001	0.97451 ± 0.00105	0.026

Table 5.32: Tag-rates for TightSecVtx, muon jets: regular tag-rate matrix observed with trigger jets.

Sample	Obs +Rate	Pred +Rate	Ratio	Systematic Error
All	0.01751 ± 0.00001	0.01795 ± 0.00001	0.97550 ± 0.00077	0.025
Sample	Obs -Rate	Pred -Rate	Ratio	Systematic Error
All	0.00608 ± 0.00001	0.00619 ± 0.00001	0.98205 ± 0.00132	0.018

Table 5.33: Tag-rates for TightSecVtx, nomuon jets: regular tag-rate matrix observed with trigger jets.

Systematic Uncertainty Summary

Table 5.34 summarizes the sources and magnitudes of the aggregate systematic errors assessed. In addition, the statistical uncertainties are computed bin-by-bin in the mistag matrix code.

	RomaNN(0.0)		RomaNN(0.2)		RomaNN(0.4)		TightSecVtx			
	μ	No μ	μ	No μ	μ	No μ	μ		No μ	
Systematic	+rate	+rate	+rate	+rate	+rate	+rate	+rate	-rate	+rate	-rate
Different Sample	2.2%	1.3%	3.6%	3.0%	4.8%	5.1%	2.1%	6.9%	4.9%	1.5%
$\sum E_T$ dependence	1.9%	2.2%	2.9%	2.3%	3.8%	2.2%	3.5%	3.3%	1.9%	2.9%
Trigger Jets	3.6%	1.8%	6.0%	2.0%	8.6%	2.4%	9.3%	2.6%	2.5%	1.8%
Total	4.6%	3.1%	7.6%	4.3%	10.6%	6.1%	10.2%	8.1%	5.8%	3.7%

Table 5.34: Fractional systematic uncertainties assessed on the tag rates predicted by the mistag matrices.

5.3.5 Mistag Rate Uncertainties

In this section we illustrate the error estimation performed for the mistag rate measurements. First, let's reiterate from section 5.3.1, that the mistag rate is obtained from the tag rate of RomaNN on jets dominated by light flavor, correcting for the heavy flavor contribution based on the difference between the positive and mistag rates of the TightSecVtx tagger:

$$rate^{mistag}_{RomaNN} = rate^{TotalTag}_{RomaNN} - (rate^{+}_{SecVtx} - \alpha\beta rate^{-}_{SecVtx}) \times \frac{\epsilon^{heavy}_{RomaNN;data}}{\epsilon^{heavy}_{SecVtx;data}}.$$

We want to approximate $\frac{\epsilon^{heavy}_{RomaNN;data}}{\epsilon^{heavy}_{SecVtx;data}}$ with $\frac{\epsilon^{b}_{RomaNN;data}}{\epsilon^{b}_{SecVtx;data}}$, and with some algebra we find $\frac{\epsilon^{heavy}_{RomaNN;data}}{\epsilon^{heavy}_{SecVtx;data}} = \frac{\epsilon^{b}_{RomaNN;data}}{\epsilon^{b}_{SecVtx;data}} \times RateFactor$,

where $RateFactor = \frac{1+\frac{(f_c)(\epsilon^{c}_{RomaNN})(ScaleFactor^{c}_{RomaNN})}{(f_b)(\epsilon^{b}_{RomaNN})(ScaleFactor^{b}_{RomaNN})}}{1+\frac{(f_c)(\epsilon^{c}_{SecVtx})(ScaleFactor^{c}_{SecVtx})}{(f_b)(\epsilon^{b}_{SecVtx})(ScaleFactor^{b}_{SecVtx})}}$.

Using W+bb and W+cc MC we find the efficiency for b-tagging and c-tagging to be as listed in Table 5.35.

OperatingPoint	ϵ^b	ϵ^c	$\frac{\epsilon^c}{\epsilon^b}$
UltraLooseRomaNN	42.5	13.5	0.32
LooseRomaNN	37.6	9.8	0.26
TightRomaNN	32.5	6.9	0.21
TightSecVtx	37.4	10.2	0.27

Table 5.35: Tagging efficiencies for b and c flavor for RomaNN and SecVtx.

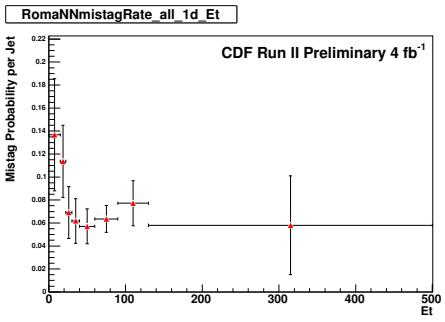
Using $\frac{f_c}{f_b} = 2$ as expected from a generic jet sample, and taking the c-tag scalefactors to be the same as the b-tag scalefactors (with two times the uncertainty), we find the $RateFactor$ to be for the three operating points as listed in Table 5.36.

OperatingPoint	NoMuon Jets $RateFactor$	Muon Jets $RateFactor$
UltraLooseRomaNN	1.07±0.11	1.07±0.05
LooseRomaNN	0.99±0.09	0.99±0.05
TightRomaNN	0.93±0.08	0.93±0.04

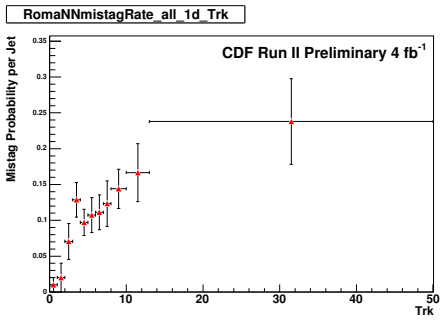
Table 5.36: $RateFactor$ for three RomaNN operating points.

In summary, for jets where we have a $rate^b$ estimate, we get a relative error of 11% for jets with muons and 15% for jets without muons. But for jets where $rate^b$ does not exist, we take $rate^{mistag}_{RomaNN} = \frac{rate^+_{RomaNN}}{2} \pm \frac{rate^+_{RomaNN}}{2}$, giving a 100% error.

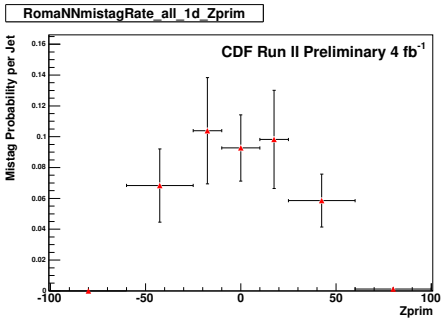
Figs. 5.28, 5.31, 5.34 show respectively for UltraLooseRomaNN, LooseRomaNN, and TightRomaNN, the 1D-projection of the mistag probability per jet in a jet sample obtained from W $\rightarrow e\nu$ and 2 jets event candidates in data. It is the union of the two set of jets: jets with muons (Figs.5.29, 5.32, 5.35), and jets without muons (Figs.5.30, 5.33, 5.36). For this particular sample, the relative error on the aggregate mistag rate is 31%, 46%, 80% for UltraLooseRomaNN, LooseRomaNN, and TightRomaNN respectively. The relative errors are larger for the tighter operating points because the mistag rate got smaller while the absolute error remains similar.



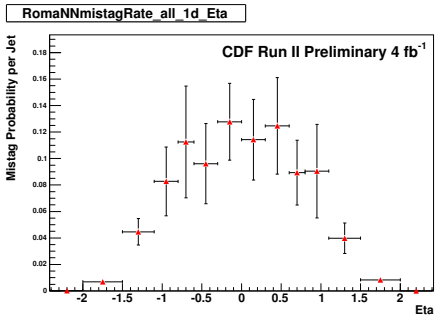
(a) Mistag probability per jet as a function of Jet E_T.



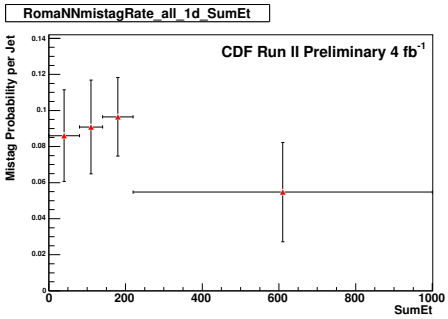
(b) Mistag probability per jet as a function of track multiplicity.



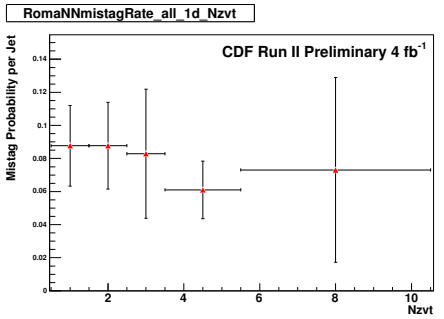
(c) Mistag probability per jet as a function of z of primary vertex.



(d) Mistag probability per jet as a function of Jet Eta.

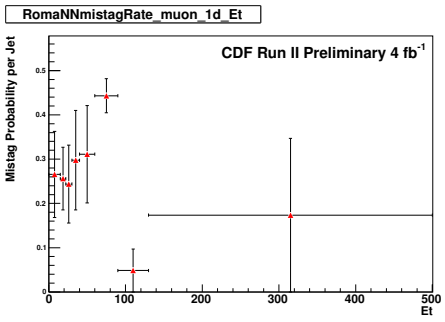


(e) Mistag probability per jet as a function of sumET of the event.

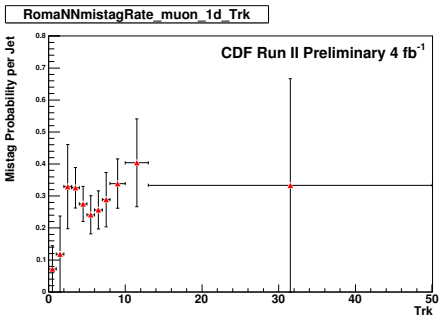


(f) Mistag probability per jet as a function of number of primary vertices.

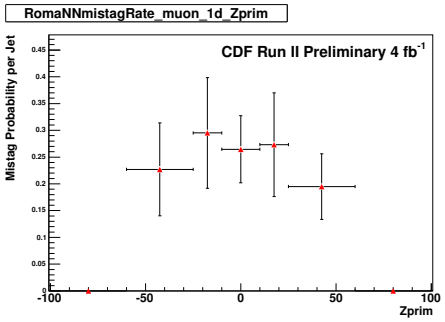
Figure 5.28: Mistag probability per jet as a function of matrix variables in a jet sample, obtained from W $\rightarrow e\nu$ and 2 jets event candidates in data, for UltraLooseRomaNN.



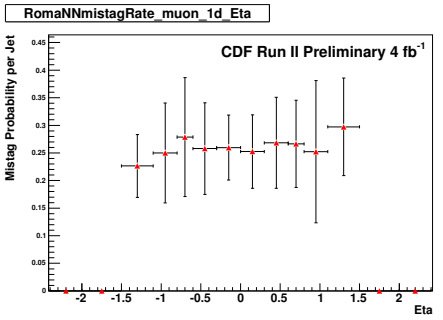
(a) Mistag probability per jet as a function of Jet E_T.



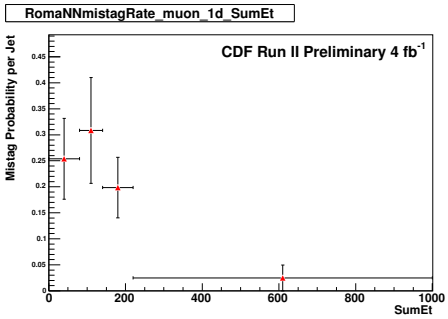
(b) Mistag probability per jet as a function of track multiplicity.



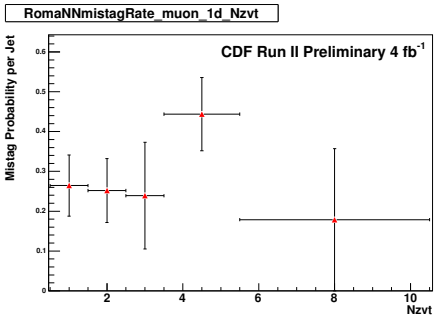
(c) Mistag probability per jet as a function of z of primary vertex.



(d) Mistag probability per jet as a function of Jet Eta.

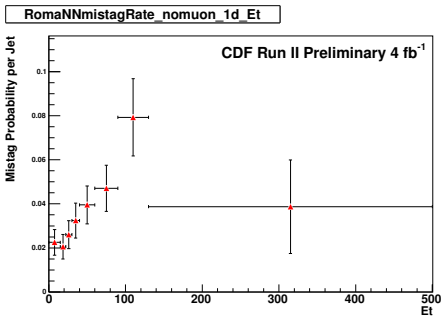


(e) Mistag probability per jet as a function of sumET of the event.

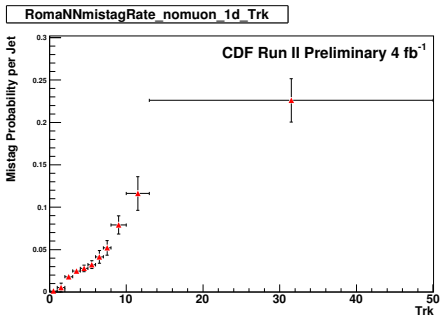


(f) Mistag probability per jet as a function of number of primary vertices.

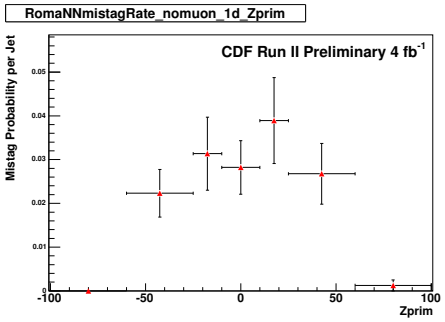
Figure 5.29: Muon-jets Mistag probability per jet as a function of matrix variables in a jet sample, obtained from W $\rightarrow e\nu$ and 2 jets event candidates in data, for UltraLooseRomaNN.



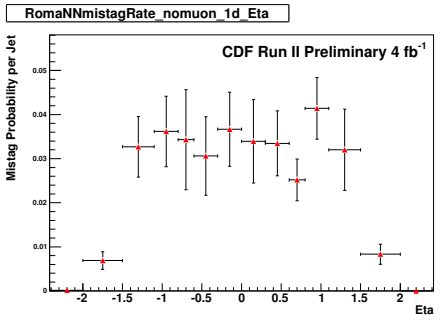
(a) Mistag probability per jet as a function of Jet E_T.



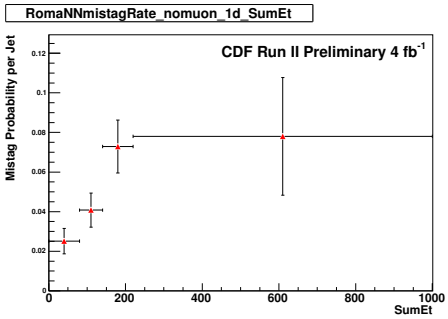
(b) Mistag probability per jet as a function of track multiplicity.



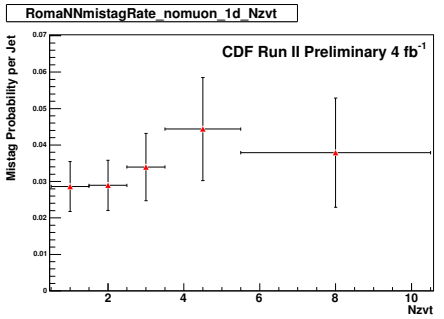
(c) Mistag probability per jet as a function of z of primary vertex.



(d) Mistag probability per jet as a function of Jet Eta.

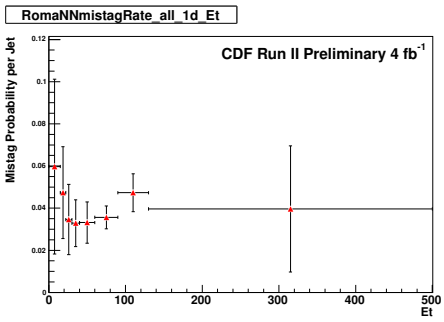


(e) Mistag probability per jet as a function of sumET of the event.

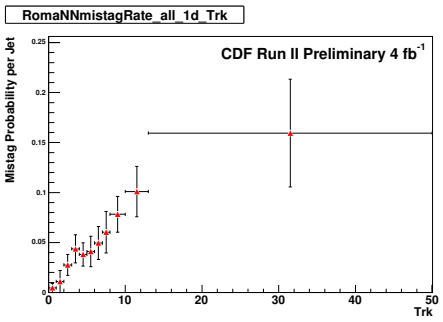


(f) Mistag probability per jet as a function of number of primary vertices.

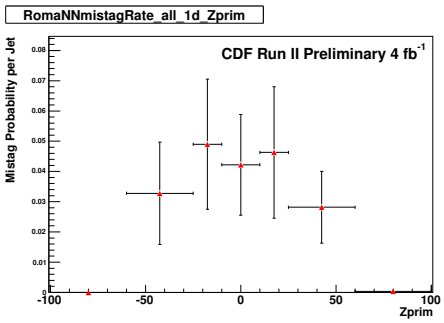
Figure 5.30: NoMuon-jets Mistag probability per jet as a function of matrix variables in a jet sample, obtained from W $\rightarrow e\nu$ and 2 jets event candidates in data, for UltraLooseRomaNN.



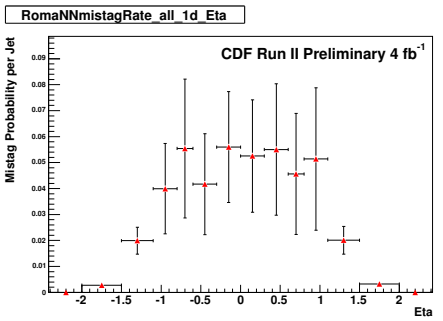
(a) Mistag probability per jet as a function of Jet E_T.



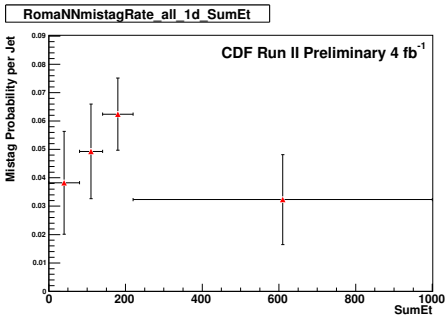
(b) Mistag probability per jet as a function of track multiplicity.



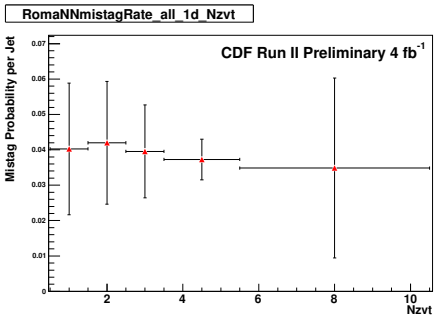
(c) Mistag probability per jet as a function of z of primary vertex.



(d) Mistag probability per jet as a function of Jet Eta.

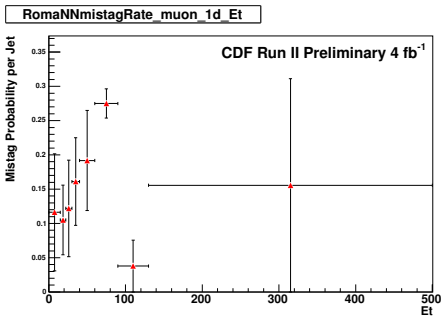


(e) Mistag probability per jet as a function of sumET of the event.

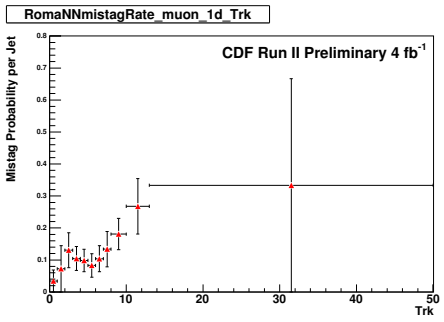


(f) Mistag probability per jet as a function of number of primary vertices.

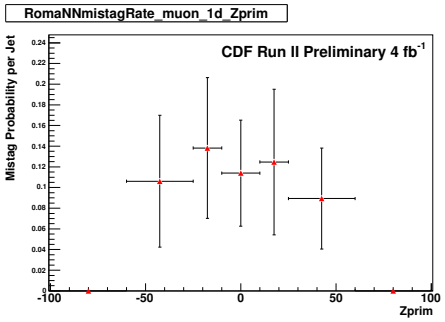
Figure 5.31: Mistag probability per jet as a function of matrix variables in a jet sample, obtained from W $\rightarrow e\nu$ and 2 jets event candidates in data, for LooseRomaNN.



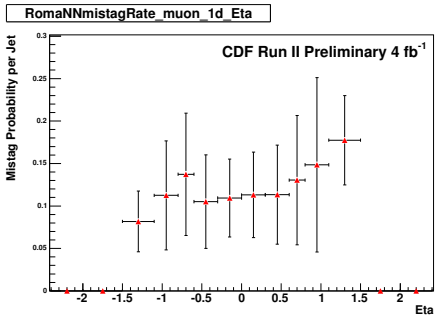
(a) Mistag probability per jet as a function of Jet E_T.



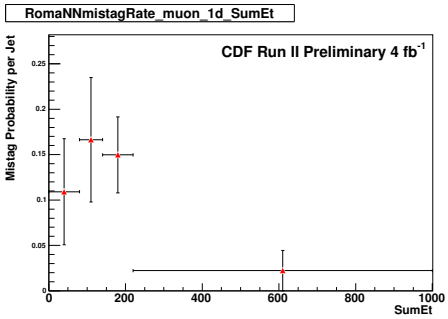
(b) Mistag probability per jet as a function of track multiplicity.



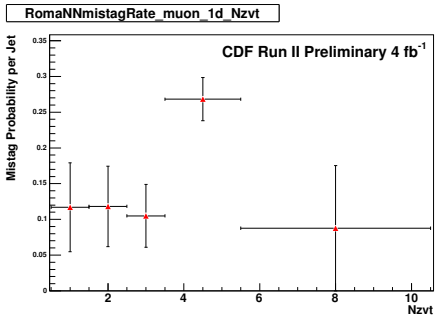
(c) Mistag probability per jet as a function of z of primary vertex.



(d) Mistag probability per jet as a function of Jet Eta.

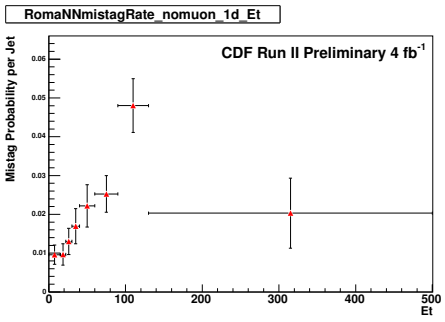


(e) Mistag probability per jet as a function of sumET of the event.

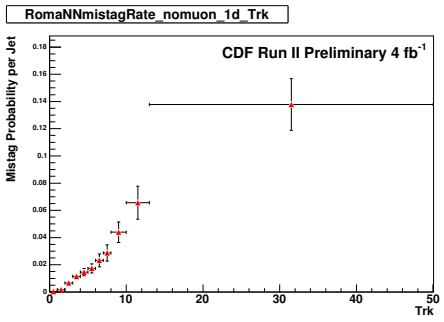


(f) Mistag probability per jet as a function of number of primary vertices.

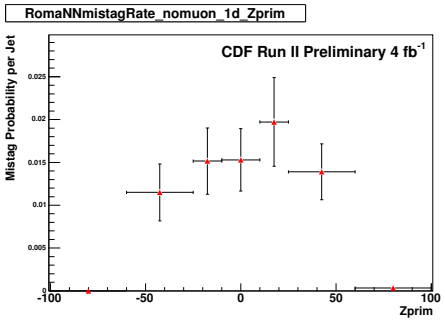
Figure 5.32: Muon-jets Mistag probability per jet as a function of matrix variables in a jet sample, obtained from W $\rightarrow e\nu$ and 2 jets event candidates in data, for LooseRomaNN.



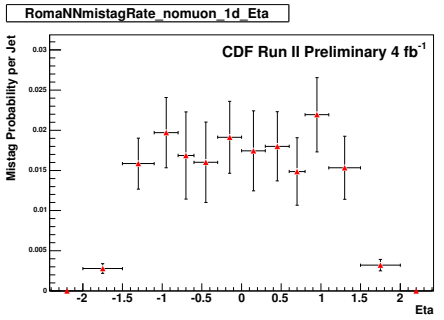
(a) Mistag probability per jet as a function of Jet E_T.



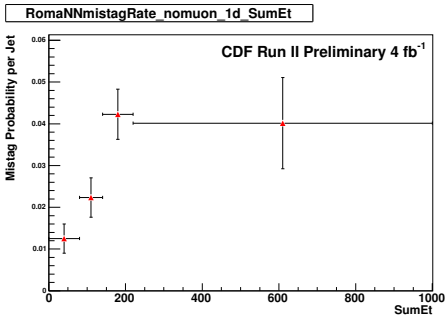
(b) Mistag probability per jet as a function of track multiplicity.



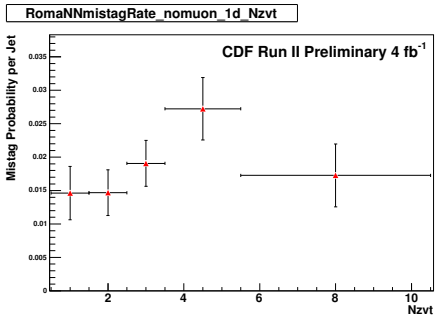
(c) Mistag probability per jet as a function of z of primary vertex.



(d) Mistag probability per jet as a function of Jet Eta.

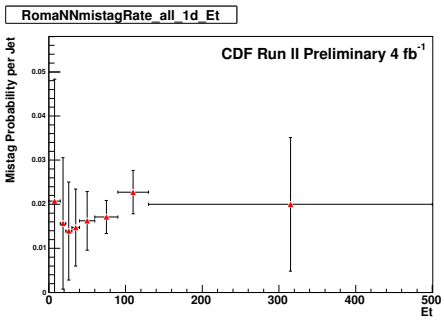


(e) Mistag probability per jet as a function of sumET of the event.

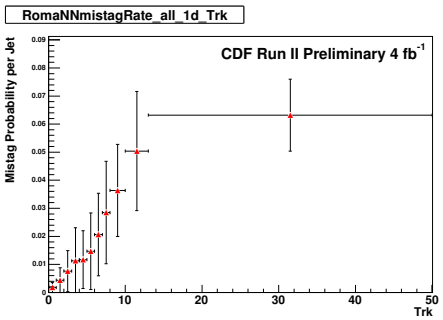


(f) Mistag probability per jet as a function of number of primary vertices.

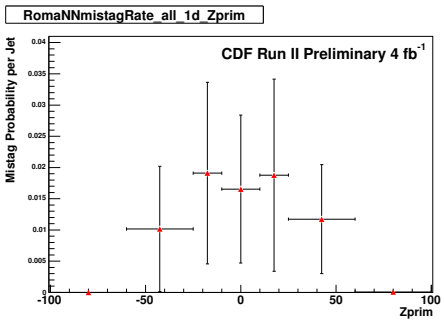
Figure 5.33: NoMuon-jets Mistag probability per jet as a function of matrix variables in a jet sample, obtained from W $\rightarrow e\nu$ and 2 jets event candidates in data, for LooseRomaNN.



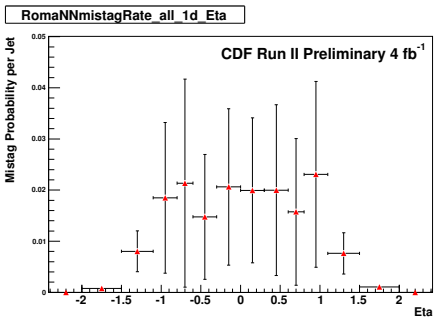
(a) Mistag probability per jet as a function of Jet E_T.



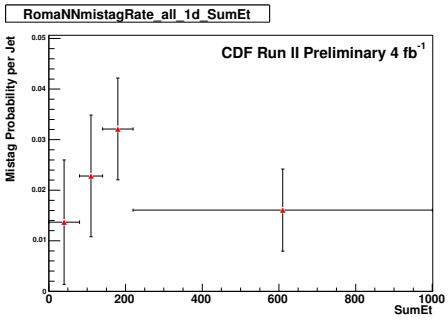
(b) Mistag probability per jet as a function of track multiplicity.



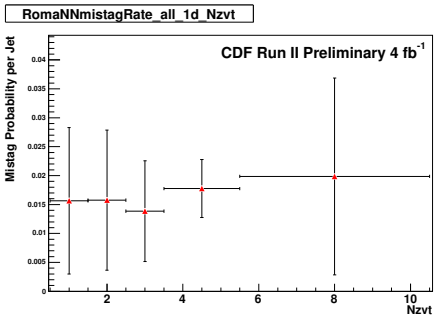
(c) Mistag probability per jet as a function of z of primary vertex.



(d) Mistag probability per jet as a function of Jet Eta.

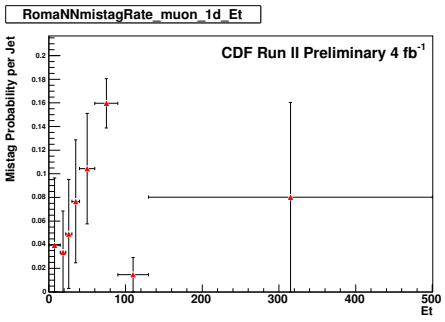


(e) Mistag probability per jet as a function of sumET of the event.

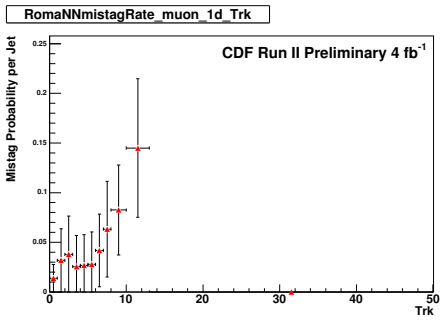


(f) Mistag probability per jet as a function of number of primary vertices.

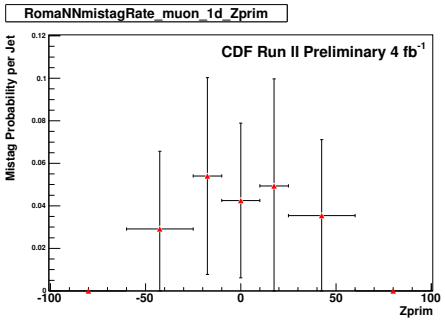
Figure 5.34: Mistag probability per jet as a function of matrix variables in a jet sample, obtained from W $\rightarrow e\nu$ and 2 jets event candidates in data, for TightRomaNN.



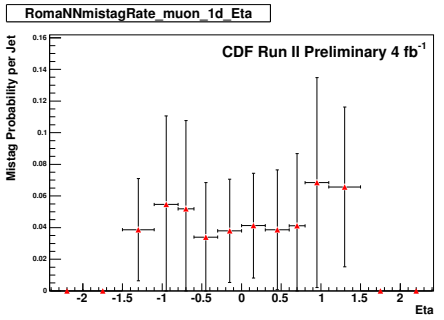
(a) Mistag probability per jet as a function of Jet E_T.



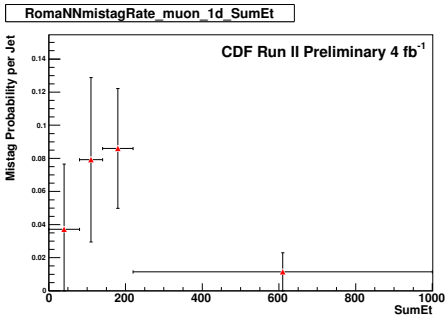
(b) Mistag probability per jet as a function of track multiplicity.



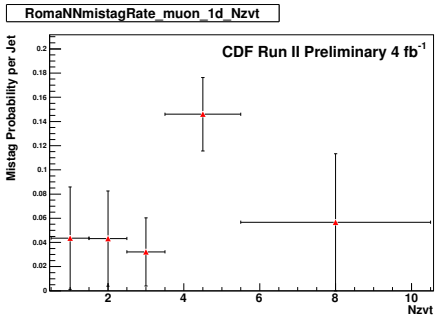
(c) Mistag probability per jet as a function of z of primary vertex.



(d) Mistag probability per jet as a function of Jet Eta.

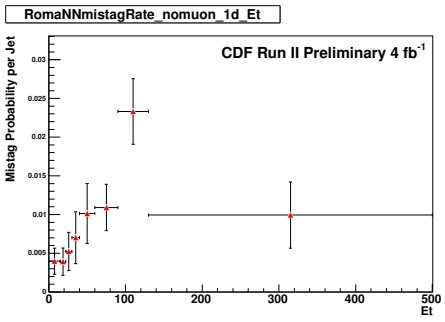


(e) Mistag probability per jet as a function of sumET of the event.

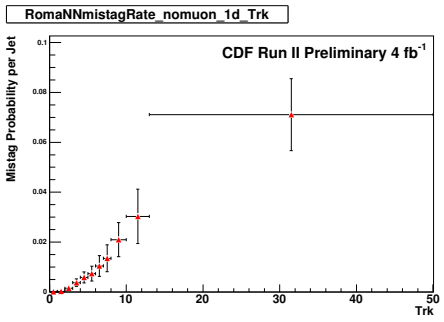


(f) Mistag probability per jet as a function of number of primary vertices.

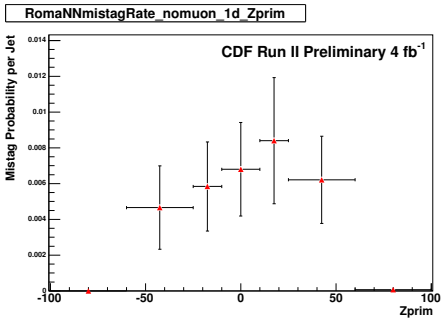
Figure 5.35: Muon-jets Mistag probability per jet as a function of matrix variables in a jet sample, obtained from W $\rightarrow e\nu$ and 2 jets event candidates in data, for TightRomaNN.



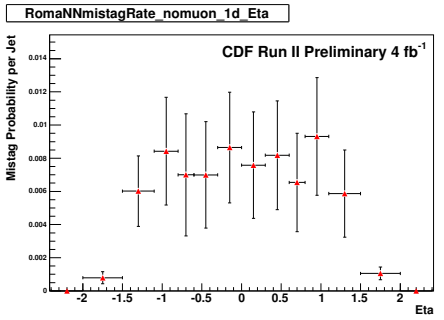
(a) Mistag probability per jet as a function of Jet E_T.



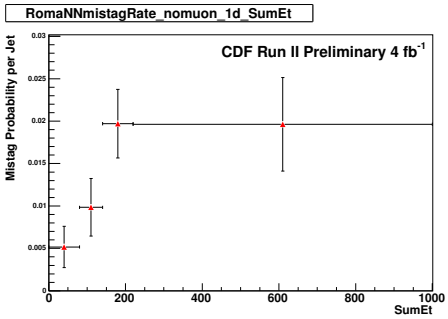
(b) Mistag probability per jet as a function of track multiplicity.



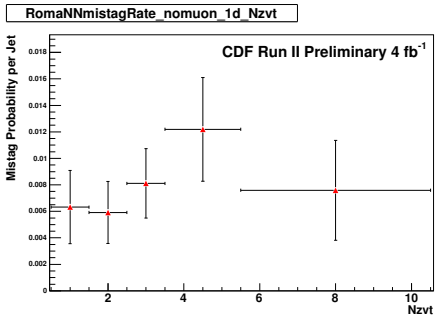
(c) Mistag probability per jet as a function of z of primary vertex.



(d) Mistag probability per jet as a function of Jet Eta.



(e) Mistag probability per jet as a function of sumET of the event.



(f) Mistag probability per jet as a function of number of primary vertices.

Figure 5.36: NoMuon-jets Mistag probability per jet as a function of matrix variables in a jet sample, obtained from W $\rightarrow e\nu$ and 2 jets event candidates in data, for TightRomaNN.

5.3.6 Misidentification Rate Measurement Conclusion

We have measured the overall tag rates for the RomaNN tagger. With these overall tag rates, we estimate the mistag rate with a correction for the contribution from heavy flavor based on the difference between the positive and negative tag rates of the SecVtx tagger.

The tag rates are represented by a tag rate matrix of the same style as was used in SecVtx. As we find different tag rates for jets with and without muons, we describe their tag rates separately.

The mistag rate error depends on the jet sample used, and is calculated jet-by-jet using the mistag matrices. As an example, the aggregate relative error of the mistag rates for the RomaNN tagger in a particular sample of jets obtained from electron plus 2 tight-jets is 31%, 46%, 80% for UltraLooseRomaNN, LooseRomaNN, and TightRomaNN respectively.

5.4 Comparisons of RomaNN with SecVtx

We briefly compare RomaNN and SecVtx b-taggers in respect of the selection efficiency and mistag rate. As seen in Fig. 5.37, for the Tight and Loose operating points, the two taggers behave essentially the same. But the advantage of RomaNN is that it has an UltraLoose operating point, which allows for greater b-jet acceptance. Using the UltraLoose Roma tagger in the double-tag selection provides a 55% increase in acceptance over the standard double-tag selections using Tight SecVtx.

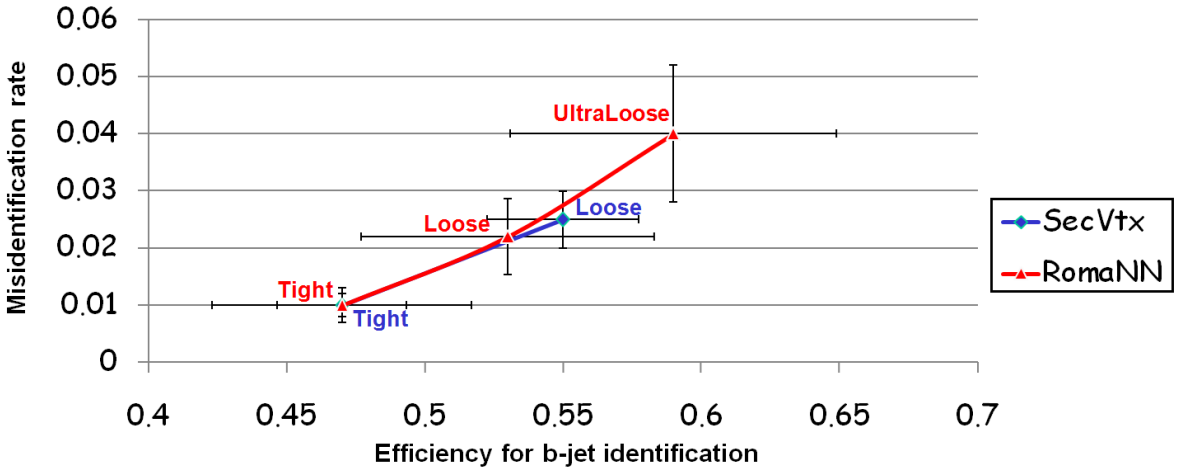


Figure 5.37: Comparison of the RomaNN and SecVtx b-taggers.

Chapter 6

Using The Top Quark Pair Production Cross Section As Verification and Calibration

In chapter 5 we described how a new algorithm to separate b-jets from the other flavors has been adapted to become another standard CDF b-tagging algorithm, and had its efficiency and misidentification rate calibrated. This chapter describes how those calibrations are validated using the pair production cross section of top quarks, a quantity already well measured at CDF, as a standard candle to further our confidence in the calibrations of the new b-tagging algorithm.

6.1 Top Quark Pair Production Cross Section Measurement with Tight RomaNN

The pair production cross section of top quarks is a quantity that has been measured previously [50] using the SecVtx algorithm, to identify the b-jets coming from the decay products of the top quarks. This measurement is a test application of RomaNN

by using it in place of SecVtx, using the operating point that is the most similar to the previous measurement to avoid additional complications.

The previous measurement used SecVtx in the "tight" operating point, and in this measurement we use the RomaNN in the "tight" operating point, since from Fig. 5.37 we see that they have similar efficiencies and mistags.

This cross section is calculated with the formula:

$$\sigma_{t\bar{t}} = \frac{N_{data} - N_{bkg}}{A \cdot \epsilon \cdot L} \tag{6.1}$$

where, N_{data} is the amount of collected data in the signal region, N_{bkg} is the predicted background content, A is the acceptance of $t\bar{t}$ events before any application of b-tagging, ϵ is the b-tag efficiency, and L is the luminosity. This quantity is sensitive to the calibrations of b-tag efficiency and misidentification rate, since it directly involves the efficiency, and uses the misidentification rate in predicting the background content.

Since this measurement shares a lot of feature with the WZ search, a lot of details are shown in chapter 7 instead.

To measure the cross section, we construct a likelihood based upon the data yields, and the predicted yields as a function of cross section. We extract the measured value and statistical uncertainty by minimizing this likelihood. Systematic uncertainties are calculated by varying the parameters one by one and re-performing the measurement.

6.1.1 Event Level Selection

From the events stored by the triggers in Table 7.1, we select events consistent with containing a W boson decay to $e\nu$ or $\mu\nu$. We require a single electron or a muon with $P_T \geq 20$ GeV/c in the pseudo-rapidity ranges of electron $|\eta| < 1.1$ (CEM), muon $|\eta| < 0.6$ (CMUP) and $0.6 < |\eta| < 1.0$ (CMX), where the selection criteria

from [18],[19] are applied. We require missing transverse energy $\not\!\!E_T \geq 20$ GeV and at least three jets with $E_T \geq 20$ and $\eta < 2.0$, where the reconstruction described in chapter 4 is performed. We require at least one jet to be identified as a b-jet candidate (tagged) by the Tight RomaNN algorithm described in chapter 5. In addition, we require the events to have $H_T \geq 230$ GeV, where H_T is the scalar sum of the transverse components of the lepton, $\not\!\!E_T$, and the jets.

6.1.2 Signal and Background Content Estimation

The signal and background content estimation method is described in section 7.3. Figure 6.1 and Table 6.1 show the signal and background estimates used in our top quark pair production cross section measurement utilizing 4.3 fb^{-1} of collected data.

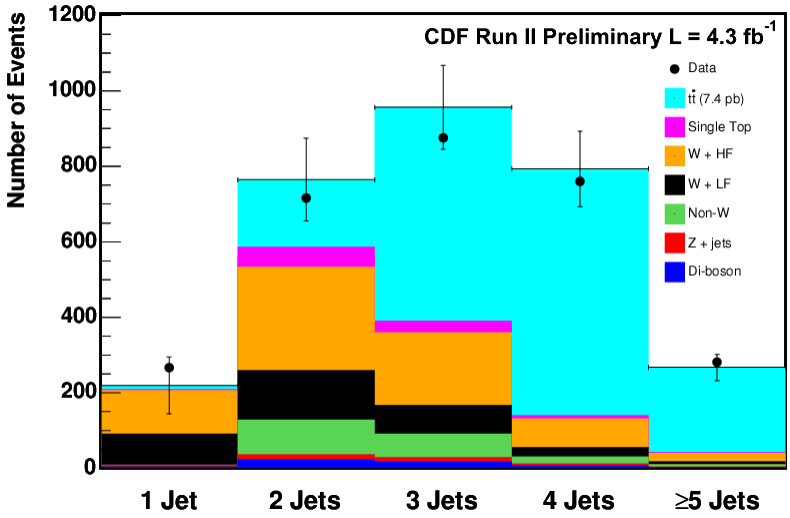


Figure 6.1: Predicted and observed for ≥ 1 Tag, HT ≥ 230 GeV, and $\not\!\!E_T$ 20 GeV, showing good agreement across jet multiplicities.

6.1.3 Calculating the Cross Section

With the background estimate in hand it would appear straightforward to calculate the cross section. But because the background estimate is dependent on the top

Process	1jet	2jets	3jets	4jets	≥5jets
Pretag Data	6411	7785	4617	2080	633
Top	10.3 ± 1.6	175.0 ± 25.3	557.6 ± 80.0	644.1 ± 91.3	221.9 ± 32.0
WW	3.4 ± 1.2	18.5 ± 4.2	13.0 ± 2.8	5.3 ± 1.2	1.8 ± 0.4
WZ	1.0 ± 0.2	5.3 ± 0.8	3.9 ± 0.6	1.7 ± 0.3	0.5 ± 0.1
ZZ	0.1 ± 0.0	0.6 ± 0.1	0.8 ± 0.1	0.4 ± 0.1	0.1 ± 0.0
Single Top s-channel	1.3 ± 0.2	26.9 ± 3.0	15.4 ± 1.7	4.4 ± 0.5	1.0 ± 0.1
Single Top t-channel	0.4 ± 0.1	25.4 ± 2.9	15.3 ± 1.6	4.3 ± 0.4	0.7 ± 0.1
Z+jets	3.6 ± 1.9	12.2 ± 3.3	11.7 ± 2.5	4.8 ± 1.0	1.4 ± 0.3
Wbb	50.0 ± 16.0	139.7 ± 44.3	101.2 ± 32.4	40.5 ± 13.7	12.6 ± 4.6
Wcc	34.9 ± 12.3	80.6 ± 27.6	64.3 ± 22.0	27.1 ± 9.7	8.7 ± 3.3
Wcj	31.5 ± 11.1	54.6 ± 18.7	28.3 ± 9.7	9.0 ± 3.2	2.3 ± 0.9
Mistags	81.6 ± 64.9	132.1 ± 48.3	76.6 ± 25.2	26.1 ± 10.5	7.1 ± 3.8
Non-W	47.6 ± 15.3	104.5 ± 31.8	61.8 ± 18.5	18.8 ± 16.0	6.9 ± 6.5
Total Prediction	265.7 ± 77.2	775.3 ± 110.9	949.9 ± 110.3	786.4 ± 98.5	265.0 ± 34.4
Observed	267	716	876	760	281

Table 6.1: Predicted and observed for ≥ 1 Tag, HT ≥ 230 GeV, and $\not\!\!E_T$ 20 GeV, showing good agreement across jet multiplicities.

quark pair production cross section, extracting the cross section is not so simple. To do that we construct a Poisson likelihood where the background's dependence on the signal estimate is taken into account. The likelihood is:

$$-2 \cdot lnL = -2 \cdot (N_{data} \cdot ln(D \cdot \sigma_{t\bar{t}} + B(\sigma_{t\bar{t}})) - ln(N_{data}!) - (D \cdot \sigma_{t\bar{t}} + B(\sigma_{t\bar{t}}))) \quad (6.2)$$

where $D = A \cdot \epsilon \cdot L$ is the denominator of equation 6.1 , N_{data} is the amount of measured data, and $B(\sigma_{t\bar{t}})$ is the background estimate for a given top quark pair production cross section. The likelihood is calculated for several values of the cross section and the resulting points are fit to a second order polynomial. The minimum of this curve is taken as the measured value. The result for our selection, $H_T \geq 230$ GeV and $\not\!\!E_T \geq 20$ GeV, is $\sigma_{t\bar{t}} = 6.88 \pm 0.29_{stat}$ pb, and the fit is shown in Figure 6.2. This result assumed a top quark mass of 172.5 GeV.

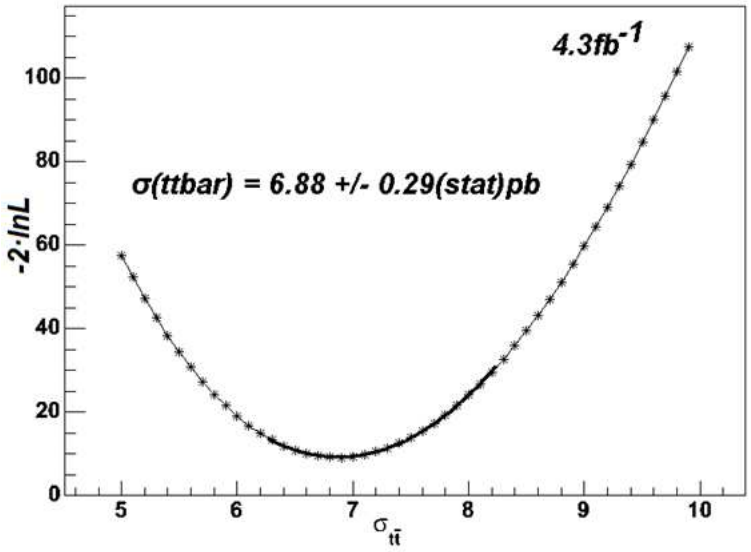


Figure 6.2: Likelihood Curve For Measured Cross Section

6.1.4 Systematics

Systematic uncertainties in our measurement are calculated by varying a given parameter within its uncertainty and redoing the entire measurement. The sources of systematic errors are explained in more detail in [31]. The total systematic error is 0.94 pb.

6.1.5 Top Quark Pair Production Cross Section Measurement with Tight RomaNN Conclusion

Assuming a top quark mass of 172.5 GeV and using ≥ 1 Tight RomaNN Tagged events from 4.3 fb^{-1} of data, we find the top quark pair production cross section to be $\sigma_{t\bar{t}} = 6.88 \pm 0.29_{stat} \pm 0.94_{sys}$ pb.

That this measurement is consistent with the combination of CDF top quark pair production measurements that assumed a top quark mass of 172.5 GeV, at $\sigma_{t\bar{t}} = 7.50 \pm 0.48$ pb [32], is a sign that the calibrations for the efficiencies and

misidentification rates were accurate for the Tight RomaNN.

6.2 Top Quark Pair Production Cross Section Measurement with UltraLoose RomaNN

The search for the WZ shares many of the event level selection requirements as the measurement of the top quark pair production cross section. They share the same lepton selection, missing transverse energy requirement, and the same jet requirements. They differ in that the WZ search uses only events with exactly two jets both tagged, whereas the measurement of the top quark pair production cross section uses events with three or more jets at least one of which is tagged.

The requirement of both jets tagged (double tagged) in the WZ search increases the signal-to-background ratio, and within the double tagged frame work we want to increase the signal acceptance. To do that we use the UltraLoose operating point. Therefore we verify that the pair production cross section of top quarks can be measured correctly with the UltraLoose RomaNN also.

Essentially repeating section 6.1, and because the UltraLoose operating point provides a tagged sample less pure in b-jets, we need to use even tighter cuts. We raise the H_T requirement to $H_T \geq 250$ GeV, and raise the $\not\!\!E_T$ requirement to $\not\!\!E_T \geq 25$ GeV we obtain a cleaner sample.

The signal and background estimate are shown in Figure 6.3, and the cross section is measured to be $\sigma_{t\bar{t}} = 8.9 \pm 0.4_{stat}$ pb. This is not in agreement with the combination of CDF top quark pair production measurements that assumed a top quark mass of 172.5 GeV, at $\sigma_{t\bar{t}} = 7.50 \pm 0.48$ pb [32]. So we conclude that the b-tag scalefactor measured for UltraLoose RomaNN using the dijet sample in chapter 5 requires an additional calibration to extrapolate to this lepton + $\not\!\!E_T$ + jets sample.

So instead of using the top quark pair production cross section as a verification, we use it as a calibration point. We will apply a simple multiplicative factor to b-tag

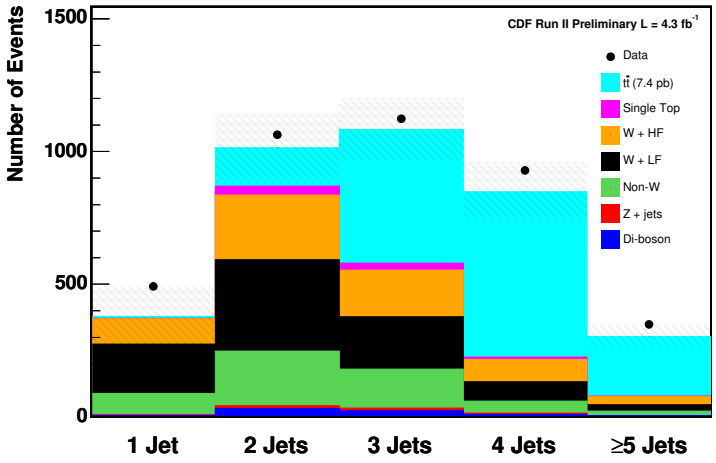


Figure 6.3: Predicted and observed for ≥ 1 Tag, showing disagreement across jet multiplicities.

scalefactor measured for UltraLoose RomaNN using the dijet sample, and with the multiplicative factor such that our measurement of the top quark pair production cross section comes out matching the combination of CDF top quark pair production measurements. We found this multiplicative factor to be 1.155 ± 0.065 (where the ± 0.065 covers the ± 0.48 pb uncertainty on our calibration point). Using this multiplicative factor, the signal and background estimate are shown in Figure 6.4, and the cross section is measured to be $\sigma_{t\bar{t}} = 7.5 \pm 0.4_{stat}$ pb.

We then proceed to the sample used for the WZ search, where we require at least two jets tagged. Because having both jets tagged significantly improves the b-purity, we can eliminate the H_T requirement and revert the $\not\!\!E_T$ requirement to $\not\!\!E_T \geq 20$ GeV. We obtain a signal and background estimate as shown in Figure 6.5. It shows excellent agreement between data and prediction. And using this double tagged sample, the cross section is measured to be $\sigma_{t\bar{t}} = 7.6 \pm 0.4_{stat}$ pb, again in excellent agreement with the combination of CDF top quark pair production cross section measurements.

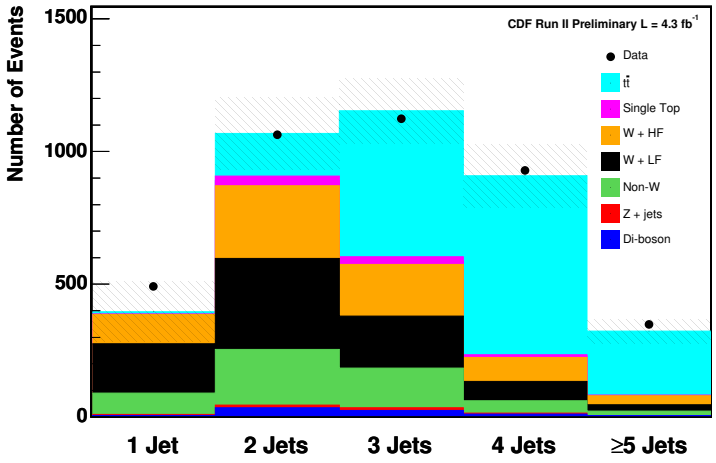


Figure 6.4: Predicted and observed for $\geq$ 1 Tag, with the calibration factor of 1.155 ± 0.065 applied, showing good agreement across jet multiplicities.

Having confidence in the calibration of the new tool, we went ahead to search for the WZ, described in chapter 7.

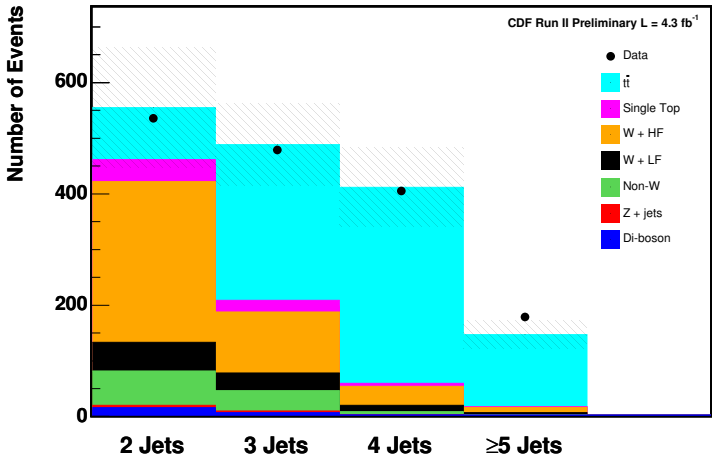


Figure 6.5: Predicted and observed for $\geq$ 2 Tag, with the calibration factor of 1.155 ± 0.065 applied, showing good agreement across jet multiplicities.

Chapter 7

Composition of Selected Data Sample and Discrimination of WZ From Other Processes

The search for a Standard Model Z boson produced in association with a W boson is a signature based analysis, where the W boson decays into a lepton (l) and a neutrino (ν) and the Z boson decays into two b quarks ($b\bar{b}$) that hadronize to form two b-jets. We have now briefly described the reconstruction of the leptons, missing transverse energy and jets in chapter 4. We have discussed in more detail in chapter 5 my own work on calibration of the efficiency and false positive rate for a new algorithm to identify b jets with higher efficiency. In this chapter, we describe the estimate of the composition of the data sample selected with this new b jet identification algorithm, where the method used requires the calibrations from chapter 5. We also describe the development of a multivariate discriminant used to improve the sensitivity of our search for WZ.

Trigger Level	ELECTRON_CENTRAL_18	MUON_CMUP_18	MUON_CMX_18
Level 1	$E_T \geq 8$ GeV $P_T \geq 8$ GeV/c $E_{HAD}/E_{EM} \leq 0.125$	$P_T \geq 6$ GeV/c $P_T \geq 4.1$ GeV/c CMP stub	$E_T \geq 6$ GeV/c $P_T \geq 8.3$ GeV/c
Level 2	$E_T \geq 16$ GeV $P_T \geq 8$ GeV/c $E_{HAD}/E_{EM} \leq 0.125$	$P_T \geq 8.3$ GeV/c CMU stub CMP stub	$P_T \geq 10.1$ GeV/c CMX stub
Level 3	$E_T \geq 18$ GeV $P_T \geq 9$ GeV/c $E_{HAD}/E_{EM} \leq 0.125$ Lshr ≤ 0.4 $\Delta Z_{CES} \leq 8$ cm	$P_T \geq 18$ GeV/c $\Delta X_{CMU} \leq 20$ cm $\Delta X_{CMP} \leq 20$ cm	$P_T \geq 18$ GeV/c $\Delta X_{CMX} \leq 10$ cm

Table 7.1: High-P_T Lepton Trigger Requirements.
ELECTRON_CENTRAL_18, MUON_CMUP_18 and MUON_CMX_18 trigger paths require events that meet the listed criteria., where E_{HAD}/E_{EM} is the ratio of energies deposited hadronic and electromagnetic calorimeters, CMU/P/X stub indicate hits in the respective chambers, Lshr (Lateral Shower Sharing) [39] is a measure of how closely the energy deposition in the adjacent towers match the expected value from the test beam electron shower, $\Delta Z_{CES}(\Delta X_{CMU/P/X})$ are the distances between the COT extrapolated track and the shower(stub) position as measured by the CES(CMU/P/X).

7.1 Data Collection Samples

7.1.1 Data Samples

We use CDF Run II data in the run range 138425–274055 collected over the period 02/2002-03/2009. The events recorded are triggered by the high-P_T lepton trigger paths as summarized in Table 7.1, these are events which contain an electron(muon) with energy(momentum) greater than 18 GeV(/c) meeting a minimal set of quality criteria. Table 7.2 lists the calibration factors for the lepton identification efficiencies and the trigger efficiencies used in the analysis [18], [19]. [38]. The total integrated luminosity is 4.3 fb^{-1}.

Our WZ signal model comes from the Higgs Discovery Group Diboson Monte Carlo (MC) samples generated with Pythia [36] simulated with a GEANT-based

Lepton type	scale factor	trigger efficiency
CEM	0.977 ± 0.005	0.962 ± 0.004
CMUP	0.898 ± 0.008	0.914 ± 0.007
CMX	0.973 ± 0.009	0.937 ± 0.010

Table 7.2: Lepton identification Scale Factors and Trigger Efficiencies.

model [40] of the detector response, and reconstructed in the same way as the data.

Our background models are composed of a number of components. The W and Z plus light-flavor and heavy-flavor jet processes are modeled using ALPGEN [41], which is a generator that has been developed to model W or Z production with additional jets, and showered with PYTHIA. Likewise, the single-top contribution is modeled using parton-level events generated by MadEvent [42], which is a generator that has been developed to model single top production, and showered through PYTHIA. The rest of the background processes, including the $t\bar{t}$, WW, and ZZ processes were generated with PYTHIA. For backgrounds involving a top quark, the top mass was set to 172.5 GeV/c^2.

7.2 Event Level Selection

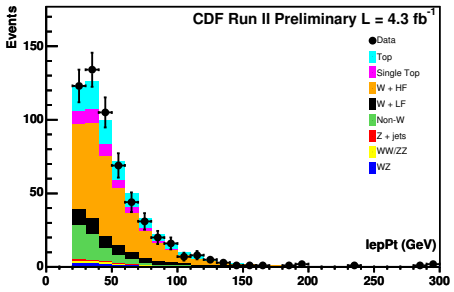
From the events stored by the triggers described in Table 7.1, we select events consistent with containing a W boson decay to $e\nu$ or $\mu\nu$. We require a single electron or a muon with $P_T \geq 20$ GeV/c in the pseudo-rapidity ranges of electron $|\eta| < 1.1$ (CEM), muon $|\eta| < 0.6$ (CMUP) and $0.6 < |\eta| < 1.0$ (CMX), where the selection criteria from [18],[19] are applied. We require missing transverse energy $\not\!\!E_T \geq 20$ GeV and exactly two jets with $E_T \geq 20$ and $\eta < 2.0$, where the reconstruction described in chapter 4 is performed. We require both jets to be identified as b jet candidates (tagged) by the UltraLoose RomaNN algorithm described in chapter 5.

7.3 Signal and Background Content Estimation

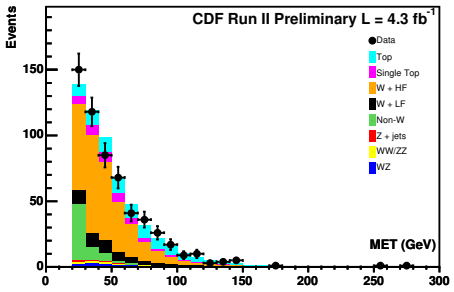
We estimate the composition of our selected data sample in a sequence of steps, in a method (the CDF jargon calls it method II) used by many physics analyses at CDF [43]. First, we estimate the contribution from processes that have precise theoretical predictions for their production cross section and kinematics. These processes include our signal WZ, as well as other diboson production (WW, ZZ), single top quark production, and top quark pair production. Secondly, we estimate the contribution from multi-jet processes (non-W) with a data-based method. At this point, the remaining events are assumed to be from W boson production with associated jets. In this way we avoid using the imprecise theoretical predictions for the production cross section of W boson with associated jets. We then estimate the contribution from W boson production with associated heavy flavor jets (W+hf), where the jets will be tagged with high efficiency. Finally, we estimate the contribution from W boson production with associated light flavor jets that have been mis-tagged. We describe each of these steps in the following sections. Table 7.3 provides the estimate of the sample composition. Figure 7.1 compares the data with the predicted sample composition for the transverse energy of the lepton, jets, and the missing transverse energy.

7.3.1 Electroweak (including WZ) and Top Content

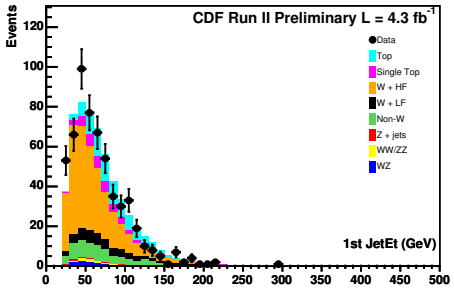
Several electroweak processes contribute to the selected data sample such as WZ, WW, ZZ, Z+jets, and top events. They exist in the sample because each process can produce signatures consistent with a lepton and neutrino and a number of jets, arising both from correctly identified and mis-identified objects in the CDF detector. We estimate their contribution to the composition of the sample before (pretag) and after tagging (tag). The CDF detector only accepts some of the events produced in collisions due to the geometrical limitations of the subdetectors. In addition, the



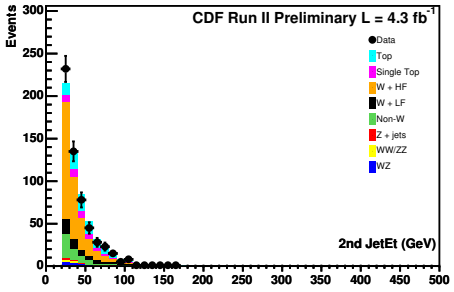
(a) P_T of the lepton.



(b) Missing Transverse Energy.



(c) E_T of the leading Jet.



(d) E_T of the second Jet.

Figure 7.1: Predicted and observed kinematic distributions, normalized to equal area.

efficiencies of reconstruction and identification are not 100%. Therefore, the number of expected events is as follows:

$$N_{ewk}^{pretag} = \sigma \cdot A \cdot \int \mathcal{L} \cdot dt \tag{7.1}$$

$$N_{ewk}^{tag} = \sigma \cdot A \cdot \epsilon \cdot \int \mathcal{L} \cdot dt \tag{7.2}$$

where σ (Table 7.4) is the theoretical cross sections, $\int \mathcal{L} \cdot dt$ is the total luminosity, A is the pre-tagged selection acceptance derived from Monte Carlo, and ϵ is the tagged selection efficiency.

We apply the event selection criteria from section 7.2 to simulated samples to estimate the acceptance A for these processes. We apply the calibration factors and trigger efficiencies in Table 7.2 to correct for deficiencies in the simulation.

Table 7.4 shows the theoretical cross sections used along with their uncertainties [44],[45],[46],[47],[48]. For backgrounds involving a top quark, the top mass was set to 172.5 GeV/c^2.

7.3.2 Non-W Based Background Estimate

The background from copious QCD multi-jet production (without a W boson, so we call it non-W) comes from cases where a jet is mis-identified as a lepton thus firing a trigger, and where jet energies mis-measurement result in significant missing transverse energy. A particular problem is multi-jet production where some of the jets are from b hadrons, as the semileptonic decays of b hadrons produce electrons or muons (albeit inside jets so most will not pass the isolation criteria for leptons), and neutrinos that result in missing energy.

We model this non-W background from a control data sample that has the same event selection criteria as our search except that some of the lepton identification criteria are reversed. Since this data sample is enriched in multi-jet production, we can use this sample to obtain a model of the missing transverse energy used to fit for the QCD fractions, as well as a model of the other kinematic variables used in the multivariate analysis described in section 7.5.

The next step in the background estimation is to determine the contribution of non-W in the pretag data sample. We make an assumption that the data contains the electroweak content shown in Table 7.3, W+jets (W+bb, W+cc, W+c, W+lf) events, and non-W events. We perform a fit of the $\not\!\!E_T$ distribution to obtain the fraction in the pretag data sample that are from non-W events. We take the data pretag sample, and constrain for the small electroweak content estimated in section 7.3.1, then using MC template for W+jets and the template for the non-W, fit

for the fractions of W+jets events and non-W events using a binned likelihood. The results of the fits for this analysis are shown in Figures 7.2.

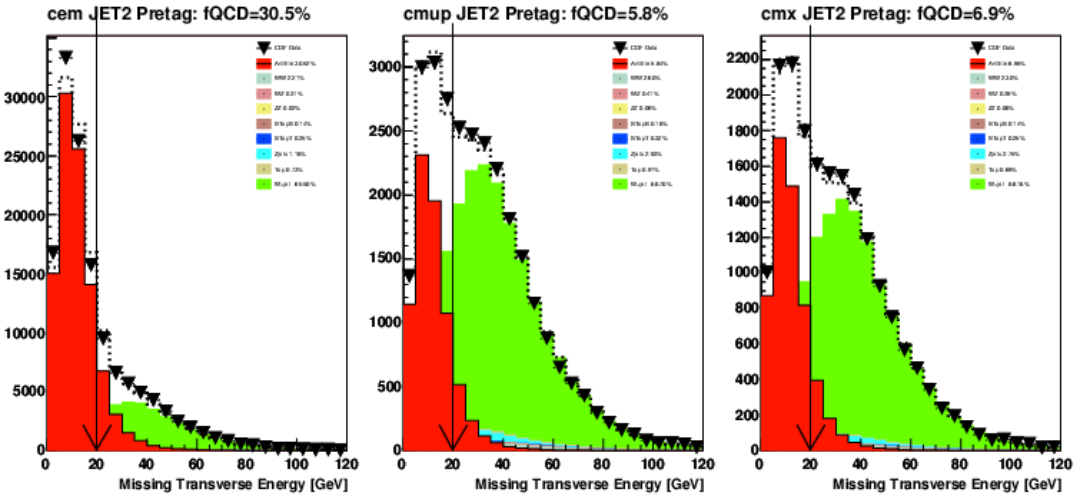


Figure 7.2: Fits for the non-W contribution in the pretag data sample, for CEM/CMUP/CMX triggers.

Once the pretag QCD fraction is obtained the normalization is given as:

$$N_{QCD}^{pretag} = F_{QCD} \cdot N^{pretag} \tag{7.3}$$

The tagged QCD fraction fit is performed after the W+jets (W+bb, W+cc, W+c, W+lf) normalizations are obtained. For the QCD template we use the same non-W templates as used in the pretag fits. The other template used is the tagged W+jets template, which is composed of the properly normalized W+bb, W+cc, W+c, W+lf plus the electroweak content weighted by the tagging rates (SF and mistag rate). The results of the fits for this analysis are shown in Figures 7.3.

Once the tagged QCD fraction is obtained the normalization is given as:

$$N_{QCD}^{tag} = F_{QCD} \cdot N^{tag} \tag{7.4}$$

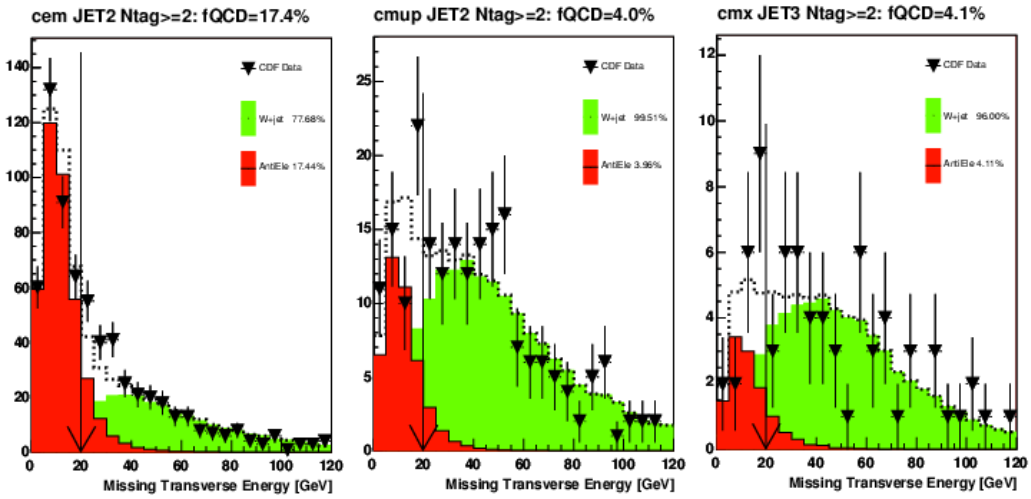


Figure 7.3: Fits for the non-W contribution in the tagged data sample, for CEM/CMUP/CMX triggers.

A 30% uncertainty is taken on the QCD fractions, obtained from studying the difference in the QCD fractions arising from using different QCD models. [50].

7.3.3 W + Heavy Flavor

Having accounted for electroweak, top quark pair production, and multi-jet production, the remaining events are assumed to be from W boson production with associated jets. The number of W+jets events before tagging is given by equation 7.5.

$$N^{pretag}_{W+Jets} = N^{pretag} \cdot (1 - F^{pretag}_{QCD}) - N^{pretag}_{ewk} - N^{pretag}_{t\bar{t}} \tag{7.5}$$

For the tagged estimate, the W+jets sample is broken down into two categories: heavy and light flavor, these two processes produce a tagged jet very differently and therefore requires different treatments in calculating the normalization.

The number of W+jets events after tagging is calculated separately for events with and without heavy flavor jets. For W+light flavor jets, we use the mistag matrix that was calibrated in chapter 5, as described in the next section. For W+heavy

flavor jets (including processes W+bb, W+cc, Wc), we avoid the imprecise prediction for the theoretical cross section by instead using a data driven estimate for the fraction of W+heavy flavor in W+jets.

The contribution from the W+heavy flavor jets is calculated by equation 7.6.

$$N_{W+hf}^{tag} = (N^{pretag} \cdot (1 - F_{QCD}^{pretag}) - N_{ewk}^{pretag} - N_{t\bar{t}}^{pretag}) \cdot f_{HF} \cdot K \cdot \epsilon \qquad (7.6)$$

where f_{HF} is the fraction of events with jets matched to heavy flavor quarks, K is a correction to the Monte Carlo heavy flavor fraction called the "K-factor", and ϵ is the tagging efficiency, which is calibrated using the scale factor described in chapter 5 and using the top quark pair production cross section in chapter 6.

The f_{HF} and b-tag efficiency are calculated from simulated ALPGEN samples for W+bb, W+cc, and W+c. The K-factor is a correction for missing higher-order effects in ALPGEN, and has been measured to be $K = 1.5 \pm 0.3_{stat+sys}$ [51].

7.3.4 Mistags

The final contribution to our selected event sample is from W+light flavor jets. This sample is called mistags because the light flavor jets are mis-identified as b-jets (mis-tagged). The predicted number of background events from W + light flavor (W+lf) processes is:

$$N_{W+lf}^{tag} = \frac{N^{mistag}}{N^{pretag}} \cdot (N^{pretag} - N_{t\bar{t}}^{pretag} - N_{QCD}^{pretag} - N_{W+hf}^{pretag} - N_{ewk}^{pretag}) \qquad (7.7)$$

Where N^{mistag} is the number of mistags in the data sample predicted using the mistag matrix, described previously in chapter 5. The predicted amount of pretag $t\bar{t}$, QCD, W+hf, and Electroweak background events is subtracted from the total pretag sample leaving an estimate for the W+lf fraction. The predicted number of mistagged W+lf events is the W+lf fraction multiplied by the predicted amount of mistagged events from the pretag data.

The mistags of the UltraLoose Roma tagger are much higher than if we had used the Tight SecVtx tagger instead. This is the trade off made in gaining additional b-jet acceptance, which is a reasonable thing to do in the double-tag category because the background had previously a negligible amount of mistags.

7.4 Systematic Uncertainties on WZ Yield

In the selected data sample with 4.3 fb^{-1} of integrated luminosity, the yield of WZ signal events is estimated to be 9.9 ± 2.0 events, using the theoretical cross section from table 7.4. The uncertainty includes the systematic sources described in table 7.5 and is dominated by the uncertainty on the b-tag scale factor. Recall that our sample requires two b-tags, so the uncertainty on the signal yield is proportional to twice the uncertainty on the b-tag efficiency.

1. The jet energy scale (JES) uncertainty is estimated by shifting the JES of the WZ MC sample by $\pm 1\sigma$ as determined by the CDF Joint Physics Group. The difference from the nominal acceptance is taken as the systematic uncertainty.

2. ISR and FSR systematic uncertainty are estimated by changing the parameters related to ISR and FSR from their default values to values constrained by data in a study of ISR in $Z/\gamma \rightarrow \mu^+\mu^-$ events. Half of the difference between the two shifted samples is taken as the systematic uncertainly.

3. PDFs uncertainties are evaluated using the uncertainties calculated from the CTEQ6M/5L [52] and MRST72/75 [53] sets, using standard re-weighting method recommended by CDF Joint Physics Group [54].

4. The b-tagging scale factor uncertainty as described in section 5.2, as well as from the uncertainty in the calibration factor using $t\bar{t}$ in section 6.2.

5. Luminosity measurement contributes an overall 6% rate uncertainty.

Process	Number of Events
Pretag Data	73861
Pretag (after MET cut) WZ	257.8 ± 26.3
Pretag (after MET cut) WW	1680.5 ± 149.9
Pretag (after MET cut) ZZ	27.2 ± 3.3
Pretag (after MET cut) Z+jets	1476.9 ± 131.7
Pretag (after MET cut) Top	531.9 ± 66.1
Pretag (after MET cut) Single Top s-channel	111.4 ± 10.3
Pretag (after MET cut) Single Top t-channel	196.0 ± 15.1
Pretag (after MET cut) Wbb	1939.9 ± 390.7
Pretag (after MET cut) Wcc	4227.6 ± 865.0
Pretag (after MET cut) Wcj	4750.3 ± 976.6
Pretag (after MET cut) Mistags	43204.1 ± 4380.4
Pretag (after MET cut) Non-W	15457.4 ± 4107.2
Tagged WZ	9.9 ± 2.0
Tagged WW	5.1 ± 1.7
Tagged ZZ	1.0 ± 0.2
Tagged Z+jets	3.9 ± 1.2
Tagged Top	93.6 ± 19.1
Tagged Single Top s-channel	28.3 ± 5.1
Tagged Single Top t-channel	11.2 ± 2.2
Tagged Wbb	209.4 ± 70.8
Tagged Wcc	37.4 ± 14.3
Tagged Wcj	42.0 ± 16.1
Tagged Mistags	51.3 ± 18.9
Tagged Non-W	61.9 ± 18.6
Total Prediction	555.0 ± 107.3
Observed	536

Table 7.3: Predicted and observed for $\geq$ 2 UltraLooseRomaNN Tag, and $\not\!\!E_T \geq 20$ GeV

Process	Cross Section
WZ	3.96 ± 0.06 pb
WW	12.4 ± 0.25 pb
ZZ	1.58 ± 0.05 pb
Z+Jets	787.4 ± 50 pb
$t\bar{t}$	7.4 ± 0.8 pb
Single Top t-channel	1.98 ± 0.08 pb
Single Top s-channel	0.88 ± 0.05 pb

Table 7.4: Theoretical Cross Sections [44],[45],[46],[47],[48] For Electroweak and Top Content (with $m_t = 172.5$ GeV/c^2).

JES	ISR/FSR	PDF	b-tag SF	Lumi	Total
0.3%	5.3%	0.4%	19.8%	6%	21.4%

Table 7.5: Systematic uncertainties on signal yield.

7.5 Artificial Neural Network

The selected data sample contains 536 events, with a small predicted contribution from WZ of 9.9 ± 2.0 events. In this section we describe the development of a multivariate discriminant to improve the separation between signal and background.

For this analysis, we used TMVA [55] to construct an artificial neural network discriminant function. We selected a set of 10 kinematic observables, and trained on all of their 1023 combinations to look for the function with the best separation power and with the fewest number of variables used.

We used the same 10000 events to optimize each artificial neural network discriminant function, and the same 30000 events to evaluate its expected separation power. Figure 7.4 shows the best separation power as a function of the number of input observables. We then selected an artificial neural network discriminant function with three input variables: M_{bb}, $|\not\!\!E_T|$, and $\mathrm{P_T}$Imbalance. Figure 7.5 shows the signal and background distributions for the three input variables, and Figure 7.6 shows in this data sample the same variables for the predicted and observed distributions.

M_{bb}: the invariant mass calculated from the two tagged jets.

$|\not\!\!E_T|$: the magnitude of the missing transverse energy.

$\mathbf{P_T}$ Imbalance: the difference between the scalar sum of the p_T of all measured objects and the $\not\!\!E_T$. Specifically, it is calculated as
$p_T(\mathrm{jet}_1) + p_T(\mathrm{jet}_2) + p_T(\mathrm{lep}) - \not\!\!E_T$.

After training is completed, a check against overtraining is done by comparing the training sample shape to that for a test sample which was not used in training. Overtraining occurs when the input parameters are optimized to too few data points, leading to an increase in the neural network's performance on the training sample, but a decrease in performance when tested on any other sample. When overtraining occurs, the training and testing sample shapes diverge [55]. Figure 7.7 shows our

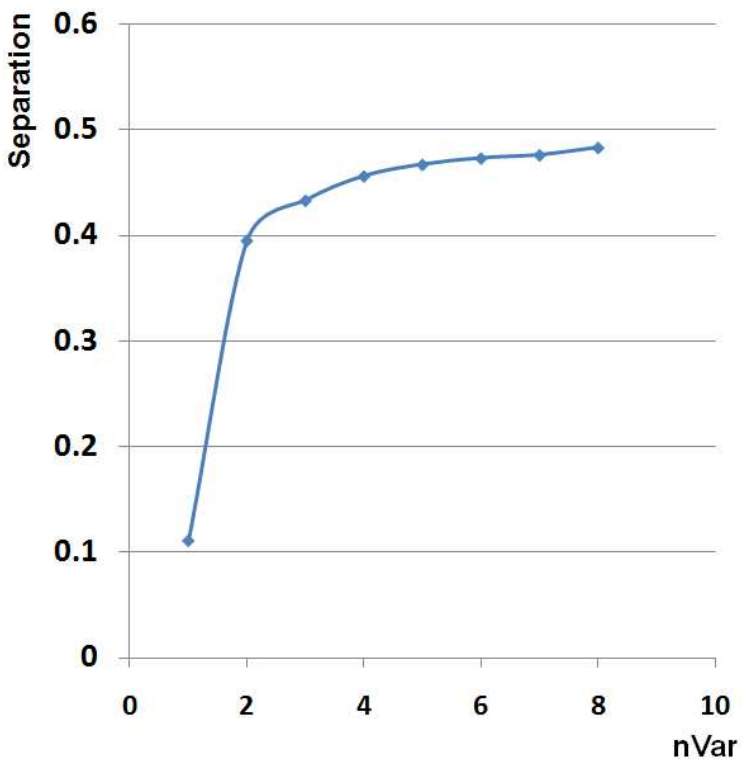


Figure 7.4: Best separation power as a function of the number of input observables.

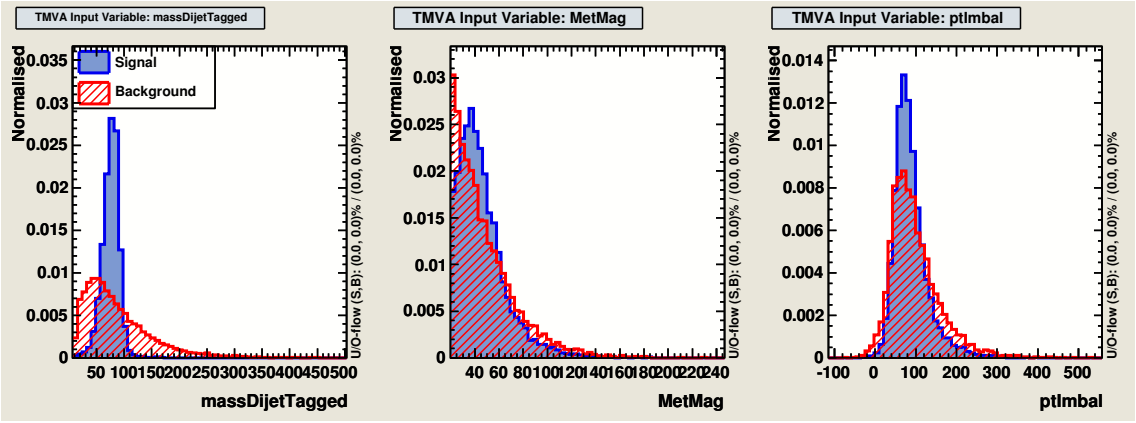
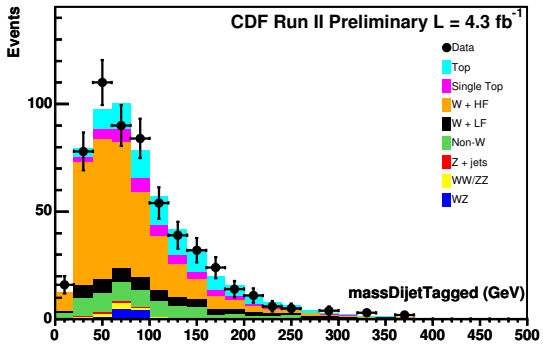


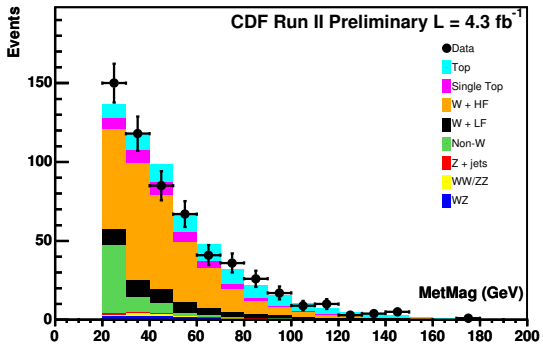
Figure 7.5: Simulated distributions for M_{bb}, MET, and P_T imbalance for WZ signal and background, normalized to unit area.

overtraining check, and we conclude that our neural network discriminant is not overtrained since the responses of the training sample and the test sample are in good agreement.

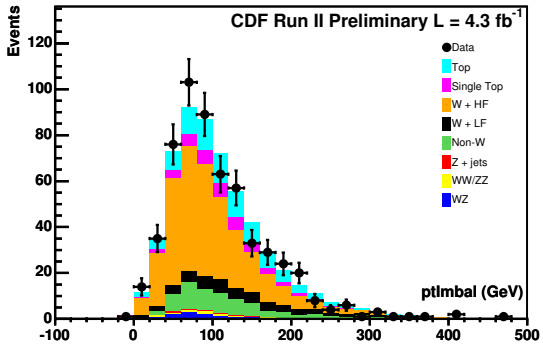
Now we look at the output of each individual background component to see how well they are separated against the signal. Figure 7.8 shows the neural network output for each signal and background process. We see that the neural network does a very good job in distinguishing the signal from non-W and top, because their underlying physics processes are vastly different from that of electroweak diboson production. It does progressively worse on single top and W+hf, because their final state mimics very much the WZ final state where the W boson decays into a lepton (l) and a neutrino (ν) and the Z boson decays into two b quarks ($b\bar{b}$) that hadronize to form two b-jets.



(a) M_{bb}.



(b) $\not\!\!E_T$|.



(c) P_T Imbalance.

Figure 7.6: Predicted and observed neural network input distributions, normalized to equal area.

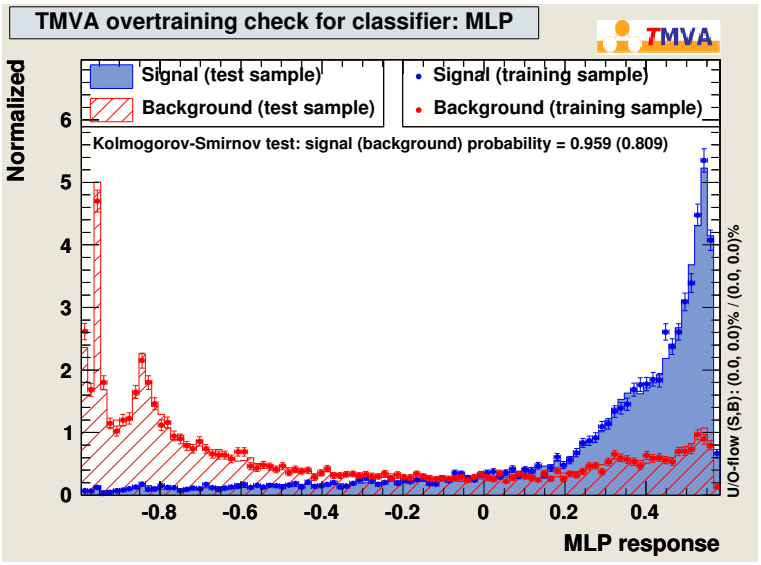


Figure 7.7: Simulated distributions for the neural network output for WZ signal and background, normalized to unit area. The solid histogram is the test sample and the points are for the training sample.

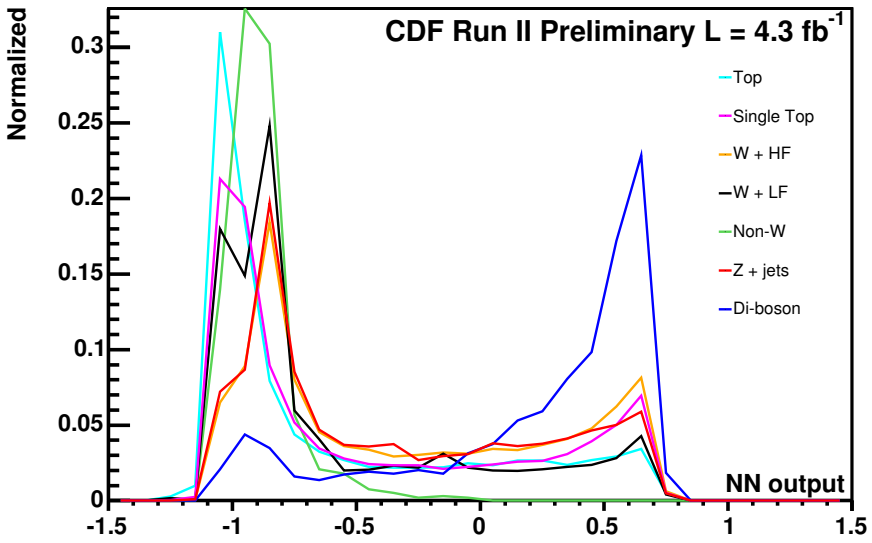


Figure 7.8: Simulated distributions for the neural network output for WZ signal and the individual background processes, normalized to unit area.

Chapter 8

Results

In this chapter we find the data result of our search for WZ production, with the improved b-jet identification algorithm from chapter 5 and the event selection and multivariate discriminant described in chapter 7.

The multivariate discriminant is shown for the data in Figure 8.1. In the absence of a significant excess of signal WZ events, we set a 95% confidence level (C.L.) upper limit on its cross section. We use the CDF limit calculation program "mclimit"[56], which we described briefly in the next section.

8.1 Limit Calculation

Given a histogram with N bins, and n_k observed events in the kth bin, $k = 1, 2, \ldots, \mathrm{N}$, the Poisson probability of obtaining the observed result is

$$\prod_{k=1}^{\mathrm{N}} \frac{e^{-(s\epsilon_k+b_k)}(s\epsilon_k + b_k)^{n_k}}{n_k!} \tag{8.1}$$

where s represents the number of signal events expected to be produced (which is the the cross section times integrated luminosity), ϵ_k is the acceptance (where $s\epsilon_k$ is the number of signal events expected to be observed) for the kth bin, and b_k is the expected number of background events for the kth bin. All the ϵ_k and b_k have

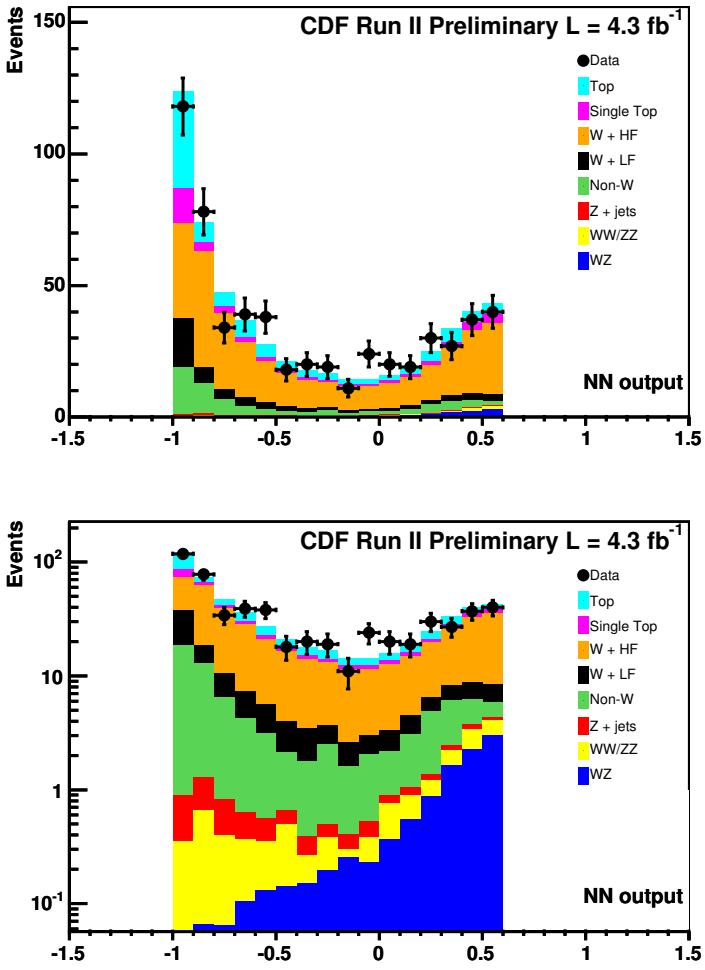


Figure 8.1: Output distribution of the neural network discriminant for data measurement and of the full method2 prediction, in linear and log scales.

uncertainties and these uncertainties are in the limit calculation called "nuisance parameters".

The systematic uncertainties on signal acceptance and the method2 uncertainty in background estimations are taken as normalization nuisance parameters. In addition, the JES is taken as a shape nuisance parameter for the signal, see Fig. 8.2.

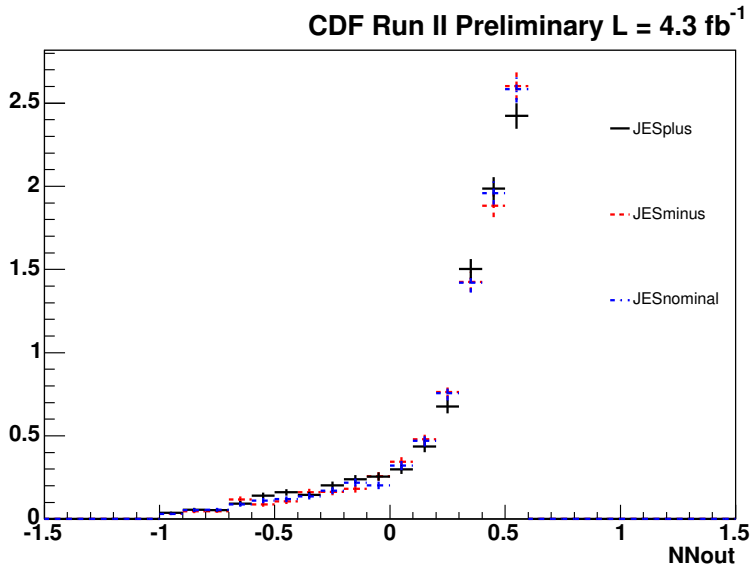


Figure 8.2: Comparison of output distribution of the neural network discriminant for signal with jet energy scale $\pm 1\sigma$.

Following the Bayesian approach, we put our knowledge about the uncertainties, which may be correlated, into a probability distribution function called the joint nuisance prior. The joint nuisance prior can be written as

$$\pi(\epsilon_1, b_1, \epsilon_2, b_2, \ldots, \epsilon_N, b_N) d\epsilon_1 db_1 d\epsilon_2 db_2 \cdots d\epsilon_N db_N \ . \tag{8.2}$$

Assuming that we have no prior information about the number of signal events expected to be produced s except that it must be non-negative (flat prior), the p.d.f. for s is proportional to

$$\iint\limits_{(2N)}\pi(\epsilon_1, b_1, \epsilon_2, b_2, \ldots, \epsilon_N, b_N)\left[\prod_{k=1}^{\mathrm{N}}\frac{e^{-(s\epsilon_k+b_k)}(s\epsilon_k+b_k)^{n_k}}{n_k!}\right]d\epsilon_1 db_1 d\epsilon_2 db_2\cdots d\epsilon_N db_N \tag{8.3}$$

where we have 2N integrals (two for each of the N bins, with one the for uncertainty on signal yield and the other for the uncertainty on background yield) to be performed averaging over the information about uncertainties. This averaging over the information about uncertainties is called marginalizing, and this p.d.f. is now a marginalized posterior.

These 2N integrals can be done using Monte Carlo integration. We use the joint nuisance prior to generate (including all correlations) M random $(\epsilon_1, b_1, \epsilon_2, b_2, \ldots, \epsilon_N, b_N)$ vectors. The 2N integrals then turn into a manageable sum.

Given this ensemble, our marginalized posterior for s then becomes

$$p(s|n) = \frac{1}{\mathcal{N}\cdot \mathrm{M}}\sum_{i=1}^{\mathrm{M}}\left[\prod_{k=1}^{\mathrm{N}}\frac{e^{-(s\epsilon_{ki}+b_{ki})}(s\epsilon_{ki}+b_{ki})^{n_k}}{n_k!}\right] \tag{8.4}$$

where the normalization constant $\mathcal{N}$ is given by

$$\mathcal{N} = \int_0^\infty \frac{1}{\mathrm{M}}\sum_{i=1}^{\mathrm{M}}\left[\prod_{k=1}^{\mathrm{N}}\frac{e^{-(s\epsilon_{ki}+b_{ki})}(s\epsilon_{ki}+b_{ki})^{n_k}}{n_k!}\right] ds \tag{8.5}$$

Given that we observed n events, a 95% confidence level (C.L.) upper limit on the signal s_{u} is defined with $\int_0^{s_{\mathrm{u}}} p(s|n)\, ds = 0.95$, the integral of $p(s|n)$ from 0 to some s_u such that the integral is 0.95.

8.2 Expected and Observed Limit

We find the expected limit by performing pseudo-experiments. For each pseudo-experiment, we randomly populate the histogram according to the background only expectation, and compute the upper limit. We define the expected limit by taking

the median value of the pseudo-experiments. The median value of these pseudo-experiments give a limit of 3.9x the Standard Model prediction, and 68 out of 100 experiments gave an upper limit between 2.8x and 5.6x the Standard Model prediction. On the data, we observe an upper limit of 3.9x the Standard Model prediction. Table 8.1 shows the expected and observed 95% C.L. upper limit on $\sigma(p\bar{p} \rightarrow WZ \rightarrow l\nu bb)$.

We then compare the likelihood of the signal+background hypothesis to the likelihood of the background only hypothesis, using the Poisson probability from Eqn. 8.2. A positive value from -2lnQ means that the background only hypothesis describes the data better, whereas a negative value means that the signal plus background hypothesis describes the data better. Since there is a large amount of overlap between the results for experiments with signal and without signal present, it shows that this analysis has little sensitivity to the presence of a signal from WZ.

We observed -2lnQ = -0.43 (Fig. 8.3), meaning that the signal plus background hypothesis describes the data slightly better, and is quite consistent with both hypotheses. This observed -2lnQ corresponds to a P-Value of 0.40, meaning that there is a 40% probability of the background fluctuating to give a value of -0.43 or lower.

-2 σ	-1 σ	median	+1 σ	+2 σ	Observed
2.0	2.8	3.9	5.6	8.0	3.9

Table 8.1: Expected and observed 95% C.L. upper limit on $\sigma(p\bar{p} \rightarrow WZ \rightarrow l\nu bb)$.

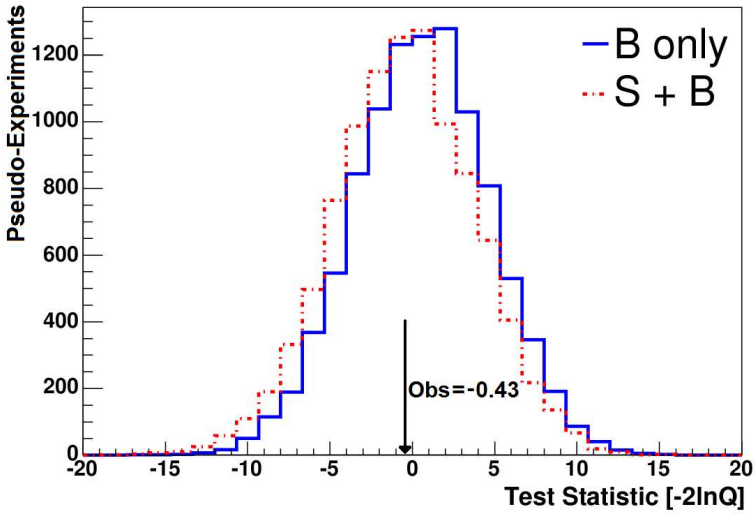


Figure 8.3: The distributions of -2lnQ in simulated S+B and B-only pseudo-experiments, assuming a Standard Model WZ signal. The value of -2lnQ observed in the data is -0.43, indicated with a black arrow, corresponds to a P-Value of 0.40.

Chapter 9

Conclusions

We have presented the results of a search for Standard Model (SM) Z boson produced in association with a $W^{\pm}$ boson, called the WZ search, in the $l\nu b\bar{b}$ final state. We found good agreement with the standard model background predictions and no significant excess of events from WZ production in a total of 4.3 fb^{-1} of data. We therefore set a 95% C.L. upper limit on the WZ cross section. We find

$$\sigma(p\bar{p} \rightarrow W^{\pm}Z) < 3.9 \times \text{SM (95\% C.L.)}$$

with a median expected limit of 3.9 $\times$ Standard Model prediction.

The identification of b-jets is crucial for this search, and my calibration of the RomaNN b-jet identification algorithm provided a 55% increase in tagging efficiency for each b-jet candidate, as compared to the SecVtx algorithm. This improvement is transferable to all the low-mass Higgs searches at CDF. For example, it will increases the signal efficiency of the present CDF WH searches [57] by 15% in the double-tag category.

The $WZ \rightarrow l\nu b\bar{b}$ search is a search for an established SM process with a known production cross-section, in a final state not observed previously. This final state is similar to the $WH \rightarrow l\nu b\bar{b}$ search, in that both look for the W to decay leptonically, and for the Z boson and Higgs boson to decay into a pair of bottom quarks. But the WH search is different because the Higgs mass is unknown, and it may not even

exist. Overall the Z boson is predicted to be five times more abundant in WZ than the Higgs boson ($m_H = 120$ GeV/c^2) in WH. But the different kinematics arising from the Z boson having a smaller mass reduces the expected sensitivity gain to a factor of 2.3 times as shown in Table 2.3. Thus the WZ search is an important step toward the WH search, since it is a similar search, but for an established process to which we have a better sensitivity.

This search is the first time that WZ has been searched for in the $l\nu b\bar{b}$ final state. In the future, we expect to gain a factor of 1.8 from improvements already tested in the CDF WH search (such as loose leptons and forward leptons, neural network b-jet specific energy correction, and including events with 3-jets) [57]. These will bring the median expected 95% C.L. upper limit to 2.2 $\times$ Standard Model WZ prediction. This is in agreement with the estimated 2.3 times sensitivity gain of WZ over WH as shown in Table 2.3, since the CDF $WH \rightarrow l\nu b\bar{b}$ search with the same amount of data (4.3 fb^{-1}) has a median expected 95% C.L. limit of 4.6 $\times$ Standard Model prediction for $m_H = 120$ GeV/c^2 [58].

From a median expected 95% C.L. limit of 3.9 $\times$ Standard Model WZ prediction, plus a factor of 1.8 from the improvements already tested in WH, and another factor of $\sqrt{2}$ from doubling the data, we predict that CDF can exclude at 95% C.L. around 1.5 $\times$ Standard Model prediction for $WZ \rightarrow l\nu b\bar{b}$. On combination with D0, assuming that both experiments have the same sensitivity, we will gain another factor of $\sqrt{2}$ to obtain a median expected Tevatron 95% C.L. upper limit of 1.1 $\times$ Standard Model WZ prediction.

The $WH \rightarrow l\nu b\bar{b}$ search at CDF with the same amount of data (4.3 fb^{-1}) has for $m_H = 120$ GeV/c^2 a median expected 95% C.L. limit of 4.6 $\times$ Standard Model prediction [58]. This WH analysis includes improvements not implemented in this thesis, but uses a combination of b-tagging algorithms that can still improve its signal efficiency by 15%, namely by using the UltraLoose Roma tagger to identify both b-jets coming from the Higgs boson as described in this thesis. In the future, with a

factor of 2 increase in sensitivity from doubling the dataset and combination with D0, the median expected Tevatron 95% C.L. upper limit for $WH \rightarrow l\nu b\bar{b}$ would be 2.1 $\times$ Standard Model prediction.

While that does not allow exclusion by itself, it is important to remember that the low-mass Higgs boson search contains additional significant contributions from $ZH \rightarrow \nu\nu b\bar{b}$ and $ZH \rightarrow llb\bar{b}$. These improve the median expected CDF 95% C.L. upper limit from 4.6 $\times$ Standard Model prediction for $WH \rightarrow l\nu b\bar{b}$ alone to 2.7 for all channels combined. Note that the full impact of the improved b-tagging discussed in this thesis has not been evaluated for the other Higgs channels yet.

As of November 2009 with approximately 4 fb^{-1} analysed by all channels and both experiments, the median expected 95% C.L. upper limit for the Higgs boson ($m_H = 120$ GeV/c^2) is 2.1 $\times$ Standard Model prediction [59]. The dataset is expected to double by the end of Tevatron operations in 2011. The expectation becomes 1.5 $\times$ Standard Model Higgs boson ($m_H = 120$ GeV/c^2) prediction with double the dataset. For a slightly less challenging $m_H = 115$ GeV/c^2, the present median expected 95% C.L. upper limit for the Higgs boson is 1.78 $\times$ Standard Model prediction [59]. The expectation becomes 1.26 $\times$ Standard Model Higgs boson ($m_H = 115$ GeV/c^2) prediction with double the dataset.

Adding more data alone does not allow the Tevatron to make any 95% C.L. exclusions for the Higgs boson in the low-mass category. Therefore development and implementation of further enhancements to these searches, at a challenging level of 26% for $m_H = 115$ GeV/c^2 and 50% for $m_H = 120$ GeV/c^2, such as the improved b-tagging discussed in this thesis, will be essential to excluding the Standard Model Higgs boson at the Tevatron.

After a hunt spanning over 40 years, the next few years should prove or disprove the existence of the standard model Higgs boson, thanks to the expected doubling of the Tevatron datasets, and the expected performance of the searches at the Large Hadron Collider [60]. Should the Higgs boson be finally discovered, we can claim for

the first time throughout human history to understand of how mass comes about. Though the Higgs mechanism enables the Standard Model to have good predictivity, there is no guarantee that the Higgs boson exists until we have observed it. If we end up excluding the existence of the Higgs boson, it will tell us that our world has mechanisms even more elegant. And that would be wonderful.

Bibliography

[1] P.W. Higgs, Phys. Rev. Lett. 13, 508 (1964).

[2] G. Abbiendi *et al.* [The LEP Working Group for Higgs Boson Searches], Phys. Rev. Lett. B565:61-75 (2003) [hep-ex/0306033v1].

[3] M. Spira *et al.*, "HDECAY: a Program for Higgs Boson Decays in the Standard Model and its Supersymmetric Extension".
`http://people.web.psi.ch/spira/hdecay/`

[4] S. Weinberg, "Mixing Angle in Renormalizable Theories of Weak and Electromagnetic Interactions", Phys. Rev. D 5, 1962-1967 (1972).

[5] M. Veltman, 1977b, Acta Phys. Pol. B 8, 475 (1977).

[6] The LEP Electroweak Working Group.
`http://lepewwg.web.cern.ch/LEPEWWG/`

[7] Aglietti *et al.*, "Tevatron-for-LHC Report: Higgs" [hep-ph/0612172].

[8] C. Amsler *et al.* (Particle Data Group), "Review of Particle Physics", Phys. Lett. B 667, 1 (2008).

[9] T. Aaltonen *et al.* , "Neural Network b-jet energy correction for WHNN analysis", CDF internal note 9858 (2009).

[10] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. 98, 161801 (2007).

[11] S.-C. Hsu *et al.* , "Updated Measurement of WZ production in the trilepton plus MET channel", CDF internal note 8910, (2007).

[12] T. Aaltonen *et al.* (CDF Collaboration),"Measurement of the WW Production Cross Section and Search for Anomalous WWgamma and WWZ Couplings in ppbar Collisions at sqrt(s) = 1.96 TeV", (2009) [hep-ex/0912.4500].

[13] T. Aaltonen et al. (CDF Collaboration), "Measurement of the WW+WZ Production Cross Section Using the lepton+jets Final State at CDF II", Phys. Rev. Lett. 104, 101801 (2010).

[14] Fermilab Beams Division, Tevatron Run 2 Handbook (1998).

[15] F. Abe *et al.* (CDF Collaboration), "The CDF detector: an overview", Nucl. Instr. Meth. A271, 387-403 (1988).

[16] D. Acosta *et al.* (CDF Collaboration), "The performance of the CDF Luminosity Monitor", NIMa 494 (1), p.57-62, Nov 2002.

[17] Robert G. Wagner, "Electron Identification for Run II: Algorithms", CDF internal note 5456 (2003).

[18] T. Spreitzer *et al.*, "Electron Identification in Offline Release 6.1.2", CDF internal note 7950 (2006).

[19] Simona Rolli, Lubomir Lovas, "Muon ID, Reconstruction and Trigger Efficiencies and Scale Factors for Period 9-12 data", CDF internal note 9085 (2007).

[20] B. Flaugher, J. Mueller, "A Guide to JETCLU: The CDF Jet Clustering Algorithm", CDF internal note 1814 (1992).

[21] Monica D'Onofrio, Kenichi Hatakeyama, Nick Van Remortel, "Validation of the Jet Energy Scale using dijet and photon+jet balance", CDF internal note 8926 (2007).

[22] D. Jeans *et al.*, "A new b & c jet identification algorithm", CDF internal note 8451 (2006).

[23] Physics Information Technologies GmbH.
`http://www.phi-t.de/`

[24] J. Marriner, "Secondary vertex fit with mass and pointing constraint (CTVMFT)", CDF Internal Note 2724 (1994).

[25] D. Bortolello *et al.*, "ID Efficiency for Intermediate pT Muons", CDF Internal Note 7197 (2005).

[26] J. Keung *et al.*, "Binary Mode RomaNN b-Tagger Efficiencies and Scalefactors up to Period 23", CDF Internal Note 9898 (2009).

[27] T. Wright *et al.*, "SecVtx B-Tag Efficiency Using Muon Transverse Momentum", CDF internal note 7448 (2005).

[28] Jean-Francois Arguin, Pekka K. Sinervo, "*b*-jet Energy Scale Uncertainty From Existing Experimental Constraints", CDF internal note 7252 (2004).

[29] S. Budd *et al.*, "Tight, Loose and Ultratight SECVTX Tag Rate Matrix with 1.2 fb^{-1}", CDF internal note 8519 (2006).

[30] J. Freeman *et al.*, "SecVtx Mistag Matrices for 2.2 fb^{-1} of Data Up to P13", CDF internal note 9280 (2008).

[31] J. Keung *et al.*, "Measurement of the Top Pair Production Cross Section using RomaNN with 4.3 fb^{-1}", CDF internal note 9913 (2009).

[32] E. Thomson *et al.* (CDF Collaboration), "Combination of CDF top quark pair production cross section measurements with up to 4.6 fb^{-1}", CDF internal note 9913 (2009).

[33] S. Grinstein, D. Sherman, "SecVtx Scale Factors and Mistag Matrices for the 2007 Summer Conferences", CDF internal note 8910 (2007).

[34] S. Dube, S. Somalwar, "Medium ET Electron identification efficiency and scale-factors using DY electrons", CDF internal note 8321 (2006).

[35] L. Cerrito, A. Taffard, "A Soft Muon Tagger for Run II: Summer-04 version", CDF internal note 7122 (2004).

[36] T. Sjstrand, P. Edn, C. Friberg, L. Lnnblad, G. Miu, S. Mrenna and E. Norrbin, "High-Energy-Physics Event Generation with Pythia", Computer Phys. Commun. 135 (2001) 238 (LU TP 00-30) [hep-ph/0010017].

[37] J. Freeman and S. Grinstein, "SecVtx Mistag Asymmetry Study through Period 13", CDF internal note 9277 (2008).

[38] Bo-Young Han, Veronique Boisvert, "Trigger Efficiencies for the High Et Central Electron in the Gen6 data", CDF internal note 8629 (2006).

[39] Robert G. Wagner, "Electron Identification for Run II: Understanding and Using Lshr", CDF internal note 6249 (2003).

[40] S. Agostinelli *et al.* (GEANT 4 Collaboration), "GEANT 4 - a simulation toolkit", NIM A 506 (2003) 250-303.

[41] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A.D. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions", JHEP 0307 (2003) 001 [hep-ph/0206293].

[42] Fabio Maltoni, Tim Stelzer, "'MadEvent: Automatic Event Generation with MadGraph", JHEP 0302 (2003) 027 [hep-ph/0208156]

[43] T. Schwarz *et al.*, "Method II For You", CDF internal note 9185 (2008).

[44] J.M. Campbell and R.K. Ellis, "An update on vector boson pair production at hadron colliders", FERMILAB-Pub-99/146-T [hep-ph/9905386].

[45] D. Acosta *et al.* (CDF Collaboration), "First measurements of inclusive w and z cross sections from run ii of the Tevatron collider", Phys. Rev. Lett. 94 (2005) 091803 [hep-ex/0406078].

[46] J. Campbell and R. K. Ellis, "Next-to-leading order corrections to w + 2jet and z + 2jet production at hadron colliders", Phys. Rev. D65 (2002) 113007 [hep-ph/0202176].

[47] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason and G. Ridolfi, "The t anti-t cross-section at 1.8 Tev and 1.96 Tev: A study of the systematics due to parton densities and scale dependence", J. High Energy Phys. 04 (2004) 068 [hep-ph/0303085].

[48] B. W. Harris, E. Laenen, L. Phaf, Z. Sullivan and S. Weinzierl, "The fully differential single top quark cross section in next-to-leading order qcd", Phys. Rev. D66 (2002) 054024 [hep-ph/0207055].

[49] S. Budd *et al.*, "Estimation and modeling of non-W background for single-top searches", CDF internal note 8489 (2006).

[50] T. Schwarz *et al.*, "Measurement of the Top Pair Cross Section with 2.7 fb^{-1} in SecVtx Tagged Lepton Plus Jets Events", CDF internal note 9392 (2008).

[51] C. Cully, *et al.*, "Calibration of Heavy-Flavor Production in W + 1 Jet Data", CDF internal note 9187 (2008).

[52] J. Pumplin *et al.*, "New generation of parton distributions with uncertainties from global QCD analysis", JHEP 0207:012 (2002) [hep-ph/0201195].

[53] A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, "Partons and alpha_S from precise deep inelastic scattering and Tevatron jet data", (2001) [hep-ph/0110215].

[54] CDF Joint Physics Group.
`http://www-cdf.fnal.gov/internal/physics/joint_physics/`

[55] TMVA - Toolkit for Multivariate Data Analysis.
`http://tmva.sourceforge.net/`

[56] MClimit - Limit Calculator
`https://plone4.fnal.gov:4430/P0/phystat/packages/0711001`

[57] T. Aaltonen *et al.* (CDF Collaboration),"Combined WH $\rightarrow$ lvbb search at CDF with Neural Network and Matrix Element techniques", Proceedings of Science, (HCP2009)061.

[58] T. Aaltonen *et al.* (CDF Collaboration),"Search for the Standard Model Higgs Boson Production in Association with a W Boson using 4.3/fb", CDF public note 9997 (2009).

[59] The TEVNPH Working Group, "Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with 2.1 - 5.4 fb^{-1} of Data", (2009) [hep-ex/0911.3930].

[60] V. Drollinger, "Review of Higgs Physics at the LHC", CMS note CMS-CR-2004-012 (2004).