A MEASUREMENT OF THE TOP QUARK'S CHARGE

By

Zeynep Günay Ünalan

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Physics

2007

ABSTRACT

A MEASUREMENT OF THE TOP QUARK'S CHARGE

By

Zeynep Günay Ünalan

The top quark was discovered in 1995 at the Fermilab National Accelerator Laboratory (Fermilab). One way to confirm if the observed top quark is really the top quark posited in the Standard Model (SM) is to measure its electric charge. In the Standard Model the top quark is the isospin partner of the bottom quark and is expected to have a charge of +2/3. However, an alternative "exotic" model has been proposed with a fourth generation exotic quark that has the same characteristics, such as mass, as our observed top but with a charge of -4/3. This thesis presents the first CDF measurement of the top quark's charge via its decay products, a W boson and a bottom quark, using $\approx 1fb^{-1}$ of data. The data were collected by the CDF detector from proton anti-proton ($p\bar{p}$) collisions at $\sqrt{s} = 1.96$ TeV at Fermilab. We classify events depending on the charges of the bottom quark and associated W boson and count the number of events which appear "SM-like" or "exotic-like" with a SM-like event decaying as $t \rightarrow W^+b$ and an exotic event as $t \rightarrow W^-b$. We find the p-value under the Standard Model hypothesis to be 0.35 which is consistent with the Standard Model. We exclude the exotic quark hypothesis at an 81% confidence level, for which we have chosen a priori that the probability of incorrectly rejecting the SM would be 1%. The calculated Bayes Factor (BF) is 2×Ln(BF)=8.54 which is interpreted as the data strongly favors the Standard Model over the exotic quark hypothesis.

To my grandparents

Şaziye and Kenan Akpolat

Zehra and Adem Günay

ACKNOWLEDGMENTS

Lewis for setting an example of an enthusiastic physicist and providing toys that kept my son busy while I was trying to work from home. I guess I should also thank people passing by in the corridor and their laughter for reminding me that I should take a break when most stressed from problematic code and crashing jobs.

Being a part of a huge collaboration was a different and enlightening experience for me after a master's thesis on which I worked alone. I feel obliged to thank all the members of the CDF collaboration who took part in designing and running the CDF experiment and the people in the accelerator division for maintaining beam in the Tevatron for long hours.

I am grateful to all my friends at MSU and at Fermilab for being my surrogate family during the last 5 years. I would like especially thank Selcuk Cigangir who tried hard (and mostly succeeded) to bring the Turkish community at Fermilab together that made us breathe the atmosphere away from home and to my beloved friend Melek Yaylaci in Istanbul for her encouragement through her emails and phone calls.

I wish to express my deepest appreciation to my parents, Yilmaz and Tülin Gunay for their love, encouragement and for sparking in me the love of knowledge and science. Finally I wish to thank my C++ expert D∅-husband Rahmi for all his help and understanding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xvii

CHAPTER 1

Theory

> And the heaven, We built it with might, and We will
> surely be extending it. And We have spread out the earth:
> How excellently We do spread out! And all things we have
> created by pairs, that you may reflect.
>
>   Adh-Dhariyat(The winds that scatter, 47-49) Quran

1.1   Introduction

In 1897 in Cambridge, England, J.J. Thomson experimented on cathode rays and showed that they were indeed particles which were much smaller than an atom and had a negative electric charge. Those particles came to be called electrons and the view of the atom as a featureless, structureless, indivisible particle was shattered forever. Then in the early 1900's came Rutherford's discovery of the atom's nucleus and Chadwick's discovery of the neutron. Accelerator experiments revealed that the electron, proton and neutron were only the first three of a long list of fundamental particles. By the early 1960's a hundred or so types of particles had been identified. Physicists realized that their previous understanding was not sufficient to explain the particles being discovered. In 1964, Gell-Mann's and Zweig's quark theory solved these problems. They found that all these particles could be explained by a few types of yet smaller objects which Gell-Mann called quarks. One revolutionary part of the

quark theory is that one has to assign the quarks fractional electric charges of $+2/3$ and $-1/3$ in units of the proton charge. The theory of quarks is now part of the Standard Model (SM) that describes all particles and the interactions between them. The SM has gradually expanded in scope and gained increasing acceptance with new evidence from particle accelerators. One exciting confirmation of the SM was the discovery of the predicted top quark at the Fermi National Accelerator Laboratory (Fermilab) in 1995 by the CDF and DØ experiments. Since then measurements of the top quark's properties have been one of the primary aims of both experiments. The goal of this research is to measure the top's electric charge using data from CDF.

1.2 The Standard Model

A more detailed view of the Standard Model (SM) can be found in [1, 2]. A brief overview of how the SM works is given in this section. Three of the four fundamental forces (the electromagnetic, weak and strong) and the behavior of all known subatomic particles through these forces is described within a single theoretical framework called the Standard Model. According to the SM, all matter is built up from spin 1/2 point-like particles, called fermions. Fermions come in two types, quarks and leptons, and each type occurs in 3 families. Each family is the same in every respect except for the masses of the particles. There are 6 quarks, 2 in each family as shown in Figure 1.1. For every particle, there is also a corresponding anti-particle with a reversed charge resulting in a total of 12 leptons and 12 quarks. All stable matter is made up of the first and lightest family of quarks, the up (u) and down (d) quarks. Protons and neutrons consist of triplets of the u and d quarks: a proton is made up of 2 up quarks and 1 down (uud) quark and a neutron of 2 downs and 1 up (udd). The heavier quarks (charm, strange, bottom and top) also form particles similar to the proton and neutron, however these are unstable and decay very rapidly. The heavier quarks can only be produced and observed in high energy collisions. Like the quarks the

lightest of the leptons, called the electron, is in every atom. But the muon (200 times heavier than the electron), and the tau (3490 times heavier), can only be observed as the product of high energy collisions or in cosmic rays.

FERMIONS					
Leptons Spin=1/2			**Quarks Spin=1/2**		
Flavor	Mass (GeV/c^2)	Electric charge	Flavor	Mass (GeV/c^2)	Electric charge
electron neutrino	$<7*10^{-9}$	0	up (u)	0.005	+2/3
electron (e)	0.000511	-1	down (d)	0.01	-1/3
muon neutrino	<0.0003	0	charm (c)	1.5	+2/3
muon (μ)	0.106	-1	strange (s)	0.2	-1/3
tau neutrino	<0.03	0	top (t)	170	+2/3
Tau (τ)	1.7771	-1	bottom (b)	4.7	-1/3

Figure 1.1. The three families of fermions, their masses and electric charges. Each family is designated with a different color.

Having placed the particles in the framework of the Standard Model, we can now talk about how the model explains the interactions between all these particles. The SM is a relativistic quantum field theory and is consistent with both quantum mechanics and special relativity. In this field theory, the classical concept of a force finds its new definition: Particles can change their identity and be created or destroyed through the emission and absorption of spin 1 particles, called bosons. Figure 1.2 lists the bosons under the force they are associated with along with their masses and charges. The Lagrangian of each set of mediating bosons is invariant under a transformation called a gauge transformation, so these mediating bosons are referred to as gauge bosons. The symmetry groups that are embedded in the mathematical formulation of the Standard Model Lagrangian make the model a unified framework to describe the quantum field theories of electromagnetism, the weak and the strong forces.

Quantum electrodynamics (QED) is a quantum field theory that can generate Maxwell's equations from a relativistic quantum theory and was developed by a number of physicists in the early 1900's. The symmetry group of the theory is the $U(1)$

BOSONS					
Unified Electroweak Spin=1	Mass (GeV/c²)	Electric Charge	Strong Spin=1	Mass (GeV/c²)	Electric Charge
Photon (γ)	0	0	Gluon (g)	0	0
W⁻	80.22	-1			
W⁺	80.22	1			
Z⁰	91.180	0			

Figure 1.2. Force carriers, their masses and electric charges.

gauge group with the massless photon as its generator. During the 1960's Sheldon Glashow, Abdus Salam, and Steven Weinberg independently discovered that they could construct a gauge-invariant theory of the weak force, provided that the electromagnetic force is also included. The electroweak theory they proposed unified electromagnetism with the weak force at high energy scales in an overall $SU(2) \times U(1)$ gauge symmetry. The $SU(2)$ group, representing the weak force part, has three generators and that means three massless gauge bosons to mediate the weak force. However, the short range of the weak force indicates that it is carried by massive particles. In order to make a gauge invariant theory work for the weak nuclear force, theorists had to produce heavy gauge bosons in such a way that wouldn't destroy the consistency of the quantum theory. The method they came up with is called "spontaneous symmetry breaking", where massless gauge bosons acquire mass by interacting with a scalar field called the Higgs field. The scalar field interactions mix up the three massless gauge bosons of $SU(2)$ with the one massless gauge boson of $U(1)$, and out of the mixture, comes three massive gauge bosons, now called the W^+, W^- and Z, and one massless gauge boson, called the photon. The W and Z bosons mediate the weak force by changing the flavor of leptons and quarks, while the photon mediates the electromagnetic force by changing the charge of particles. The electroweak theory

with its $SU(2) \times U(1)$ symmetry group later became part of the $SU(3) \times SU(2) \times U(1)$ Standard Model that also includes the strong force. $SU(3)$ is the gauge group of the theory of the strong interactions that is also known as Quantum Chromodynamics (QCD). There are 8 gluons that carry the strong force by changing the quark quantum number named color. Each flavor of quark can take on three possible color values, conventionally called red, green, and blue. One interesting property is that the carriers of the strong force, gluons, also carry color charge. This causes quite different results than what we are familiar with in QED. As two quarks separate, the gluons form narrow tubes of color charge and the force experienced by one quark remains constant regardless of its distance from the other quark. Since energy is equal to force times distance, as the quarks separate the total energy increases linearly with distance unlike in QED. At some point the vacuum becomes so energetic that new quark anti-quark ($q\bar{q}$) pairs are created from it. This collection of new quarks (q) and antiquarks ($\bar{q}$) then are rearranged into pairs (mesons) or triplets (baryons) of quarks that are color-neutral hadrons.[1]

To date, almost all experimental tests of the three forces described by the Standard Model have agreed with its predictions. The Standard Model predicted the existence of the W and Z bosons, the gluon, the top quark and the charm quark before these particles had been experimentally observed. So far the predicted properties of these SM particles have also been confirmed experimentally. Despite the Standard Model's success it is not a complete theory of fundamental interactions, primarily because it does not include the gravitational force but also because there are still many fundamental questions left unanswered. For example, why are there three types of quarks and leptons? Why do the particles masses span at least 11 orders of magnitudes (See Figure 1.3)? Is there some pattern to their masses? Are there more types of particles to be discovered at yet higher energies? Are the quarks and leptons really fundamen-

---

[1]The general name for all color-neutral combination of quarks is "hadron".

5

tal, or do they also have substructure? What particles form the dark matter in the universe? As the field of particle physics continues to push the boundary of the high energy frontier, we are coming closer to the answers of these questions and extending our understanding of the most fundamental aspects of nature. One way of doing that is by measuring the fundamental properties of the top quark and checking if it indeed behaves as the SM predicts.
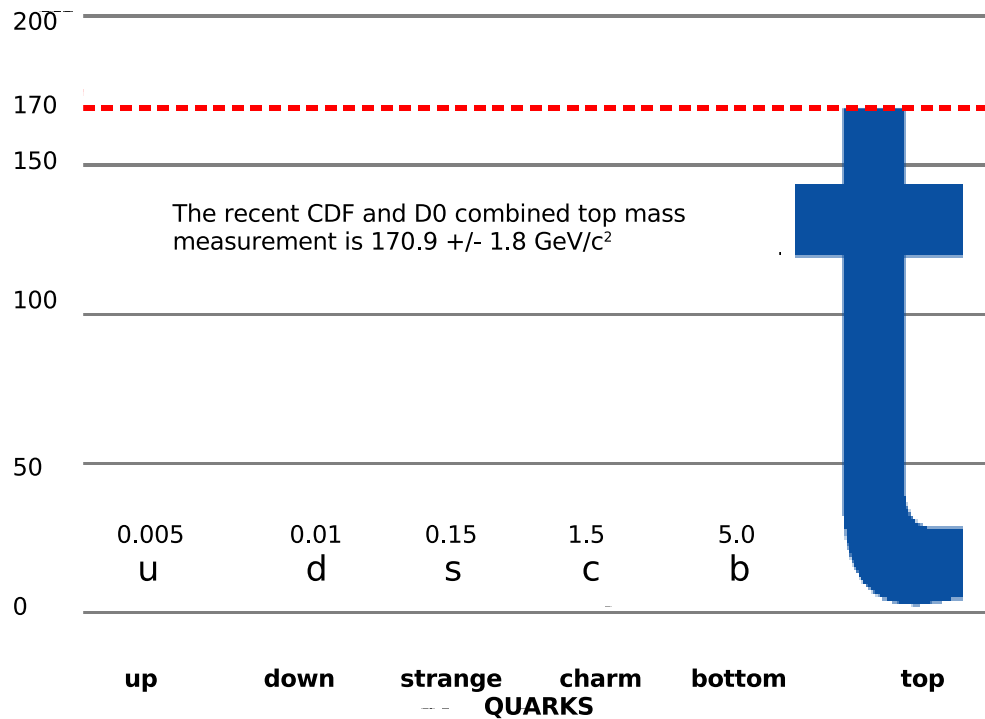
Figure 1.3. A scale showing the masses of the fermions and bosons.

1.3 Top Quark

By the mid-1970's the up, down, charm and strange quarks were all well established. With the discovery of the Upsilon particle at Fermilab in 1977, the fifth quark, bottom, took its place in the Standard Model. However, the theory suggested that the bottom quark should have an $SU(2)$ partner named top. The CDF and DØ collaborations at Fermilab began a top quark search that came to a successful conclusion in February

1995 [3, 4, 5]. Since then, the study of the top quark has been a primary focus of both experiments. The top quark is the most massive fundamental particle in the Standard Model, with a mass approximately twice that of the W and Z bosons and 35 times that of the next most massive fermion, the b quark (See Figure 1.3).
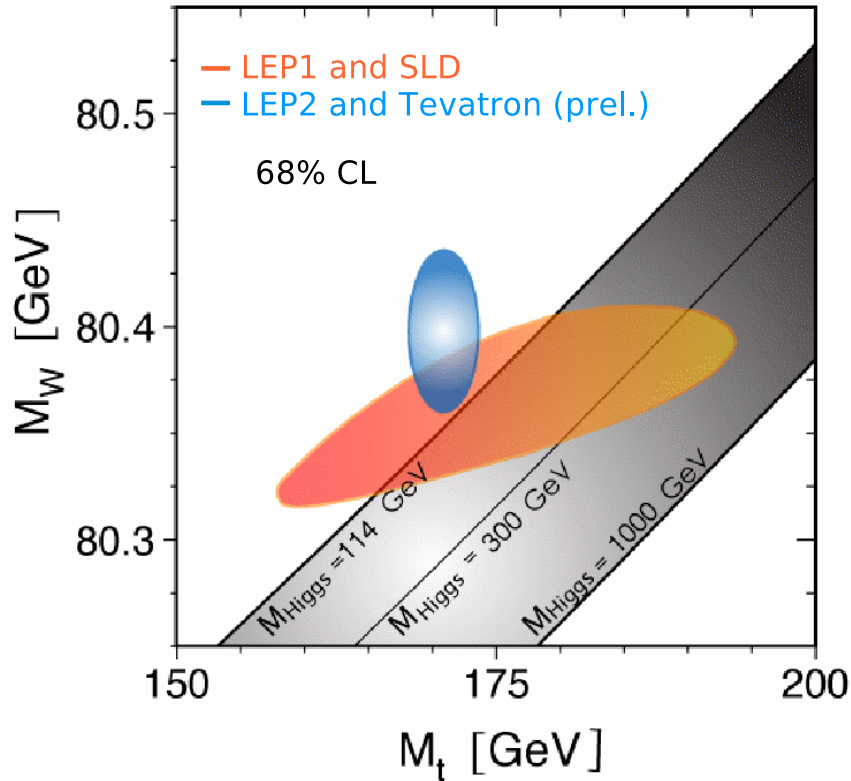
Figure 1.4. Higgs boson mass as a function of top quark and W boson mass. The red ellipse shos the result using electroweak measurements at LEP and SLD, while the blue ellipse shows the most recent result based on all available experimental data, including the CDF result. Diagonal lines represent Higgs boson masses at 114, 300 and 1000 GeV/c^2 based on theoretical constraints and direct experimental searches. The fact that the intersection between the blue ellipse and the red one is very small, and does not include any gray region, indicates a lighter Higgs mass than than the one obtained at LEP II (114 GeV).

The top quark gains its heavy mass through its large coupling to the Higgs boson. Actually the Higgs coupling is the source of the masses of all fermions and the quanta

7

of the field, the Higgs boson, has a mass that is constrained by the masses of the top quark and the W boson as shown in Figure 1.4. Therefore, measuring the top mass precisely is important for finding the undiscovered Higgs boson. Because of its large mass, top has a short lifetime of 10^{-24} seconds which is shorter than the hadronization time scale of 10^{-23} s. This means the top quark decays before it hadronizes and therefore can give its spin and charge information to its decay products.

1.4 Top Quark's Production and Decay

At the current Fermilab Accelerator (Tevatron) energy of 1.96 TeV, top-antitop ($t\bar{t}$) quark pairs are produced with a cross section of around 7 picobarns[2] (pb) for a top mass[3] of 175 GeV/c^2. This means that just 1 out of every 10^{10} collisions at Fermilab contains a top quark. The main production mechanism of top quarks at the Tevatron is pair production via quark-antiquark annihilation ($q\bar{q} \rightarrow t\bar{t}$) which happens 85% of the time. The remaining 15% of the time $t\bar{t}$ is produced by gluon-gluon fusion. The leading order Feynman diagrams for the $t\bar{t}$ production mechanisms are shown in Figure 1.5.

The top quark interacts primarily by the strong force but can only decay via the weak force. Almost 100% of the time the top quark decays to a W boson and a bottom quark ($t \rightarrow Wb$) due to the large coupling between the top and bottom quark, V_{tb}, via the Cabibbo-Kobayashi-Maskawa (CKM) matrix. The CKM matrix is a unitary matrix that describes the probability of a transition from one quark flavor to another quark [1, 2].

The decay of $t\bar{t}$ events can be classified based on the decay modes of the W boson. A W decays to either a pair of quarks ($q\bar{q}$) or a lepton-neutrino ($\ell\nu$) pair resulting

---

[2] 1 barn (b)= 10^{-28} m^2, 1 picobarn (pb) = 10^{-40} m^2

[3] In high energy physics, the masses are given in units of energy per c^2 ($E = m \cdot c^2$), mostly in MeV/c^2, GeV/c^2, TeV/c^2.
MeV=10^6 eV , GeV= 10^9 eV , TeV= 10^{12} eV.
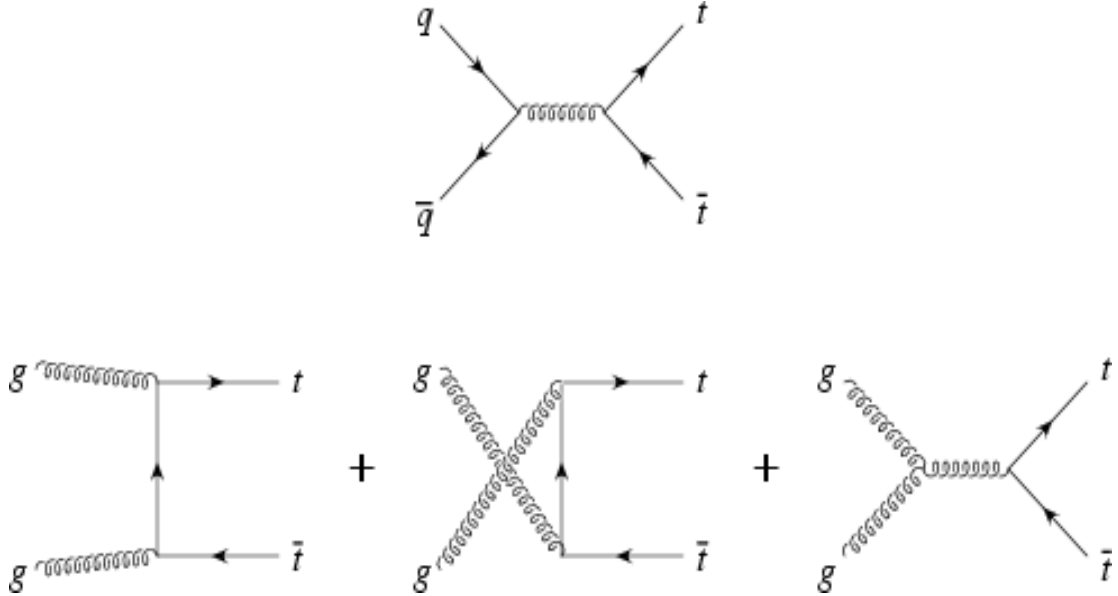
Figure 1.5. The leading order diagrams for $t\bar{t}$ production at the Tevatron.

in three categories of decay channels. All possible $t\bar{t}$ decay modes are represented in Figure 1.6 where the fermions from $W^-(W^+)$ are shown along the y(x) axis. The area of each region in the figure is proportional to the branching fraction of the designated decay mode. If both of the W's produced in the $t\bar{t}$ event decay to lepton-neutrino ($\ell\nu$) pairs, the decay channel is called the dilepton channel. Decays of W's to tau (τ) leptons are generally not included in top analyses due to the difficulty of identifying the hadronic decay of the tau leptons. The signature for a dilepton event in the detector is two leptons (an electron or muon), a large amount of missing transverse energy, $\vec{\not{E}}_T$, coming from the undetected neutrinos and two or more jets[4]. This decay mode only occurs 5% of the time but is the cleanest mode due to the small amount of background processes that can mimic a dilepton $t\bar{t}$ event signature. Backgrounds for the dilepton channel come from $b\bar{b}$, WW, $Z \rightarrow \tau\tau$ and Drell-Yan ($Z, \gamma \rightarrow l\bar{l}$) production.

In the case when both W's decay to $q\bar{q}$ pairs, it is called the all-hadronic channel.

---

[4]A jet is a spray of particles in detector coming from hadronization of a quark or gluon.
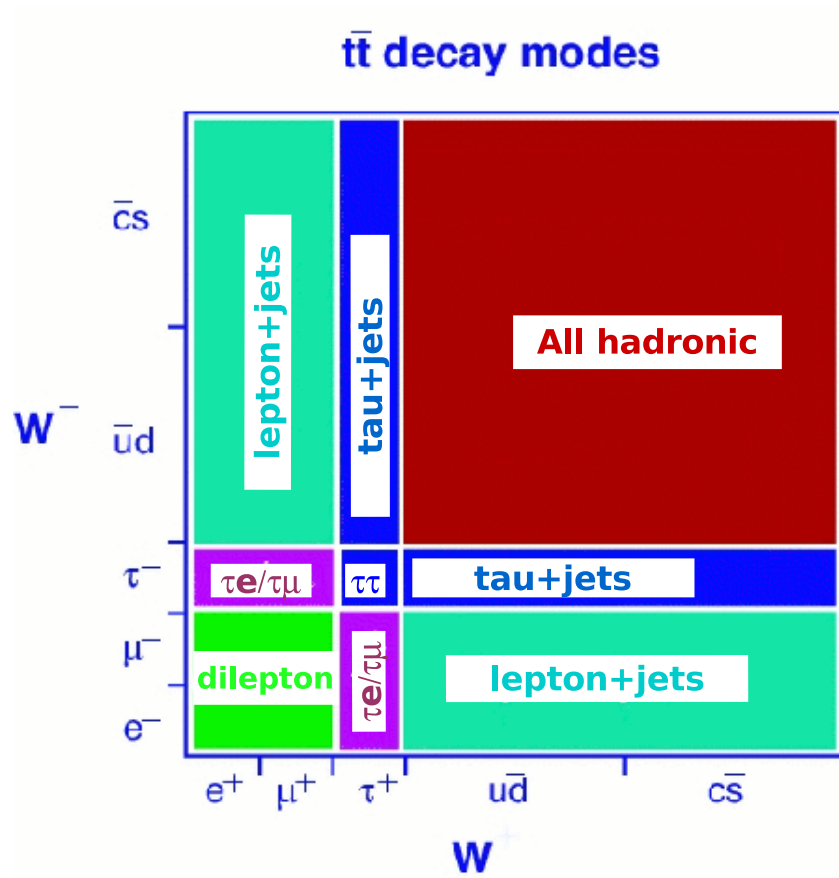
Figure 1.6. Representation of $t\bar{t}$ decay modes.
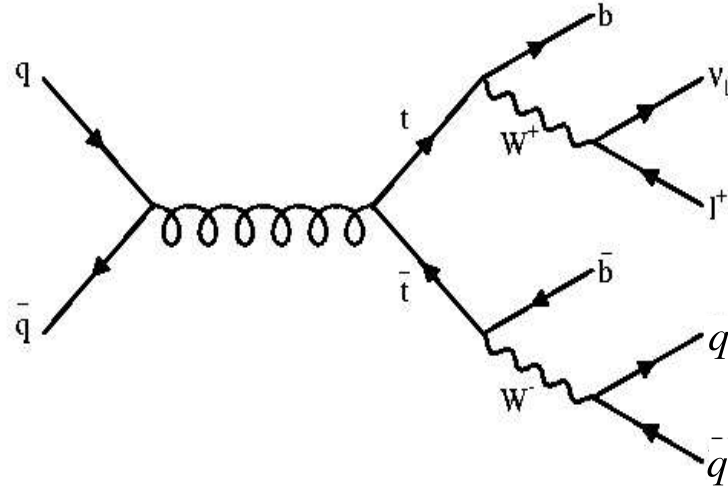
Figure 1.7. The tree-level Feynman diagram for top quark production by $q\bar{q}$ annihilation and Standard Model top decay to the lepton+jets channel.

Such a $t\bar{t}$ decay will have six or more jets in the event, two from b quarks and four light quarks from the W decays. The all-hadronic channel is the hardest channel to see the $t\bar{t}$ signal in due to the huge amount of background coming from QCD multijet production processes. However this decay channel has the largest branching fraction at 44% so more events than the dilepton channel. The last decay mode is the lepton+jets channel which occurs 30% of the time. In this channel one W decays to a lepton-neutrino pair and the other W decays hadronically to two light quarks. The topology of a lepton+jets event is a single electron or muon, some amount of $\vec{\not{E}}_T$ from the neutrino, and four or more jets, two from the b's and two from the W. The tree-level Feynman diagram of a lepton+jets event is shown in Figure 1.7. The background in the lepton+jets channel is substantially smaller than the all-hadronic channel, but there are still backgrounds from generic QCD events with a fake W boson, W+multijet production, WW events and top events where only one top is

11

produced. The dominant background however is from W+multijet production which can be suppressed by the identification of the bottom (b) quarks in the event. There are several methods for tagging b quarks at CDF which will be described in Chapter 4.

This thesis result uses the lepton+jets $t\bar{t}$ sample where both b quarks are tagged, also called the "double-tagged lepton+jets sample". In the end, this thesis result is combined with a result from the dilepton channel for a final CDF top charge measurement. Chapter 8 is dedicated to briefly explaining the measurement in the dilepton channel.

1.5   Possible Techniques for Measuring the Top's Charge

There are mainly two techniques that can be used to determine the electric charge of the top quark:

- by measuring the strength of the electromagnetic coupling via photon radiation in $t\bar{t}$ events.

- by using the charges of the top decay products.

The first technique is based on the direct measurement of the top quark electromagnetic coupling through photon radiation in $t\bar{t}$ events. A top quark can radiate photons (γ) during its production and/or its decay. All possible Feynman diagrams for $t\bar{t}$ production by gluon-gluon fusion and by $q\bar{q}$ annihilation are shown in Figures 1.8 and 1.9 respectively. In radiative top production, ($p\bar{p} \rightarrow t\bar{t}\gamma$), the cross-section is expected to be proportional to the square of the top charge. The situation is more complicated in the case of radiative top decay, i.e, $p\bar{p} \rightarrow t\bar{t}$ followed by $t \rightarrow Wb\gamma$, since a photon can also be radiated from the b quark or W boson. The Feynman diagrams for $t\bar{t}$ decay associated with a photon can be seen in Figure 1.10. To measure the top charge with this method, we need to distinguish between the radiative
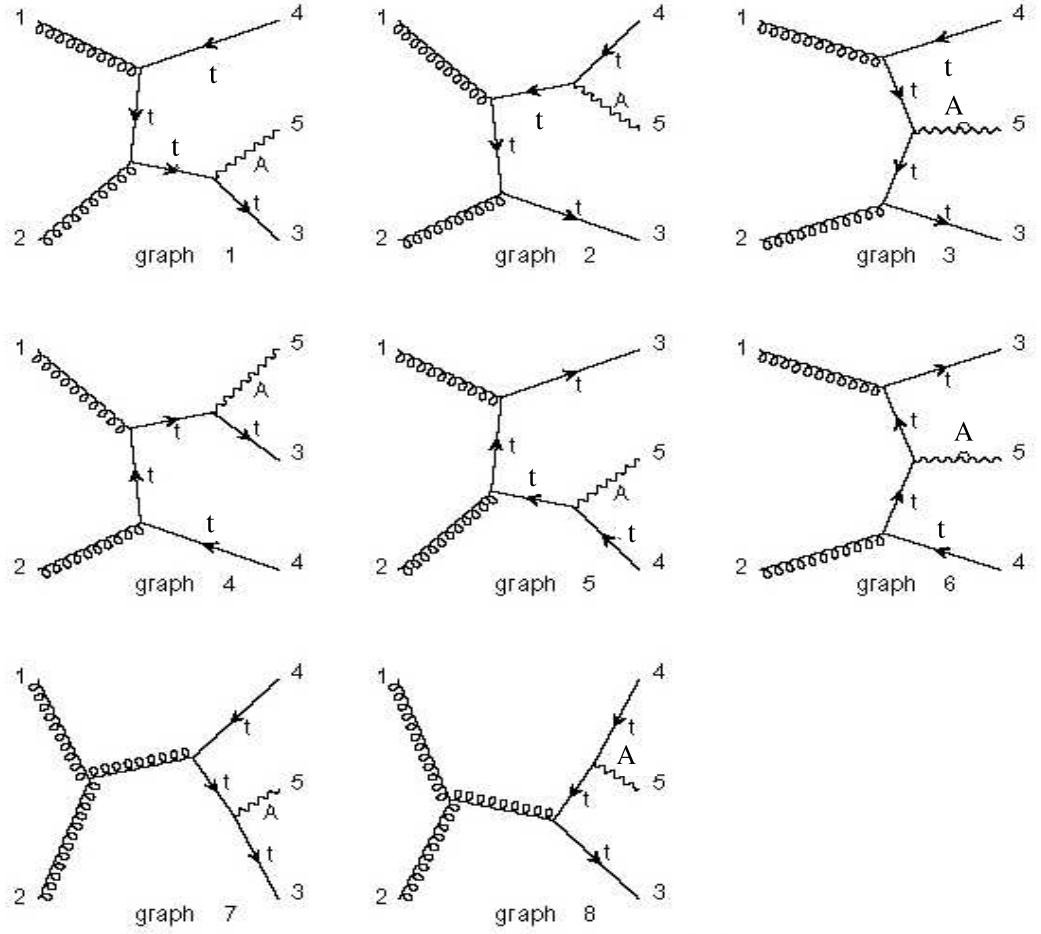
Figure 1.8. Feynman diagrams for $t\bar{t}\gamma$ production by gluon-gluon fusion
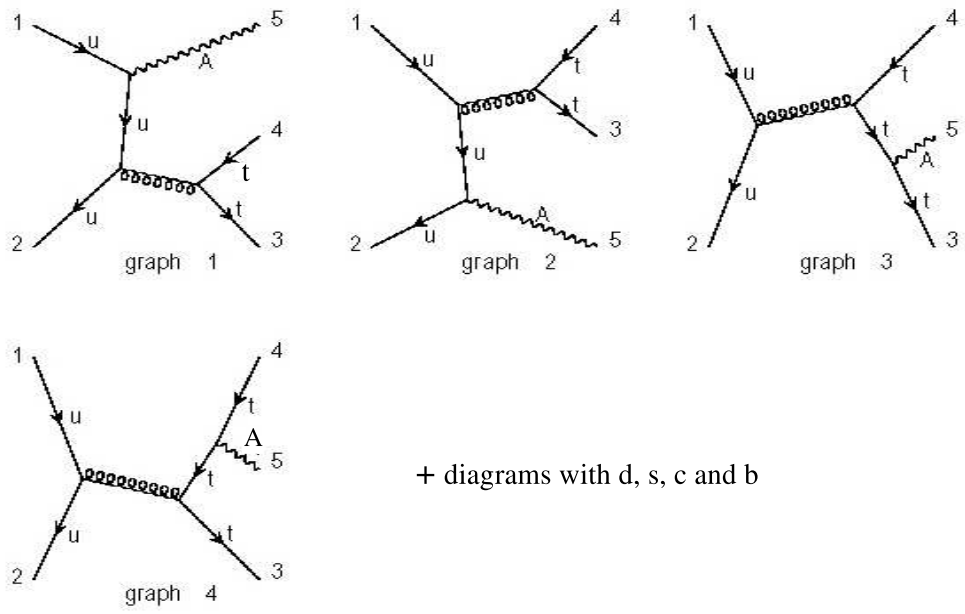
Figure 1.9. Feynman diagrams for $t\bar{t}\gamma$ production by $q\bar{q}$ annihilation
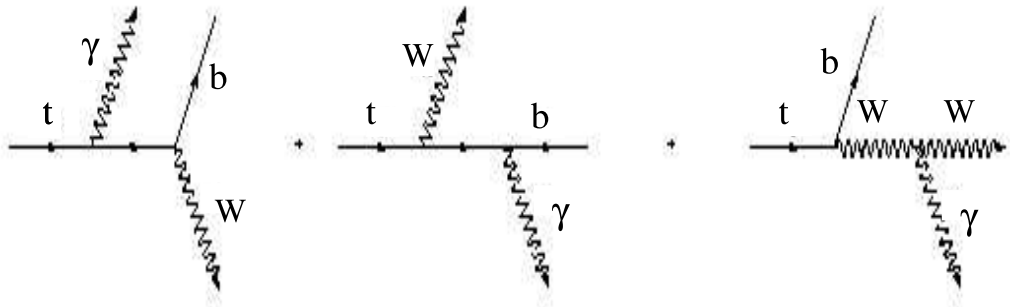
Figure 1.10. Feynman diagrams for $t \rightarrow Wb\gamma$

processes sensitive to the top charge and other radiative processes with the same experimental signature. At the Tevatron, $q\bar{q}$ annihilation dominates, so there is a huge irreducible background coming from radiative processes. Studies done by U. Baur et all. [6] show that CDF needs around $20\,fb^{-1}$ of data to measure the top charge at 95% confidence level using $t\bar{t}\gamma$ events. However, this technique is the suggested method for measuring the top charge at the Large Hadron Collider (LHC) where gluon-gluon fusion dominates.

The second technique, using the top decay products, is more promising at the Tevatron and is what is used for this thesis result. In the Standard Model the top (t) quark decays into a $W^+(Q_{W^+} = +1)$ and a bottom (b) quark ($Q_b = -1/3$) while the anti-top ($\bar{t}$) quark decays to a $W^-(Q_{W^-} = -1)$ and an anti-bottom ($\bar{b}$) quark ($Q_{\bar{b}} = +1/3$). Adding the charges of the decay products together gives the t ($\bar{t}$) a charge of +2/3 (-2/3). However, if the decay of the observed top is such that it decays to a W^- and a b quark, then the charge of t ($\bar{t}$) would be -4/3 (+4/3) and would not correspond to the Standard Model top quark. Such a possibility has been put forward in reference [7]. In the Standard Model, top is the $SU(2)$ partner of the left-handed b quark and the right-handed b is a singlet [1, 2]. In reference [7] a fourth generation of quarks is introduced and the right-handed b is allowed to mix with a heavy quark (Q_1) of charge -1/3 whose doublet partner(Q_4) has charge -4/3. In this model, Q_4 has a mass around 174 GeV/c^2 ($\approx$ the mass of the observed top quark) while the left-handed top quark is heavier with a mass of 274 ± 40 GeV/c^2. According to this scenario, our observed top is the exotic Q_4 quark and the Standard Model top would be the one at 274 ± 40 GeV/c^2. Below is shown the left-handed SU(2) doublets and how the right-handed singlet, b_R, is modified to mix with the fourth generation of quarks (Q_1, Q_4) where Θ_b is the mixing angle [7].

$$Standard Model \rightarrow \begin{pmatrix} t \\ b \end{pmatrix}_l \quad b_R$$

15

$$ExoticModel \rightarrow \begin{pmatrix} t \\ b \end{pmatrix}_l \quad \begin{pmatrix} Q_1 cos\Theta_b + bsin\Theta_b \\ Q_4 \end{pmatrix}_R$$

We will determine the charge of the top quark observed at the Tevatron using its decay products and check if it is the Standard Model top with charge 2/3 or the exotic Q_4 quark introduced above.

1.6 Overview of the Thesis

Chapter 2 is dedicated to the description of the Fermilab accelerator complex and the CDF detector without which this measurement could not have been done. How top events are selected out of millions of collisions at the CDF detector is explained in Chapter 3. Chapter 4 describes in detail how we determine the performance of the top charge analysis and each of the three essential ingredients for measuring the top's charge.

1. The charge of the W

2. The flavor of the b-jet: To determine if the b-jet is coming from a b or $\bar{b}$ quark

3. The correct Wb pairing: To assign the b-jet with the correct W to ensure that the b-jet and W come from the same top decay branch.

Chapter 5 is a detailed study to check how well the b flavor tagging method works in data. Chapter 6 describes how we take care of the non-top events, called background, that end up in our event sample despite the top event selection. Chapter 7 discusses the sources of systematic uncertainties and the effect of each on our measurement. Chapter 8 summarizes similar studies done in the dilepton channel. The total amount of expected background and signal is given in Chapter 9. The next chapter explains the statistical methods we chose for quoting our results. Chapter 11 is dedicated to the final result using both the lepton+jets and dilepton channels.

CHAPTER 2

Experimental Apparatus

> A machine that was powerful enough to accelerate
> particles to the grand unification energy would have to
> be as big as the Solar System and would be unlikely to be
> funded in the present economic climate.
> Stephen Hawking (1942-..), a British theoretical physicist.

2.1   Tevatron

The Tevatron is a large superconducting-magnet synchrotron[1] located at Fermi National Accelerator Laboratory (Fermilab) in Batavia, IL where bunches of protons (p) and anti-protons ($\bar{p}$) are accelerated and then made to collide. The counter rotating beams of protons and anti-protons collide with a center of mass energy of $\sqrt{s} = 1.96$ TeV at two locations where the CDF and DØ detectors have been placed. The Tevatron is housed in a tunnel with a radius of 1 km. In order to create energetic particle beams, Fermilab uses a chain of accelerators. The Tevatron is actually the name of the main circular ring, but is commonly used to refer to the entire chain of accelerators that are shown in Figure 2.1. The accelerator complex and the CDF detector used to collect the data for this thesis are explained below.

---

[1]A synchrotron is a type of circular accelerator where the strength of the magnetic field that keeps the beam in its orbit and the radio frequency found inside the accelerating region are synchronized to the beam momentum and revolution frequency.

2.1.1 Proton source

The Cockcroft-Walton pre-accelerator provides the first stage of acceleration. It takes hydrogen ions off a Cesium target and produces 750 keV H^- ions every 66 milliseconds. The linear accelerator, Linac, is approximately 500 feet long and receives the H^- ions from the Cockcroft-Walton pre-accelerator before accelerating them to an energy of 400 MeV. This is accomplished by RF cavities[2] carrying particles along a wave of electromagnetic radiation. The beam is also focused at this point using quadrapole magnets. The Booster receives the beam of H^- ions from the Linac and strips the electrons off, leaving bare protons (H^+ ions).

The Booster is a synchrotron that uses magnets to bend the beam of protons in a circular path. It is composed of a series of 75 magnets arranged around a 75 meter radius circle where the protons are accelerated to 8 GeV by the use of 18 RF cavities along the circle. Then the protons are sent to a larger synchrotron, the Main Injector, which is about half a kilometer in radius. The Main Injector is composed of 18 accelerating RF cavities and can accelerate 8 GeV protons from the Booster to 150 GeV every 22 seconds. The Main Injector also takes part in anti-proton ($\bar{p}$) production and acceleration that will be explained in the following subsection. The proton and anti-proton beams are injected into the Tevatron from the Main Injector which accelerates the beams to 980 GeV.

---

[2]An RF cavity is a gap placed into the beam pipe across which an electric field is applied. They are electrically resonant structures with a natural frequency in the radio frequency (RF) range [8].

Accumulator (8GeV)
Debuncher (8 GeV)

Linac

Booster

Switchyard

8 GeV
Inj

Main Injector
150 GeV

A0

Tev Extractions
Collider Aborts

Target

Recycler
8 GeV

CDF
and Low Beta

F0

p Abort

p

Tevatron
1 TeV

p̄

E0

C0

D0 Detector
and Low Beta

p Abort
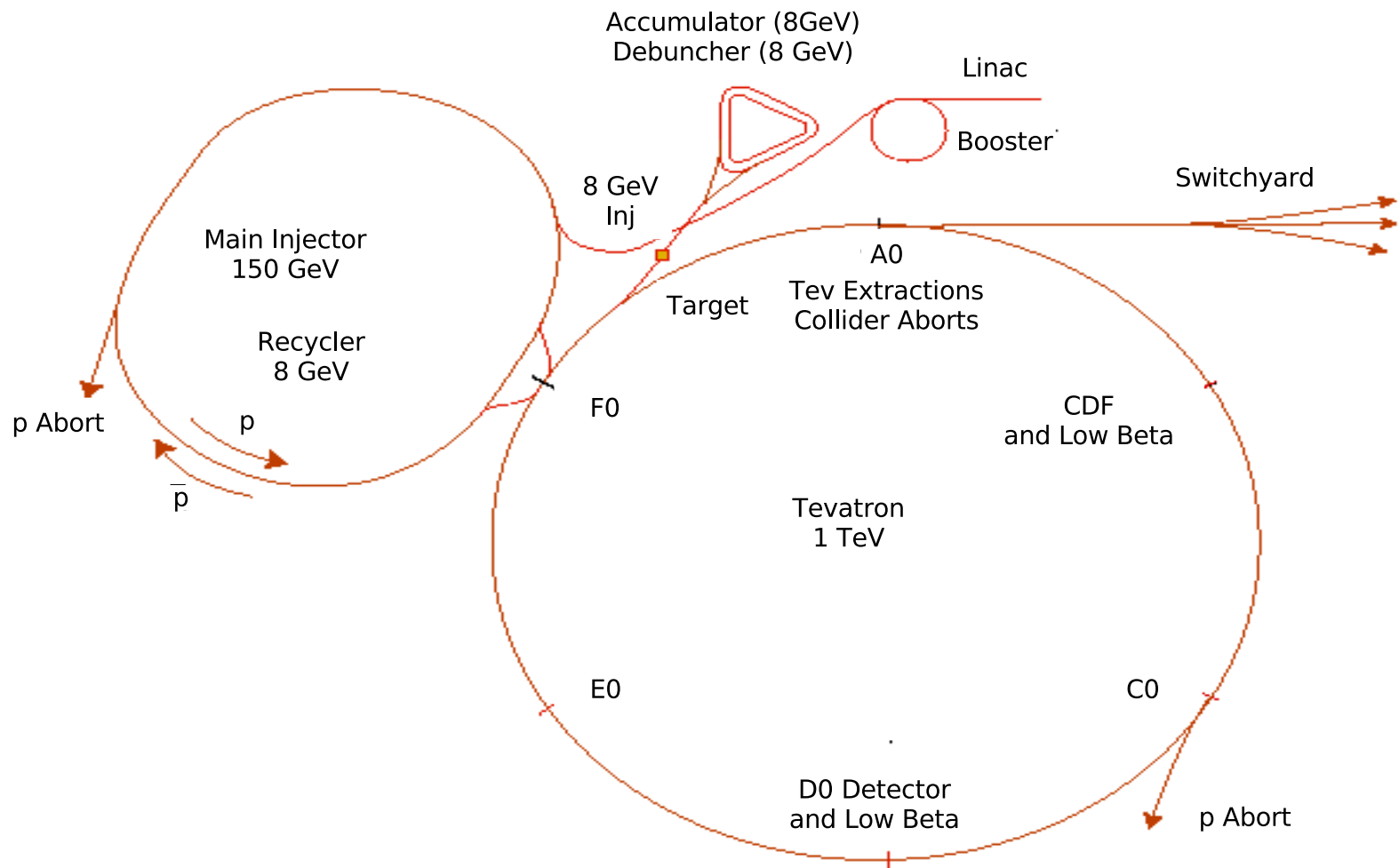
Figure 2.1. Diagram of the Tevatron accelerator complex.

2.1.2 Anti-proton Source

To produce anti-protons, the Main Injector sends 120 GeV protons to the anti-proton source, where they collide with a nickel target. The collisions produce a wide range of secondary particles including many anti-protons. The secondary particles are sent through a magnetic field and the particles with different masses and charges curve at different radii. The anti-protons are selected by negative expected curvatures that have a wide range of momenta, averaging around 8 GeV. The Debuncher, a rounded triangular synchrotron with a mean radius of 90 meters, comes into play next. It takes the anti-protons from the target by RF manipulation and performs stochastic cooling [8] to cool the beams transverse energy and decrease the momentum spread. Since anti-particles annihilate with particles, a storage system named the Anti-proton Recycler was built inside the Main Injector tunnel to keep anti-protons away from matter by keeping them in a circular orbit with the use of magnets. When a sufficient number of anti-protons has been produced, they are sent to the Main Injector for acceleration and then injected into the Tevatron.

2.1.3 Collisions at the Tevatron

The Tevatron is a circular synchrotron, and contains superconducting magnets, made of a niobium titanium alloy, cooled to 4.1 Kelvin and 8 RF cavities. Once all the protons and anti-protons are loaded into the Tevatron, both beams are ramped to 980 GeV. The same set of magnets and RF fields are used for the acceleration of the beams but the protons travel clockwise while the anti-protons travel in counter clockwise direction since they have opposite charges. The beams are squeezed in both beam directions using focusing magnets located on either side of both the CDF and D0 detectors and made to collide at these locations. Ideally the collisions would take place at the center of the detectors, but actually are distributed as a gaussian around the centers.

The Tevatron can sustain both beams for hours, called a "store". Both the proton and anti-proton beams in the Tevatron are divided into 36 bunches, each containing billions of particles at the beginning of a store. Each bunch is separated by electro-static separators so that beam crossings occur every 396 nanoseconds. Each product of a $p\bar{p}$ interaction is called an "event". Most of the collision events are from "glanc-ing" blows, with almost all the energetic particles moving along the beam pipe in both directions. The glancing collisions are called "Minimum-Bias" or "Minbias" events. Only a few hard collisions occur which produce energetic particles going off at large angles. These hard collisions are the ones that will be detected by the CDF detector.

At collider experiments, the particle flux produced by the accelarator is called luminosity (L). Both cross-section and luminosity are measures of collision rate and therefore measures of the amount of data collected. However, the cross-section covers the physics of the particle interaction whereas luminosity depends on the properties and performance of the accelerator. At the Tevatron, the "instantaneous luminosity", also called the peak luminosity, depends on the total number of protons and antipro-tons $(N_p, N_{\bar{p}})$, the number of bunches of each type (B), the frequency of bunch revo-lution (f_0) and the cross-sectional area of the bunches (σ^2) as given in the following equation:

$$L = \frac{N_p N_{\bar{p}} B f_0}{4\pi\sigma^2} \tag{2.1}$$

The instantaneous luminosity is not constant through out a store but falls exponen-tially with time due to the momentum spread in the transverse plane of the beam direction and particle losses from collisions. The best peak luminosity obtained at the Tevatron to date is $2.8 \times 10^{32} cm^{-2} s^{-1}$. The total luminosity or "integrated lumi-nosity" is obtained by the total particle luminosity integrated over time. Figure 2.2 shows the integrated luminosity gathered by the CDF detector, up to 2007, during the Run II data taking period that started in 2002. A summary of the Tevatron parameters for Run II are shown in Table 2.1. More information on the Tevatron can

be found at [8, 9].
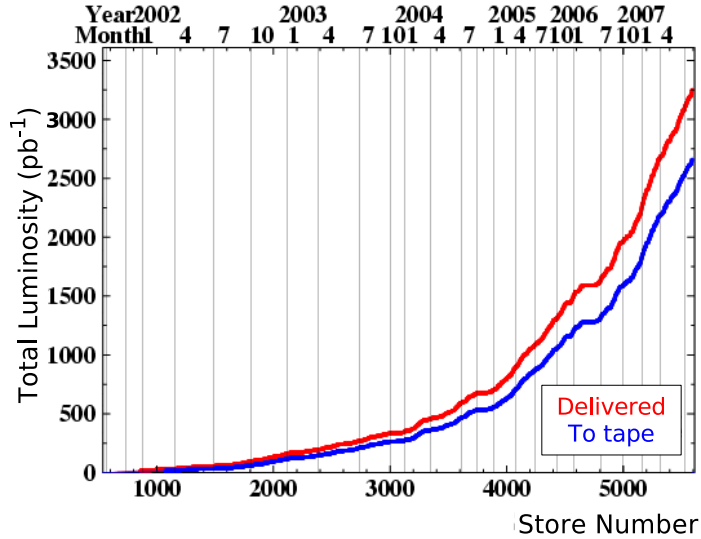
Figure 2.2. Total luminosity gathered by the CDF detector as of 2007. The red curve is the luminosity delivered by the Tevatron and the blue curve is the luminosity written to tape by CDF.

parameter	Run II
Number of bunches	36
bunch length [m]	0.37
bunch spacing[ns]	396
protons/bunch (N_p)	2.7×10^{11}
antiprotons/bunch $(N_{\bar{p}})$	3.0×10^{10}
interactions/crossing	2.3
integrated luminosity $[pb^{-1}]$	2000
peak luminosity $[cm^{-2}s^{-1}]$	2.8×10^{32}

Table 2.1. Tevatron parameters during the Run II data taking period.

2.2 The CDF Detector

The CDF detector is a multi-purpose solenoidal detector, about the size of a 3-story house, designed to identify particles produced from $p\bar{p}$ collisions at the Tevatron. CDF is specifically used to identify and measure the energy and momentum of electrons,

muons, photons and jets. An isometic view of the CDF detector can be seen in Figure 2.3.

The CDF detector consists of several subsystems arranged coaxially around the beamline. The innermost layers are tracking chambers, followed by electromagnetic and hadronic calorimeters and then muon chambers. The tracking chambers are located inside of a 1.4 T axial magnetic field. The magnetic field causes the trajectory of a charged particle to bend within the tracking chambers and the curvature of the trajectory is used to measure the momentum of the particle. The calorimeters that are outside the tracking system and magnetic field are used to measure the energy of electrons, photons and jets. The muon chambers that are wire and gas detectors are placed outside the calorimeter and used to identify muons. This chapter continues with the basics of particle detection. Then we will introduce the CDF coordinate system and give brief descriptions of each detector component. A more complete description of CDF can be found elsewhere [10, 11, 12].

2.2.1   Basics of Particle Detection at CDF

Information from each detector subsystem is combined for particle identification. The particles are recognized by their electronic signals and their detection relies on the way they interact with the detector material's nuclei and electrons. Therefore, only those particles that undergo strong or electromagnetic interactions can be detected. The neutrino, which is free from both interactions, is inferred from missing transverse energy. Both photons and electrons deposit energy in the electromagnetic calorimeter, but photons, like other uncharged particles, do not leave a signal in the tracking chambers. Hadrons (such as protons and pions $\pi^{\pm}$) start their energy deposit in the electromagnetic calorimeter, but will be absorbed fully only in the outer hadronic calorimeter. The only particles that traverse the entire detector are muons and they leave signals in the outer muon detectors. A basic representation of particle detection

EM Calorimeter

Hadron calorimeter

Plug Calorimeter

Forward Calorimeter

Muon Systems

Silicon Detectors

Solenoid Magnet

Tracking Drift Chamber (COT)
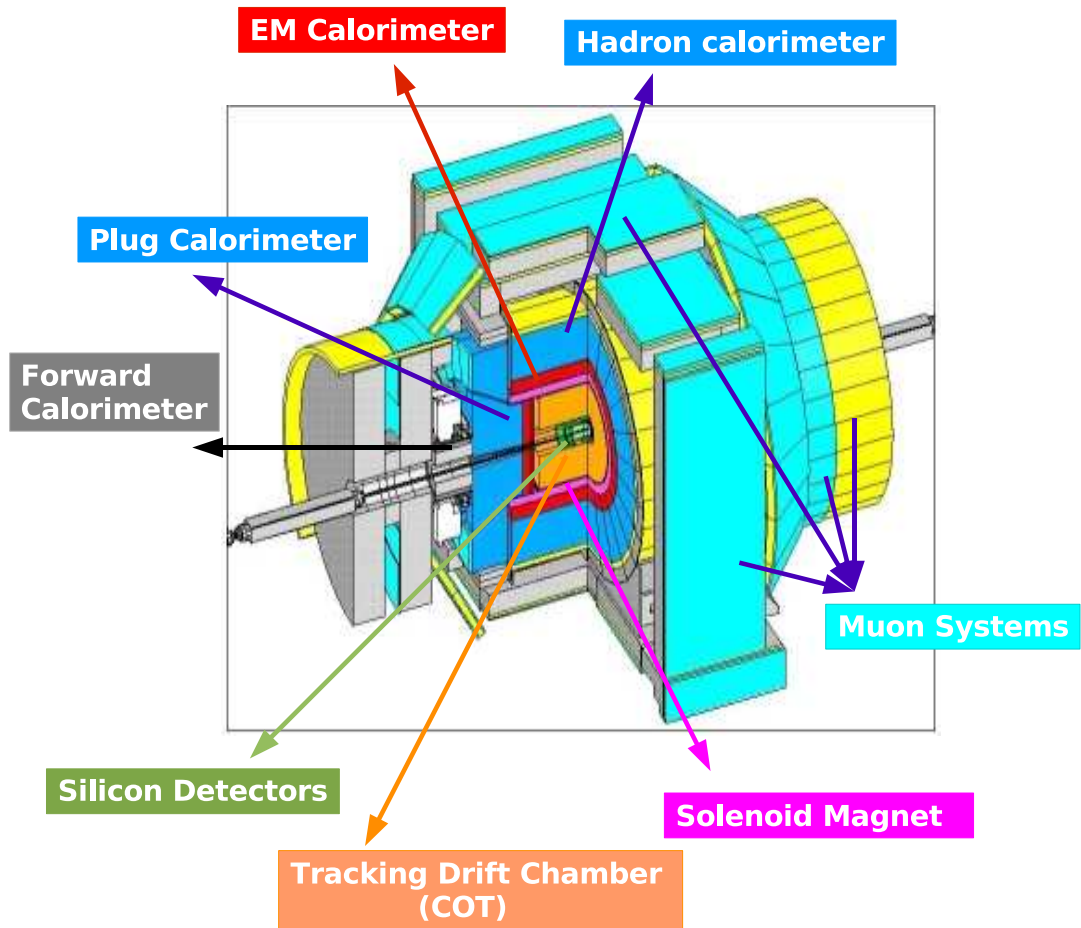
Figure 2.3. Isometric view of the CDF detector.

is shown in Figure 2.4.
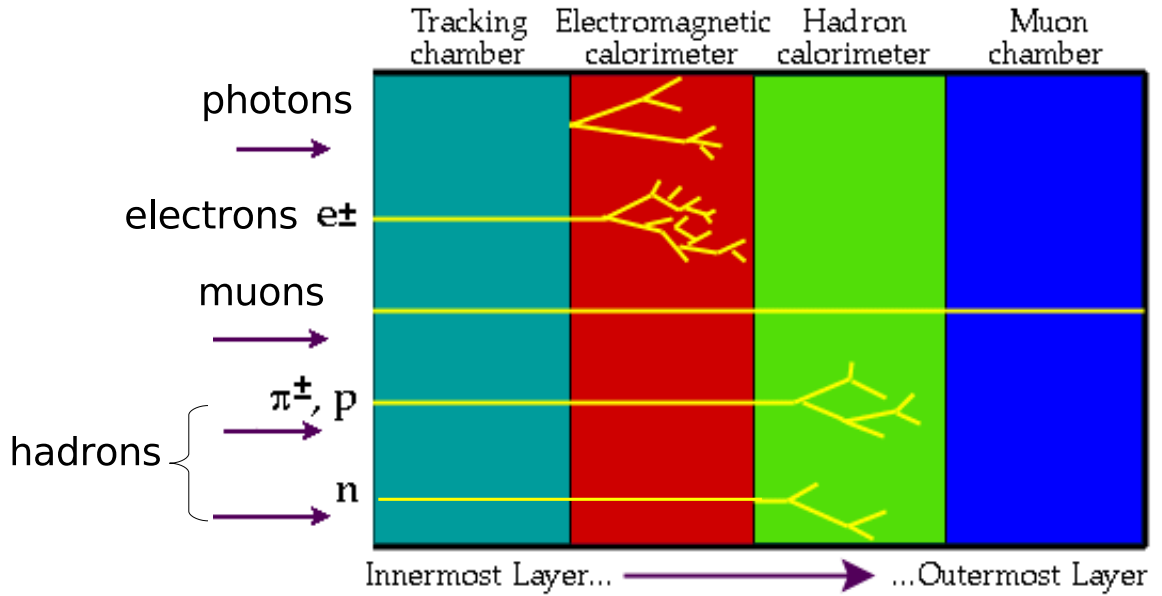
Figure 2.4. Basic representation of particle detection

2.2.2 CDF Coordinate System

The CDF detector is forward-backward symmetric about the transverse plane through the interaction region. The z-axis is defined to lie along the proton beam direction (from west to east). The variable ϕ is the azimuth around the z-axis and θ is the polar angle relative to the z-axis. Because θ is not a Lorentz invariant variable, the rapidity (Y) is used instead. Equation 2.2 is the expression for rapidity where E and p_z are the energy and z component of the momentum of the particle respectively. The rapidity simplifies to pseudorapidity (η) in high energy collisions where the momentum of the particle is much greater than its mass. Equation 2.3 defines η as a function of θ. As can be seen in Figure 2.5 $\theta = 90^0$ corresponds to $\eta = 0$ and the lower the θ the higher the η value.

$$Y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{2.2}$$

$$\eta = -\ln(tan(\frac{\theta}{2})) \tag{2.3}$$

25
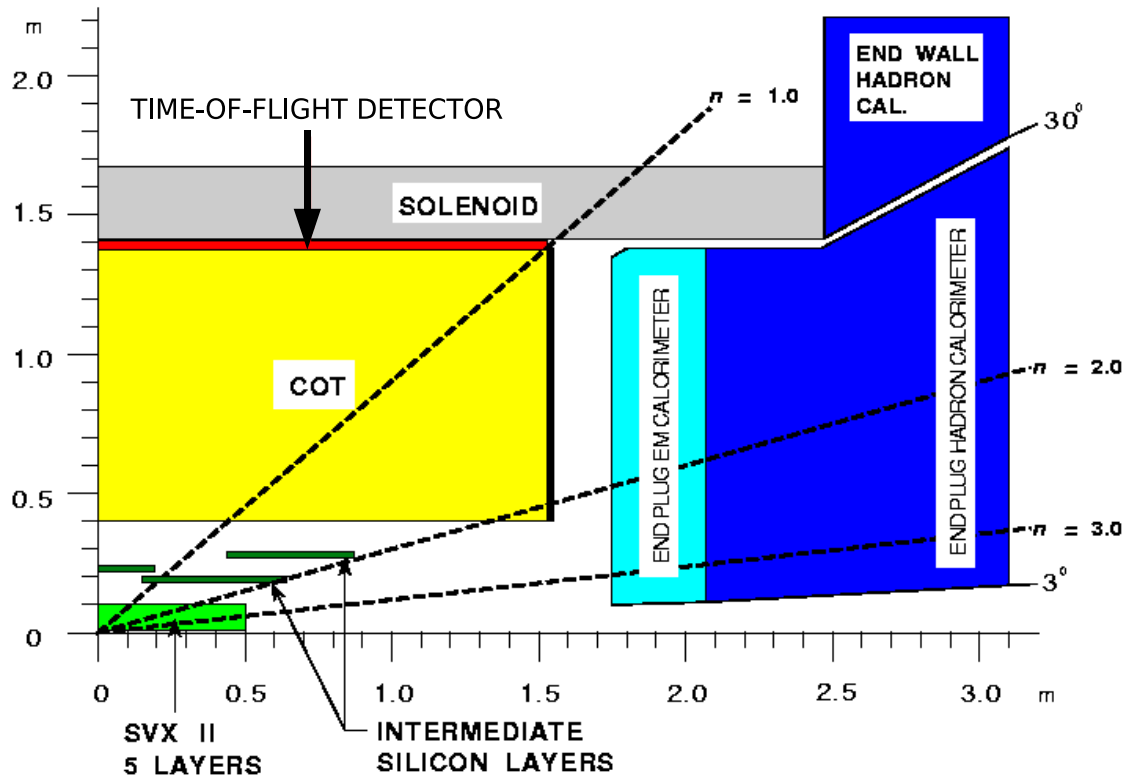
Figure 2.5. A quadrant of the longitudinal cross-section of the CDF detector.
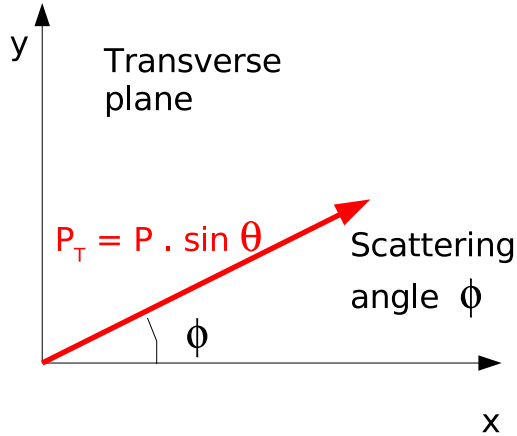
Figure 2.6. The projection of the particle's momentum onto the transverse $(x - y)$ plane. The beam is in the direction of the z-axis that is out of the page. The angle θ is between z-axis and the momentum of the particle.

At CDF the direction of an outgoing particle is represented by a point in $\eta - \phi$ space. We can not accurately measure the momentum along the z-axis (P_z) since we can not know which parton from proton was collided with which parton from the anti-proton and also particles are lost down the beampipe. So the variable used instead is the momentum transverse (P_T) to the beam.

$$P_T = P \cdot sin(\theta) = \sqrt{P_x{}^2 + P_y{}^2} \tag{2.4}$$

where P_x and P_y are the momentum components in the x and y axes respectively. Figure 2.6 represents the P_T of a particle's momentum vector in the detector's coordinate system. Similarly, transverse energy is the energy perpendicular to the beam direction.

$$E_T = E \cdot sin(\theta) \tag{2.5}$$

2.2.3 The Tracking Detectors

The CDF tracking systems sit inside a 1.4 T magnetic field which is created by a 4.8 m long, 1.5 m radius solenoid. The precision reconstruction of a track's momentum is acquired using the curvature of the particle's track in the magnetic field which is

pointing in the z direction. The tracking system includes three silicon detectors and the Central Outer Tracker (COT).

2.2.3.1   THE SILICON DETECTOR

The silicon detector is at the center of CDF and is the first detector that particles traverse after the initial collision. It consists of three subsystems: Layer∅∅ (L∅∅), the Intermediate Silicon Layer (ISL) and the Silicon Vertex Detector (SVXII). All three subsystems use the same principle for a charged particle's position measurement, basically the ionization signal left behind in a semiconductor. When a charged particle passes through the depletion region of a biased p-n semiconductor junction, ionization produces electron-hole pairs and the electric field causes them to drift in opposite directions. By segmenting the p and n sides of the junction into strips and reading out the charge deposition left on every strip, one can measure the path length traversed in the detector by the charged particle and find its position. The SVXII and ISL detectors are made of double-sided silicon-microstrips. In the double-sided detectors the p side has strips parallel to the z-direction called axial strips, and the other side has strips at an angle with respect to the z-direction called the stereo strips. The stereo strips provide measurements in the $r - \phi$ plane. Together with the axial information a 3 dimensional reconstruction of each track is possible. This information is readout from electronic chips that are mounted on the silicon sensors.

The SVXII detector is comprised of three cylindrical barrels, as shown in Figure 2.7, placed end to end, each is 29 cm in length. Each barrel provides the support frame for five layers of double-sided silicon microstrip detectors arranged in concentric rings starting at a radius of 2.4 cm from the beamline and moving out to 10.6 cm. The three SVXII barrels cover the interaction point where the beams collide and provides standalone track information in the pseudo-rapidity range of $1 < |\eta| < 2$. The design parameters of the SVXII detector are summarized in Table 2.2.

Number of layers	5
Number of ϕ wedges	12
Number of barrels	3
Barrel length	$29cm$
Ladder length	$20.9cm$
Radius: Layer 0	$2.44cm$
Radius: Layer 4	$10.6cm$
Stereo angle	$90, 90, 1.2, 90, 1.2^0$
$r - \phi$ pitch	$60, 62, 60, 60, 65\mu m$
$r - z$ pitch	$141, 125, 5, 60, 141, 65\mu m$

Table 2.2. Design parameters of the SVXII detector. The numbers in the last three rows of the second column correspond to each layer from inward to outward.

The ISL detector is located concentrically outside the SVXII as shown in Figure 2.8. In the central region ($|\eta| < 1$) the ISL has one layer of silicon at a radius of 22 cm. In the region $1 < |\eta| < 2$ there are two layers of silicon at radii of 20 cm and 28 cm. The additional coverage provided by the ISL aids in linking SVXII hits to COT tracks in the region $|\eta| < 1$. Furthermore, in the $|\eta| > 2$ region where there is little COT coverage the combination of the ISL and SVXII provides tracking in three dimensions. The η coverage of each silicon system can be seen in Figure 2.9. The design of the ISL is identical to the SVXII as far as the data acquisition and their power supplies are concerned, but longer strips with a wider readout pitch are used for the ISL since it sits outside of the SVXII. Being further away from the beam, the ISL is also less affected by radiation damage. The double-sided ISL detector has one side with axial strips at a pitch (the distance between the strips) of 55 μm and stereo strips at a 1.2^0 angle with respect to the z direction at a pitch of 73 μm on the other side.

The Layer∅∅ (L∅∅) is a single-sided high-voltage silicon detector that sits directly on the beam pipe, just 1.5 cm away from the beamline. It provides the position measurement closest to the interaction point and covers $1 < |\eta| < 2$. L∅∅ was added to improve CDF's impact parameter[3] (d_0) resolution. Layer∅∅ compensates for

---

[3]The distance between a track and the beam axis in the $r - \phi$ plane at the closest approach.
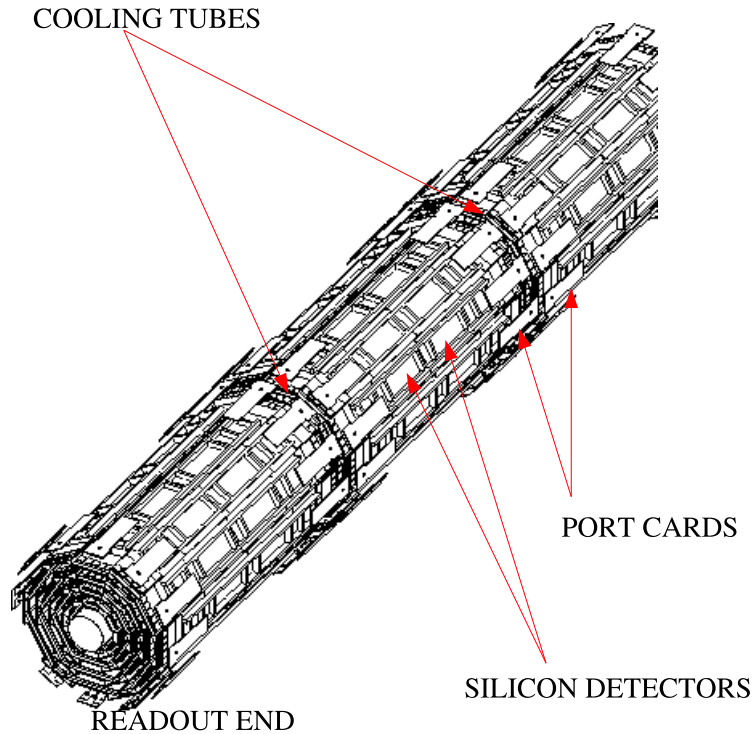
Figure 2.7. Portrait of SVX barrels.

radiation damage to layers of the SVXII. It also helps to resolve any ambiguities in matching COT tracks to SVXII tracks in the dense-track environment that is expected at high instantaneous luminosities. More information on the CDF silicon detectors can be found at reference [13].

2.2.3.2 THE CENTRAL OUTER TRACKER

The Central Outer Tracker (COT) is a 3 m long cylindrical open-cell drift chamber located outside the CDF silicon detectors. It provides tracking data between 40 cm and 132 cm radially from the beam pipe. The chamber is filled with a gaseous mixture of argon, ethane and alcohol. It uses the signal information from electrons ionized in a gas to calculate the spatial position of the ionizing particle.

The COT contains 96 sense wire layers, which are grouped into 8 superlayer (See Figure 2.10). The superlayers are divided into two types of cells arranged in alter-
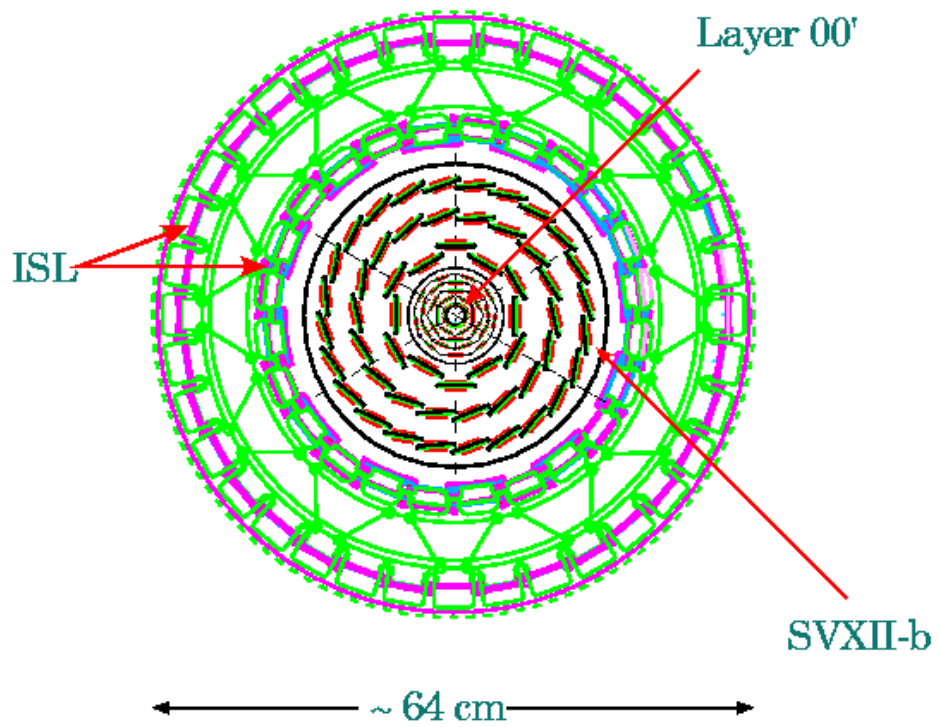
Figure 2.8. The silicon tracking detectors projected on the $r - \phi$ plane.
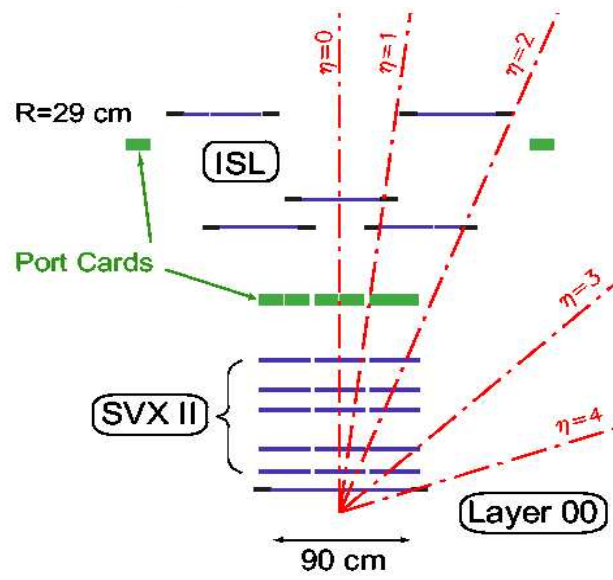
Figure 2.9. The silicon tracking detectors projected on the $r - z$ plane.
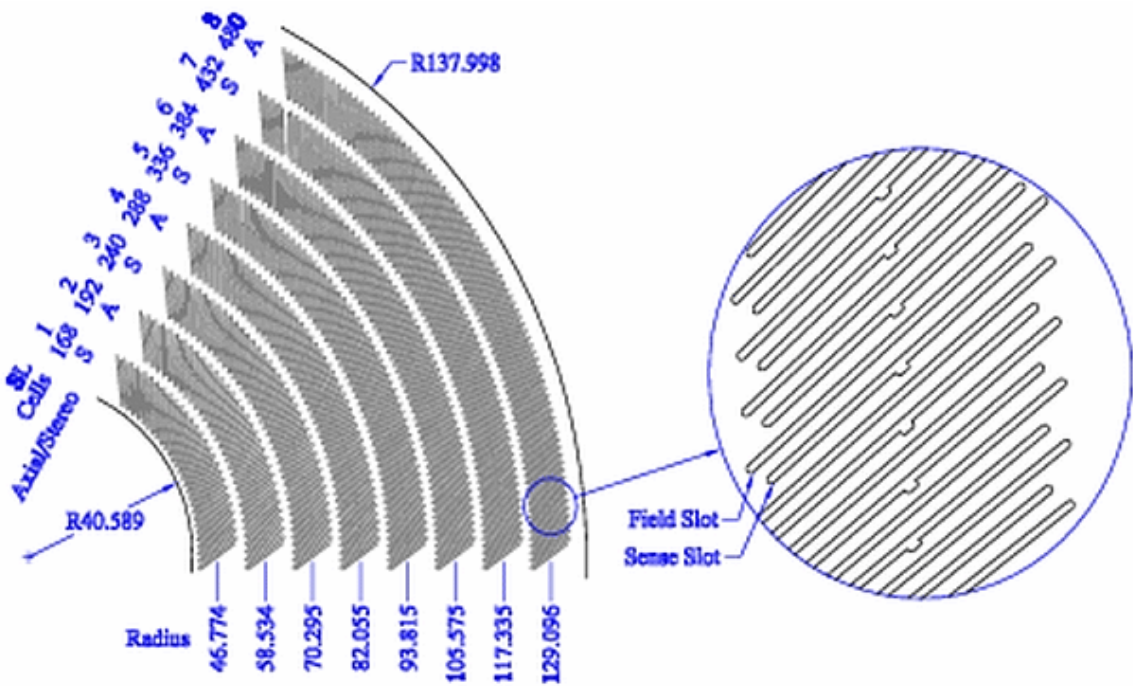
Figure 2.10. The layout of the wire planes on a COT endplate.
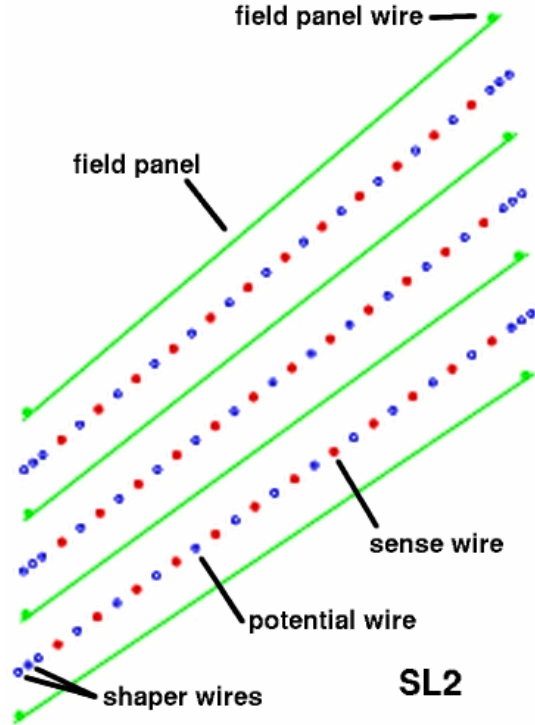
Figure 2.11. The cross-section view of the COT cells.

nating rows of axial (parallel to beamline) and stereo (2-degree offset from parallel) wires. The five axial layers are cells of 12 sense wires interleaved with potential wires as shown in Figure 2.11 and provide tracking information in the $r - \phi$ plane, while the stereo layers are cells of 6 sense wires that provide tracking information in the $r - z$ plane. A 3-dimensional track reconstruction is achieved by combining the information from all the cells.

When a charged particle passes through the COT chamber, it deposits a number of ionized particles that drift toward and hit the sense wires in the electric field created by the potential wires. These ionized particles are distributed along the track's trajectory and the signals on the sense wires are processed by COT electronics that provide hit time and charge deposition information from each wire. To determine the moment at which a charged particle passes through the COT, a time of flight detector (TOF) made out of scintillating counters is used. The TOF is located between the COT and

the solenoid as can be seen in Figure 2.5. With a resolution of 100 ps, the TOF system also provides particle identification for low momentum tracks. The momentum and hit resolutions of the COT are 0.0017 GeV/c and 140 μm respectively.

2.2.4 Calorimetry

The measurement of particle energies is done using calorimetry. As particles pass through the detector, they lose energy as they interact with the detector material and form showers. The energy absorbed by the calorimeter is measured by the use of scintillators that are connected to photomultiplier tubes (PMTs). The amount of light collected by the PMTs is proportional to the amount of energy lost by the particle. Different particles interact differently with the detector material. High energy electrons lose their energy primarily through the radiation of photons, called bremsstrahlung and form electromagnetic showers. However, for muons the electric field of the atoms in the calorimeter is generally not large enough to change its direction and start a shower in the calorimeter. So muons generally pass straight through the calorimeter depositing very little of their energy. Hadrons also are not deflected by the electric fields produced by the atoms of the calorimeter material. However, when a hadron passes sufficiently close to a nucleus, there are residual strong interactions between the hadron and the protons in the nucleus. These interactions result in a variety of processes that produce additional particles and slow down the initial high energy incoming particle producing a shower of particles called a "jet" and leaving energy in the calorimeter. The energy of a particle after traversing a distance x in the calorimeter material is given by $E_x = E_o e^{-x/x_0}$ where x_0 is the radiation length in the case of electromagnetic showering and nuclear interaction length in the case of hadronic showering, E_o is the energy of the particle before entering the calorimeter. The CDF calorimeter consists of electromagnetic and hadronic parts that are made of alternating layers of absorbing and detecting material as depicted

in Figure 2.12. Lead and iron layers are used at CDF as the absorbing material for the electromagnetic and hadronic calorimeters respectively. The reason behind using a much denser material for the electromagnetic calorimeter is to ensure that the electrons are all absorbed in the electromagnetic portion of the calorimeter that is placed before the hadronic portion.
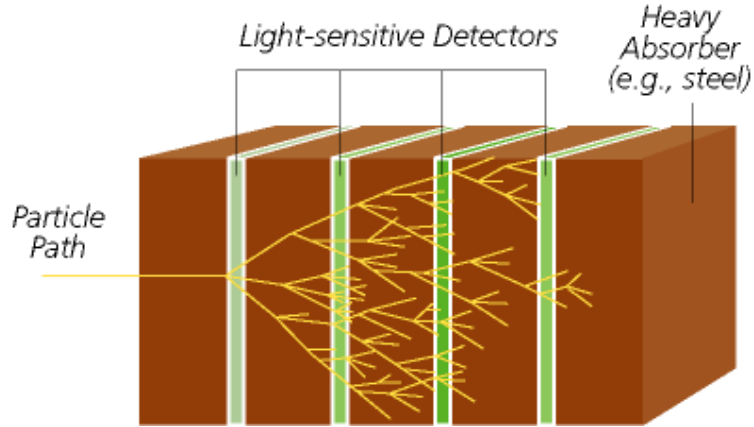
Figure 2.12. A representation of a calorimeter structure where scintillators are sandwiched between absorbing calorimeter material.

The CDF electromagnetic calorimeter system, placed just outside the tracking chambers and magnetic field, is made up of two subsystems: The Central Electromagnetic Calorimeter (CEM) and the Plug Electromagnetic Calorimeter (PEM). Similarly the CDF hadronic calorimeter system has central (CHA) and plug (PHA) parts. Figure 2.13 is a drawing of CDF with the calorimeter systems labeled. The central calorimeters CEM and CHA are cylindrical shaped detectors filling the radius from 1.5 m to 3.0 m and z from -2.5 to 2.5 m. Each calorimeter piece has a support structure called a wedge.

Figure 2.14 shows the structure of a single central wedge with alternating layers of absorbing materials and scintillators. The light produced in the scintillators are collected by acrylic lightguides and transmitted to the PMTs. The CEM is organized into 24 wedges in phi each with 10 towers which are readout separately. The CHA
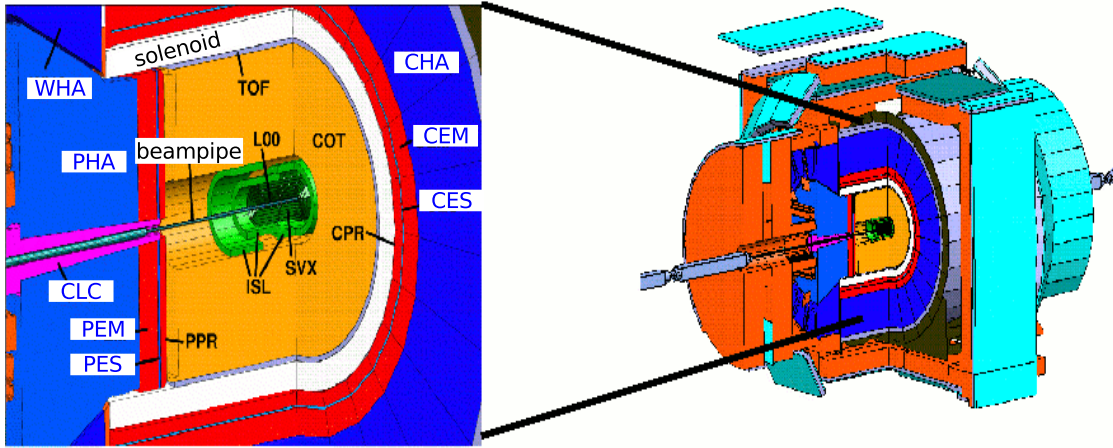
Figure 2.13. The CDF detector with the calorimeter systems labeled.

sits just outside of the CEM and is designed to match the geometry of the CEM.

System	Acronym
Central Electromagnetic Calorimeter	CEM
Plug Electromagnetic Calorimeter	PEM
Central Hadronic Calorimeter	CHA
Plug Hadronic Calorimeter	PHA
Wall Hadronic Calorimeter	WHA
Central Preradiator	CPR
Plug Preradiator	PPR
Central Electromagnetic Shower Max detector	CES

Table 2.3. The parts of the CDF calorimeter and their acronyms.

To enable a more precise measurement of the transverse profile of an electromagnetic shower, a proportional strip and wire chamber, called the Central Electromagnetic Shower Max detector (CES), is embedded in the CEM. This chamber consists of cathode strips running in the azimuthal direction and anode wires running in the z direction, enabling a 3 dimensional measurement of the showers. This is useful for the precision matching of tracks to electromagnetic clusters. The Central PreRadiator (CPR), placed before the central calorimeter helps to distinguish electrons from
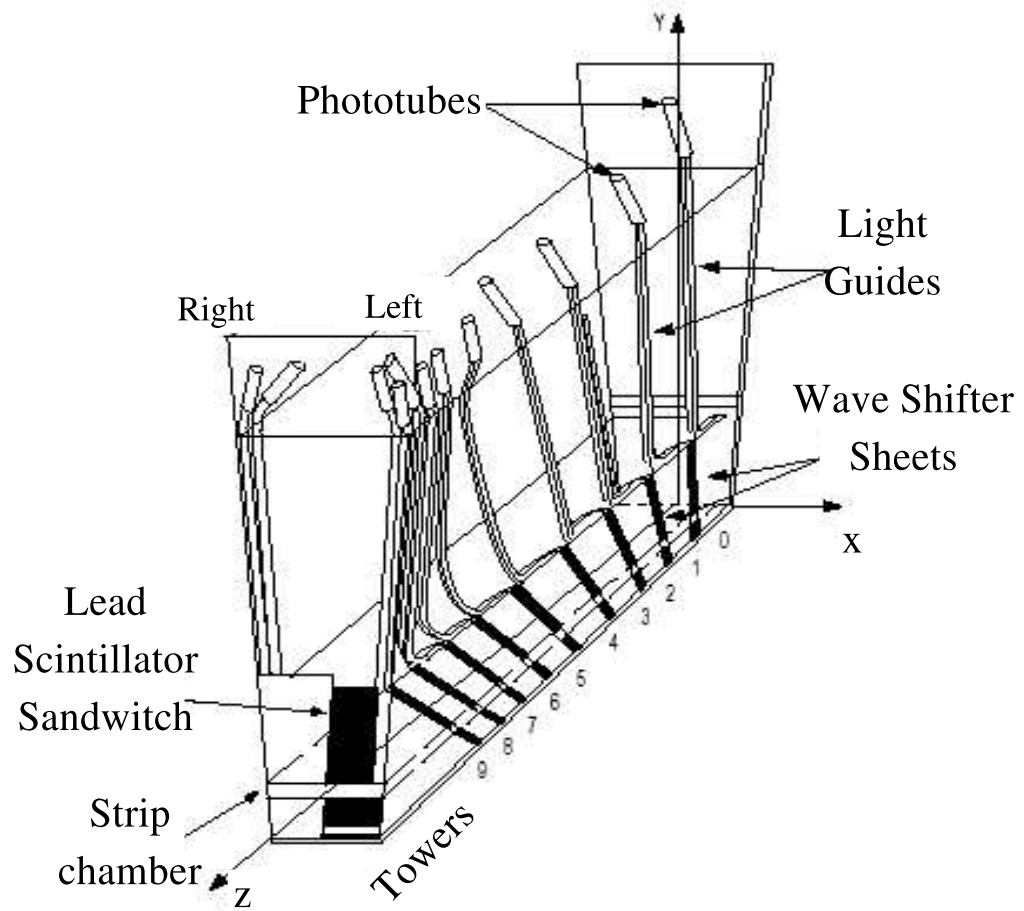
Figure 2.14. Diagram of a single calorimeter wedge.

System	detector type	# of towers	Coverage η	Energy Resolution (%)	Thickness
CEM	Pb/scintillator	478	$|\eta| < 1.1$	$13.5/\sqrt{E_T} \oplus 3$	$18X_0$
PEM	Pb/scintillator	480 per plug	$1.1 < |\eta| < 3.6$	$16/\sqrt{E_T} \oplus 1$	$21X_0$
CHA	Fe/scintillator	384	$|\eta| < 0.9$	$50/\sqrt{E_T} \oplus 3$	$4.5\lambda_0$
WHA	Fe/scintillator	288	$0.7 < |\eta| < 1.2$	$75/\sqrt{E_T} \oplus 4$	$4.5\lambda_0$
PHA	Fe/scintillator	432 per plug	$1.2 < |\eta| < 3.6$	$80/\sqrt{E_T} \oplus 5$	$7\lambda_0$

Table 2.4. Properties of the CDF calorimeter systems. The energy resolutions are for electrons for the electromagnetic calorimeter and for isolated pions for the hadronic calorimeters. The percentages after $\oplus$ are the average uncertainty due to tower calibrations. λ_0 implies interaction lengths and X_0 implies radiation lengths.

hadrons. Electrons deposit some energy in the CPR due to their interaction with the solenoid coil. Hadrons, on the other hand, are less likely to interact with the coil and should leave little or no energy.

The plug calorimeters have analogous components to the central system; electromagnetic and hadronic calorimeters and a preradiator detector. The plug calorimeters cover the region from $|\eta| > 1.1$ to $|\eta| < 3.6$. Like the central calorimeters, the Plug Electromagnetic Calorimeter (PEM) is followed by the Plug Hadronic Calorimeter (PHA). Each PEM has 480 towers, organized into 12 tower groups in η and each PHA has 432 towers organized into 11 towers. The Wall Hadronic Calorimeter (WHA) fills the gap between the hadronic plug and central calorimeter and extends the hadronic calorimeter coverage to the endwall region ($0.8 < |\eta| < 1.2$). The properties, such as the number of towers, η coverage and resolution, are summarized for each calorimeter component in Table 2.4 with the names of each acronym listed in Table 2.3.

Each calorimeter is made up of multiple individual cells, over whose volume the absorbed energy is integrated. These cells are aligned to form towers typically along the direction of the incident particle. Each calorimeter tower consists of an electromagnetic shower counter followed by a hadronic calorimeter. The top drawing in Figure 2.15 is a lego plot of the energy deposits detected in each cell of the calorimeter from an event with both an electron and a muon.

The calorimeters were originally calibrated using a test beam of known particle types but they are still calibrated periodically using radioactive sources as well as with Xenon and LED light flashers. The energy measured in the calorimeter is corrected for differences in response, non-linearities and time-dependent changes.

2.2.5 Muon systems

Because muons are more massice than electrons, they do not initiate electromagnetic showers. They also do not interact strongly so do not shower in the hadronic calorime-

3-SEP-1997   10:30 Run 58796   Event 417

An e-mu event

Neutrino

Electron   Muon

Jet1

Jet2

φ   η

Proton

Antiproton
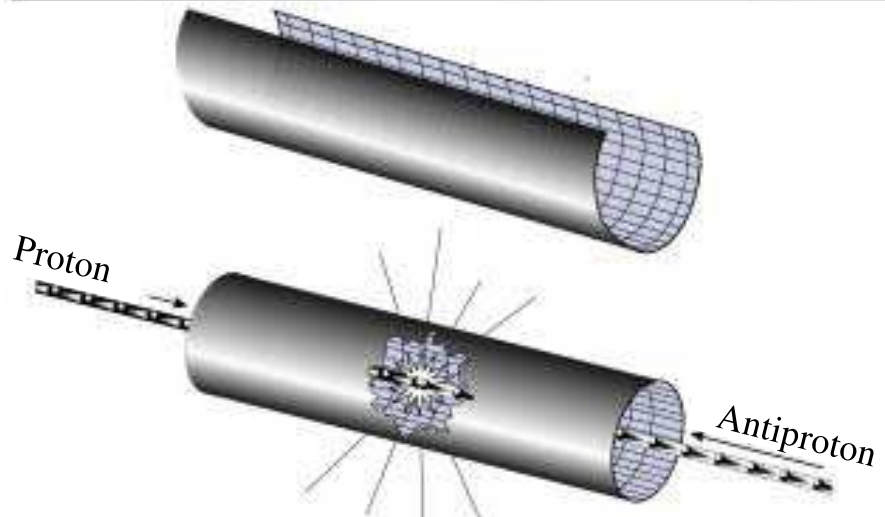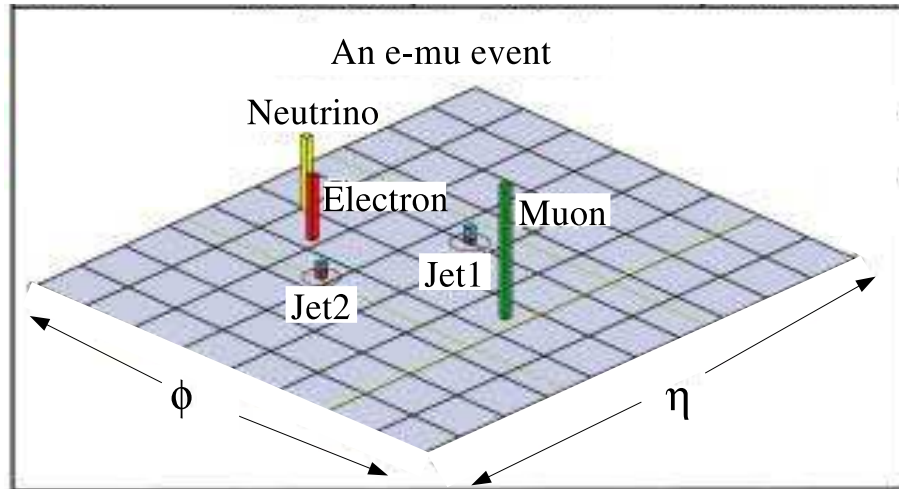
Collision in Calorimeter

Figure 2.15. Calorimeter towers for an electron and a muon shown in $\eta - \phi$ space of the calorimeter, obtained by rolling out the cylindrical calorimeter.

ter either (however muons with transverse momenta of 1 GeV/c or more will deposit a small fraction of their energy in the calorimeters due to ionization). Therefore, the muon system is placed at the outer most part of the CDF detector. The CDF muon system consists of four detectors: The Central Muon Detector (CMU), the Central Muon Extension Detector (CMX), Central Muon Upgrade Detector (CMUP) and Intermediate Muon Detector (IMU=BMU+BSU+TSU). A view of CDF with all the muon systems labeled can be seen in Figure 2.16.
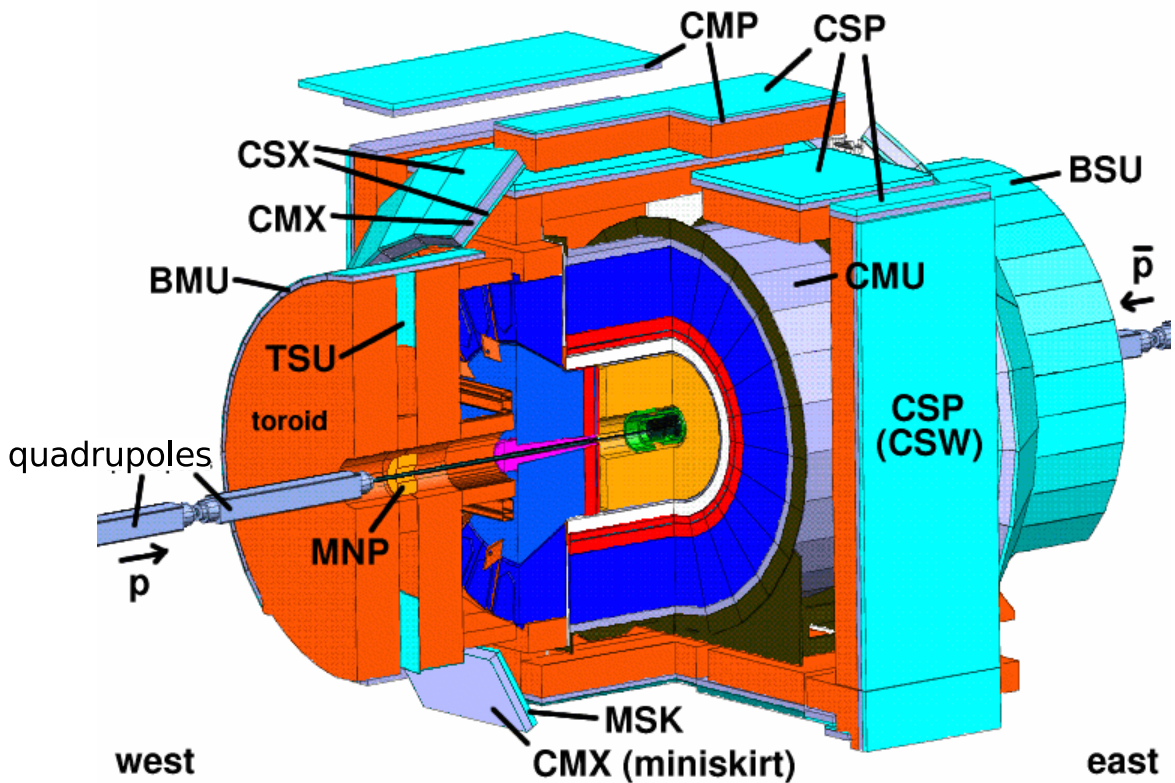
Figure 2.16. The CDF detector with the muon systems labeled.

The CMU and CMX detectors are capable of detecting muons with $P_T > 1.4$ GeV/c while the CMP and IMU can detect muons with $P_T > 2.2$ GeV/c and $1.4 \leq P_T \leq 2.0$ GeV/c respectively. If a hadron reaches the CMU chambers, it is generally a pion or kaon. About 1% of all pions and 2 to 4% of all kaons can fake a muon signal in the CDF detector. The CMP is less sensitive to these hadrons since it is
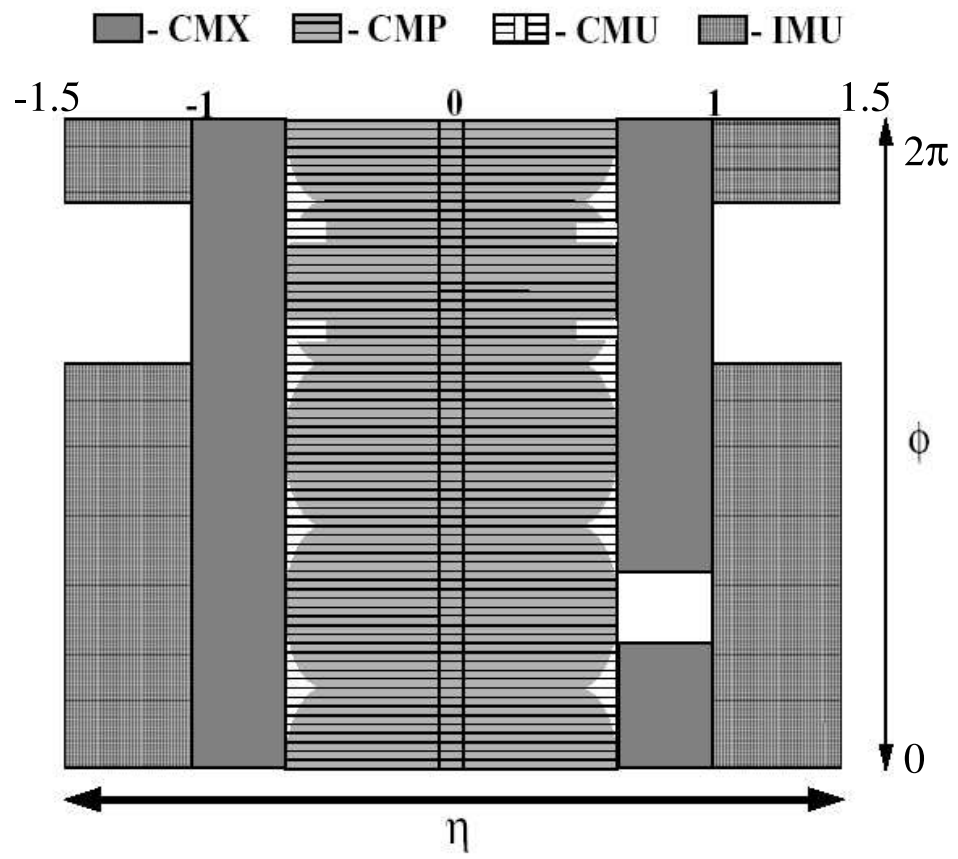
Figure 2.17. CDF muon coverage.

Parameter	CMU	CMP	CMX	IMU
η range	≤ 0.6	≤ 0.6	$\geq 0.6 \leq 1.0$	$\geq 1.0 \leq 1.5$
Drift tube cross-section (cm)	2.68×6.35	2.5×15	2.5×15	2.5×8.4
Drift tube length (cm)	226	640	180	363
Max drift time(ns)	800	1400	1400	800
Number of drift tubes	2304	1076	2208	1728
Scintillation counter dimensions (cm)	-	$2.5 \times 30 \times 320$	$1.5 \times 30 - 40 \times 180$	$2.5 \times 17 \times 180$
Number of scintillator counters	-	269	324	864
Multiple scattering resolution	12cm/p GeV/c	15cm/p GeV/c	13cm/p GeV/c	13-25cm/p GeV/c
Pion interaction lengths	5.5	7.8	6.2	6.2-20
Minimum detectable muon P_T GeV/c	1.4	2.2	1.4	1.4-2.0

Table 2.5. Design parameters of the muon chambers

placed behind an additional 60 cm of steel. The CMU and CMP cover the central region up to $|\eta| < 0.6$. The muons in the central region $0.6 \leq |\eta| \leq 1$ and the muons in the forward region $1 \leq |\eta| \leq 2.0$ can be identified by the CMX and IMU detectors respectively. The ϕ coverage of the muon system is not complete as can be seen in Figure 2.17. The CMU and CMP cover 84% and 63% of the solid angle respectively while the CMX chambers, located on the east and west ends of the CDF detector, cover 71% of the solid angle.

The CMU detector is a wire and argon-ethane gas detector, while the CMP, CMX and IMU are made up of both drift cells and scintillator plates. The inner and outer surfaces of the CMP are lined with scintillator plates. Similarly, the CMX and IMU drift tube chambers are backed by scintillator counters. Each counter is readout by a single phototube. More detailed information, such as the number of scintillators and drift tubes, can be found in Table 2.5.

Muons ionize the argon-ethane gas as they pass through the drift chambers and the ionized particles are gathered by the sense wires. By using the hits on the sense wires the path of the muon track segment, called a stub, can be found. A muon is reconstructed if such a stub is found in one of the muon systems and can be extrapolated back to a COT track. Muons reconstructed in both the CMU and CMP chambers are called CMUP muons. There are also muons that are reconstructed only in the CMU or the CMP muon chambers due to gaps in the muon chambers coverage. These muons are called CMU-only and CMP-only respectively. Because the CMU/CMP coverage is only up to $|\eta| < 0.6$ and there are gaps in ϕ a muon can miss the central muon detector but have a track in the COT. Therefore, a high P_T track without a corresponding muon stub can also be a muon candidate, called a CMIO (Central Minimum Ionizing Object).

2.2.6 Luminosity Measurement and CLC detector

At hadron collider experiments the beam luminosity can be measured using the process of inelastic proton anti-proton ($p\bar{p}$) scattering that has a cross-section (σ_{in}) of roughly 60 mb. The rate of inelastic $p\bar{p}$ interactions is expressed by:

$$\mu \cdot f_{bc} = \sigma_{in} \cdot L \qquad (2.6)$$

where L is the instantaneous luminosity, f_{bc} is the rate of bunch crossings and μ is the average number of $p\bar{p}$ interactions per bunch crossing.

At CDF Cherenkov Luminosity Counters (CLC), long conical isobutane gas-filled counters, are used for the luminosity measurement. The CLC is actually designed to measure μ accurately. There are two CLC detectors at CDF one located in the direction of the proton beam and the other in the direction of the anti-proton beam. There are 48 CLC cones per side, arranged in 3 layers of 16 cones each. Figure 2.18 shows the location of one of the CLCs in a quadrant view of the CDF detector. At the ends of the detectors, there are light collectors that reflect light into photomultiplier tubes (PMTs) that are protected from the magnetic field with an iron shield. Since the CLC detectors point to the interaction point, they are sensitive to the primary particles coming directly from inelastic collisions, while rejecting softer secondary particles. Figure 2.19 shows the passage of a particle through a Cherenkov cone with Cherenkov angle θ, which is 3.1^0 for a primary particle at CDF. Considering the dependence of Cherenkov radiation to θ, to the length of the Cherenkov cone and the efficiency of the PMTs, a single $p\bar{p}$ collision is expected to generate a PMT signal of around 100 photoelectrons.

2.2.7 Triggers and Data Acqustion System

The $p\bar{p}$ beams cross each other in the center of the CDF detector approximately 2.5 million times every second and almost every time a collision occurs several additional
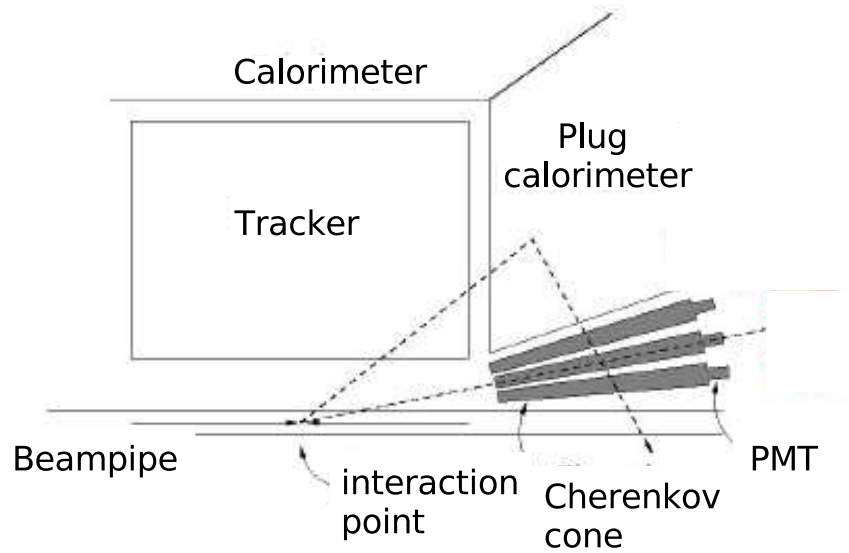
Figure 2.18. The quadrant view of the CDF detector that shows the location of one of the CLCs. The amplitude of the signal from the PMTs is proportional to the luminosity.
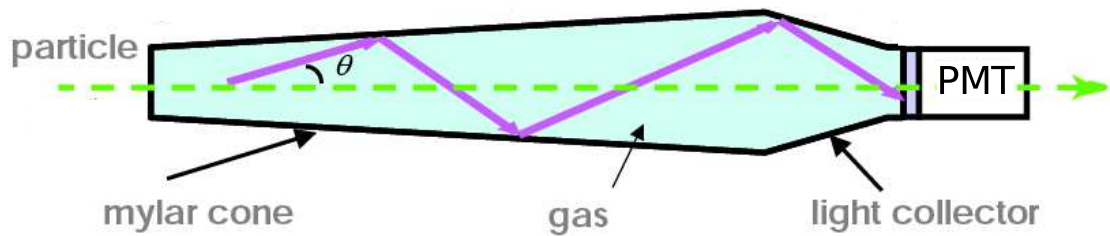
Figure 2.19. The passage of a particle through a Cherenkov cone.

particles are produced. However, for practical reasons CDF limits the number of collisions that will be stored for further analysis on magnetic tape to ≈ 75 events per second. It is the trigger system that picks these saved events out of the data flow. The time allowed to decide if an event is interesting or not depends on the separation between the accelerator bunches which was designed to be 132 ns. The CDF trigger architecture was built in such a way that it would be "deadtimeless", meaning the trigger system should make a decision before the next collision occurs. CDF has implemented a 3-stage (3-level) trigger system to accomplish this. The data flow through the trigger system is shown in Figure 2.20. Each level provides sufficient rate reduction for the next level to have minimal deadtime. Each trigger level is more sophisticated than the previous one and requires more processing time than the previous level. Level-1 and Level-2 triggering is implemented with custom electronics while Level-3 is implemented as a PC farm. A brief description of each level is given below:

- Level-1

  At the first level of triggering only quick pattern recognition and filtering algorithms are used. Calorimeter, muon and COT tracking information are all available at this stage. There are around 60 Level-1 triggers and they may involve various combinations of leptons, tracks and energy depositions in the calorimeter. The eXtremely Fast Track (XFT) processor identifies tracks by the patterns of hits that are left on the wires in the COT after a collision. A limitation of the current XFT design is that at high luminosities, the trigger rate is dominated by fake tracks that are incorrectly identified because of the large number of overlapping low momentum tracks produced in inelastic $p\bar{p}$ collisions. Also at Level-1, calorimeter towers are combined to form trigger towers that have $\delta\eta \times \delta\phi$ of about $0.2^0 \times 15^0$. The tracks from the XFT are matched to calorimeter trigger towers and/or muon tracks in the muon chambers. Level-
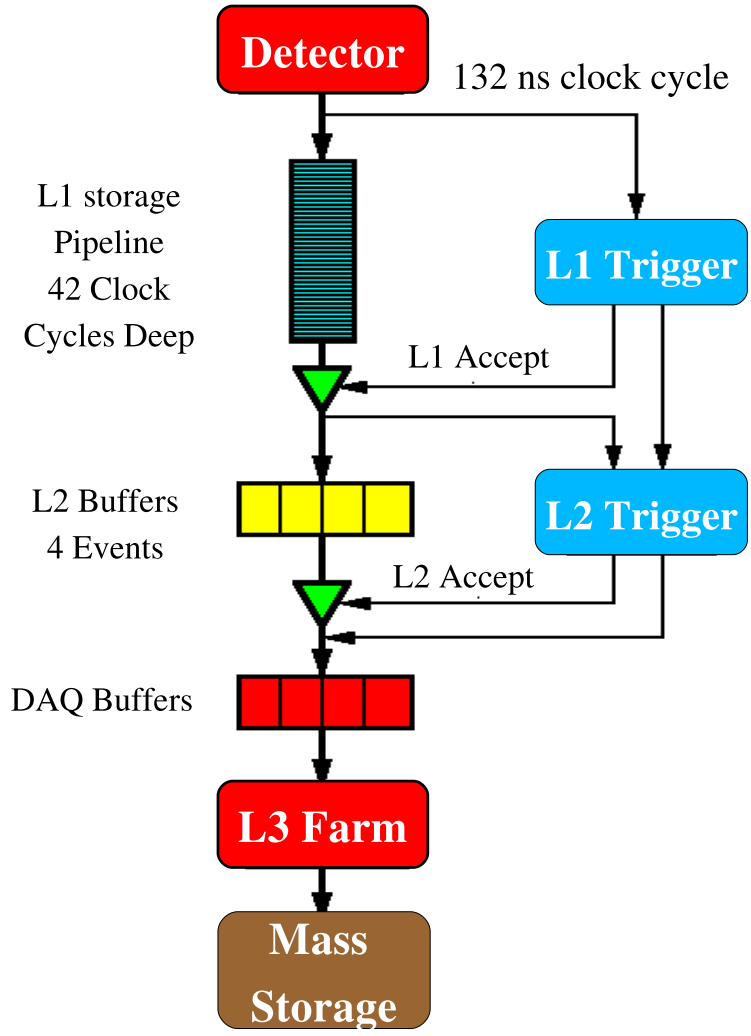
Figure 2.20. Functional block diagram of the CDF data acqusition system.

1 also uses information about the total missing energy above a certain energy threshold or the total energy deposited in the calorimeter. The number of events selected by Level-1 and sent to Level-2 (the output rate) is around 50 kHz allowing 5.5 μs for Level-1 to decide if the event should be passed to the next level.

- Level-2

When an event passes the Level-1 trigger, it is then stored in one of four available Level-2 buffers. The same trigger objects are reconstructed as in Level-1 but with more information added. A clustering alogrithm is available at this stage for the calorimeter enabling reconstruction of jets. Shower information from the CES detectors is also added that helps to separate jets from electrons and photons. The Silicon Vertex Tracker (SVT) provides information about tracks' impact parameters. The tracking information from Level-1 is combined with SVT information on impact parameters to trigger on decays of B hadrons[4]. All of this takes approximately 20 μs to make a decision and corresponds to an output rate of 1 kHz. If all four Level 2 buffers are occupied when a Level-1 accept is issued, the coming event is lost and some deadtime is introduced.

- Level-3

Level-3 is a PC farm of around 350 computers. It uses full event reconstruction software to assemble the data fragments that come from Level-2 into higher level objects such as jets, taus, muons etc. The decision of whether the event will be saved is done by a Level-3 filtering mechanism which classifies the events according to analyses purposes. The accepted events are sent to the Consumer Server Logger (CSL). The CSL writes the data to disk which is later transfered to magnetic tape for permanent storage. The triggers applied at Level-3 reduce

---

[4]A hadron that contains a bottom (b) quark.

the output rate to around 100 Hz. The output rate at each trigger level is summarized in Table 2.6.

Trigger Level	Output Rate (Hz)
Level-1	50000
Level-2	1000
Level-3	100

Table 2.6. Output rate at each trigger level.

2.2.8 Monte Carlo Event Generation and Detector Simulation

Monte Carlo (MC) event generators and detector simulations are tools that are widely used in high energy physics. This analysis makes extensive use of MC event samples to help measure the top charge and systematic uncertainties. It is with event generators that possible outcomes of high energy beam collisions can be generated. While generating the physics of $p\bar{p}$ collisions, the theoretical knowledge of QCD and electroweak physics, in addition to information like a particle's mass, lifetime and decay channels are used. For this analysis $t\bar{t}$ MC events were generated with PYTHIA v6.2 [14]. The PYTHIA MC program is based on leading order QCD matrix elements for the hard scattering process followed by coherent parton[5] shower evaluation and hadronization to simulate gluon radiation and fragmentation. In order to construct a realistic simulation, one needs to convert these partons into hadrons. This needs to be included when generating collision events. Since these processes take place at a low momentum transfer scale, for which the strong coupling is large and perturbation theory is not applicable, non-perturbative QCD needs to be applied. Perturbative QCD processes like gluon or photon radiation from initial or final state objects, called initial and final state radiation respectively, should also be included. PYTHIA accomplishes these with an interface to the CTEQ5L parton distribution functions[6]

---

[5]A constituent of a hadron (a gluon or a quark).

[6]The parton distribution functions give the probability density to find partons in a hadron with a certain longitudinal momentum fraction of the proton's momentum and momentum transfer.

[15]. It also has an interface with the decay algorithm EvtGen [16] to properly simulate heavy flavor (bottom and charm) quark decays. We also used the HERWIG MC [17] event generator for calibrating the jet charge and some systematics studies. Like PYTHIA, HERWIG employs leading order matrix elements for the hard parton scattering, followed by parton showering.

In MC, particles are generated according to simple distributions and then put through repeated random processes to describe the theoretical complications. To make the parton level MC and detector data comparable, a detailed detector simulation program is needed, on top of the MC generators, that models the passage of particles through the detector. The reconstruction algorithms, some of which will be described in the next chapter, are applied to MC generated final state particles. The simulation of particle tracks is performed with the GEANT3 program [18]. For the COT , a GARFIELD [19] simulation is used that inputs COT parameters. The calorimeter uses a parametrization of the GFLASH [20] simulation package interfaced with GEANT3. Further details of the CDF simulation can be found in reference [21].

CHAPTER 3

Event Selection

    Scientific principles and laws do not lie on the surface
  of nature. They are hidden, and must be wrested from nature
  by an active and elaborate technique of inquiry.

            John Dewey (1859-1952), an American philosopher.

The selection of top events is done by identifying the final state objects from the decay of the top quark. As explained in the first chapter, top decays to a b quark and a W boson. The W can then decay either to two quarks (e.g $W \to u\bar{d}$) or to a lepton and a neutrino (e.g $W \to e\nu_e$). This thesis focuses on the 30% of the $t\bar{t}$ events where one W decays hadronically and the other decays leptonically called the lepton+jets events. These events have a signature that is characterized by a lepton (e or μ), two b quarks, two light quarks and a neutrino. We do not use tau leptons because it is hard to distinguish taus in our detector. A candidate lepton+jets top event is shown in Figure 3.1 as observed in the CDF tracking and calorimeter system . Since top events are energetic and central (away from the beamline), we restrict ourselves to central leptons with $|\eta| < 1$ and large transverse energy, E_T. In this chapter, the lepton triggers used to select the high momentum electron and muon events will be presented. Then, the requirements imposed on the lepton+jets events after reconstruction to acquire the top sample will be described. The remainder of the

chapter is dedicated to explaining how we do jet reconstruction, jet energy corrections, the missing energy calculation and identification of bottom quarks.
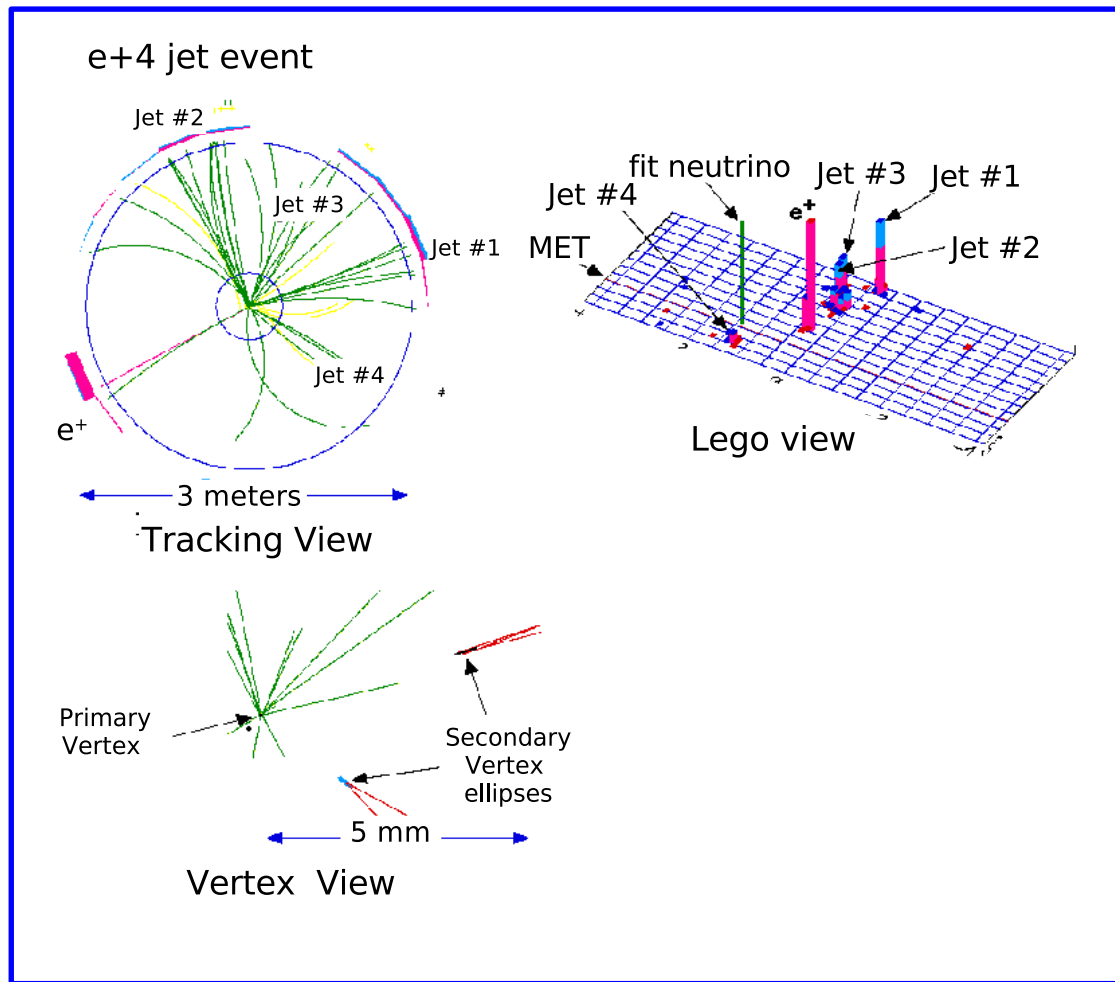
Figure 3.1. The tracks and the calorimeter towers of objects from a lepton+jets top event.

3.1 Trigger Requirements

The primary selection of the top events is performed online (during data collection) at the trigger level. Fortunately, an electron or a muon with large E_T provides a highly efficient way to trigger on the top events. There are certain topological and kinematic requirements applied at each trigger level.

3.1.1 Triggering Electrons

The main detector characteristics of a central electron candidate event is a track in the COT and energy deposition in the calorimeter, most of which is in the electromagnetic portion. Below are the electron requirements at each trigger level:

- Level-1

  Level-1 requires a CEM cluster with $E_T > 8$ GeV and an XFT track pointing to this cluster with at least 10 (or 11) COT hits in 3 (or 4) superlayers. The XFT track must have a momentum of $P_T > 8.34$ GeV/c (a detailed description of E_T and P_T for an electron can be found in Section 3.2.1). If the CEM tower energy has $E_T \geq 14$ GeV, then the ratio of hadronic energy to electromagnetic energy, $E_{\mathrm{had}}/E_{\mathrm{em}}$, is required to be less than 0.125.

- Level-2

  At Level-2 online jet clustering is performed. The clustering starts with finding trigger towers[1] with $E_T > 8$ GeV, called seed towers. Then surrounding towers, adjacent to the seed tower in η, are added and the ratio of hadronic to electromagnetic energy is required to be $E_{\mathrm{had}}/E_{\mathrm{em}} < 0.125$. The CEM cluster is required to have $E_T > 16$ GeV. The XFT track must point to the seed tower of the cluster.

- Level 3

  At Level-3 we require a CEM cluster with $E_T > 18$ GeV and $E_{\mathrm{had}}/E_{\mathrm{em}} < 0.125$ matched to a COT track with $P_T > 9$ GeV/c. The lateral energy, L_{shr}, which is a comparison of the lateral shower profile in the calorimeter cluster with that of test beam electrons, is required to be less than 0.4 and the centroid of the

---

[1]The calorimeter towers are summed into trigger towers of 15^0 in ϕ by approximately 0.2 in η. One trigger tower is generally made up of two calorimeter towers.

CES cluster must agree with the extrapolated track position to within 8 cm in z ($|\delta z| < 8$).

3.1.2 Triggering Muons

Muons are identified by matching muon stubs in the muon chambers with a reconstructed COT track and requiring that little energy be deposited in the calorimeter along the trajectory of the particle. The muon trigger requirements for each trigger level are described below.

- Level 1

  The Level-1 muon trigger requires a track segment in the CMU with $P_T > 6$ GeV/c in coincidence with hits in the CMP, called a CMUP muon, or a track segment in the CMX with $P_T > 10$ GeV/c in coincidence with scintillators placed on both sides of the chambers. The scintillator coincidence is required to occur in a narrow time window centered about the interaction time in order to reduce the rate from particles not associated with the primary interaction.

- Level 2

  At Level-2 the P_T requirement for CMUP is increased from 6 to 9 GeV/c. The muon segments are matched to COT tracks. No additional requirement is imposed at Level-2 for CMX muons, so all CMX events passing Level-1 are sent to Level-3.

- Level 3

  The P_T requirement on the COT track for both CMUP and CMX muons is increased to 18 GeV/c at Level-3. A match in a $r - \phi$ window of 10 cm between COT tracks and muon stubs in the CMU is required. This requirement is 20 cm for CMP stubs and 10 cm for CMX stubs.

3.2 Offline Reconstruction and Selection of Leptons

So far the events are distinguished only by a high E_T electron or a high P_T muon. Offline, after a full event reconstruction is performed, we require the event to have a lepton which is isolated from jet activity. Electron events are selected by requiring one isolated electron with transverse energy $E_T > 20$ GeV in the central electromagnetic calorimeter. Similarly, muons are required to be isolated with transverse momentum $P_T > 20$ GeV/c in the central region ($\eta < 1$). We restrict ourselves to events where the leptons and other top decay products are well measured in calibrated regions of the CDF detector. Next, in the offline, large amounts of missing transverse energy, $\vec{E}_T$, that indicate the presence of a neutrino is required. In this way, we acquire a sample that is enriched in W bosons decaying to a lepton and a neutrino. Then the selection requirements (cuts) are optimized further to extract $t\bar{t}$ events from the W sample by requiring three or more jets in the event. For this thesis, we use data gathered at CDF between March 2002 and September 2005 that corresponds to 695 pb^{-1} of integrated luminosity. Below are the offline lepton selections first for electrons and then for muons along with a detailed description of each selection variable.

3.2.1 Electron Reconstruction and Selection

- $E_T > 20$ GeV where $E_T = E * sin(\theta)$

  E_T is the transverse electromagnetic energy where E is the total electromagnetic energy deposited by the electron in the CEM cluster and θ is the polar angle of the COT track pointing to the seed tower of the cluster. An electromagnetic cluster is made up of a number of towers. The towers with $E_T^{\mathrm{em}} > 2$ GeV are called seed towers. The towers adjacent in pseudorapidity to the seed tower are added to the seed tower in energy to form a cluster. The clusters are ordered in a list. To prevent multiple counting of the same tower in several clusters, towers

that have been allocated to a cluster are removed from the list. The addition of towers continues until a maximum cluster size is reached. The maximum cluster is defined to have no more than three towers in pseudorapidity corresponding to 0.33 and one tower in azimuth corresponding to 5^0.

- $P_T > 10$ GeV/c

  P_T is the projection of the electron's momentum as measured by the COT track associated to the electron onto the transverse $x - y$ plane. The transverse momentum of the COT track is measured by its curvature in the magnetic field.

- $E_{\mathrm{had}}/E_{\mathrm{em}} < 0.055 + 0.00045 \times E_{\mathrm{total}}$

  The ratio of the hadronic energy (E_{had}) in the CHA to electromagnetic energy (E_{em}) in the CEM must be less than 0.055 plus a factor that is used to correct the total cluster energy (E_{total}). The additional factor compensates for the inefficiency of the $E_{\mathrm{had}}/E_{\mathrm{em}}$ selection at high energies since the high energy electrons are more likely to produce showers that leak into the hadronic calorimeter.

- $E/P < 2$

  A bremsstrahlung photon can be radiated when an electron passes through the calorimeter. This photon deposits its energy in the calorimeter while stealing from the electron's track momentum. This effect is corrected for by requiring the ratio of the electromagnetic calorimeter energy (E) to the COT track momentum (P) to be less than 2 for electrons with $E_T < 100$ GeV or $P_T < 50$ GeV/c. This requirement is removed for high P_T tracks.

- Fiduciality

  Fiducial cuts are applied to ensure that the electron candidate is away from calorimeter boundaries and the energy is well measured. The fiducial volume

for electrons covers 85% of the solid angle in the $|\eta| < 1.0$ region. A shower position from the shower max detector (CES) is used to make sure that the electron is measured in the trusted regions of the detector. In order to match a track to an electromagnetic cluster each track in the event is extrapolated, assuming helical motion in the magnetic field, to the plane of the CES. The electron track is required to be within 21 cm in the $x - \phi$ view of the shower position of the CES ($x_{CES} < 21$ cm). The track is also required not to be in a region ($z_{CES} < 9$ cm) where the two halves of the central calorimeter meet. Also if the electron track falls in the region $z_{CES} > 230$ cm, it is removed since the outer half of the last CEM tower has significant leakage into the hadronic calorimeter.

- $-3.0 < Q \times \delta x < 15$ cm; $\delta z < 3$ cm

  The CES is also used to reject possible hadron contamination by requiring the CES cluster to match with a track. The position of the CES clustes (δz) should agree to within 3.0 cm in z and must satisfy $-3.0 < Q \times \delta x < 15$ cm where Q is the charge of the track and δx is the separation in the $r - \phi$ plane between the CES cluster centroid and the extrapolated track position. The cut is asymmetric to account for photon radiation which does not bend in the magnetic field and causes a shift in the position of the CES cluster compared to what we would expect from the electron track alone.

- $\chi^2_{\text{strip}} < 10$

  The distribution of electron energy deposited on each strip of the CES detector is compared to the shower profile extracted from test beam electrons. The comparison is done in the $r - z$ view of the shower profiles using a χ^2 fit. The result from the fit is required to be less than 10.

- $z_{\text{vertex}} < 60$ cm

The distance between the intersection of the track with the beam axis in the $r - z$ plane should be within 60 cm. The interaction point at CDF is a Gaussian distribution centered around $z = 0$ and has a width of 26 cm. With the 60 cm constraint, we require the track to be within 2σ of the interaction point.

- Not a conversion

  The aim is to reject electrons that come from $e^- e^+$ conversion of photons. Photon conversions are identified by the presence of another track of opposite sign near the electron candidate. An electron candidate is flagged as a conversion if the $r - \phi$ separation between the two tracks is ≤ 2 mm and the difference in their polar angle is $|\delta cot\theta| < 0.04$. If a candidate electron is consistent with a photon conversion it is rejected. However, if a third track can be combined with the positron(e^+) to form a photon conversion, the situation is most likely due to a high energy bremsstrahlung photon emitted by the initial electron and the candidate electron is accepted to be real. The efficiency for conversion removal is $88 \pm 4\%$ meaning we reject $\approx 90\%$ of conversion electrons.

- Track quality cuts

  We require a good three dimensional COT track that points to an electromagnetic cluster. To ensure the electron track is well reconstructed, the track must contain at least 3 axial track segments and 2 stereo segments with at least 5 hits on each track segment type.

- Isolation < 0.1

  Isolation is defined as $Iso_{\text{ele}} = (E_T^{\text{cone}} - E_T^{\text{ele}})/(E_T^{\text{ele}})$ where E_T^{cone} is the calorimeter energy deposited in a cone with radius of 0.4 cm (see jet cone in Figure 3.3) around the electron cluster centroid and E_T^{ele} is the calorimeter energy of the electron cluster. Isolation should be less than 0.1 or in other words the E_T in the cone should be less than 10% of the cluster energy. This cut is imposed to

isolate electrons from extra hadronic activity and to reduce the possibility of electrons from semileptonic decays[2]. The isolation is corrected for any energy that leaks into the neighboring calorimeter wedge outside of the cone.

3.2.2 Muon Reconstruction and Selection

More stringent requirements regarding the amount of electromagnetic and hadronic energy associated with muon candidates are applied after reconstruction than when they were triggered on. Muons can be mimicked by hadrons that shower unusually late or not at all and manage to escape the detector, so some of the selection requirements below are applied to discriminate muons from charged hadrons.

- $P_T > 20$ GeV/c

  Muons have the same basic requirements as in the electron case but the COT track is required to have $P_T > 20$ GeV/c instead of the 10 GeV/c for electrons.

- Fiduciality

  As a fiducial requirement, the radius of a muon track at the point the track leaves the COT must be greater than 140 cm to guarantee that the muon passes through all 4 axial layers of the COT.

- $E_{\mathrm{em}} < \mathrm{maximum}(2.0 \text{ or } 2.0 + 0.0115 \times (P - 100))$ GeV

  High P_T muons do not deposit substantial amounts of energy in the electromagnetic calorimeter. The deposited energy must be less than the maximum of 2 GeV or $2 + 0.0115 \times (P - 100)$ GeV where P is the momentum of the muon. This requirement removes minimum ionizing particles from the muon sample.

- $E_{\mathrm{had}} < \mathrm{maximum}(6.0 \text{ or } 6.0 + 0.0280 \times (P - 100))$ GeV

---

[2]A semileptonic decay is a b or c quark decay that includes a lepton (e.g. $b \to \ell \nu_\ell X$).

To minimize hadronic background, the energy deposited in the hadronic calorimeter is required to be less than 6 GeV or $6 + 0.0280 \times (P - 100)$ GeV whichever is the maximum value.

- $E_{\text{em}} + E_{\text{had}} > 0.1$ GeV for stubless muon.

  The CMIOs (stubless muons) are required to have non-zero energy deposition in the calorimeter.

- $|\delta x|_{\text{CMU}} \leq 3$ cm, $|\delta x|_{\text{CMP}} \leq 5$ cm, $|\delta x|_{\text{CMX}} \leq 6$ cm

  The distance between the muon stub and the extrapolated COT track ($|\delta x|$) must be less than 3.0 cm in the CMU, 5.0 cm in the CMP and 6.0 cm in the CMX. The cut on $|\delta x|$ is looser for CMP and CMX because muons that reach these detectors transverse more material than CMU muons and therefore experience greater deflection due to multiple scattering.

- $z_{\text{vertex}} < 60$ cm

  As for electrons, the z position of the muon track is required to be within 60 cm from the center of the detector.

- Cosmic veto

  Since cosmic rays do not originate from $p\bar{p}$ collisions, they can be removed by requiring a small impact parameter (d_0). The impact parameter is the distance between the track and the beam axis in the $r - \phi$ plane at the closest approach as shown in Figure 3.2. If there are no silicon hits on the track, the d_0 must be ≤ 0.2 cm. If hits from the silicon vertex detector are attached to the track, the requirement on d_0 is more stringent with $d_0 < 0.02$ cm since the track resolution is greatly improved. Most cosmic rays leave back-to-back tracks in the muon chambers of the CDF detector. So the signature of a cosmic ray is a large

61
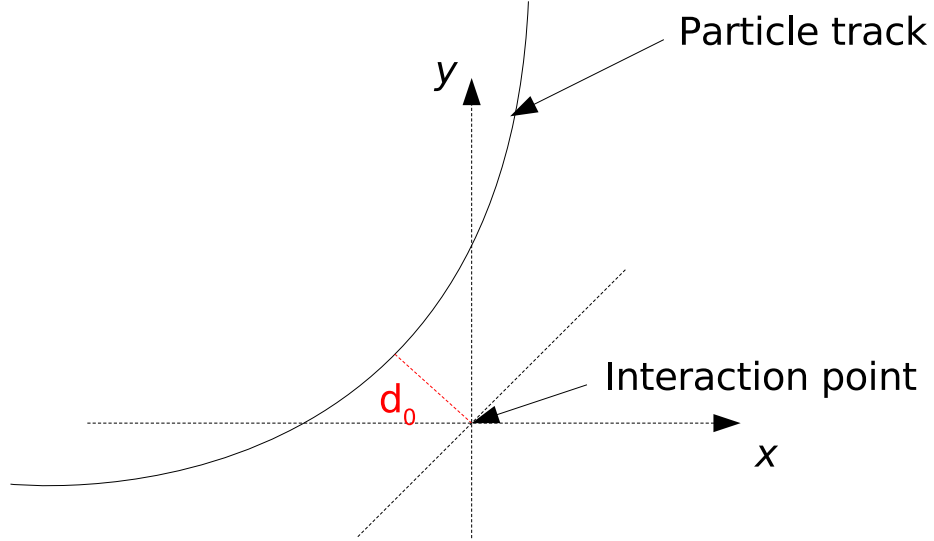
Figure 3.2. Impact parameter d_0 is the distance of closest approach of the particle track to the interaction point.

separation in ϕ between a reconstructed muon and any other muon stub in the detector.

- Track quality cuts

  A muon track is required to have at least 3 axial and 3 stereo superlayers with at least 5 hits on each superlayer (axial/stereo).

- Isolation < 0.1

  Like the electron candidates, muon candidates are also required to be isolated. Specifically the E_T (both electromagnetic and hadronic) in a cone with radius of 0.4 (see Figure 3.3) around the muon track must be less than 10% of the muon P_T. Isolation for muons is defined as $Iso_\mu = (E_T^{\rm cone} - E_T^\mu)/(P_T^\mu)$ where $E_T^{\rm cone}$ is the calorimeter energy deposited in the jet cone around the muon cluster and E_T^μ is the energy of the tower associated with the muon track. A small correction to the muon momentum (P_T^μ) is applied (curvature corrections) due to chamber misalignment.
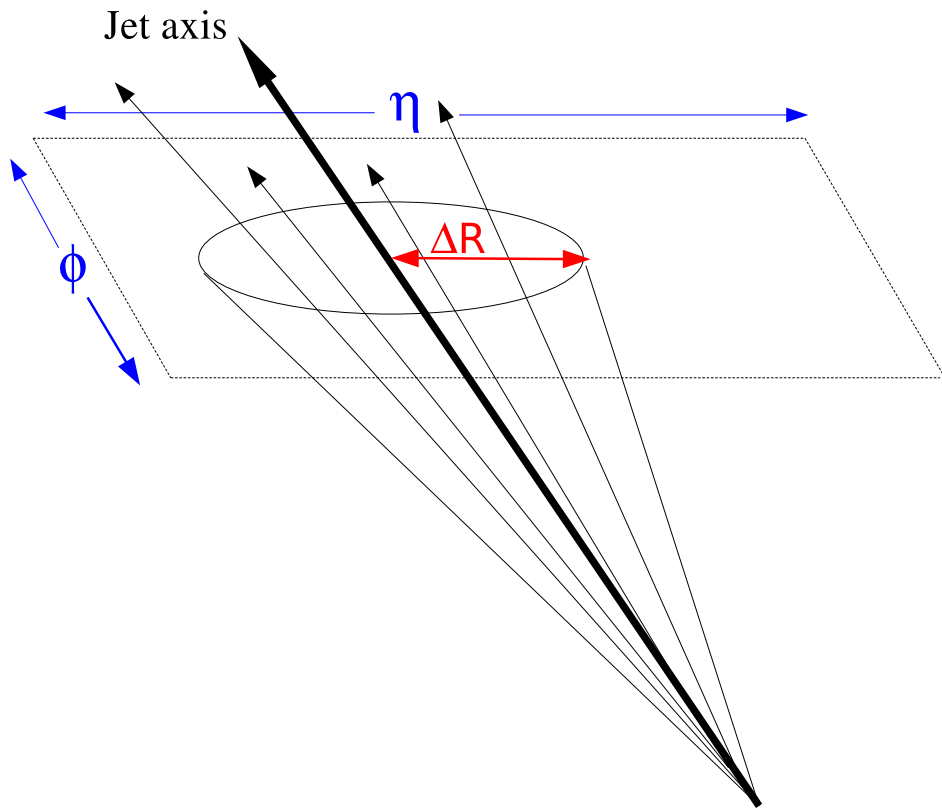
Figure 3.3. A jet cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.
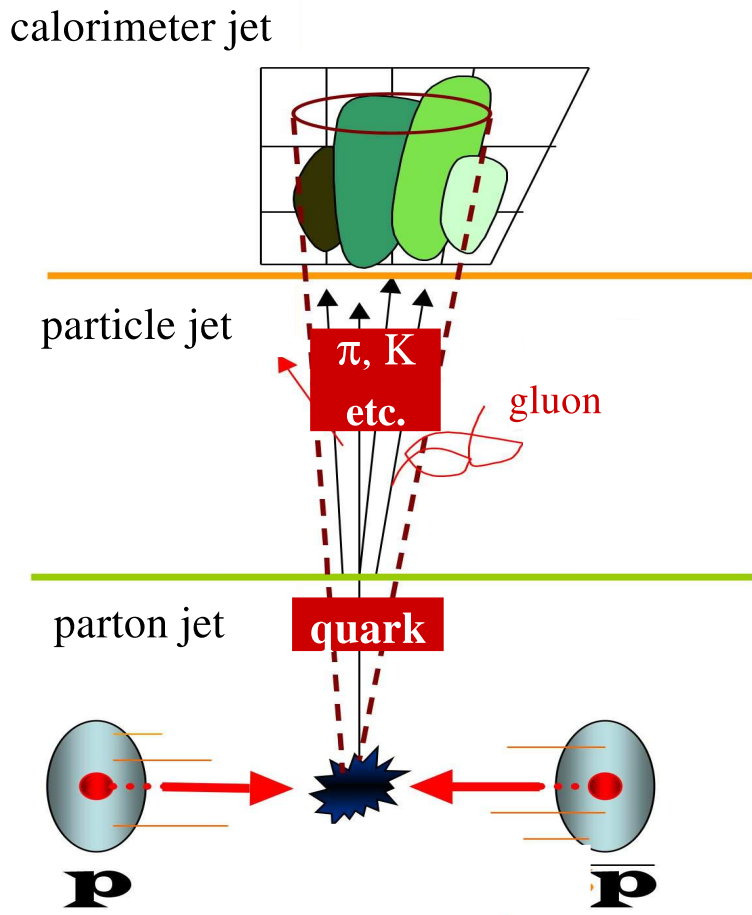
Figure 3.4. The initial parton experiences fragmentation leading to the creation of energetic particles. What we observe in the calorimeter is a shower of particles, called a "jet", that is collimated in the direction of the initial parton.

3.3 Jet Reconstruction and Corrections

A jet is formed from the scattering of a strongly interacting parton (a quark or gluon). The initial parton experiences fragmention leading to the creation of energetic colorless particles such as pions, kaons, and other hadrons that are emitted spatially collimated along the initial parton direction. See Figure 3.4. In the detector, jets are observed as clusters of energy located in adjacent calorimeter towers. The energy of the initial parton can be approximated by summing the tower energies within a cone of specified size. This procedure is called jet clustering.

3.3.1 Jet Clustering

Jets for this analysis are reconstructed using a cone algorithm called JETCLU [22]. Jets coming from top events usually have higher momenta and therefore are more collimated than QCD jets so a smaller cone size ($\Delta R = 0.4$) is used. Jet clustering starts by creating a list of calorimeter towers above a fixed E_T threshold of 1 GeV, called seed towers. Seed towers are then ordered in decreasing E_T. A cluster is formed by clumping together adjacent seed towers within a particular cone radius. If there is any tower left outside of the reconstructed cones, a new cluster is started with it. The η and ϕ of the centroid (jet axis) of each cluster is then calculated as:

$$\phi^{\text{cluster}} = \sum_{i=0}^{N_{\text{tow}}} \frac{E_{T(\text{i})}\phi_{(\text{i})}}{E_T^{\text{cluster}}} \tag{3.1}$$

$$\eta^{\text{cluster}} = \sum_{i=0}^{N_{\text{tow}}} \frac{E_{T(\text{i})}\eta_{(\text{i})}}{E_T^{\text{cluster}}} \tag{3.2}$$

where $E_{T(\text{i})}$ is the transverse energy of tower i and the sum is taken over the number of towers in the cluster. ϕ_i and η_i are the azimuth and rapidity of tower i. E_T weighted centroids are found for all clusters. Then, using the centroids as seed towers, new cones are generated as explained above. The iterative procedure is

repeated until the jet axis is stable in $\eta - \phi$ space, meaning the centroid of the energy within the cone is aligned with the geometric axis of the cone. In some cases two clusters can overlap. If the sum of the energies in shared towers exceeds 75% of the energy of the smaller cluster, the two clusters are merged into one cluster. Otherwise, towers in the overlapping region are assigned to the nearest jet in $\eta - \phi$ space. The approximate jet energy resolution is $0.1 \times E_T + 1.0$ GeV. The final product from jet clustering is our reconstructed jet whose raw energy (E_{raw}) is the scalar sum of tower energies.

$$E^{\text{raw}} = \sum_{i=0}^{N_{\text{tow}}} E_i \tag{3.3}$$

3.3.2 Jet Energy Scale (JES) Determination

Jets are reconstructed from energies measured in calorimeter towers as explained in the previous section. However this energy does not exactly correspond to the energy of the parton the jet came from. Jets may be mismeasured due to a variety of effects like loss of energy in cracks between detector components, loss of energy outside the cluster cone, contributions from the underlying event, multiple interactions and absolute energy scale. Jet energies are corrected back to the particle level and then from there to the parton level by adjusting the measured energy for all of the above affects, called the Jet Energy Scale (JES) corrections [23]. The corrected energy (E_T^{corr}) can be expressed by the following equation with each variable described below:

$$E_T^{\text{corr}} = (E_T^{\text{raw}} \times f_{\text{rel}} \times f_{\text{time}} \times f_{\text{scale}} - E_T^{\text{MI}}) \times f_{\text{abs}} - E_T^{\text{UE}} + E_T^{\text{OC}} \tag{3.4}$$

- Relative Jet Energy Scale (f_{rel})

  This JES correction is applied to remove the η dependence of the calorimeter. It accounts for non-uniformities and uninstrumented regions of the central, wall and plug calorimeters. f_{rel} is extracted from dijet studies using PYTHIA MC where one jet is central and well measured and the second jet is required to

balance the first jet, the so called "dijet-balancing" approach. Jets are corrected for any variation in the response as a function of detector η. The relative JES is flat up to $|\eta| = 0.2$ but different corrections are found for $|\eta|$ from 0.2 to 0.7.

- Raw Jet Energy Scale (f_{scale})

  The raw JES accounts for the non-linear response of the calorimeter. The CDF calorimeters respond differently to particles of various energies. The relevant correction factor (f_{scale}) is determined by using back-to-back photon+jet events where the separation in ϕ between the photon jet and the other jet is 180^0. The photon-jet sample is a good sample for studying the jet energy scale since photon energy is measured precisely by the electromagnetic calorimeter. The average P_T imbalance in a photon-jet event is around 3% but the uncertainty for the raw jet energy scale is inflated to 5% to cover the possibility of mismeasurement of the photon energy and the possibility of low energy gluon radiation that can occur in photon+jet events.

- Time Dependent (f_{time})

  The response of the calorimeter towers changes with time due to calorimeter deterioration and effects such as aging of the phototubes. The raw energy is corrected by f_{time} to take care of these effects.

- Absolute Jet Energy Scale (f_{abs})

  The absolute JES corrects calorimeter jets back to particle jets and accounts for the fragmentation effects. The calorimeter simulation derives the absolute correction over a wide range of jet P_Ts. The relevant correction factor (f_{abs}) is obtained using dijet PYTHIA events with Tune A [24, 25] parameters. Tune A refers to the values of parameters describing multiple-parton interactions and initial state radiation[3] which have been adjusted to reproduce the energy

---

[3]A massless virtual gauge boson (i.e. photon or gluon) that is radiated before a hard collision.

67

observed in the region transverse to the highest E_T jet in jet data. For the determination of f_{abs}, the two highest E_T jets in MC are matched to the calorimeter jets and the ratio between the particle and calorimeter jet P_T is calculated. The amount of the absolute JES correction as a function of jet P_T is shown in Figure 3.5.

$$f_{\text{abs}} =< \frac{P_T^{\text{particle}}}{P_T^{\text{calorimeter}}} > \tag{3.5}$$
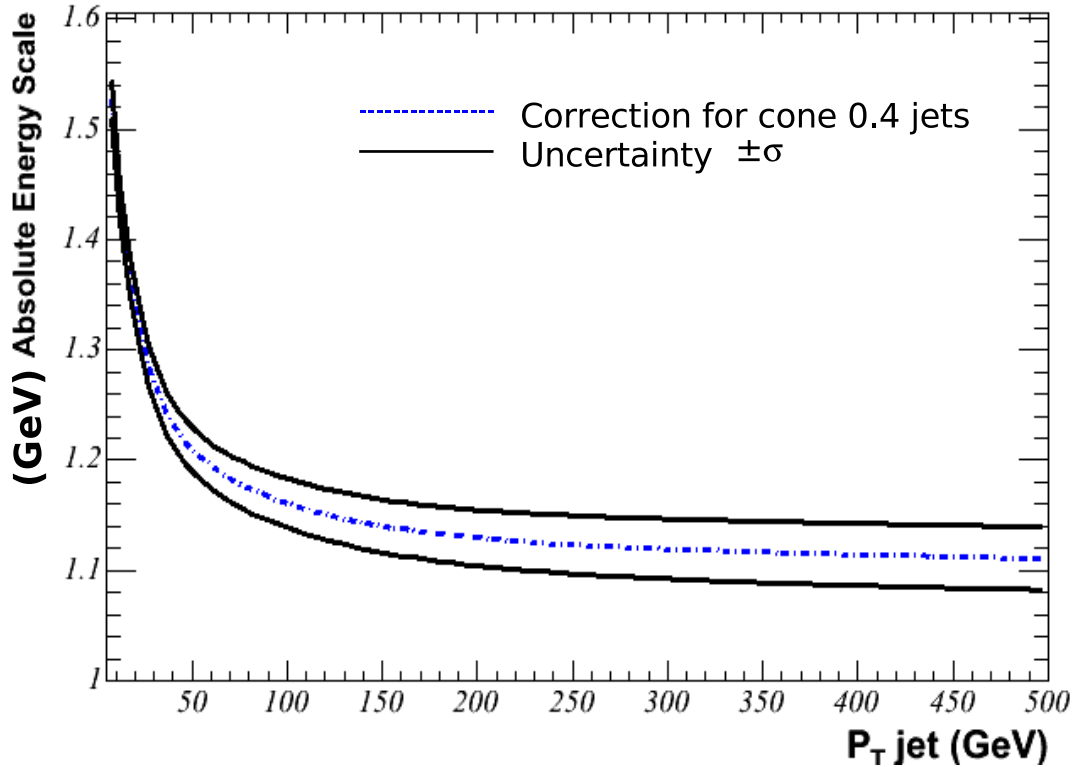
Figure 3.5. Absolute JES correction as a function of jet P_T.

- Underlying Event (E_T^{UE})

  In order to go from the particle jet energy to the parton energy, the contributions from beam-beam remnants, referred to as the underlying event, must be

subtracted from the total jet energy. See Figure 3.6. The contribution from the underlying event (E_T^{UE}) is estimated from minbias events in a similar way to the absolute JES calculation, but this time the particle jet is matched to the parent parton. For a jet cone of 0.4 the correction is around 3 GeV.
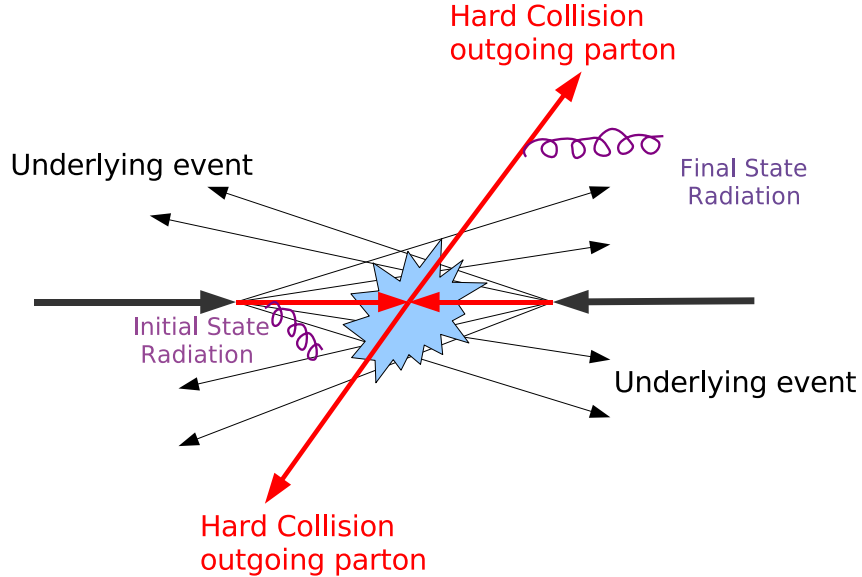
Figure 3.6. An underlying event consists of beam-beam remnants plus initial and final state radiation.

- Multiple Interactions (E_T^{MI})

  This correction takes into account multiple $p\bar{p}$ interactions that can occur in the same bunch crossing. On average one $p\bar{p}$ interaction is expected at a luminosity $L = 1 \times 10^{30}$ cm^{-2}s^{-1} with 36 bunches. However, the number of interactions goes up as the luminosity increases, i.e up to 3 interactions when $L = 2 \times 10^{32}$ cm^{-2}s^{-1}. In case there are multiple interactions, there is a contribution (E_T^{MI}) that should be subtracted from the total energy. The correction is parametrized by the number of primary vertices observed in the event and amounts to 0.36 GeV per additional vertex.

- Out-of-cone corrections (E_T^{OC})

Another correction needed to correct from particle to parton energy is out-of-cone energy. Part of the original parton's energy can be deposited outside of the jet cone due to fragmentation and final state gluon radiation[4]. Out-of-cone energy is evaluated by comparing the energy flow outside the jet cone up to a radius of 1.3. Using PYTHIA MC it was found that it is generally 15% of the total jet energy. For example a 10 GeV jet needs a 1.3 GeV (13%) correction as can be seen in Figure 3.7.
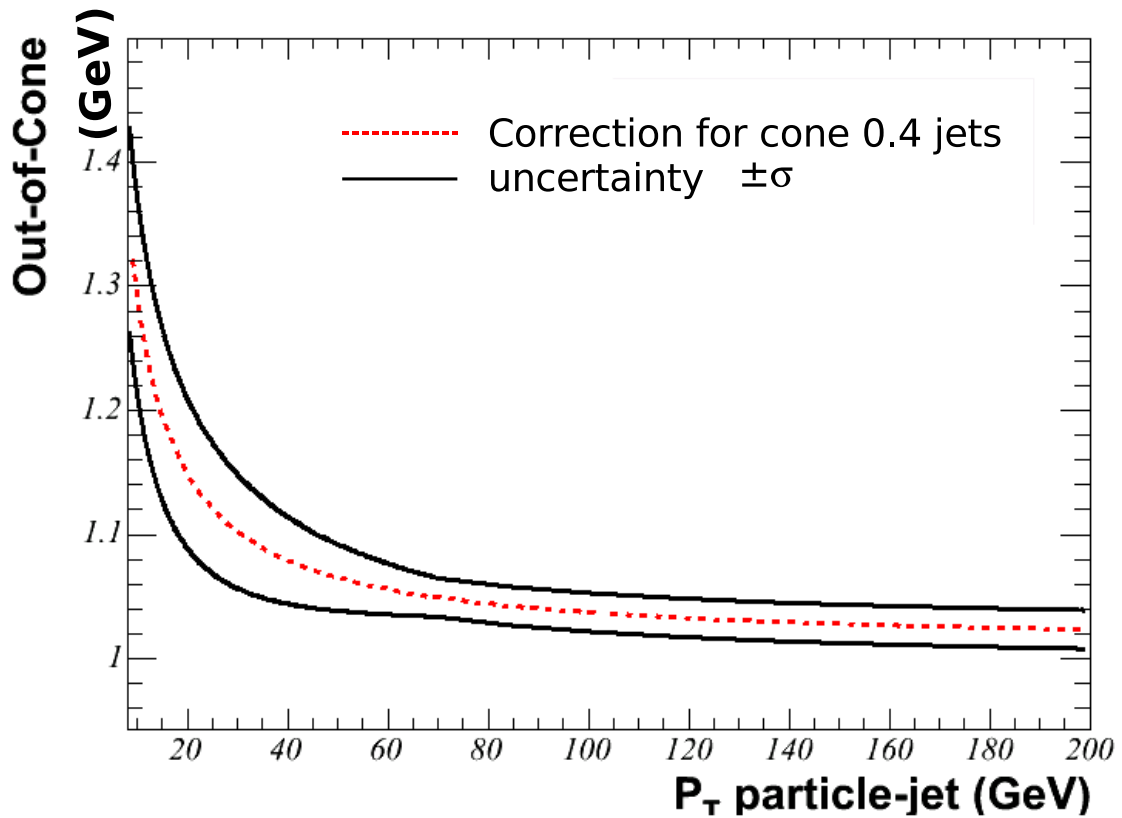
Figure 3.7. Out-of-cone JES correction as a function of jet P_T.

---

[4]Photons or gluons radiated from final state particles.

3.4 Missing Energy

Neutrinos escape the detector unnoticed since they only interact through the weak force. We can infer the existence of a neutrino by using momentum and energy conservation in the transverse plane. The raw missing transverse energy $(\vec{\not{E}}_T)$ is calculated from the negative of the vector sum of the transverse energy in all towers of the calorimeter, as shown in Equation 3.6, where $\hat{n}_i$ is the unit vector pointing from the event vertex to the i$^{\text{th}}$ calorimeter tower.

$$\vec{\not{E}}_T^{\text{raw}} = -\sum_i E_T^i \hat{n}_i \tag{3.6}$$

The $\vec{\not{E}}_T$ is corrected if there is a muon in the event since muons do deposit some energy in the calorimeter. This is done by adding the transverse momentum of the muon track $(\vec{P}_{T,\mu})$ into the above equation and subtracting the energy of the muon $(\vec{E}_{T,\mu})$ from the corresponding calorimeter towers. Similar corrections are done for minimum ionizing tracks with $P_T > 10$ GeV/c that pass loose matching requirements with a muon stub or extrapolate to regions not covered by muon detectors. The raw $\vec{\not{E}}_T$ is also corrected for the jet energy mismeasurements that were covered in the previous section. The corrected $\vec{\not{E}}_T$ can be expressed as in Equation 3.7 where $\vec{P}_{T(\text{jet})}^{\text{corr}}$ and $\vec{E}_{T(\text{jet})}^{\text{raw}}$ are the corrected momentum and raw energy of the jets respectively.

$$\vec{\not{E}}_T^{\text{corr}} = \vec{\not{E}}_T^{\text{raw}} - \left(\sum_{\text{muon}} \vec{P}_{T,\mu} - \sum_{\text{muons}} \vec{E}_{T,\mu} \right) - \left(\sum_{\text{jets}} \vec{P}_{T(\text{jet})}^{\text{corr}} - \sum_{\text{jets}} \vec{E}_{T(\text{jet})}^{\text{raw}} \right) \tag{3.7}$$

3.5 Identifying b quarks

Identifying or tagging the b quarks in top events is crucial for the event selection. The main background to top events is a W boson produced in association with a few jets. Only a few percent of the W+jets background events contain a b or c quark jet

but there are two jets originating from b quarks in every $t\bar{t}$ event. Thus tagging the b-jet can significantly reduce the amount of background.

There are three methods used at CDF to differentiate b quarks from other quarks. The first method looks for a low momentum lepton coming from the semileptonic decay of a heavy flavor hadron ($b \to \ell\nu_\ell X$ or $b \to c \to \ell\nu_\ell X$). The momenta of these additional leptons from bottom quarks in top events are smaller than (or softer), typically a few GeV/c, than leptons from W decays. Thus, this method is called the Soft Lepton Tagger (SLT)[26]. The discussion of the SLT is left for Section 4.4 which describes b flavor taggers. For the top charge analysis we will identify b-jets by using the second method, called Secondary Vertex Tagging (SecVtx), that makes use of the long lifetime of B hadrons and will be described in the following subsection. The last method is the jet probability method that considers the impact parameter of each track within the jet and constructs a probability that a given jet is consistent with coming from a zero-lifetime source. The details of this method can be found elsewhere [27].

3.5.1   Secondary Vertex Tagger

The SecVtx [28] algorithm takes advantage of the long lifetime of B hadrons. The b quark hadronizes almost immediately (on the order of 10^{-24} seconds) to form a jet of particles. Included in this jet are B mesons (such as $B^0, B^{+/-}, B_s^0$) or B baryons (e.g. Λ_B, Ξ, Σ). A B hadron from a $t\bar{t}$ event travels a macroscopic distance away from the primary interaction point before decaying into several charged and neutral particles. The distance before decaying can be up to a few millimeters.

The secondary vertex algorithm selects tracks with a large impact parameter (d_0) and reconstructs a vertex from these tracks. If the reconstructed vertex is displaced in the transverse direction from the primary vertex, it is called a secondary vertex (see Figure 3.8). The ability to distinguish tracks coming from the secondary vertex
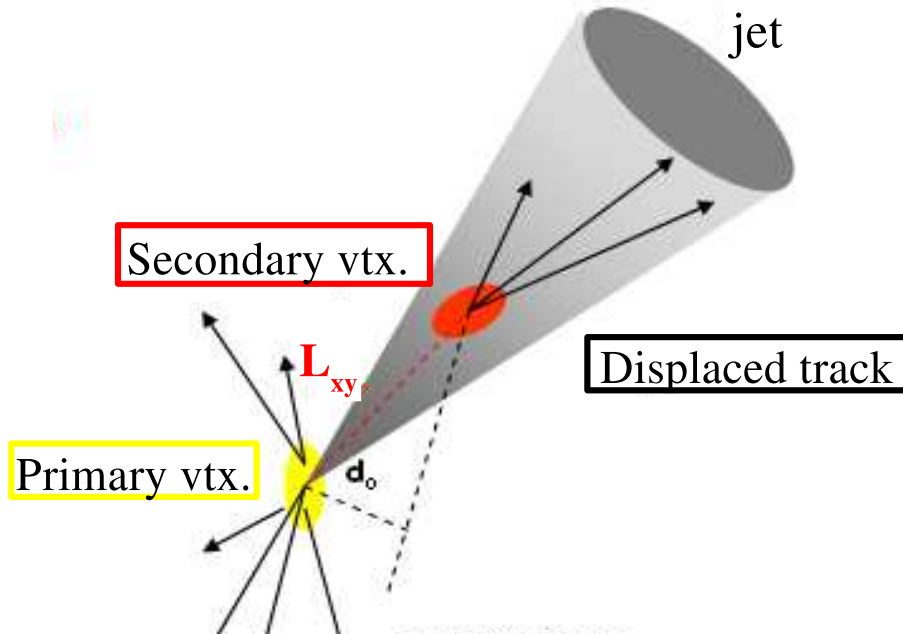
Figure 3.8. A drawing showing the primary vertex and secondary vertex with its associated displaced tracks.

apart from the primary vertex depends on a good determination of the primary vertex which will be described in the next section.

3.5.2 Finding the Primary Vertex

The primary vertex location in a given event can be found by fitting well measured tracks to a common point of origin. The first step is to choose tracks which pass a set of requirements. The tracks should have $P_T > 0.5$ GeV/c. They should be within 1 cm of the beamline ($d_0 < 1$ cm) with an impact parameter significance $S_{d_0} \equiv |d_0/\sigma_{d_0}| \leq 4$ where σ_{d_0} is the uncertainty on the impact parameter's position. For a COT track to be considered in the primary vertex algorithm, it should have at least 2 stereo or 2 axial superlayers with at least 6 hits on each of the stereo and axial superlayers. The algorithm begins with an input vertex chosen using the beamline position information. The tracks are ordered by decreasing P_T, keeping a maximum of 50 tracks. A χ^2 for each track with respect to the vertex is then calculated and

tracks with a $\chi^2 > 20$ are removed. The tracks that are left are then fit again and the removal of tracks with $\chi^2 > 20$ is repeated until a vertex with no tracks over the χ^2 cut is found. The uncertainty in the coordinates of the primary vertex is 10 - 20 μm and strongly depends on the z position of the vertex and the number of tracks. At high luminosities, more than one collision can occur resulting in multiple event vertices. For events with a high P_T lepton the vertex closest to the highest P_T lepton is chosen as the primary vertex. For events with low momentum leptons, the vertex with the highest total scalar sum of transverse momentum of the associated tracks is used.

3.5.3 Finding the Secondary Vertex

The secondary vertex (SecVtx) algorithm [28] first looks for displaced vertices with three or more tracks pointing to them. Every track in the secondary vertex must have an impact parameter significance $S_{d_0} > 2.5$ and $P_T > 0.5$ GeV/c with at least one track having $P_T > 1$ GeV/c. If this fails, the algorithm searches for two track vertices with tighter requirements on the track quality with $S_{d_0} > 3$ and $P_T > 1$ GeV/c with at least one track having $P_T > 1.5$ GeV/c. The estimated uncertainty on the track's impact parameter σ_0 is largely determined by the SVXII detector resolution which is currently around 50 μm. The tracks are then associated with a jet. As can be seen in Figure 3.8 the distance in the transverse plane between the primary and the secondary vertex is called L_{xy}. If the secondary vertex is significantly displaced from the primary vertex ($S_{L_{xy}} \equiv |L_{xy}/\sigma_{L_{xy}}| > 3$) the jet is tagged. L_{xy} is a signed distance that is positive if the vertex is on the same side as the jet and negative if it is on the opposite side. A negative L_{xy} value is indicative of background.

The SecVtx tagging method is independent of the type of B decay involved. B hadron decay channels are numerous and most of them involve neutral particles whose trajectories can not be reconstructed. Sometimes a B hadron decays to a charm

hadron which in turn decays to several particles producing a tertiary vertex making the kinematics of the decay more complex. The disadvantage of the SecVtx method is that charm hadrons can not be distinguished from B hadrons. Even though charm hadrons have a shorter lifetime than B hadrons they are tagged with a relatively high efficiency of around 10%. Thus the SecVtx tagger is actually a heavy flavor (bottom or charm quark) tagger. The efficiency of tagging b-jets using the SecVtx algorithm depends strongly on the jet kinematics. Studies were carried out at CDF to determine the b-tagging efficiency as a function of jet E_T. Figure 3.9 shows the SecVtx tag efficiency for b-jets from top decays as a function of jet E_T and $|\eta|$. As can be seen in the figure there is a loose and tight version of the SecVtx algorithm, each with slightly different track selection requirements resulting in different tagging efficiencies and purities. The comparison of the track selection cuts is shown in Table 3.1. As a result of optimization studies which will be presented in the next chapter, we decided to use the loose SecVtx tagger.

SecVtx Tagger	Loose	Tight
χ^2 threshold for primary vtx. finding	20	10
First pass for secondary vtx. finding	track $P_T > 0.5$ GeV $S_{d_0} > 2.5$	track $P_T > 1$ GeV $S_{d_0} > 2$
Second pass for secondary vtx. finding	$S_{d_0} > 3$	$S_{d_0} > 2.5$
track $S_{L_{xy}}$	> 3	> 7.5

Table 3.1. Track selection criteria used in the Loose and Tight SecVtx algorithms.

Studies at CDF revealed that the b-tagging efficiency from Monte Carlo samples differs from CDF data as a result of higher tracking efficiency in the MC simulation [28]. The ratio of the tagging efficiency in MC over the data, averaged over jet E_T, is known as the b-tagging scale factor and is used to correct the tagging efficiency in MC. It is also possible to mistakenly tag a light quark (u, d, s) jet, which are then called mistagged jets. Mistags are caused mostly by random overlap of tracks which are displaced from the primary vertex due to tracking errors. The mistag rate is parametrized as a function of jet variables like jet E_T, track multiplicity, η, ϕ and is
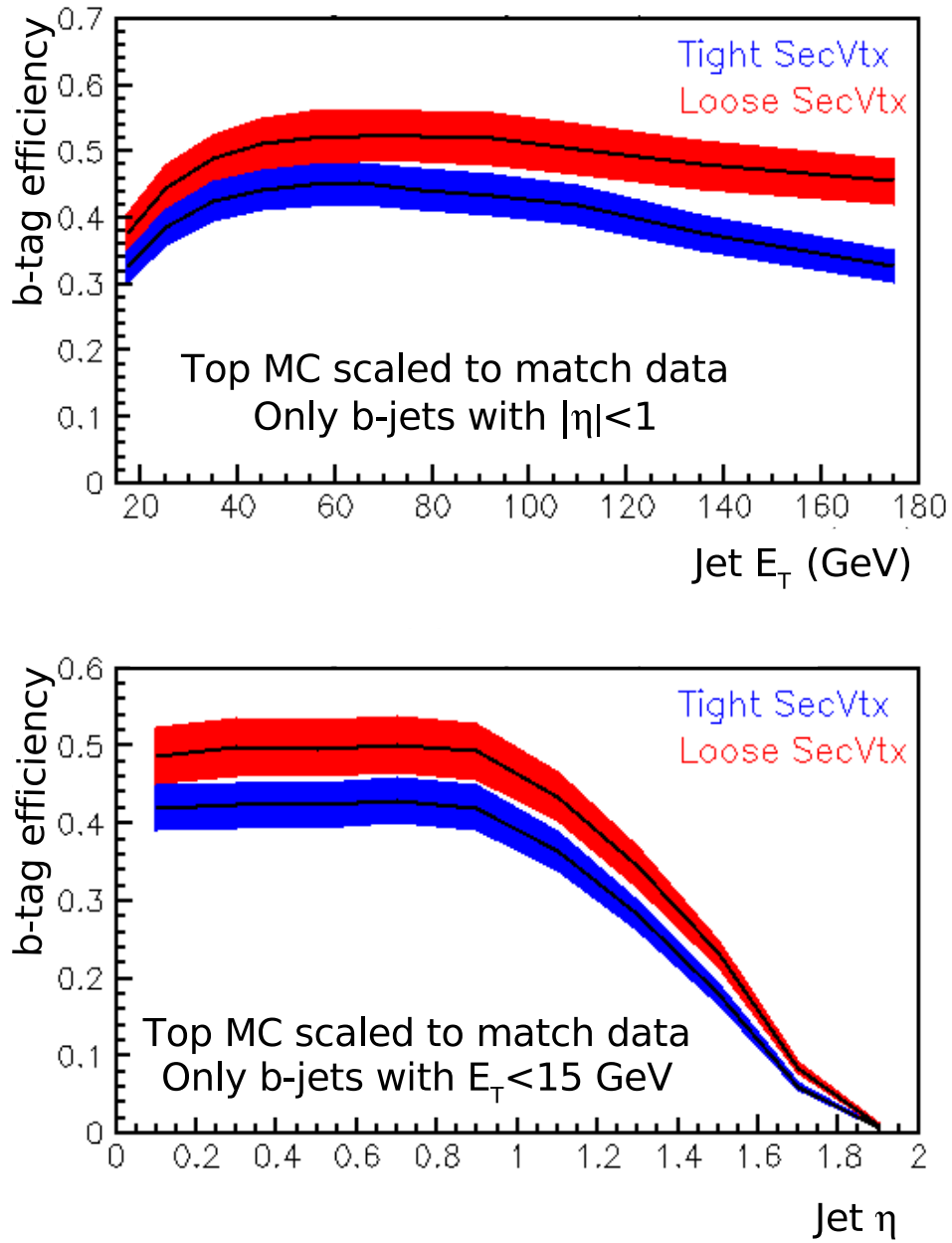
Figure 3.9. The *b*-tagging efficiency using the Tight/Loose SecVtx algorithm for *b*-jets from top decays as a function of jet E_T (left) and $|\eta|$ (right).
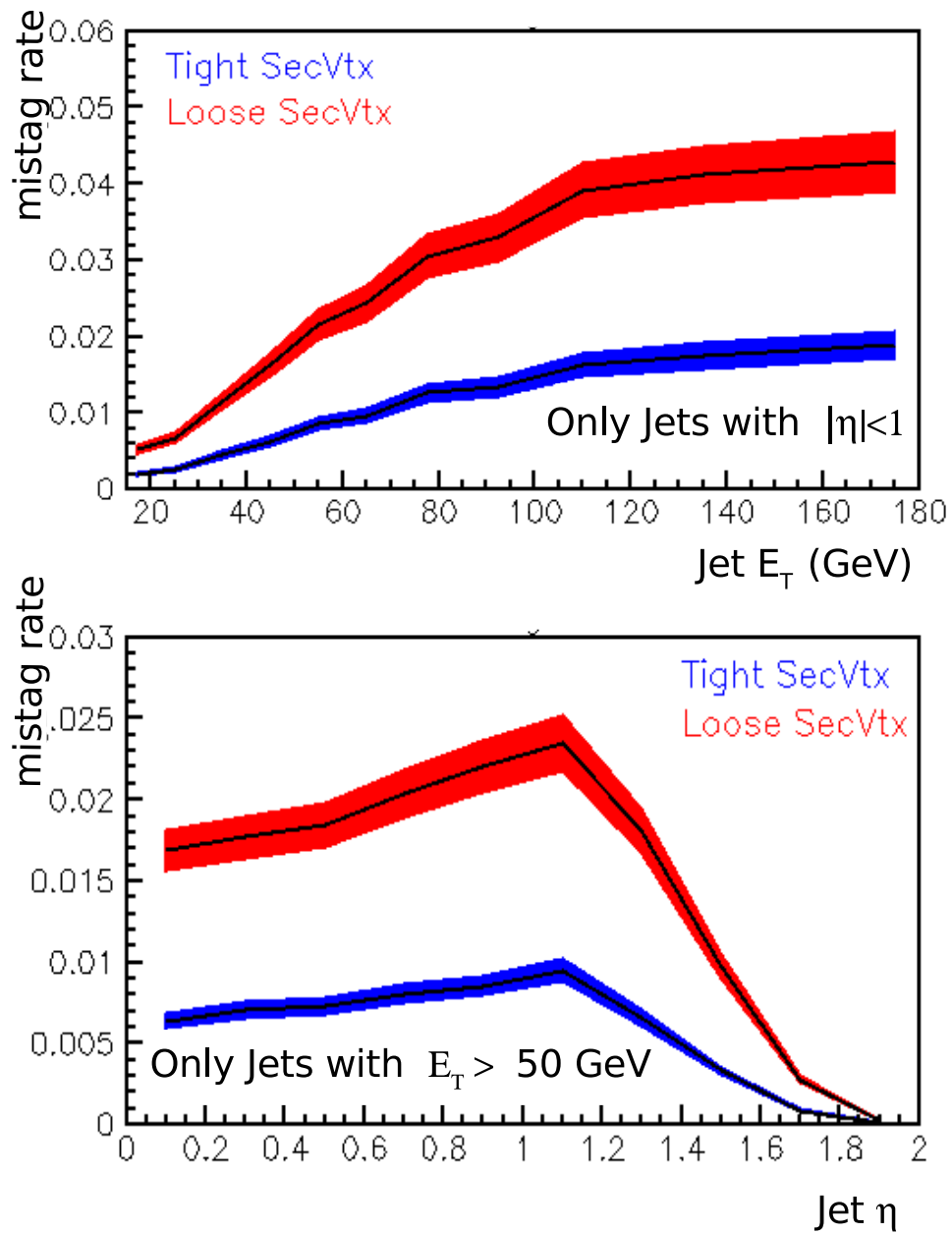
Figure 3.10. Mistag rate in data as a function of jet E_T (left) and $|\eta|$ (right).

actually a matrix [29]. Due to imperfections in the tracking simulation, the mistag rate in MC is generally smaller than that observed in data. This affects the number of events where the assigned b jets are not actually from b quarks. Figure 3.10 shows the mistag rate in data as a function of jet E_T and $|\eta|$. The comparison of mistag rates as well as b-tag efficiencies between the loose and tight SecVtx tagger for a central jet at a typical E_T of 60 GeV is summarized in Table 3.2.

SecVtx Tagger	b-tag Efficiency	Mistag Rate
Loose	$\approx 52\%$	$\approx 2.5\%$
Tight	$\approx 45\%$	$\approx 1\%$

Table 3.2. The b-tag efficiency and mistag rate for a central jet with $E_T = 60$ GeV from a top event.

CHAPTER 4

Determining Top's Charge via its Decay Products

4.1 Introduction

As explained in the first chapter, the charge of the top quark can be measured directly through electromagnetic couplings using $t\bar{t}\gamma$ events [6]. However, that measurement needs more data than CDF will ever accumulate. Instead, since the top quark decays before it can hadronize its charge information is passed to its decay products and enables us to determine the top's charge from the charges of the W boson and b quark. However this is not enough because we have two top quarks (a t and $\bar{t}$) in each event, so there are two Ws and two b-jets. The remaining question becomes what is the correct association or pairing between the Ws and bs. Getting the correct association, along with the charge information of the decay products, will enable us to answer if top decays into a W^+ and a b or a W^+ and a $\bar{b}$ (See Figure 4.1). The first case (and its charge conjugate) $t \rightarrow W^+ b$ ($\bar{t} \rightarrow W^- \bar{b}$) corresponds to a top charge of $+2/3$ $(-2/3)$ as expected in the Standard Model, while the second case $t \rightarrow W^+ \bar{b}$ ($\bar{t} \rightarrow W^- b$) corresponds to a top charge of $-4/3$ $(+4/3)$. In summary we need three

ingredients to establish the charge of the top quark:

1. The charge of the W

2. The flavor of the b-jet: whether it is coming from a b or $\bar{b}$ quark

3. The correct Wb pairing
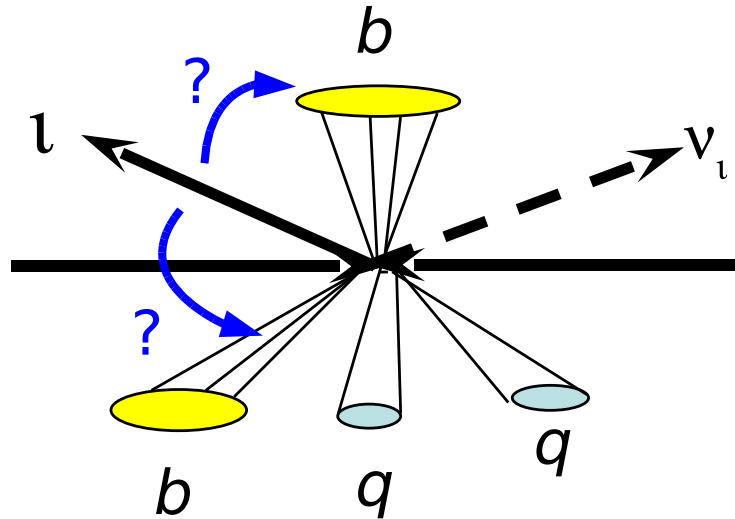
Figure 4.1. To establish the charge of the top quark, we must determine which b-jet is associated with the lepton from the W decay in the event.

The determination of the W charge is straight forward if we restrict ourselves to the leptonically decaying Ws where the W decays to a lepton and its neutrino. In the lepton+jets sample we assign the charge of the electron or muon in the event as the charge of the leptonically decaying W. We then assume that the charge of the other W (the hadronically decaying W) has the opposite charge. This chapter is dedicated to the discussion of the other two needed ingredients, namely the Wb pairing and flavor tagging the b-jets. We will present the methods used to pick the correct Wb pairing and to find the correct flavor of the b-jets and their relevant optimization studies. Before describing the methods we need to define how we quantified the performance of our methods.

4.2 Definition of Performance

In order to optimize various algorithms that are part of this analysis, we need a quantitative criteria for picking the best option. We have chosen to use a combination of two quantities: efficiency (ϵ) and purity (P). Efficiency is defined as the number of events remaining after a particular selection algorithm over the number of events available before we applied the algorithm. The purity is the number of events that are correct (based on MC information) over the number of events that the algorithm selected.

More formally, let N_+ be the number of events assigned by our analysis method to be of the $+2/3$ hypothesis (that means $\ell^+ + b$ or $\ell^- + \bar{b}$) and N_- be the number of events assigned to be of the $-4/3$ hypothesis (that means $\ell^+ + \bar{b}$ or $\ell^- + b$). Then the measured asymmetry is given by:

$$A_{\text{meas}} = \frac{N_+ - N_-}{N_+ + N_-} \qquad (4.1)$$

while the true asymmetry is given by:

$$A = \frac{N_+^t - N_-^t}{N_+^t + N_-^t} \qquad (4.2)$$

where N_+^t is the true number of events that are really $+2/3$ events by nature. Similarly N_-^t is the true number of $-4/3$ events. If we define the dilution (D) as:

$$D = \frac{N_{\text{right}} - N_{\text{wrong}}}{N_{\text{right}} + N_{\text{wrong}}} \qquad (4.3)$$

where N_{right} and N_{wrong} are the number of correctly and incorrectly assigned events, then a measurement of the true asymmetry is given by:

$$A = \frac{A_{\text{meas}}}{D} \qquad (4.4)$$

which shows why D is called the dilution. If we had no incorrectly assigned events, the dilution would be 1 and the measured asymmetry would be the same as the true

asymmetry. A dilution equal to 0 means that there is no power to distinguish between right and wrong events and a negative dilution means that we are incorrect more often than correct. It has been noted that the dilution can be considered a misnomer since we really want as high a dilution as possible (more right than wrong). The dilution is actually related to the purity (P) by $D = 2P - 1$. The dilution goes to zero when the method in question gives the wrong answer 50% of the time ($P = 50\%$).

The statistical uncertainty on the true asymmetry (A) is given by:

$$\sigma_A = \sqrt{\frac{1 - D^2 A}{\epsilon D^2 N}} \tag{4.5}$$

where N is the total number of events and ϵ is the efficiency of the assignment method. Since the uncertainty scales with $1/\sqrt{\epsilon D^2 N}$ rather than $1/\sqrt{N}$, ϵD^2 is the quantity to optimize for a given algorithm of interest. In the rest of this chapter, we present different methods tried for determining the correct Wb pairs and the correct b flavors. The decision of the optimal method is based on the comparison of ϵD^2 values and choosing the one with the largest ϵD^2.

4.3 Wb Pairing

There is a twofold ambiguity in the pairing of the W boson with the b quark in a $t\bar{t}$ event which must be resolved in order to determine the top charge. The resolution of the ambiguity relies on finding the b-jet that is associated with the leptonically decaying W, which will be called the "leptonic b-jet".

The problem of reconstructing lepton+jets events has been addressed in other top analyses, for example in the W helicity measurement [30], where various methods were explored. It was found that the best performance came from a kinematic fitter (described below in more detail) that correctly assigns the leptonic b-jet to the leptonically decaying W between 60% and 70% of the time. The 10% difference comes from different sets of requirements imposed on the fitter as will be explained in our

optimization studies for the Wb pairing. Table 4.1 shows all the tests that were performed on lepton+jets events from $t\bar{t}$ Monte Carlo to select the jet corresponding to the true leptonic b parton the largest fraction of the time. Remember that we have parton level information in the MC so "leptonic b parton" here implies the b quark associated to the leptonically decaying W.

method	Corr. b Selected
Kinematic Fitter	$60 - 70\%$
Closest b-tagged jet to the lepton	54%
Furthest b-tagged jet from the lepton	49%
Closest of two highest E_T jets to the lepton	35%
Closest jet to the lepton	30%
Randomly selected jet	25%

Table 4.1. The results of various methods tried in a previous top analysis [30] to match jets to the correct b parton. The second column represents how often each method matches a jet to the correct b parton (leptonic b-jet to the leptonic b parton).

In order to obtain the jet in the event which is most likely to be correctly matched to the b parton coming from the top quark where the W decays leptonically, the kinematic fitter constructs a χ^2 using the MINUIT [31] program. Actually the kinematic fitter calculates the top mass for each possible jet-to-parton combination in an event and its corresponding χ^2 value. For each event with 0, 1, or 2 b-tags, there are 12, 6 or 2 different permutationsin the assignment of the four highest E_T jets to the partons from the top quark decay. For this analysis, we use 2 b-tags and choose the jet assignments with the minimum χ^2 combination for the determination of the Wb pairs. The χ^2 expression is shown in Equation 4.6. The first two terms consider the difference in momenta between the fitted and measured values. Specifically the first term is the difference between the transverse momentum of the 4 jets and lepton. The second is between the fitted and measured momentum of the unclustered energy.

Each difference is divided by the uncertainty in the measured value. The last four terms are the mass differences between the W and its decay products and between the top and its decay products. Each of these terms is divided by the related decay width of the W boson or the top quark.

$$\chi^2 = \Sigma_{i=\ell,4\text{jets}} \frac{(\hat{P}_T^i - P_T^i)^2}{\sigma_i{}^2} + \Sigma_{j=\text{x,y}} \frac{(\hat{P}_j^{UE} - P_j^{UE})^2}{\sigma_j{}^2} + \frac{(M_{jj} - M_W)^2}{\Gamma_W^2} +$$
$$\frac{(M_{\ell\nu} - M_W)^2}{\Gamma_W^2} + \frac{(M_{bjj} - M_t)^2}{\Gamma_t^2} + \frac{(M_{b\ell\nu} - M_t)^2}{\Gamma_t^2}$$

(4.6)

In this fit, the transverse energy of the neutrino is defined as the negative sum of the lepton, jet and unclustered energies. The kinematic fitter assigns 5.0 and 0.5 GeV for the masses of the b partons and the light partons respectively, with the charm quark being treated as a light quark. The mass of the W (M_W) is constrained to be 80.41 GeV and its width (Γ_W) is taken as 2.12 GeV. Regarding the top mass, there are two different modes that the kinematic fitter may use. The constrained mode where the two top quark masses are both constrained to a certain value with a chosen width or the free mode where the top mass is free to float.

4.3.1   Kinematic Fitter Optimization Studies for Wb paring

As mentioned above, the kinematic fitter gave the best performance for the determination of the correct Wb pairing in lepton+jets $t\bar{t}$ events. We tried various ways to improve the performance of this fitter. As explained in Section 4.2 the performance is driven by ϵD^2 where here ϵ is the efficiency of the kinematic fitter (or $\epsilon_{\text{pairing}}$) and the purity ($P_{\text{pairing}} = (D+1)/2$) is how often the fitter finds the correct Wb pairs.

We started with investigating if a χ^2 cut on the best fit jet-to-parton assignment

improves the ϵD^2. Figure 4.2 shows the ϵD^2 values obtained for different χ^2 cuts where we kept the events with χ^2 less than the chosen cut. The figure shows that ϵD^2 is relatively flat between 8 and 15. We chose to use $\chi^2 < 9$ since it is also used by several other CDF top analyses [32]. With this cut, we find an efficiency of $56.2\pm0.3\%$ and a purity of 83.0 ± 0.3. The corresponding ϵD^2 is 0.25.
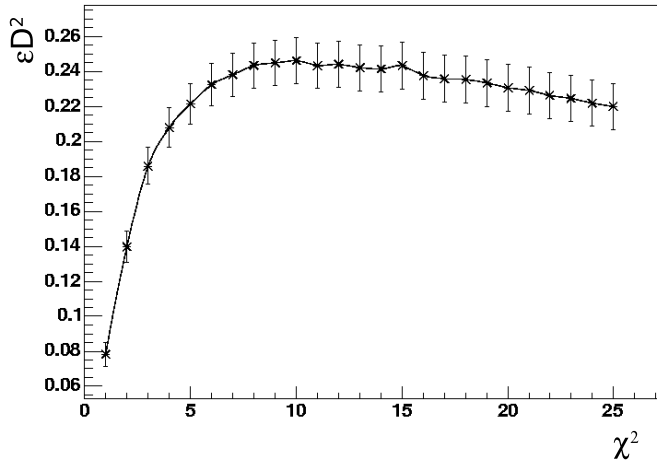
Figure 4.2. Performance of the kinematic fitter as a function of the χ^2 on double b-tagged (with the tight SecVtx algorithm) $t\bar{t}$ MC events passing the lepton+jets selection.

The χ^2 study was done using a $t\bar{t}$ MC sample where the mass of the top quark was set to 175 GeV/c^2. Remember that the kinematic fitter also places a constraint on the top mass. For the above study, we used a constraint of 175 ± 3 GeV/c^2. However, in search of improving the fitter's performance, we also tried other constraints such as 178 ± 6 GeV/c^2 or 175 ± 0 GeV/c^2. No significant change on the performance was observed by imposing different top mass constraints. We also checked the free (no constraint) mode which decreased the pairing purity by 14%. In the end, we decided to use the constrained mode of the kinematic fitter with the top mass set to 175 ± 3 GeV/c^2. As mentioned in Section 3.5, there are different versions of the SecVtx tagger that we use to tag our b-jets. To see the effect of different b-tagging requirements, we checked the pairing efficiency and purity using the loose and tight

	tight tagger	loose tagger
$\epsilon_{b-\mathrm{tag}}$ %	16	22
P_{pairing} %	83	83
$\epsilon_{\mathrm{pairing}}$ %	57	56

Table 4.2. Efficiency ($\epsilon_{\mathrm{pairing}}$) and purity ($P_{\mathrm{pairing}}$) of getting the true leptonic b using the kinematic fitter on double tagged lepton+jets events. The first column shows results using the tight SecVtx tagger and the second column shows results using the loose SecVtx tagger.

SecVtx algorithms. As shown in Table 4.2 there is a $\approx 40\%$ gain in the b-tagging efficiency using the loose version but no loss in purity so we decided to use the loose SecVtx tagger.

This analysis is using double b-tagged events (events with 2 jets tagged as bs) but we also checked the performance of the fitter on single b-tagged events where only one of the b-jets in the event is loose tagged. Although the overall efficiency (selection acceptance + b-tag efficiency) is higher for single tagged events the pairing purity decreased from 83% to 61%. An increase in the purity for single tagged events is achievable if we only accept events for which the leptonic b assigned by fitter is the tagged jet and throw the event out if it is not. Results from this study are summarized in Table 4.3 where $\epsilon_{b-\mathrm{tag}}$ and $\epsilon_{\mathrm{pairing}}$ in the table correspond to efficiencies for b-tagging and the χ^2 cut respectively. The first two rows show the results when we require exactly 4 jets in the event while the last two rows show results with 4 or more jets in the event. A "YES" in the second column indicates that the jet assigned as the leptonic b is checked for tagging and event is picked only if the jet passes the check. Therefore, the corresponding efficiencies ($\epsilon_{b-\mathrm{tag}}$) for "YES" rows are much lower. In the end we decided to use double tagged events because, even after increasing the purity of single tagged events, the overall ϵD^2 of the single tagged events was still lower than double tagged events as summarized in Table 4.4.

Apart from all of the above checks where we focused on the fraction of events for which the Wb pairing was correct, we also examined events for which the leptonic b
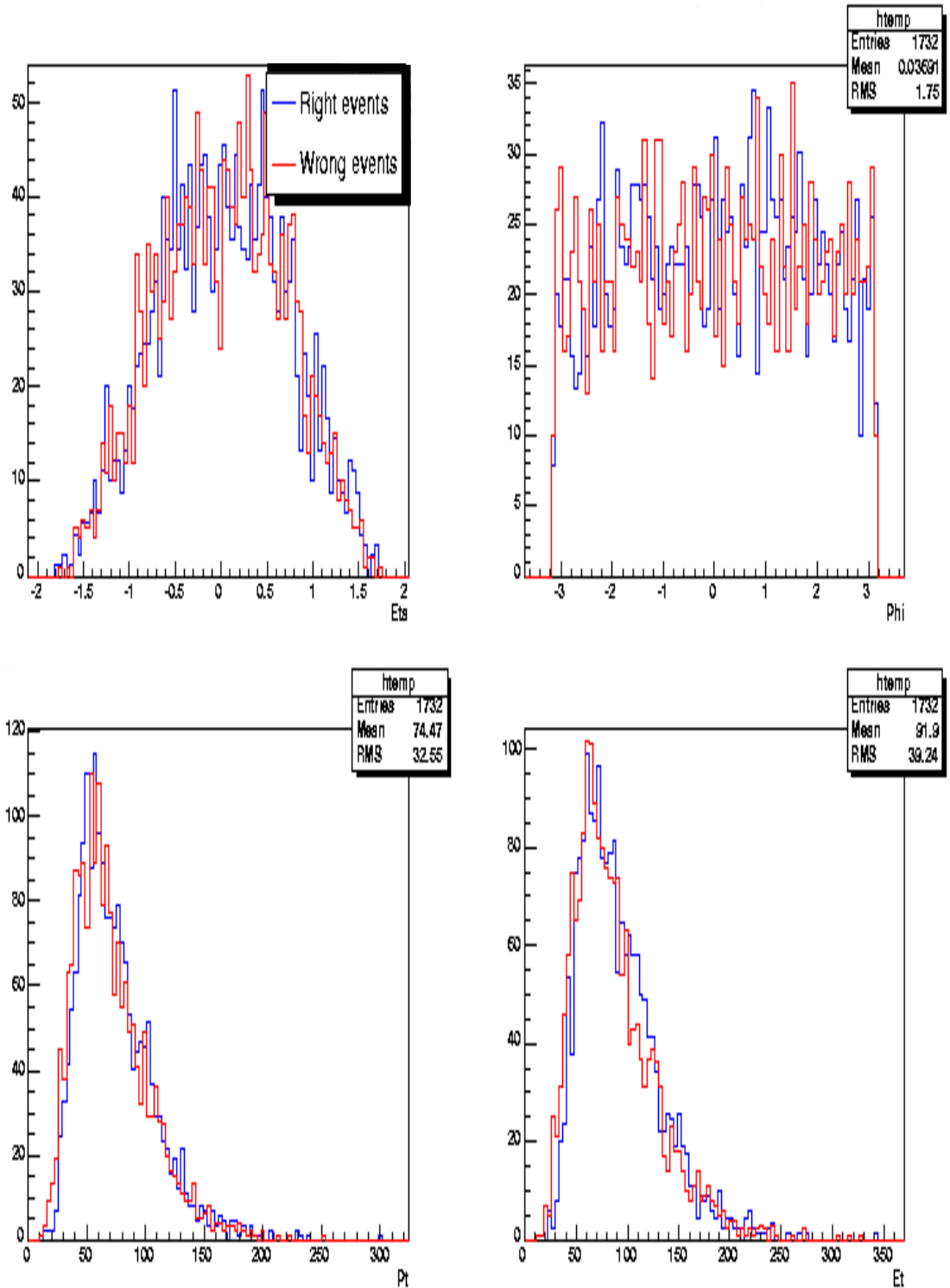
Figure 4.3. Comparison of jet kinematic variables for events where the leptonic b-jet is matched to a leptonic b parton (right events) with those events where the leptonic b-jet is not matched to a leptonic b parton (wrong events). The top two plots show the eta and phi of the jets in the event. The bottom two plots show the P_T and E_T of the jets.

# of jets	SecVtx tag check	$\epsilon_{b-\text{tag}}\%$	$\epsilon_{\text{pairing}}\%$	$P_{\text{pairing}}\%$	ϵD^2
$= 4$	N0	14.9	62.4	62.4	0.04
$= 4$	YES	4.2	17.6	77.8	0.05
$>= 4$	NO	17.5	73.3	60.7	0.03
$>= 4$	YES	4.9	20.3	76.4	0.06

Table 4.3. Performance of the kinematic fitter for single b-tagged events. The ϵD^2 in the last column is the product of $\epsilon_{\text{pairing}}$ and the square of pairing dilution D ($D = 2P_{\text{pairing}} - 1$).

	single tagged	double tagged
$P_{\text{pairing}}(\%)$	60.7	83.0
$\epsilon_{\text{pairing}}(\%)$	73.3	56.2
ϵD^2	0.03	0.23

Table 4.4. Comparison of the performance of the Wb pairing method using single tagged events with double tagged events. The tight version of the SecVtx algorithm was used for b-tagging.

assignment was incorrect, called "wrong events". For these events we checked which parton the leptonic b-jet matched to at the parton level. We expected it to be matched to the hadronic b parton which is the b parton that is associated with the hadronically decaying W most of the time. As expected we found that the leptonic b-jet matched to the hadronic b parton 75.1% of the time, one of the light quarks 14.3% of the time and did not match to any parton the remaining 10.6% of the time. In order to understand why the leptonic b actually turns out to be the hadronic b most of the time in these wrong events, we compared some kinematic variables between the two b-jets. No convenient cut was found that could separate the leptonic b-jet from the hadronic b-jet in any significant way. In the pursuit of kinematic cuts that can increase the pairing purity we also compared kinematic variables for the events where the reconstructed leptonic b matches to a leptonic b parton with those where there is no such match. No clear separation in any of the variables was observed as can be seen Figure 4.3.

As a result of our optimization studies, we decided to use events where both b-jets are tagged with the loose SecVtx tagger. The choice of the leptonic and hadronic

b-jets are based on the assignments of the kinematic fitter where we constrain the top mass to be 175 ± 3 GeV/c^2 and use a cut of $\chi^2 < 9$. The efficiency and purity found are $\epsilon_{\mathrm{pairing}} = 0.562 \pm 0.003$ and $P_{\mathrm{pairing}} = 0.830 \pm 0.003$.

4.4   Flavor Tagging the b Jet at Production

Determining whether a b-jet is from a b quark (b) or an anti-b $(\bar{b})$ quark is called "flavor tagging". The name "flavor tagging" comes from B physics because in B mixing analyses flavor tagging the B_{s} or B^0 meson at its production as well as its decay is essential. The top charge case is different in two ways:

1. The environment for top physics is different, for example the b-jets from top decay have larger P_Ts than b-jets from B hadron decay, and therefore the B physics tools may not be applicable in top's high P_T environment.

2. We are only interested in the b flavor at production.

In this section the methods used by the CDF B physics group to tag the flavor of the B meson at production will be presented. We will then discuss if they are applicable to b-jets from top decay and finally mention the method chosen for the top charge analysis. The chapter ends with a discussion of how to improve the b flavor tagging method used for top analyses.

To better understand the b-tagging methods used by the CDF B mixing group, let's quickly review their event selection and purpose. They study $b\bar{b}$ events selected by a single lepton trigger (See Figure 4.4). Both b-jets are tagged by SecVtx or some other tagging method. The event selection is such that the b-jet that includes the trigger lepton is the "signal jet" for which the flavor at production and at decay is needed. Production flavor tagger methods are used to determine the flavor of the signal B hadron at the time of its production. There are two approaches used to construct a production flavor tagger; one is a same-side flavor tagger (SST) and the
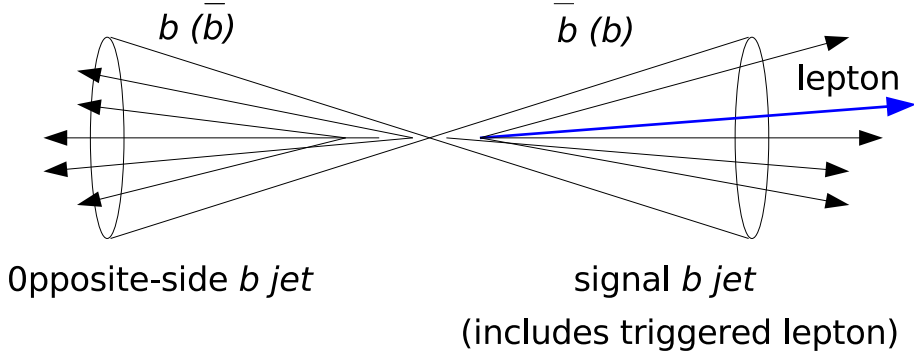
Figure 4.4. $b\bar{b}$ events selected by a single lepton trigger are used for B mixing analyses. The b-jet that includes the trigger lepton is the "signal jet" for which the flavor at production and at decay is needed.

other is an opposite-side flavor tagger(OST).

4.4.1 Same-Side Tagging (SST)

A same-side tagger is applied on the "signal jet" and uses the correlations which exist between the B meson in the jet and the charge of the first charged particle in the fragmentation chain. This correlation comes from the physics of the fragmentation process, with the creation of light quark pairs out of the vacuum and formation of new hadrons. As an example take the formation of $\bar{B}^0$ which happens when a $d\bar{d}$ pair is created and the $\bar{d}$ couples with a b quark. See Figure 4.5. If the second created quark pair is $u\bar{u}$, a negatively charged pion (π^-) is the next hadron. But if the second created quark pair is $d\bar{d}$, a π^0 is the next hadron but is neutral so does not have any tagging power. However if we keep going down the fragmentation chain, the first encountered charged hadron is again a π^-. In the case of a B^0 (Figure 4.6), the first charged particle in the fragmentation chain is a π^+. To summarize, the nearest charged π indicates the flavor of the B_0 meson (a π^- indicates a $\bar{B}^0$; a π^+ indicates a B^0). A B_{s} meson is the same as B_0 except a kaon (K) is produced instead of a pion(π). CDF uses two different SST methods which are briefly described in the following subsections.

$\overline{\mathbf{B^0}}$

b ——————— b
$\overline{d}$ $\overline{B}^0$

d

$\overline{u}$ π^-

...

b ——————— b
$\overline{d}$ $\overline{B}^0$

d

$\overline{d}$ π^0

d

$\overline{u}$ π^-

u

...
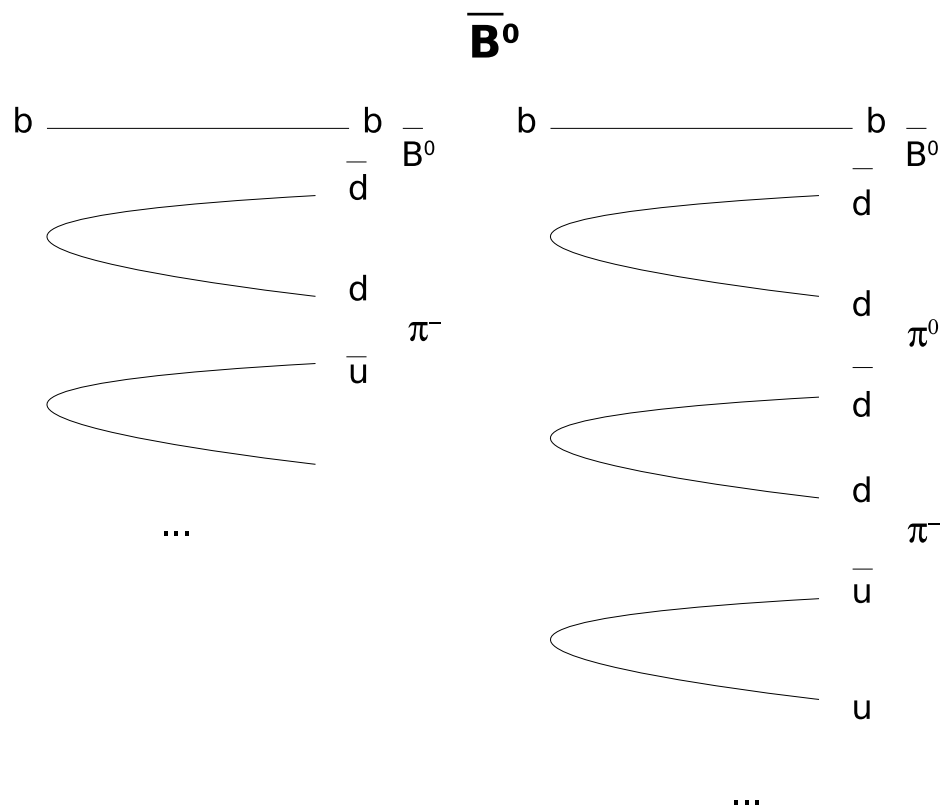
Figure 4.5. Fragmentation chain for $\bar{B}^0$.
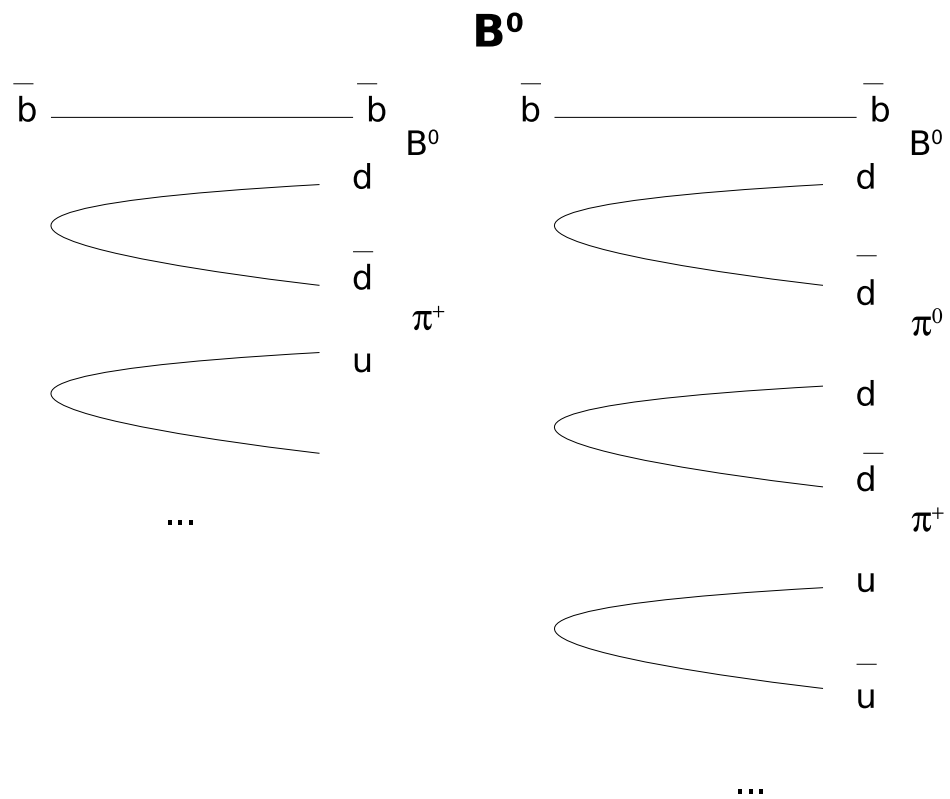
Figure 4.6. Fragmentation chain for B^0.

In SST methods one tries to find the first charged hadron in the fragmentation chain. Because the charged hadrons from the fragmentation process at the time of production have small impact parameters with respect to the primary vertex, they appear to come from the primary vertex. One wants to distinguish between the primary vertex tracks coming from the B production and the secondary vertex tracks that are coming from the B hadron decay. The minimum $P_{T,\mathrm{rel}}$ method relies on the fact that the particles produced in the fragmentation chain have small momenta transverse ($P_{T,\mathrm{rel}}$) with respect to the direction of the b quark momentum (as shown in Figure 4.7). In this method, the track with the smallest $P_{T,\mathrm{rel}}$ in a cone of radius 0.7 is chosen as the first charged particle in the fragmentation chain. Once we determine the ID of this particle, meaning identifying if it is a K or π, we can deduce the flavor of the B meson. The B physics group at CDF uses the mean energy loss rate of a charged particle in a material (dE/dx) for K-π separation.
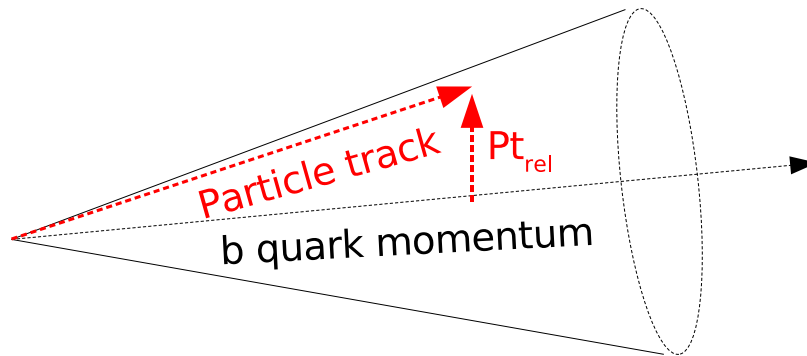
Figure 4.7. $P_{T,\mathrm{rel}}$ is the track momentum transverse to the b quark momentum.

4.4.1.2   Voting SST Method

In some cases, the minimum $P_{T,\mathrm{rel}}$ method chooses one of the daughters (decay products) of the B meson instead of the hadron from the fragmentation chain. If a track from the B meson has a small impact parameter with respect to both the primary

and the secondary vertex, the track may look like it is coming from the primary vertex. Voting SST method tries to distinguish primary tracks and secondary tracks by summing the charges of all tracks in the b-jet cone instead of using a single track. Each track's charge is multiplied by a factor that depends on the probability of it being a B daughter track.

4.4.2 Opposite Side Tagging (OST)

In order to study B hadron mixing, $b\bar{b}$ events are chosen where one of the the b-jets includes the trigger lepton and is called the "signal b". The methods described up to this point were directly applied to the signal b-jet to find its flavor at production. The second approach is to determine the flavor of the other b and infer the flavor of the signal b at production by using the fact that the two b quarks are produced with opposite flavors. The next two subsections will describe the OST methods used at CDF.

4.4.2.1 SOFT LEPTON TAGGING (SLT)

This method looks for a lepton from the semileptonic decay of the opposite side B meson ($b \rightarrow \ell\nu_\ell X$). This method is called soft lepton tagging (SLT) because the lepton from the semileptonic decay is softer (has lower P_T) than the trigger lepton on the signal b side. A b quark always decays to a negative lepton while a $\bar{b}$ quark always decays to a positive lepton ($\bar{b} \rightarrow X\ell^+$, $b \rightarrow X\ell^-$) as shown in Figure 4.8. Therefore the charge of the lepton tags the flavor of the parent b. The SLT algorithm, which was also devised for identifying b-jets from top decays, is used for finding the lepton from the semileptonic b decay. The P_T threshold for these leptons is $P_T > 2$ GeV. Similar requirements to the ones explained in Section 3.2.2 are imposed on these low P_T leptons. One main difference is that these leptons should be embedded in a jet or non-isolated. The main problem with using the soft lepton tagged b-jets is the
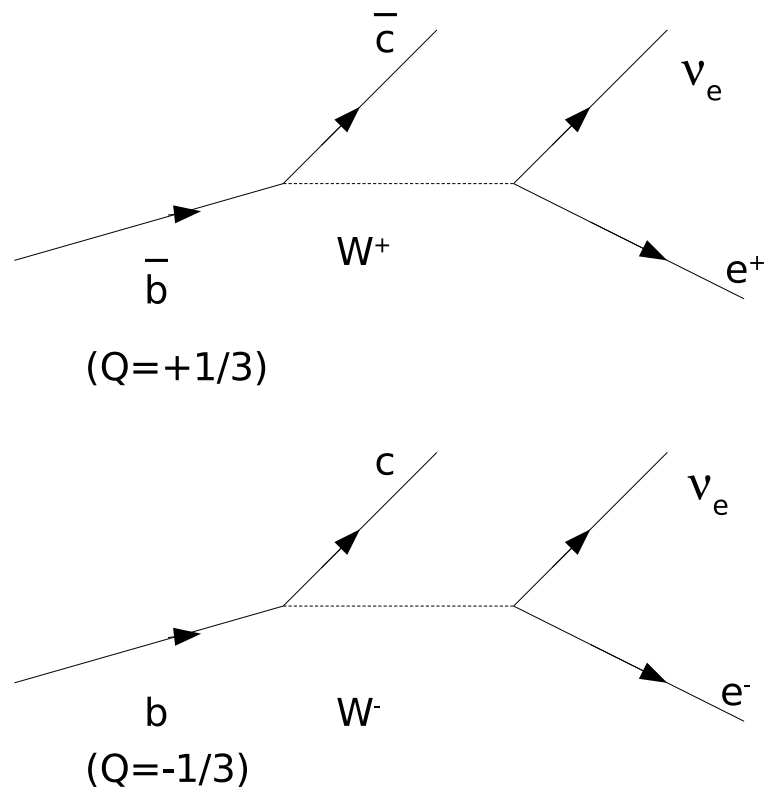
Figure 4.8. The charge of the charged lepton from a semileptonic b decay is correlated to the flavor of the b quark.

presence of $B_d\bar{B}_d$ and $B_s\bar{B}_s$ mixing. Another problem is cascade decays where the b decays to a charm which in turn decays to a lepton. ($b \to \bar{c} \to X\ell^+$ or $\bar{b} \to c \to X\ell^-$). Both the mixing and cascade decays change the sign correlation between the soft lepton and the parent b.

4.4.2.2 Momentum Weighted Jet Charge

The Jet Charge (JQ) method uses the charge information of the tracks associated with the b-jet. This idea relies on the fact that the charges of the tracks in a heavy flavor jet are correlated to the charge of the initial heavy quark the jet is coming from. The jet charge algorithm does not take a direct sum of the charges of the tracks in the jet, but instead weights each track by its momentum projected onto the jet axis as it sums over the charges of the tracks as shown in Equation 4.7.

$$JQ = \frac{\sum_{i=1}^{n} q_i (\vec{p_i}\hat{a})^x}{\sum_i^n (\vec{p_i}\hat{a})^x} \tag{4.7}$$

where q_i and p_i in the JQ calculation are the charge and momentum of track i respectively, and $\hat{a}$ is the unit vector along the jet axis. The weight factor x is used to emphasize different parts of the P_T spectrum. In the case where $x = 0$, equal weight is given to all tracks. A low x gives more weight to low momentum tracks while a high x gives more weight to high momentum tracks. The jet charge is normalized to 1 after dividing by the weighted momentum sum.

From all the flavor tagging tools presented above, we concentrated on the OST algorithms and in the end chose the weighted jet charge (JQ) algorithm as the b flavor tagger for the top charge measurement. Note that for B mixing analyses OST algorithms are applied on the opposite side jet to determine the flavor of the signal jet. In our case we have no such distinction on our b-jets. We directly apply the algorithm on the b-jet to get its flavor. The OST (SLT and momentum weighted JQ) algorithms are applicable to b-jets from top decay while the SST ($P_{T,\mathrm{rel}}$ and Voting SST) algorithms are much harder to apply. To be able to apply the SST algorithms we

need to explicitly know the type of B hadron. Even if we knew this, the particle ID of the first charged particle in the fragmentation chain is also needed. The dE/dx used for this purpose will not work for top physics due to the decrease in the resolution of dE/dx at large momenta. What can we gain from the SST studies then? We can definitely make use of the main idea of separating primary tracks that are more likely coming from the fragmentation chain than from secondary vertex tracks that are coming from the B hadron decay. Turning to the OST algorithms, let's first look at the soft lepton tagging algorithm (SLT). A drawback of using the SLT is the low branching fraction for semileptonic B decays. The SLT method is also not ideal in a high occupancy jet environment due to the difficulty of reconstructing low momentum leptons. The SLT has much lower efficiency compared to the jet charge algorithm. One other complication is that the SLT purity has to be corrected for cascade decays ($b \to \bar{c} \to X\ell^+$ or $\bar{b} \to c \to X\ell^-$) and the B flavor mixing rate. A quick study was done using SLT on top MC samples applying different P_T and vertex cuts and the purity of getting the correct b flavor is optimized up to 70%. Using the momentum weighted jet charge algorithm, we get a similar purity as in SLT but much higher efficiency, therefore a higher ϵD^2 value as will be presented in the next section.

4.5 Optimization of the Momentum Weighted Jet Charge Algorithm for the Top Charge Measurement

As mentioned above, we chose the momentum weighted jet charge (JQ) algorithm to tag the flavor of b-jets from top decays. This section presents the studies performed on lepton+jets events from top MC samples to acquire the highest purity (P_{JQ}) and more importantly, the highest ϵD^2 where ϵ is the efficiency of the jet charge algorithm (ϵ_{JQ}) and the purity ($P_{JQ} = (D+1)/2$) is how often the algorithm assigns the correct sign for a b-jet (+ for $\bar{b}$ and - for b). We declare a jet as a b-jet if the calculated jet

charge using Equation 4.7 is negative and as a $\bar{b}$-jet if it is positive.
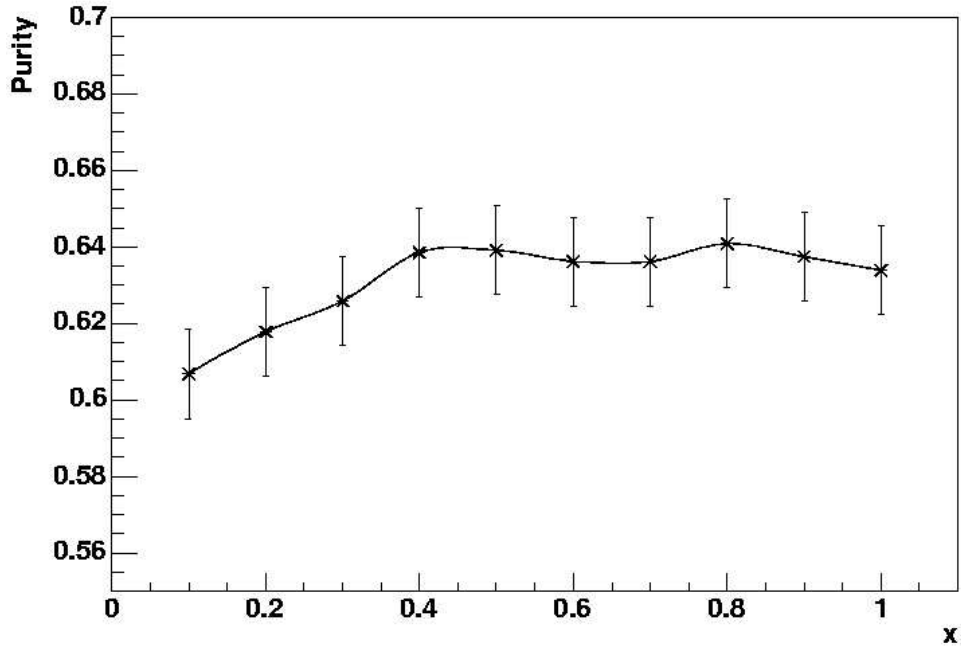
Figure 4.9. Jet charge purity obtained using different weight factors from 0.1 to 1.

The first attempt to optimize JQ was by varying the weight factor (x) in Equation 4.7. Figure 4.9 shows how the purity changes as x increases from 0.1 to 1. It is flat between 0.4 and 0.8. So we chose $x = 0.5$ and used it throughout the following studies. This choice gives equal weight to low and high momentum tracks. Other than implementing the jet charge algorithm using different weight factors, we also tried using the transverse momenta (P_T) of the tracks in the jet, instead of their total momenta (P). No significant improvement in ϵD^2 was observed. We further tried to increase ϵD^2 by imposing cuts on the quality and the number of tracks used in the jet charge algorithm. Certain requirements on the number of tracks and the tracks' momenta increased ϵD^2 but were not used due to a bias found in the jet charge algorithm between b and $\bar{b}$-jets (The JQ purity obtained using only $\bar{b}$-jets was found to be different from the purity obtained using only b-jets. Various studies are documented

in Appendix A). We tried to eliminate the charge bias by trying different P_T and impact parameter cuts on the tracks in the jet. The final selection requirements on the tracks, which reduces the charge bias, that were used for JQ are given below:

- Tracks must be in the silicon detector (SecVtx tracks)

- Jet cone size < 0.4

- Only good tracks with $|d_0| < 0.15$ cm

- Tracks with $P_T > 1.5$ GeV/c

- Number of tracks > 1

Note that we require the jet to have more than one track since the probability of a single track to carry the jet charge information is small. This requirement decreased the efficiency only by 2%. The jet charge distributions obtained from loose SecVtx b-tagged jets in top MC using the above selection requirements are shown in Figure 4.10. A jet is said to be a b-jet if the calculated JQ is negative and a $\bar{b}$-jet if it is positive.

Once we know the flavor of the b and which W boson it is associated with, we know the charge of the top quark they came from. Even though in the lepton+jets channel, there is only one lepton and thus only one side of the event where we can infer the W charge from the lepton, we can also apply the jet charge algorithm to the b-jet associated with the hadronically decaying W. Three possibilities were studied to see if the value of ϵD^2 could be improved. As the first possibility, we tried accepting events where the two b-jets (the two tagged jets) were required to have the opposite sign (OS) from each other, meaning a b and $\bar{b}$. The resulting purity was 70%, but the efficiency was reduced by $\approx 50\%$. The second option was to also accept events where both jets had the same sign. The sign assignment was decided upon by comparing the two jet charge values and assuming the jet charge algorithm gave the correct result if the leptonic b-jet charge was greater (smaller) than that of the other jet

99
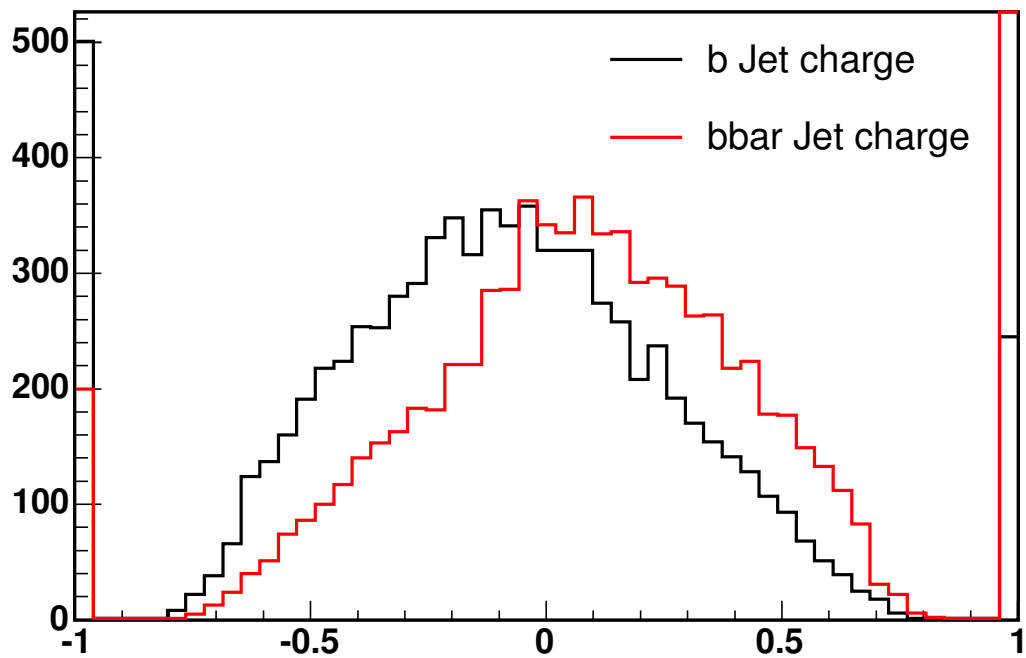
Figure 4.10. Jet charge distributions for b and $\bar{b}$ jets that are tagged using the loose SecVtx algorithm on the top MC ($M_{top} = 175$ GeV/c^2) sample. Note that entries at point +1 and -1 are two-track jets where both tracks have the same charge sign.

and was matched to a generator level $\bar{b}$ (b). We dropped this option because of its low purity of 52%. The last option was to use all the events, but each event would provide two separate top charge measurements. One corresponds to the leptonic side and the other is obtained from the hadronic side, where the assumption is that the hadronic W charge is the opposite of the one assigned to the leptonically decaying W. In this way, we can also get a measurement from the hadronic side of the event, depending on the charge of the b-jet assigned to the hadronically decaying W. The same purity is obtained for the leptonic and hadronic side (0.610 ± 0.003 for leptonic vs 0.606 ± 0.003 for hadronic) and the efficiency increases by a factor of 2.

method	$\epsilon_{JQ}\%$	$P_{JQ}\%$ (evts. passed χ^2)	ϵD^2
opposite sign (OS)	50.0 ± 0.4	70.9 ± 0.5	0.09
using pairs	98.00 ± 0.07	60.8 ± 0.3	0.06

Table 4.5. Performance of the jet charge algorithm on b-jets tagged by the loose SecVtx algorithm in the top MC sample ($M_{top} = 175$ GeV/c^2). ϵ_{JQ} and P_{JQ} are the efficiency and purity for JQ respectively. The first row is for lepton+jets events where we require both b-jets to have opposite sign. The second row is when we check the JQ performance for the leptonic and hadronic sides separately. The purity results are the average of both sides.

Among the three options presented above, we concentrated on the first and last options. Table 4.5 helps to understand why we picked the last option where we use both sides of the $t\bar{t}$ event and make two separate top charge measurements per event. By using pairs we get an $\epsilon D^2 = 0.06$. A high ϵD^2 means better performance, so one may tend to think the first option (both b-jets in the event are required to have opposite sign) that gives $\epsilon D^2 = 0.09$ is better. However to correctly compare the ϵD^2 values one must realize the 0.06 value is per Wb pair but when using both sides of the event this must be multiplied by two so we should compare $\epsilon D^2 = 0.09$ for OS to $\epsilon D^2 = 0.12$ for using pairs.

As a result of the optimization studies done with the weighted jet charge algorithm, we decided to make two separate top charge measurements per $t\bar{t}$ event applying the JQ algorithm on both the leptonic and hadronic b-jets. The weight factor in the JQ

calculation was chosen to be 0.5 and the sum is done over the tracks associated with the jet that are defined as good tracks by the tight SecVtx algorithm and have a P_T > 1.5 GeV/c. With these selections, we acquired a jet charge efficiency of 98% and purity of 61% with a corresponding ϵD^2 of 0.06.

Determining the flavor of the b-jets in a top event is the hardest and most challenging part of the top charge analysis. The purity of the method chosen for flavor tagging the b-jets is what drives the sensitivity of the analysis. To better visualize this, let's look at Figure 4.11. The figure shows the distributions for the fraction of events (f_+) that look like Standard Model assuming different purities. If the jet charge purity decreases from 0.69 to 0.60 and its uncertainty doubles, the separation between f_+ distributions for Standard Model (SM) and Exotic model (XM) diminishes considerably. There are a couple of ideas we thought of, summarized in the next section, to improve the b flavor tagging.
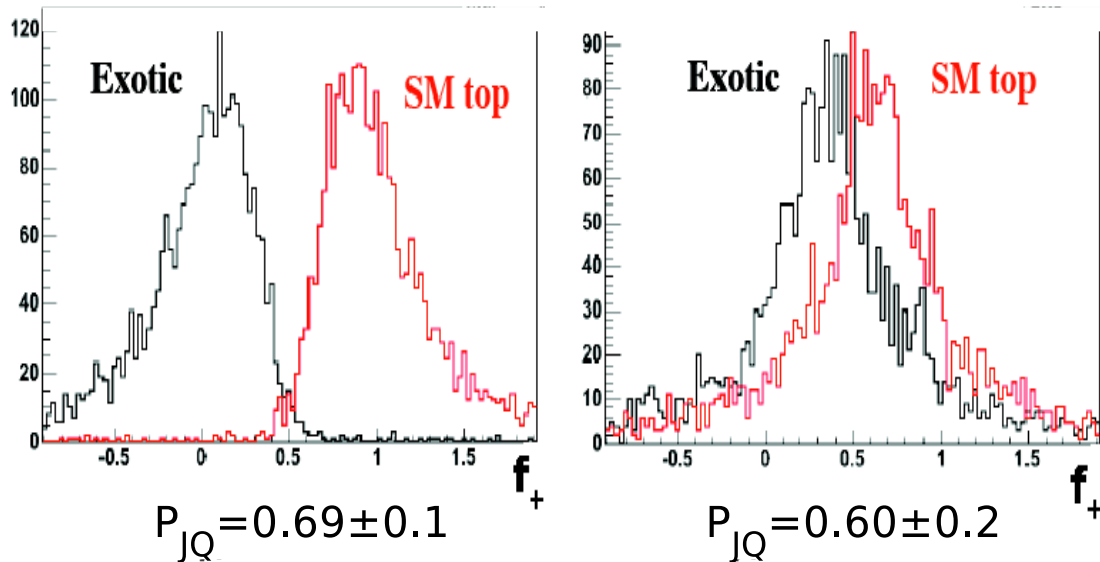
Figure 4.11. f_+ (the fraction of events that look like SM) distributions from pseudo-experiments for SM and XM generated using different jet charge purity (P_{JQ}). If SM is true f_+ should be 1, if the exotic case is true, f_+ should be 0. Refer to chapter 10 for the explanation of pseudo-experiments and how we generate these plots.

4.6 b Flavor Tagging Improvements

As shown in Figure 4.11 the top charge measurement is very sensitive to the value of the jet charge purity (P_{JQ}) and its uncertainty so knowing the P_{JQ} with precision is crucial for the top charge analysis. One promising way to improve the top charge measurement would be to improve the jet charge purity. One way to do this is to improve the momentum weighted jet charge algorithm. Remember that we use all the tracks in the jet cone that pass certain P_T and impact parameter cuts. However, only tracks from the fragmentation chain carry b flavor information. So a study, similar to the "Voting SST method" explained in 4.4.1, can be done on b-jets from the top decay to separate primary vertex tracks from secondary vertex tracks. The optimal way to flavor tag b-jets however, would be to combine all possible b flavor tagging methods like soft lepton tagging, weighted jet charge algorithm methods and Voting SST in a neural net [1].

For this thesis the strategy chosen to measure the top charge is to count events that look like the Standard Model and compare it with the number of events that look like the exotic quark model. We could use the information contained in the shape of the jet charge distributions. In other words, we can compare the data and MC JQ distributions and perform a fit. However, this approach involves some assumptions, such as believing the JQ distributions from MC represent the shape for the backgrounds accurately. A less shape dependent way to use the JQ shape information is to change our criteria for declaring a jet a b or a $\bar{b}$. Remember that we claim a jet corresponds to a b if the calculated jet charge is negative and to a $\bar{b}$ if it is positive. However, a jet with a jet charge 0.7 is more likely to be a $\bar{b}$ jet than a jet with a charge of 0.006. In other words, the jet charge purity is different in different bins. So instead of using one JQ value to decide if the jet is b or $\bar{b}$, we could divide

---

[1]A neural network is a way of combining many variables into one single variable using a model of a biological neuron [33].

the JQ distribution into several bins and calculate the jet charge purity for each bin.

CHAPTER 5

Calibration of Flavor Tagging in Data

5.1   Introduction

The biggest challenge in measuring the top charge is to tag the flavor of the b-jets
in top events. In the previous chapter we introduced the jet charge (JQ) algorithm
to identify the b-jets as b or $\bar{b}$ (flavor tag) in a high P_T environment. Using $t\bar{t}$ MC
we matched the b-jets to the b-partons and the number of b-jets for which the JQ
algorithm gave the correct charge sign over the total number of b-jets was defined as
our jet charge purity. Calculating a jet's charge (flavor) is sensitive to the details of
the fragmentation process, thus results obtained from the MC are not guaranteed to
perfectly match the results in data. Therefore, we have also studied the performance
of the jet charge algorithm directly in the data.

For the purpose of this study, we need a data sample that is enriched in b-jets. A
subset of the dijet data sample can be enriched in $b\bar{b}$ pairs by b-tagging the jets and
doing a careful event selection. In these events, one of the jets is identified by requiring
a muon, coming from the semileptonic decay of a b, to be within the jet. This jet is
referred to as the "muon-jet". Another jet that is back-to-back (180^0 away in ϕ) from
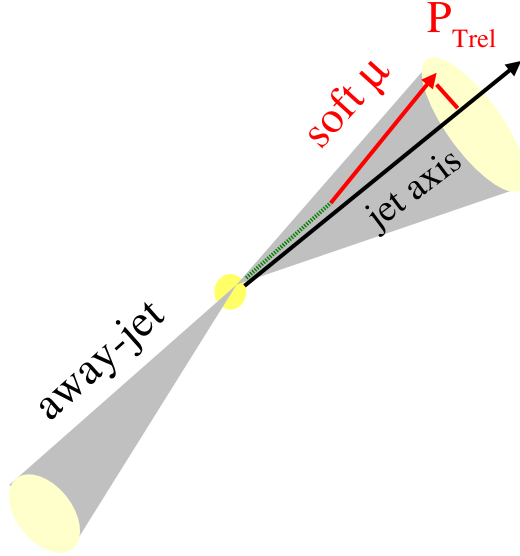
Figure 5.1. The back-to-back $b\bar{b}$ events were chosen from the dijet data sample where one of the b-jets decays semileptonically to a muon (μ). The jet charge algorithm was applied to the away-jet which is expected to have an opposite charge compared to the muon charge.

the muon-jet is required and referred to as the "away-jet" as shown in Figure 5.1. If the muon-jet and away-jet are indeed from $b\bar{b}$, they should have opposite charges. The charge of the muon-jet is taken to have the same charge as the muon. The jet charge algorithm is then applied to the away-jet which is expected to have the opposite sign compared to the charge of the muon. The observed purity ($P_{\rm obs}$) can then be determined as the number of pairs with opposite sign (OS) correlation over the total number of jet pairs.

The difficulty in this method is the determination of the fraction of events that are really from $b\bar{b}$ ($f_{b\bar{b}}$) among those events that survive the selection requirements. After selection we may still have contamination from muons that are produced by charm decays or are fakes. In addition, a light quark can be incorrectly assigned as the away-jet. In order to determine the heavy flavor content of the selected events in the dijet data we have identified two variables that are powerful in discriminating between bottom quark jets and charm or light quark jets. These variables are the transverse

momentum ($P_{T,\mathrm{rel}}$) of the muon relative to the muon-jet axis and the invariant mass of the secondary vertex (M_{vtx}) on the b-tagged away-jets. Figure 5.1 shows the $P_{T,\mathrm{rel}}$ and M_{vtx} distributions for b, c and light quark jets from dijet MC.

This section is followed by a description of the data and MC samples used for this study. Section 5.3 presents the event selection, while the method is described in detail in Section 5.4. Results are in Section 5.5. Note that the measurement is done with $b\bar{b}$ events with an average E_T of 30 GeV. However, the b-jets coming from top decays have E_Ts in the range from 60 GeV up to 150 GeV. The extrapolation of the P_{JQ} result obtained from dijet data to higher E_T jets is the largest systematic and is covered in Section 5.6. The rest of the systematic uncertainties on P_{JQ} and the conclusions can be found in Sections 5.7 and 5.8 respectively.

5.2   Data and MC Samples

For this study we have used muon-enriched data samples. In addition several dijet MC samples, which differed by the generated jet P_T threshold, and a muon-enriched MC sample were also used. All of the samples used for this study are listed in Table 5.1.

Description of samples used
Data sample
CMUP muon $P_T > 8$
Pythia dijet MC samples
with muon $P_T > 7$ GeV/c, $|\eta| < 0.8$
jet $P_T > 15$ GeV/c
jet $P_T > 18$ GeV/c
jet $P_T > 40$ GeV/c
jet $P_T > 60$ GeV/c
jet $P_T > 90$ GeV/c

Table 5.1. Data and MC samples used for the calibration of the jet charge algorithm.
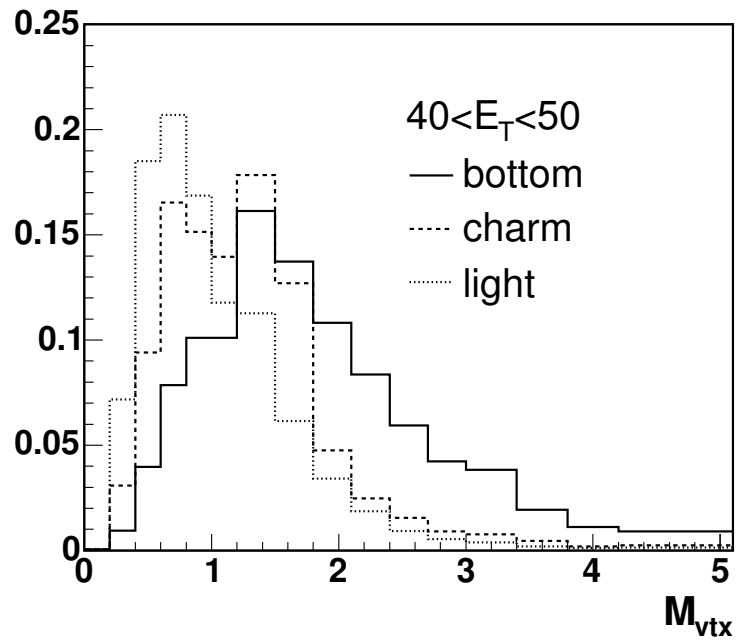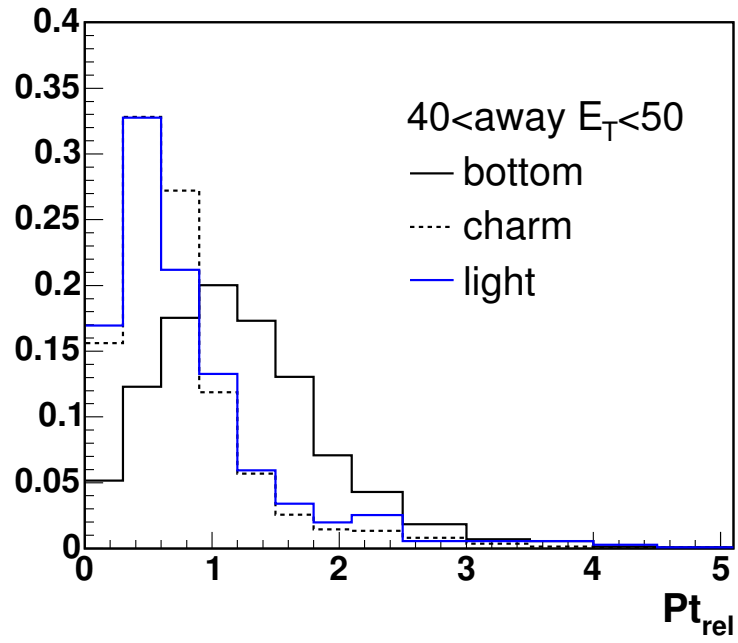
Figure 5.2. The $P_{T,\mathrm{rel}}$ and vertex mass (M_{vtx}) of b, c and light jets from dijet MC.

5.3 Event Selection

In this study the events were required to have only two high E_T jets plus a muon. The muon must be a CMUP muon matched to a jet of cone size 0.4 (the muon-jet) and the second jet (away-jet) is required to be separated in ϕ from the muon-jet by more than 2 radians. The muon-jet direction is corrected due to the shift caused by the presence of a muon. Equation 5.1 expresses the corrected momentum vector ($\vec{P}_{\text{corr}}$) where $\vec{P}_{\text{jet}}$ and $\vec{p}_\mu$ are the three momentum vector of the muon-jet and muon respectively.

$$\vec{P}_{\text{corr}} = \vec{P}_{\text{jet}} + \left(1 - \frac{2}{|\vec{p}_\mu|}\right)\vec{p}_\mu \qquad (5.1)$$

All of the requirements applied to the muon and away-jet are given below:

- Muon track $P_T > 9$ GeV/c

- Muon track $|z_0| < 60.0$ cm

- Muon CMU stub $|\delta x| < 3.0$ cm

- Muon CMP stub $|\delta x| < 3.0$ cm

- Distance to primary vertex $|z_0 - z_{\text{vertex}}| < 5$ cm

- Muon track isolation> 0.1

- Muon track must pass through every layer of the SVXII detector

- Muon jet $E_T > 9$ GeV

- Away jet $E_T > 15$ GeV

- Away jet $|\eta| < 1.5$

- Away jet must have at least two good secondary vertex tracks.

In order to enhance the b content of the sample, we require that the away-jet is tagged by the loose SecVtx tagger and the muon-jet is tagged by the tight SecVtx tagger. All of these criteria construct a dijet sample that has a high fraction of events coming from $b\bar{b}$. However, as can be seen in Table 5.2, when we study the MC we still have contamination from other types of events. In the table events are classified depending on the result of matching the jets to partons, where μ and AJ implies "muon-jet" and "away-jet" respectively. For example $\mu = b$, $AJ = c$ in the first column means the muon jet is coming from a b quark and the away-jet is coming from a c quark.

Cases		pretagged(%)	tagged(%)
bb	$\mu = b, \ AJ = b$	77	86.8
bc	$\mu = b, \ AJ = c$	2.4	2.8
bl	$\mu = b, \ AJ = \mathrm{non}b, \mathrm{non}c$	5.4	6.6
cc	$\mu = c, \ AJ = c$	8.4	1.7
cb	$\mu = c, \ AJ = b$	2.2	0.4
cl	$\mu = c, \ AJ = \mathrm{non}b, \mathrm{non}c$	3.2	0.8
fq	$\mu = \mathrm{fakes}, \ AJ = b/c/l$	1	1

Table 5.2. Classification of events in the heavy flavor enriched MC. The muon (μ) and away-jets (AJ) were matched to partons within a cone of 0.4 and classified accordingly for different cases. The fakes category includes those events where the reconstructed muon did not match a generator level muon or those where, although there is a matched muon, the jet is not from a b or c quark. The "Pretagged" column corresponds to the fraction of each case when only the away-jet is b-tagged (loose) and the "tagged" column to the case where the muon-jet is also b-tagged (tight).

Of special interest are the $c\bar{c}$ events which also have an expected opposite sign correlation between the muon charge and away-jet charge like $b\bar{b}$ events. Even though charm hadrons have shorter decay lengths, smaller impact parameters and decay products with lower P_T compared to the ones from B hadrons, they can still pass our stringent selection cuts due to their similar event topology to $b\bar{b}$ events. The performance of the jet charge algorithm is different for c-jets than b-jets. It would be beneficial to know the purity of the jet charge algorithm for c-jets (P_c). However we can not measure P_c directly from data since we do not have a method to identify c-jets in the detector. Luckily, the $c\bar{c}$ acceptance is greatly reduced by b-tagging the

muon-jet as can be seen in Table 5.2 for the cc case. There is no charge correlation between the two jets for the other cases such as cb, cq but their presence dilutes the purity. The dilution to the observed purity will be discussed in the next section.

5.4 Method

By selecting events with a pair of jets, one containing a CMUP muon and the other being located opposite to it in ϕ, we can now calculate the JQ purity. Due to the charge correlation between the muon and the away-jet in $b\bar{b}$ events we expect the event to have an opposite sign (OS) correlation, but because our JQ algorithm is not perfect some events will have the wrong assignment and give a same sign (SS) correlation. The observed purity is defined as:

$$P_{\text{obs}} = \frac{N_{OS}}{N_{OS} + N_{SS}} \tag{5.2}$$

where N_{OS} and N_{SS} correspond to the number of opposite or same sign events respectively. However, this equation needs a few corrections. First, the muon contained in the muon-jet, may not come directly from a b semileptonic decay, but from a cascade decay ($b \rightarrow c \rightarrow \mu$). This would change the charge of the muon and a SS event would be the correct flavor assignment. In addition a SS event is expected if B mixing occurs. Besides these effects the data sample is not necessarily all $b\bar{b}$ events, as already shown in Table 5.2. These "background" events weaken the purity, since no charge correlation is expected (the special case of $c\bar{c}$ was discussed in the previous section). Taking all these into consideration Equation 5.2 is modified and the observed purity, P_{obs}, is defined as:

$$P_{\text{obs}} = f_{b\bar{b}}(1 - f_{\text{casmix}})P_{JQ} + f_{b\bar{b}}f_{\text{casmix}}(1 - P_{JQ}) + f_{c\bar{c}}P_c + (1 - f_{b\bar{b}} + f_{c\bar{c}})0.5 \tag{5.3}$$

where P_{JQ} is the purity we actually want to measure, that is, the performance of the JQ algorithm on b-jets. We will refer to it as the "real purity" from now on and

this is the purity we want to compare with P_{JQ} obtained from MC in the previous chapter. Keep in mind that the observed purity is the fraction of OS events over total events (Equation 5.2). Then one might think that since $c\bar{c}$ and $b\bar{b}$ are the only OS events, P_{obs} should be equal to $f_{b\bar{b}} \cdot P_b + f_{c\bar{c}} \cdot P_c$ where $f_{b\bar{b}}$ and $f_{c\bar{c}}$ are the fraction of $b\bar{b}$ and $c\bar{c}$ events respectively and P_c is the jet charge purity for charm jets. However as we mentioned before $f_{b\bar{b}}$ has to be corrected by the fraction of events that undergo cascade decays or B mixing (f_{casmix}) that change the OS correlation to a SS correlation. So the first term in Equation 5.2 is the fraction of $b\bar{b}$ events that do not undergo cascade decays or B mixing times the purity of getting the b charge correct. In other words the purity of getting OS correlation for b-jets. The second term is the fraction of $b\bar{b}$ events that do mix or cascade decay times the purity of getting the b charge incorrect. Note that for these events if the calculated b charge is correct the two b-jets will have a SS correlation, but we want the contribution to the OS events so we must multiply the second term by the purity of getting the b charge incorrect or $1 - P_{JQ}$. The third term is the fraction of $c\bar{c}$ events times the purity of getting the c quark's charge correct, but remember that we do not have a way to determine P_c. Since the fraction of c events is small after our event selection we assume there is no correlation between the charges of the two jets and assign a purity of 0.5 for P_c. Finally the last term is the fraction of the events that do not have any charge correlation between the muon and away jet (all cases in Table 5.2 except $b\bar{b}$ and $c\bar{c}$) times the purity of getting an opposite sign correlation for these events.

5.4.1 Measuring the $b\bar{b}$ fraction

The problem now turns to measuring the $b\bar{b}$ fraction ($f_{b\bar{b}}$) in the data sample. In order to extract the b content on the muon side, a previous CDF study [34] used the $P_{T,\text{rel}}$ (diagram in Figure 5.3) distribution of the muon with respect to the jet axis to fit the data. We have chosen the same method and obtained $P_{T,\text{rel}}$ distributions for

b and c quarks from the heavy flavor enriched MC sample and for light quark and gluon jets using the dijet MC sample. The plots are shown in Figure 5.4.1 for various away-jet E_T ranges. Although some dependence on the away-jet E_T is seen for the heavy flavor, there is none for the light case.
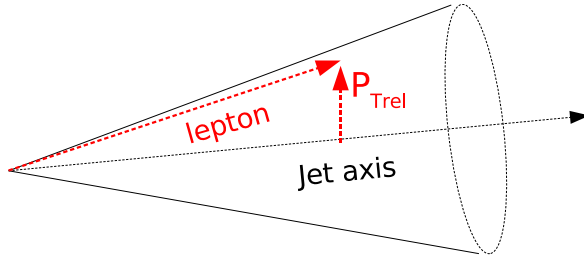
Figure 5.3. $P_{T,\mathrm{rel}}$ is the track momentum transverse to the jet axis.

For the away-jet side, we use a variable that has been shown to be a powerful discriminator of b-jets from charm and light jets in a previous CDF analysis[35]. This variable is the invariant mass of the secondary vertex (M_{vtx}) found in the jet. To calculate M_{vtx}, we start with the tracks which were assigned to the secondary vertex by SecVtx algorithm. The momentum three-vector measured by COT is turned into four-momenta by assuming that each track is a charged pion ($m_\pi = 0.1396\ \mathrm{GeV/c^2}$) to compute an energy ($E_i$) for track i. The four-momenta are then added up to get the vertex four-momentum, p_μ.

$$M_{\mathrm{vtx}}^2 = p^\mu \cdot p_\mu = (\sum_i E_i)^2 - (\sum_i \vec{p}_i)^2 \tag{5.4}$$

Distributions for b, c and light quarks were obtained from the dijet MC, where the jets have to pass the same cuts as listed in Section 5.3. In the case of light jets, only events where no heavy flavor was present were used. Figure 5.4.1 shows the M_{vtx} distributions for various away-jet E_T ranges. To produce the M_{vtx} plots we used a mixture of MC samples with different jet P_T thresholds. Instead of trying to reweight the MC to obtain the correct jet E_T spectrum or trust the heavy flavor enriched MC to reproduce it (especially at high E_T where there is a lack of statistics and the

113
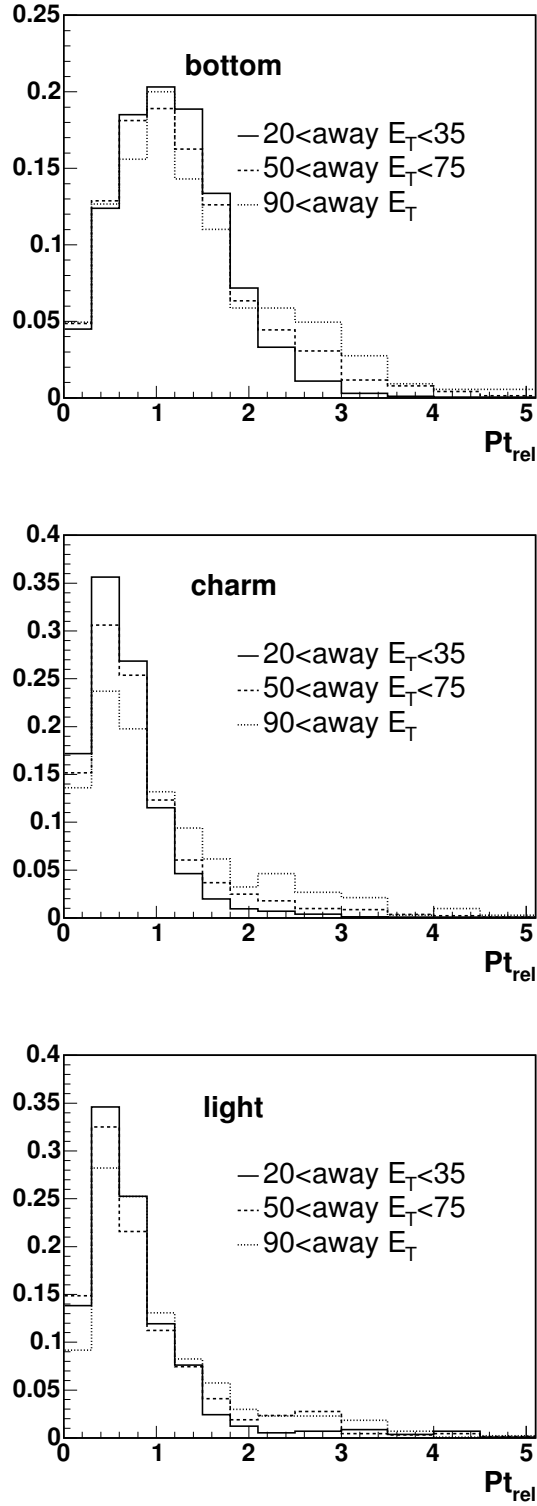
Figure 5.4. $P_{T,\mathrm{rel}}$ distributions for bottom (top plot), charm (middle plot) and light quark/gluon (bottom plot) muon-jets for various away-jet E_T ranges. The distributions were obtained from heavy flavor enriched MC for bottom and charm jets and from dijet MC for light jets.

mistag rate is larger), we decided to perform the analysis as a function of E_T with the assumption that the distributions are E_T independent in each E_T range. This is also practical, since we need to study the E_T dependence of the purity to be able to extrapolate to the high E_T b-jets coming from top decays.

5.4.1.1 FITTING PROCEDURE

The fraction of b quarks in the muon-jet (away-jet) was determined by fitting the $P_{T,\mathrm{rel}}$ (M_{vtx}) distribution from the dijet data sample to MC distributions from the different quark flavors. This fraction was measured as a function of the away-jet E_T for the various E_T bins given in Table 5.3. The distributions for b and c jets were taken from heavy flavor enriched dijet MC while generic dijet MC samples were used for the light jets. To illustrate this procedure, Figure 5.4.1.1 shows the fits, for one away-jet E_T bin, to $P_{T,\mathrm{rel}}$ for the muon-jet side and M_{vtx} for the away-jet side. In the $P_{T,\mathrm{rel}}$ case we use only b and c distributions because the $P_{T,\mathrm{rel}}$ spectrum for charm and light quarks are similar and the fitter can not distinguish between them (see Figure 5.1). For the M_{vtx} fit we use separate distributions for the b, c and light quarks. The b-fraction from the $P_{T,\mathrm{rel}}$ fit shown in Figure 5.4.1.1 is $89.5 \pm 2.7\%$ indicating that for $\approx 10\%$ of the events the muon-jet is not from a b-jet. The M_{vtx} result is $83.0 \pm 2.6\%$ indicating that for 17% of the events the away-jet is not a b-jet.

bin	E_T range (GeV)
1st	$10 \leq E_T \leq 20$
2nd	$20 \leq E_T \leq 25$
3rd	$25 \leq E_T \leq 30$
4rd	$30 \leq E_T \leq 40$
5th	$40 \leq E_T \leq 50$
6th	$50 \leq E_T \leq 60$
7th	$60 \leq E_T \leq 70$
8th	$70 \leq E_T \leq 90$
9th	$90 \leq E_T \leq 120$

Table 5.3. The nine different away-jet E_T bins used for the $P_{T,\mathrm{rel}}$ and M_{vtx} fits.
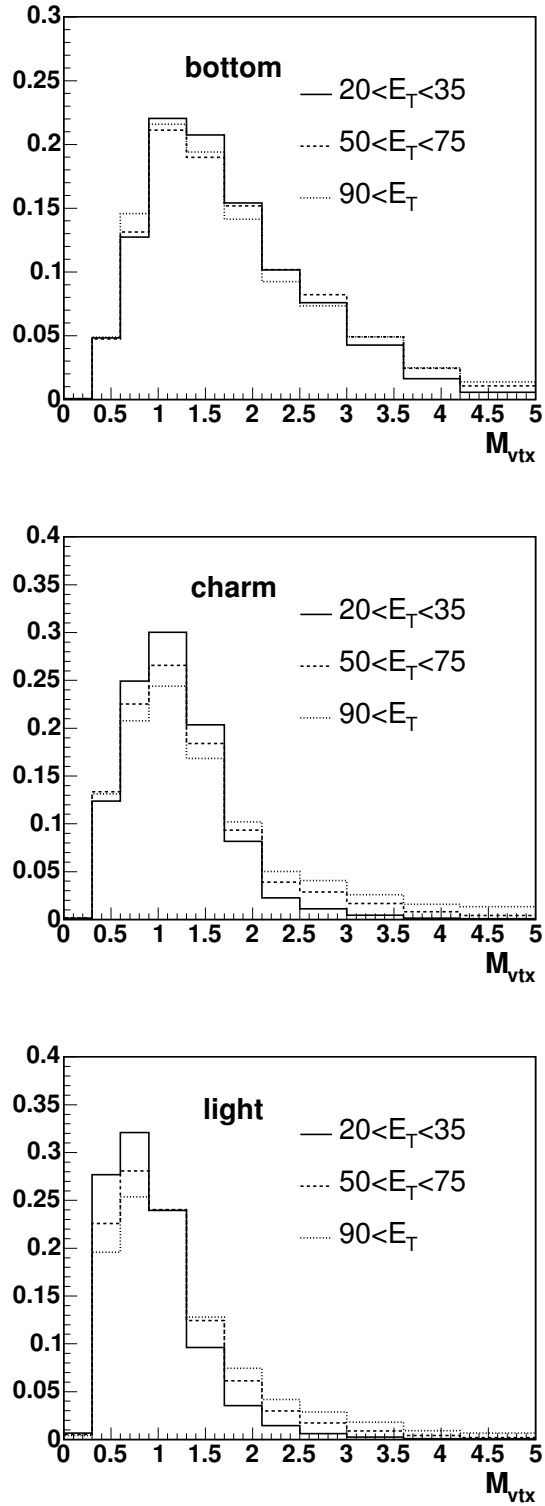
Figure 5.5. M_{vtx} distributions for bottom (top plot), charm (middle plot) and light quark/gluon (bottom plot) away-jets obtained from the dijet MC for various away-jet E_Ts.
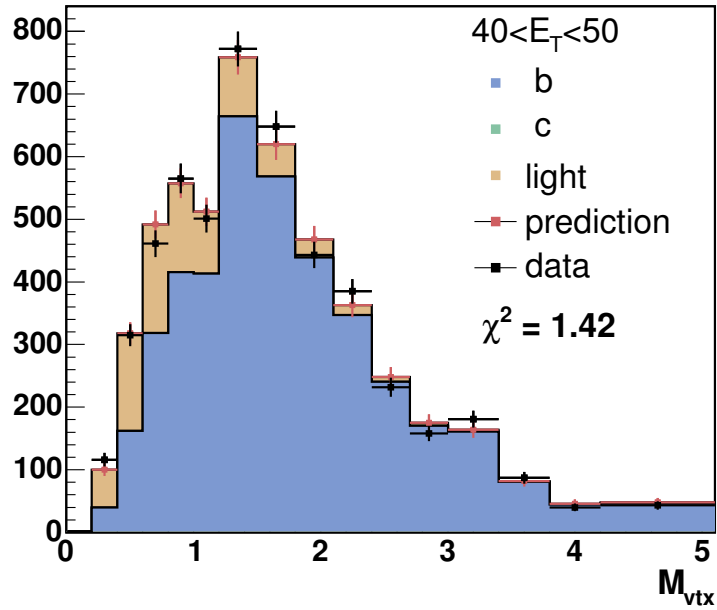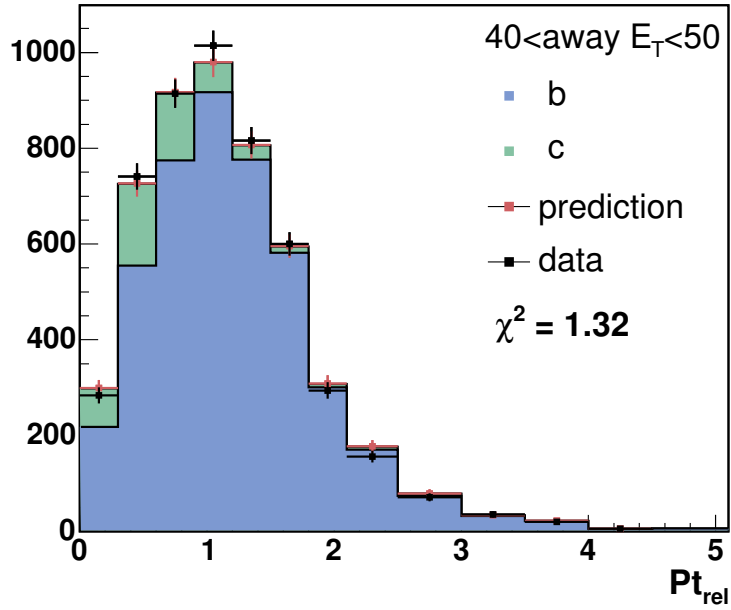
Figure 5.6. The top plot is the $P_{T,\text{rel}}$ fit obtained by fitting the data to the b and c quark distributions from MC for the events with away-jets in the E_T range of 40-50 GeV. The bottom plot is the M_{vtx} fit obtained by fitting the data to the b, c and light quark distributions from MC for away-jets with E_Ts between 40-50 GeV. The complete fit is done for nine different away-jet E_T bins.

fraction of different cases

muon jet   **away jet**

From $P_{T,rel}$ fit   From M_{vtx} fit

cc, cb, cl, fq   10.5%   Xc, Xl   17%

bb   bX: 89.5%   Xb   83%

extreme case

bc

bl
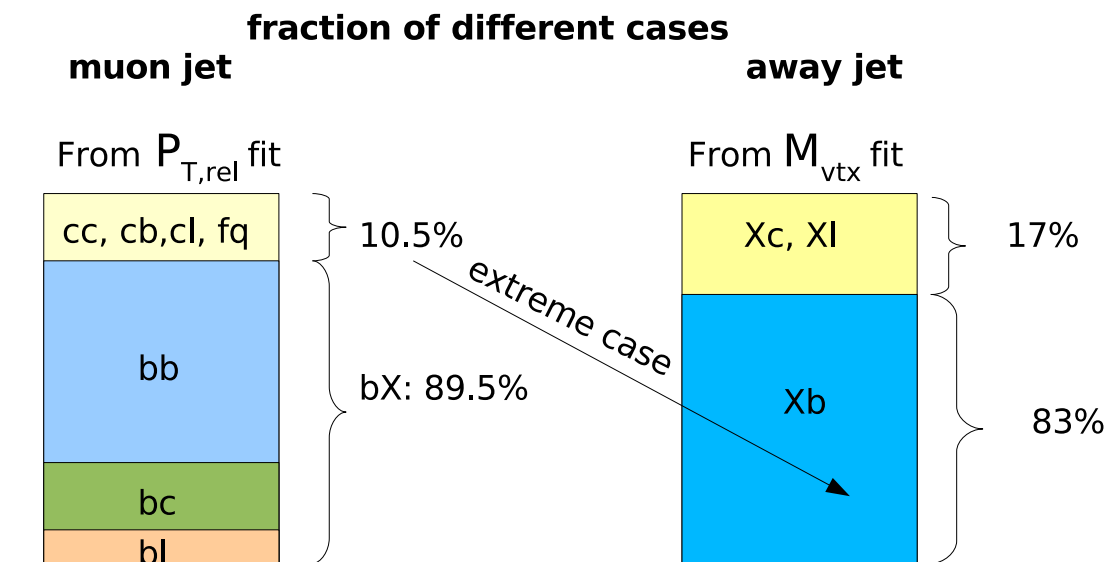
Upper limit for b$\bar{\text{b}}$ fraction = 83%
Lower limit for b$\bar{\text{b}}$ fraction = 83% -10.5%=72.5%
used b$\bar{\text{b}}$ fraction(average of two limits)= 77.8±5.9%

Figure 5.7. The fraction of different cases obtained using the $P_{T,\text{rel}}$ and M_{vtx} fits in the away-jet E_T bin of 40-50 GeV. The lower limit for the $b\bar{b}$ fraction is 72.5% if all of the non-b cases on the muon side are matched to the b-jets on the away side. The upper limit for the $b\bar{b}$ fraction is 83% if all of the muon-jets are b-jets and matched to b-jets on the away side.

In order to get the $b\bar{b}$ fraction, we need to combine the results from the fits for the b fraction in the muon-jet ($89.5 \pm 2.7\%$) with the result in the away-jet ($83.0 \pm 2.6\%$). If we assume that a muon-jet is always from a b then the maximum $b\bar{b}$ fraction is 83% becase the away-jet is a b 83% of the time. This is an upper limit since the muon-jet is not a b 10.5% of the time. The other extreme occurs when the away-jet which is a b (83%) is always matched to a non-b muon-jet (10.5%). So the lower limit on the $b\bar{b}$ fraction is $83\% - 10.5\% = 72.5\%$ (see Figure 5.7). The $b\bar{b}$ fraction is then calculated as the average of the upper and lower limits giving $77.8 \pm 5.9\%$. The uncertainty of 5.9 is calculated as the quadrature sum of 5.3, which is the difference between the average and upper/lower values for $f_{b\bar{b}}$, and the uncertainties from the fits. Note that $77.8 \pm 5.9\%$ is the $f_{b\bar{b}}$ result for the away-jet E_T bin of $40 < E_T < 50$ GeV. We repeated the same procedure for the remaining eight E_T bins listed in Table 5.3.

5.4.1.2 Mixing and Cascade corrections

As mentioned before, the muons in our data sample are not only produced from $b \rightarrow \mu$ but also from mechanisms like $b \rightarrow c \rightarrow \mu$. The latter case will change the expected muon charge and the correlation with the away-jet charge is of the same sign, instead of the opposite sign. The same effect is obtained if B mixing occurs. We have measured the amount of cascade decays and B mixing from MC samples that include both of these effects. To include all possible cases we have defined f_{casmix} used in Equation 5.3 as:

$$f_{\mathrm{casmix}} = f_{\mathrm{casOS}}(1 - f_{\mathrm{mix}}) + (1 - f_{\mathrm{casOS}})f_{\mathrm{mix}} \tag{5.5}$$

where f_{casOS} ($1 - f_{\mathrm{casOS}}$) includes only the $b \rightarrow c \rightarrow \mu$ ($b \rightarrow \bar{c} \rightarrow \mu$) case and is measured over a MC sample of non-mixing events. f_{mix} is the fraction of B mixing. The first term is for the case when the cascade decay has OS and there is no mixing. The second term is the case where the cascade decay does not result in OS but mixing occurs. Both f_{casOS} and f_{mix} were measured from MC by matching reconstructed B

hadrons to their generator level information. The measured values for f_{casOS} and f_{mix} in heavy flavor enriched dijet MC when the away-jet $E_T < 50\,\text{GeV}$ are $f_{\text{casOS}} = 0.109\pm$ 0.005 and $f_{\text{mix}} = 0.154\pm0.005$. In the high E_T region when the away-jet has $E_T > 50$ GeV, the values for f_{casOS} and f_{mix} are $f_{\text{casOS}} = 0.165\pm0.017$ and $f_{\text{mix}} = 0.147\pm0.015$. Remember these values are for the cascade and B mixing fractions in a heavy flavor enriched dijet MC sample. These fractions may change depending on how the MC models the B production rates and the semileptonic branching ratios. Therefore we also measured f_{casOS} and f_{mix} from generic dijet MC and confirmed that they are consistent with the values obtained from the heavy flavor enriched dijet MC. In the end the f_{casmix} we obtained using Equation 5.5 was 0.229 ± 0.029 for the low E_T region and 0.263 ± 0.031 for the high E_T region.
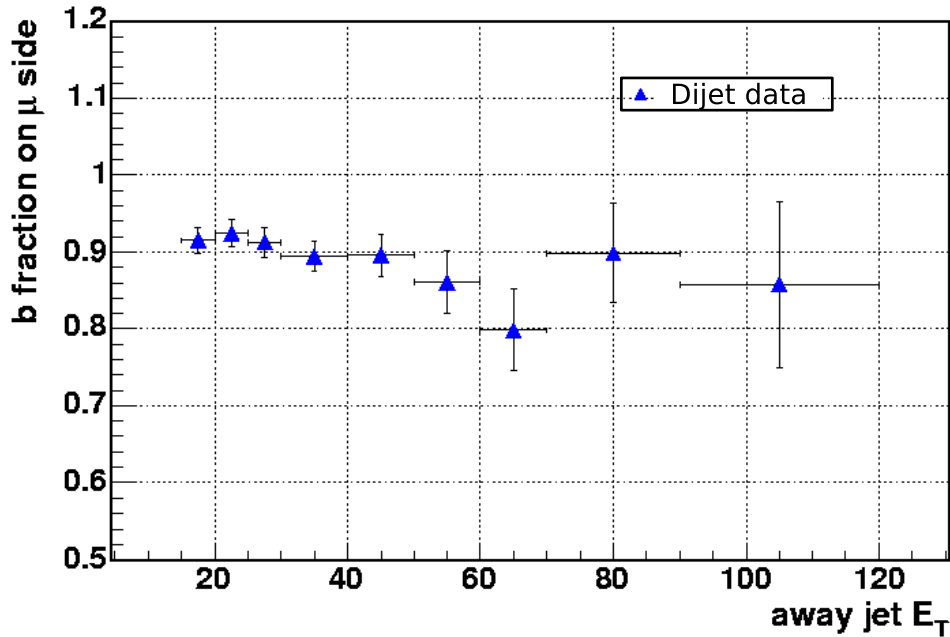
Figure 5.8. The b fraction on the muon side as a function of the away-jet E_T obtained by fitting the $P_{T,\text{rel}}$ spectrum in a muon-enriched dijet data sample using 1 fb $^{-1}$ of data.
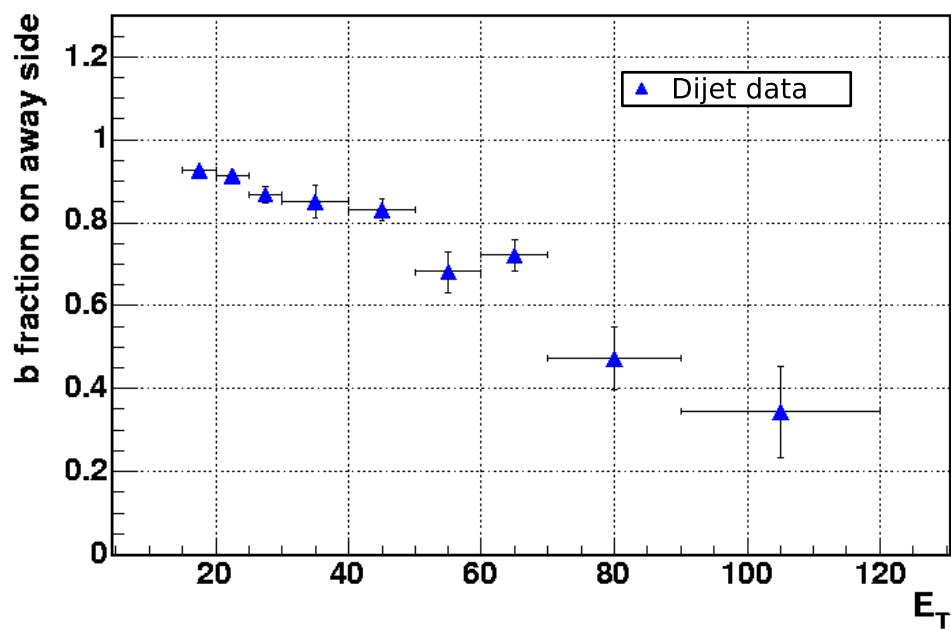
Figure 5.9. The b fraction on the away-jet side as a function of its E_T in a muon-enriched dijet data obtained by fitting its M_{vtx} spectrum.

5.5   Results

By following the previously described procedure we fit the $P_{T,\text{rel}}$ spectrum of the muon-jet for a muon-enriched dijet data sample and found the b fraction to be close to 92% when the away-jet has $E_T \approx 20$ GeV. As presented in Figure 5.8, the b fraction decreases to $\approx 85\%$ at high away-jet E_T values, where the probability of finding fake muons increases. The result of fitting the secondary vertex mass of the away-jet to find the b fraction on the away-jet side is shown in Figure 5.9 for the same data sample. In this case the b fraction decreases from 90% to 40% as the away-jet E_T increases from 20 to 100 GeV due to the fact that the probability of tagging a light jet as a b-jet increases with jet energy. We combined the b fraction on the muon side
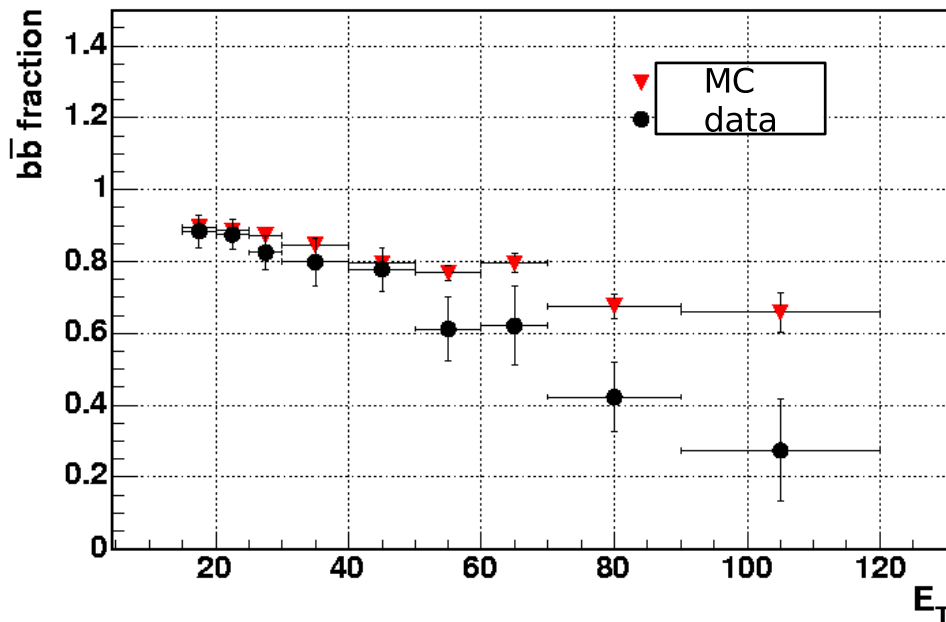
Figure 5.10. Fraction of $b\bar{b}$ events in a muon-enriched dijet data sample as a function of the away-jet E_T for data (circles) and MC (triangles).

with the b fraction on the away-jet side as described in Section 5.4.1.1 and obtained the $b\bar{b}$ fraction (f_{bb}) for each away-jet E_T bin as shown in Figure 5.10. The $b\bar{b}$ fraction

from heavy flavor enriched dijet MC is also shown in this figure and although the drop is not as large as for the data it is still present. In addition to the $f_{b\bar{b}}$ fraction we also need the observed purity (P_{obs}) before we can calculate the real purity of the jet charge algorithm. Using Equation 5.2 we calculated P_{obs} for each away-jet E_T bin. The results are shown in Figure 5.11. A decreasing trend in P_{obs} is seen, similar to $f_{b\bar{b}}$, because lower values of $f_{b\bar{b}}$ dilute the purity of the jet charge algorithm.
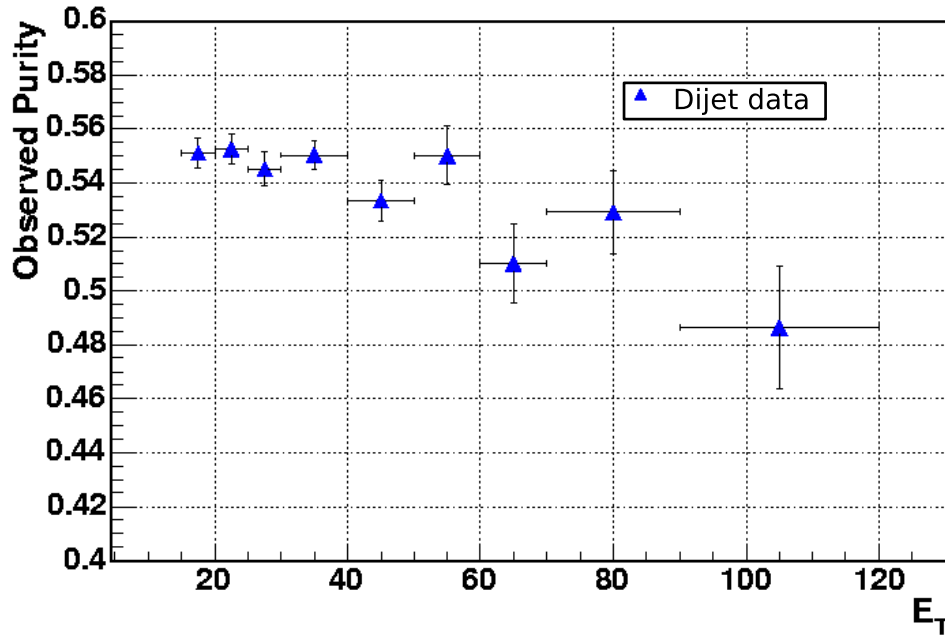
Figure 5.11. The measured purity (P_{obs}) calculated on jets in a muon-enriched dijet data sample as a function of the away-jet E_T .

The real purity of the JQ algorithm (P_{JQ}) from data was obtained by inserting the P_{obs} and $b\bar{b}$ fraction values shown in Figures 5.10 and 5.11 into Equation 5.3. The resulting P_{JQ} distribution for data and MC are shown in Figure 5.12. For the P_{JQ} from MC we calculated the purity using two different MC samples (generic dijet and heavy flavor enriched dijet) because our data are actually somewhere between these two extremes. The P_{JQ} values (red triangles) for MC in Figure 5.12 are the weighted

average between these two MCs. Figure 5.13 shows the calculated P_{JQ} for the two MC samples separately. In addition, in the MC samples the charges of the b-jets were inferred from the charges of the b partons they matched to. This is not possible in data (we don't have parton information) so as a cross-check we took the charge of the b-jets from the charge of the muon in the heavy flavor enriched dijet MC (the black triangles in Figure 5.13). This was done to look for any bias in our method of using the parton level information in MC.
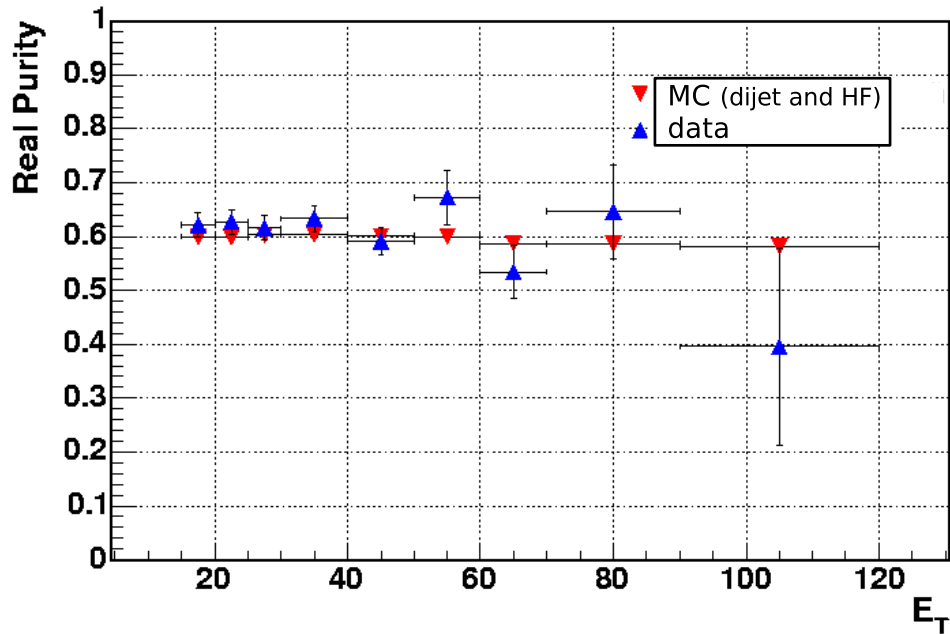
Figure 5.12. The real purity of the jet charge algorithm (P_{JQ}) as a function of the away-jet E_T for the data (blue triangles) and MC (red triangles). The purity for jets matched to b quarks in the MC samples is the weighted average between the purity in the generic dijet MC and the heavy flavor enriched dijet MC.

5.5.1   Scale Factor

To be able to use the real jet charge purity (P_{JQ}) for any sample, in particular for the high E_T b-jets in top events, we decided to present P_{JQ} as a scale factor between

124
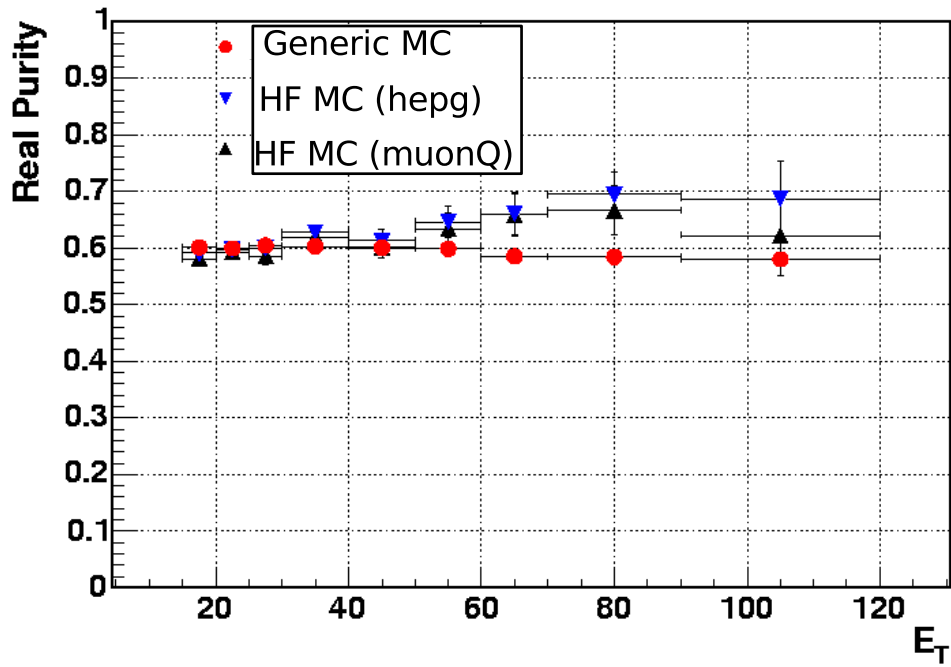
Figure 5.13. The real purity of the jet charge algorithm (P_{JQ}) as a function of away-jet E_T obtained from generic dijet MC (red circles) and heavy flavor enriched dijet MC (inverted blue triangles). The parton level (hepg) bs were matched to jets for the calculation of jet charge purity. The black triangles correspond to P_{JQ} obtained from heavy flavor enriched dijet MC but using the muon charge correlation to decide on the correct JQ assignment instead of the parton level information.

data and MC. Figure 5.14 presents the result for the jet charge scale factor (S_{JQ}) calculated from the ratio between the JQ purity in the muon-enriched data sample and the dijet MC samples. The red line is a fit to a straight line with zero slope while the blue line was a fit to a line with non-zero slope. The nominal SF_{JQ} is taken from the horizontal red line and is 1.03 ± 0.02. The blue line fit will be used to get an uncertainty on the SF_{JQ} due to the E_T dependence and will be explained in Section 5.6.
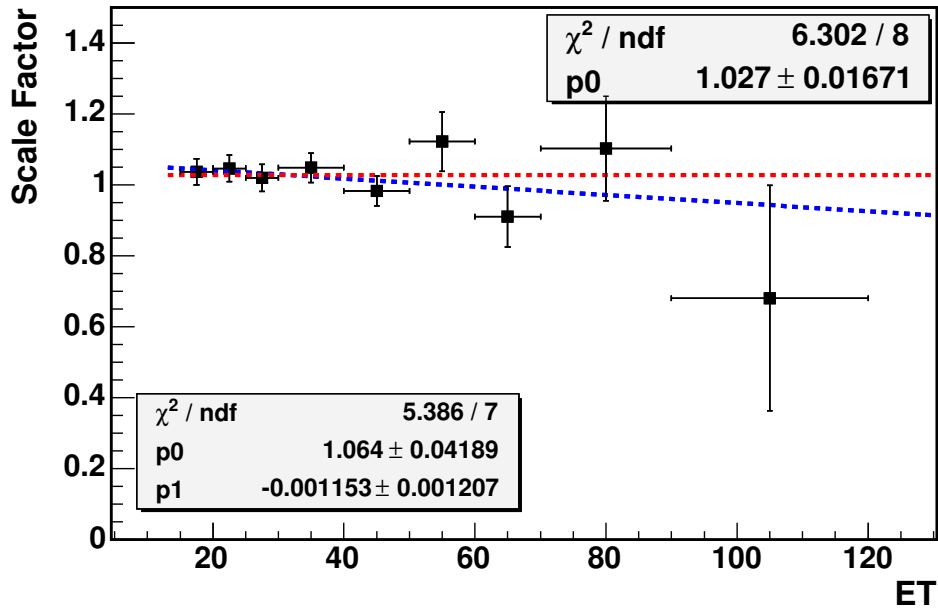
Figure 5.14. The scale factor as a function of E_T for loose SecVtx tagged jets calculated from the ratio between the JQ purity in the muon-enriched data sample and in dijet MC. The red line corresponds to a fit with zero slope while the blue line is from a fit to a line with non-zero slope.

5.5.2 Method checks

As mentioned previously, the jet charge purity (P_{JQ}) from MC was obtained by taking a weighted average between the jet charge purity in a generic dijet MC sample

and a heavy flavor enriched one. As a cross check we used the P_{JQ} from the heavy flavor enriched dijet MC only to calculate the SF_{JQ}. The resulting plot is shown in Figure 5.15 and gives a SF_{JQ} of 1.01 ± 0.02. It is consistent with the nominal SF_{JQ} from Figure 5.14. We also compared M_{vtx} distributions from the heavy flavor enriched MC with the ones from the dijet MC (Figure 5.16) and made sure that the distributions were the same. In addition the entire SF_{JQ} study was repeated using tight SecVtx tagged jets instead of loose tagged jets. The resulting plot is shown in Figure 5.17 and gives a SF_{JQ} of 1.02 ± 0.02 which is consistent with the result from using the loose SecVtx tagger. All the cross checks of the SF_{JQ} agree with our nominal value of 1.03 ± 0.02.
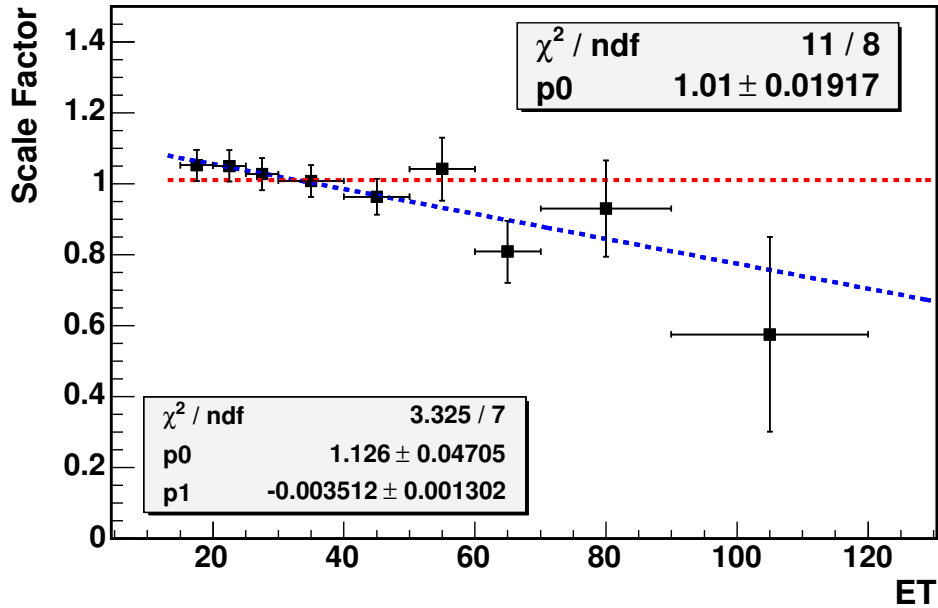
Figure 5.15. The scale factor as a function of E_T obtained from the ratio between the jet charge purity in data and the heavy flavor enriched dijet MC only.
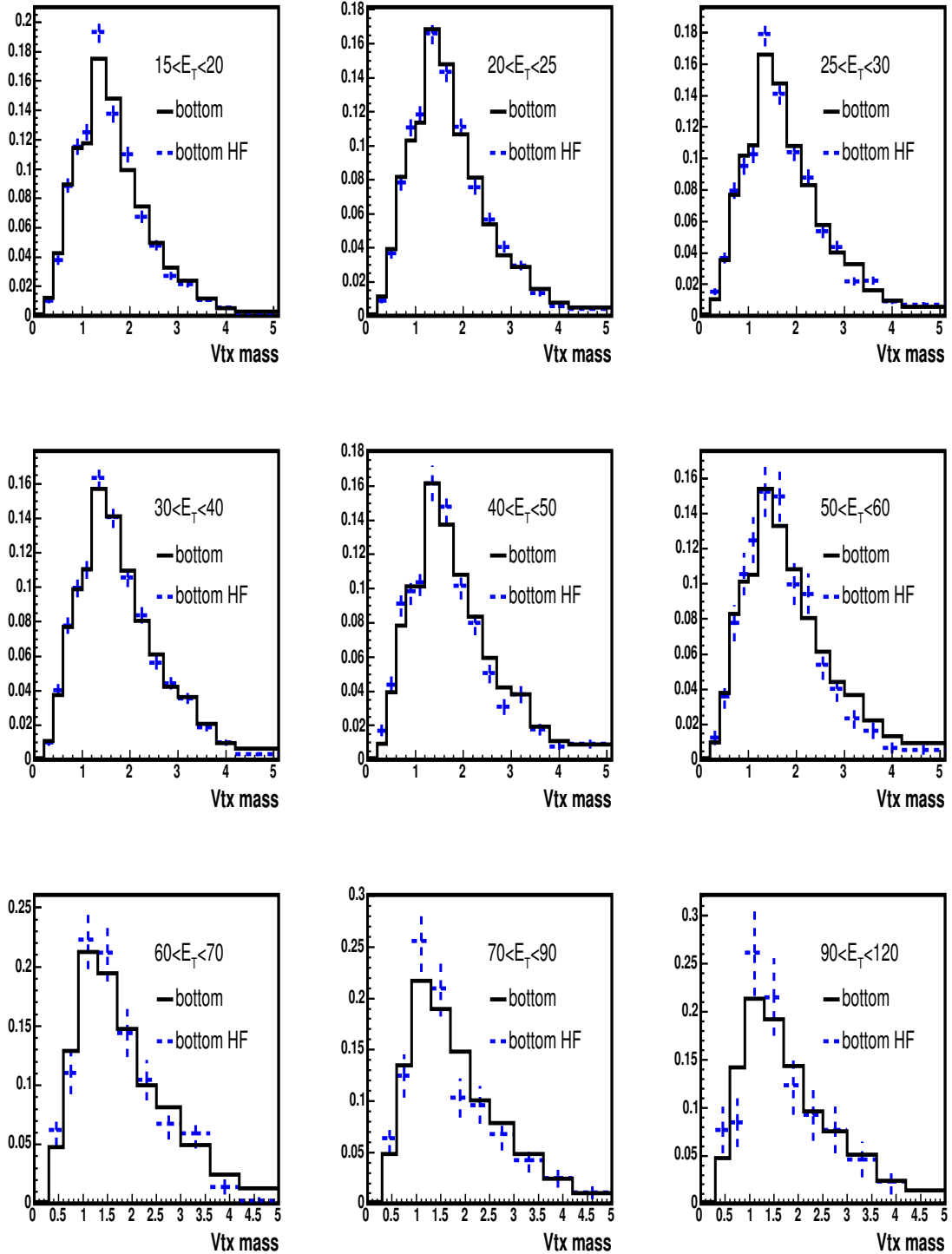
Figure 5.16. M_{vtx} distributions for tight SecVtx tagged b-jets obtained from the heavy flavor enriched dijet MC (labeled bottom HF) compared to ones from the generic dijet (labeled bottom) MC for several away-jet E_T bins.
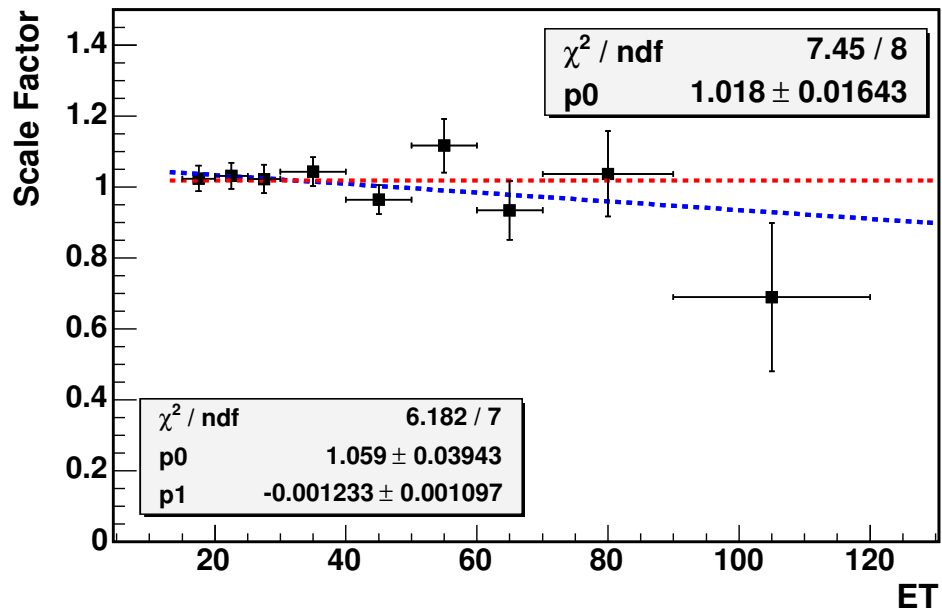
Figure 5.17. The scale factor as a function of E_T obtained from the ratio between the JQ purity using tight SecVtx tagged jets in a muon-enriched data sample and in dijet MC. The red line is a fit to a line with zero slope while the blue line is a fit to a line with non-zero slope.

5.6   Dependence on E_T, η and Number of Vertices

5.6.1 E_T Extrapolation

The calibration of the jet charge algorithm was performed on a $b\bar{b}$ enriched data sample where the E_T of the b-jets ranged from 20 GeV to 60 GeV, but the result is to be used in $t\bar{t}$ events where the b-jets have higher E_Ts as shown in Figure 5.18. Although the scale factor was assumed to be constant with E_T, an uncertainty was assigned to account for any possible dependence on E_T. To get this uncertainty, the ratio between the data and MC jet charge purities was fit as a function of jet E_T with a line of non-zero slope (the blue line in Figure 5.14) and the result was weighted by the b-jet E_T distribution in $t\bar{t}$ events from Figure 5.18. The percentage difference between this weighted scale factor and the nominal one (the red line in Figure 5.14) was added as a systematic uncertainty of 2.9%.
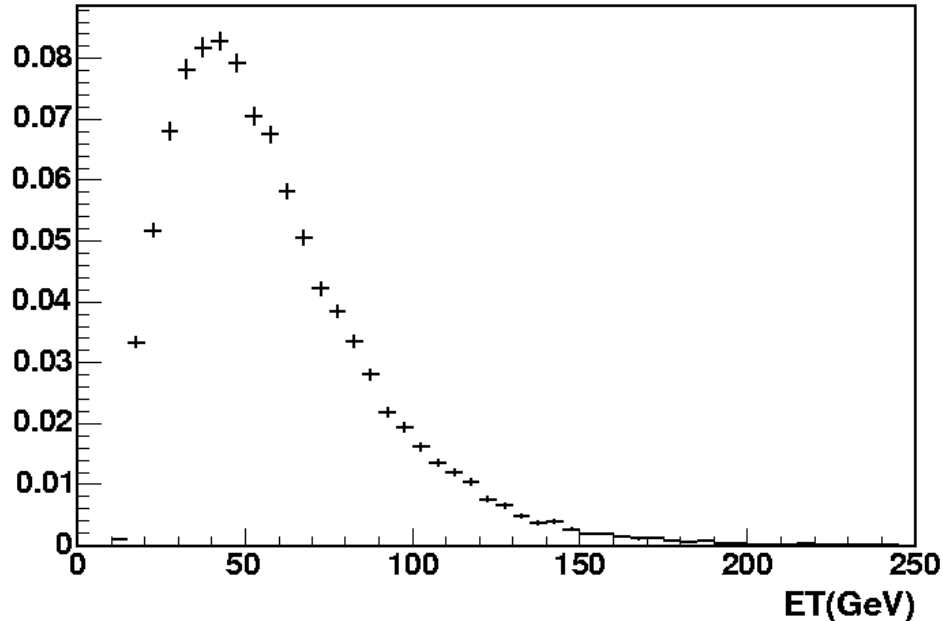
Figure 5.18. The E_T distribution for b-jets in $t\bar{t}$ events.

5.6.2 η and Number of Vertices

Not only was a dependence with respect to E_T studied but also any dependence with respect to the η of the jets and the number of primary vertices in the event. Both of these studies were done in two different away-jet E_T bins, to ensure that any dependence seen was not due to E_T. Figure 5.6.2 shows the $P_{T,\mathrm{rel}}$ and M_{vtx} distributions for b and light quark jets in 3 different η ranges when the away-jet E_T is between 45-75 GeV. The distributions show no dependence on the η of the jets. We also calculated the jet charge scale factor SF_{JQ} for different jet η ranges for two different away-jet E_T bins. No significant dependence on η is observed as can be seen in Figure 5.20. We did a similar study for the number of primary vertices in the event. Figure 5.21 shows the scale factor for two away-jet E_T bins as a function of the number of vertices in an event and shows no significant dependency.

5.7   Systematic Uncertainties

This section explains the sources of systematic uncertainties in measuring the scale factor. We distinguish between ones related to the procedure used to find the b fraction in muon-jets and those corresponding to the b fraction in away-jets. An E_T dependence uncertainty of 2.9% as described in Section 5.6.1 was also added. A summary table is presented at the end of this section.

5.7.1 $P_{T,\mathrm{rel}}$ Tag Bias

For the nominal SF_{JQ} result, the $P_{T,\mathrm{rel}}$ distributions were obtained from MC samples where the tagged b-jets were matched to b quarks. An uncertainty may have been introduced due to the fact that in data we don't know if the b-tagged jet is correctly matched to a b quark. To address this we recalculated the scale factor using $P_{T,\mathrm{rel}}$ distributions that were obtained from tagged b-jets where we did not explicitly check
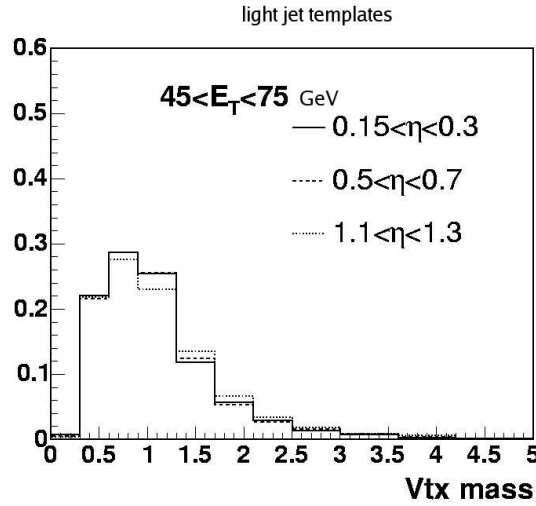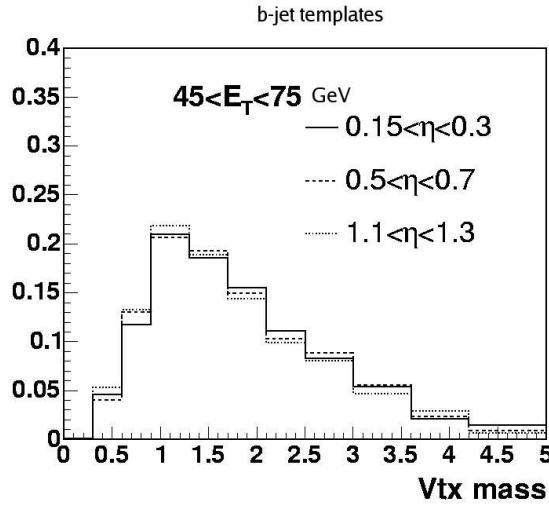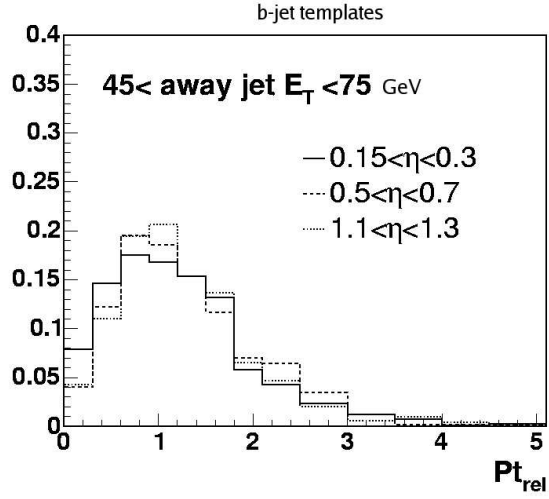
Figure 5.19. The $P_{T,\mathrm{rel}}$ distribution for bottom quarks (left plot) in three different η regions when the away-jet has an E_T between 45 and 75 GeV. The middle and right plots are the M_{vtx} distributions for the bottom and light jets respectively in three different η regions when the away-jet has an E_T between 45 and 75 GeV.
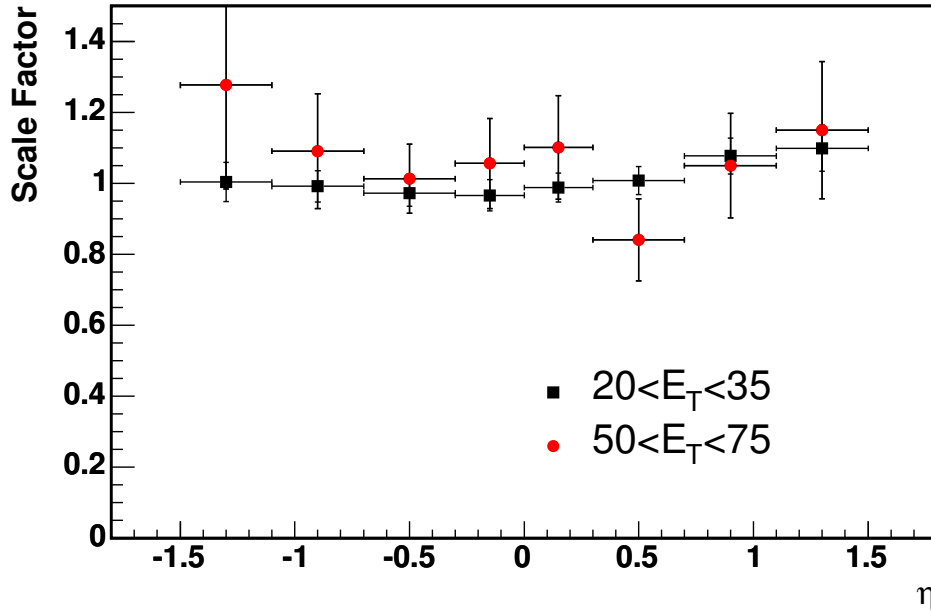
Figure 5.20. The scale factor as a function of η for away-jet E_T ranges of 20-35 GeV and 50-75 GeV.

that the tagged jet came from a b quark. The percent difference for the SF_{JQ} with respect to the nominal value was 1% and is taken as a systematic uncertainty.

5.7.2 $P_{T,\mathrm{rel}}$ non-b Bias

The $P_{T,\mathrm{rel}}$ fits were done using only two templates, one obtained from bottom jets and the other from charm jets. An additional template for light quarks was not used because the charm and light quark distributions were similar. This was discussed in Section 5.4.1 and is shown in Figure 5.2. Despite their similarity the charm and light distributions may affect the SF_{JQ} differently. We recalculated the scale factor using the light quark distribution instead of the charm one. The percent difference with respect to the nominal scale factor was 0.1% and is added as an uncertainty.
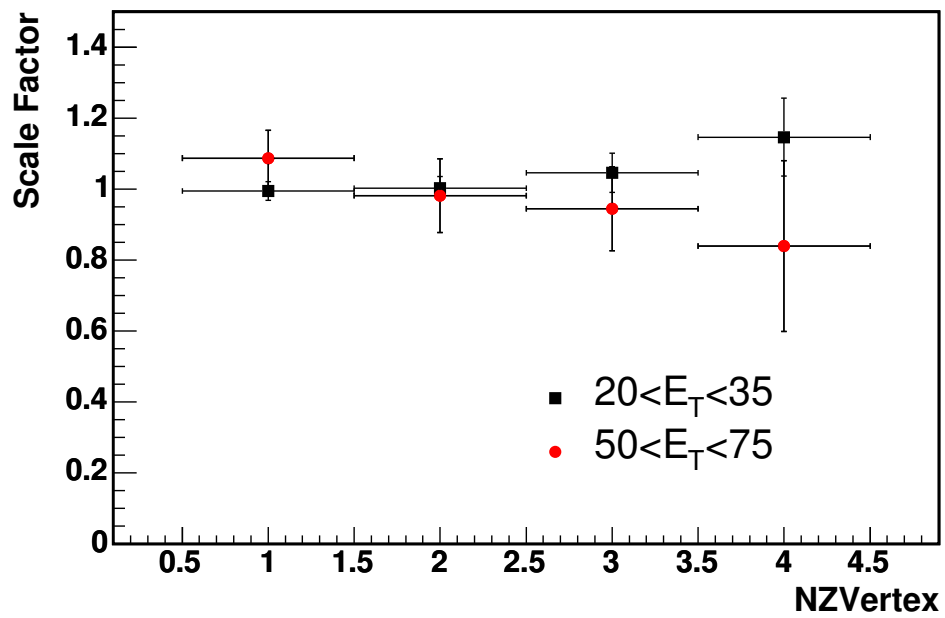
Figure 5.21. The scale factor as a function of the number of primary vertices in an event for away-jet E_T ranges of 20-35 GeV and 50-75 GeV.

5.7.3 M_{vtx} Bias

Any error in the tracking system and/or simulation will effect the M_{vtx} (invariant mass of secondary vertex) distributions since the mass is calculated by using tracks. Actually a study done at CDF using MC shows that we miss 3% of the COT tracks at the track reconstruction level. Missing tracks may cause smaller M_{vtx} values resulting in a 3.5% shift in the M_{vtx} distributions [36]. To account for this track reconstruction inefficiency, M_{vtx} distributions for bottom, charm and light jets were shifted by 5% and the scale factor was recalculated. With the conservative shift of 5%, the percent difference with respect to the nominal SF_{JQ} result was found to be 1.8% and is assigned as an uncertainty.

5.7.4 M_{vtx} Fits

Although the b fraction on the away-jet side was obtained by fitting the M_{vtx} distribution from data to three M_{vtx} distributions (one for bottom, one for charm and the other for light jets), we repeated the study using only two distributions. First we recalculated the SF_{JQ} using only the bottom and charm (bc) distributions for the M_{vtx} fit and then we recalculated using only the bottom and light (bl) distributions. The percent differences with respect to the nominal three distribution case are taken as systematic uncertainties. The results are 1.6% and 0.4% for the bc and bl cases respectively.

5.7.5 Summary

Table 5.4 lists all the systematic uncertainties on the SF_{JQ} from the sources mentioned above. The uncertainty from the E_T extrapolation discussed in Section 5.6.1 is also included. The systematics are shown as the relative uncertainty in % on the SF_{JQ}. The total systematic uncertainty on the SF_{JQ} is 3.9%.

Systematic source	Relative Syst. Uncertainty (%)
$P_{T,\mathrm{rel}}$ Tag Bias	1.0
$P_{T,\mathrm{rel}}$ non-b Bias	0.1
M_{vtx} Bias	1.8
M_{vtx} fit with bc	1.6
M_{vtx} fit with bl	0.4
E_T Dependence	2.9
Total	3.9

Table 5.4. The relative systematic uncertainties on the scale factor from the $P_{T,\mathrm{rel}}$ fit, M_{vtx} fit and E_T extrapolation. The "M_{vtx} fit with bc" implies the fit is performed using bottom (b) and charm (c) jet M_{vtx} distributions. Similarly the "M_{vtx} fit with bl" implies the fit is performed using bottom (b) and light (l) jet M_{vtx} distributions.

5.8   Conclusion

A method to calibrate the Jet Charge (JQ) algorithm, used to determine the flavor of b-jets from data, was presented in this chapter. The procedure compared a muon-enriched dijet data sample to similar MC samples. The fraction of $b\bar{b}$ events was determined by fitting the data to $P_{T,\mathrm{rel}}$ and M_{vtx} distributions and using the expected correlation between the charge of the away-jet and the charge of the muon from the muon-jet. The purity of the JQ algorithm was found after correcting for cascade decays and B mixing. The result of the study is a data to MC scale factor which is shown in Figure 5.14. The measured scale factor for loose SecVtx tagged jets is $1.03 \pm 0.02(\mathrm{stat}) \pm 0.04(\mathrm{syst})$. This scale factor will be used to correct the jet charge purity obtained from MC. The details of how we do this correction will be presented in Chapter 9.

CHAPTER 6

Backgrounds

> Last night, in private, I asked the wise old man
> To reveal to me the secret of the world.
> Softly he whispered, Hush!, in my ear:
> It's something you learn, not words you can hear.
>
> I asked, "What should I do?", He said, "That's it.
> Keep asking what to do.", I asked, "That's it?
> Is that the best that you can do?", He turned to me:
> "Truth seeker, stick to this: What should I do?"
> Rumi (1207-1273), a Turkish sufi.

6.1   Introduction

Even though the event selection presented in Chapter 3 is optimized to pick $t\bar{t}$ lepton+jets events, different non-$t\bar{t}$ processes that mimic the lepton+jets event signature may enter our analysis sample. These processes are called backgrounds and come from W+multijet and non-W events. One example of W+multijet production is W+Heavy Flavor (W+HF) events that include a W boson produced in association with jets that come from b or c quarks. W bosons can also be produced along with jets from light (u, d, s) quarks. These W+Light Flavor background events are also called mistags since they can mimic the top sample only if the light quark jets are mistakenly tagged as b-jets. Other possible W+multijet backgrounds come from diboson (WW, ZZ, WZ)

events and events that include only one top quark called single top events. The non-W background, also called QCD background, is basically multijet events that do not include the production of a W boson. Since only a small fraction of background events contain jets coming from b or c quarks, tagging both b-jets using the SecVtx algorithm significantly reduces the background contribution.

The expected contribution from each background has already been calculated for the CDF top cross-section (σ) measurement [29], so we rely directly on their background estimation which is given in the second column of Table 6.1. The expected numbers of background events (N_b) after the top charge analysis selection requirements are presented in the last column of Table 6.1. They are obtained by multiplying the background predictions from the cross-section analysis (second column) by the total efficiency of the selection requirements for the top charge analysis which is the product of the pairing and jet charge efficiencies as shown below.

$$\epsilon_{\text{total}} = \epsilon_{\text{pairing}} \cdot \epsilon_{JQ} \tag{6.1}$$

The uncertainties on each background are statistical only. For the lepton+jets channel the fraction of each background that has two jets tagged by the loose SecVtx tagger is shown as a pie chart in Figure 6.1. The left diagram shows the fractions for the cross-section analysis and the chart on the right shows how these fractions change after the selection requirements for the top charge analysis are applied.

For the top charge analysis, in addition to the number of background events, we need to know if some of the backgrounds are more likely to mimic Standard Model like (N_+) events or exotic quark like (N_-) events.

$$P_b = \frac{N_+}{N_+ + N_-} \tag{6.2}$$

We define the purity (P_b) on each background type as the number of events assigned as N_+ over the total number of events passing the selection cuts (see Equation 6.2).

background	L+J σ prediction	$\epsilon_{\text{total}} = \epsilon_{\text{pairing}} \cdot \epsilon_{JQ}$	N_b
W+HF	5.3 ± 1.2	0.15 ± 0.01	0.8 ± 0.2
QCD fakes	1.8 ± 1.0	0.20 ± 0.10	0.4 ± 0.3
Diboson	0.38 ± 0.01	0.10 ± 0.05	0.04 ± 0.02
Mistag	1.3 ± 1.2	0.08 ± 0.03	0.1 ± 0.1
Single Top	0.77 ± 0.06	0.16 ± 0.01	0.13 ± 0.01
Total	9.6 ± 2.0	-	1.4 ± 0.3

Table 6.1. Table of expected background events (N_b) for the top charge analysis in the lepton+jets (L+J) channel for 695 pb^{-1} of data. L+J σ prediction is the number of background events from the cross section analysis. Third column is the total efficiency of the top charge selection requirements.

Figure 6.1. Lepton+jets background fractions for the loose double tagged sample before and after the selection requirements for the top charge analysis.

A P_b value of 0.5 is equivalent to saying that the background events are as likely to be assigned as Standard Model top events as exotic quark events. Therefore, we sometimes refer to P_b as the charge asymmetry and a purity of 0.5 means the background is charge symmetric. This chapter describes each background source to the top events in the lepton+jets channel along with the charge asymmetry it exhibits.

6.2 $W+$Heavy Flavor

6.2.1 Production and Estimation

Heavy flavor production in association with a W boson contributes significantly to the background in the b-tagged lepton+jets sample even though $W+$Light Flavor dominates the sample before b-tagging has been applied. Remember that for the top charge analysis we tag both b-jets using the loose SecVtx algorithm. In this case, 55% of the background events are from $W+$Heavy Flavor ($W+$HF). A $W+$HF event is the result of gluon splitting as shown in Figure 6.2 where $pp \rightarrow Wg$ is followed by the gluon splitting to $b\bar{b}$, $g \rightarrow b\bar{b}$. Gluon splitting can also produce $Wc\bar{c}$ events that may mimic the $t\bar{t}$ lepton+jets event signature if one or both of the c jets is tagged as a b.

Several Monte Carlo generators are capable of performing matrix element calculations for $W+$jets. The MC generator used to model "$W + 4$ jet" events in this analysis is ALPGEN [37]. The ALPGEN MC calculates exact matrix elements at leading order along with providing proper treatment of heavy quark masses, spins and color flows. On these ALPGEN events a simulation of parton fragmentation with a shower algorithm based on the HERWIG [17] program and a decay algorithm based on EvtGen [16] are used. The estimation of the amount of $W+$HF background requires special care since imprecise values of heavy flavor fractions (the number of events expected to contain a heavy flavor quark) and/or mistakes in the detector

simulation can cause significant errors on the W+HF prediction. The number of W+jets events in the data before b-tagging is measured. The number of events we expect from other backgrounds is then subtracted off and the remaining amount is multiplied by the W+HF fraction and b-tagging efficiency to calculate the amount of W+HF in the b-tagged lepton+jets sample. For the top charge analysis we expect 0.8 ± 0.2 W+HF background events.
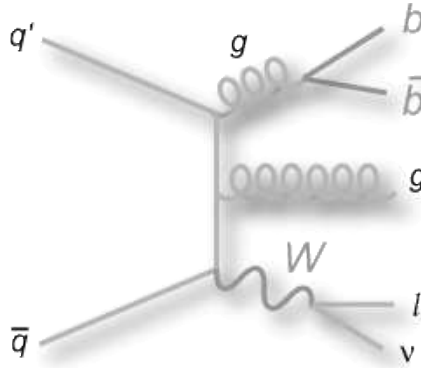
Figure 6.2. The Feynman diagram of a $W + b\bar{b}$ event that may mimic a lepton+jets $t\bar{t}$ event. The resulting gluon splits into a $b\bar{b}$ pair while the other resulting product, the W boson, decays to a lepton neutrino pair. $W + c\bar{c}$ events are produced in an identical manner.

6.2.2 Charge Asymmetry

As mentioned before, we need to make sure that there is no bias towards either the Standard Model or exotic model case for each of the backgrounds. In other words, we must ensure that the jet charge purity for each background (P_b) is 0.5. The main problem in the pursuit of measuring P_b is that once the top charge analysis cuts have been applied, such as the double b-tag requirement and χ^2 cut, there are only a few events left in the background samples. In such cases, we calculated P_b by loosening these requirements, such as using events with a single or no b-tag. The number of N_+ and N_- type events in the single b-tagged $W + b\bar{b}$ and $W + c\bar{c}$ Monte Carlo samples and their corresponding purities are shown in Table 6.2. The P_b for $W + b\bar{b}$ and $W + c\bar{c}$
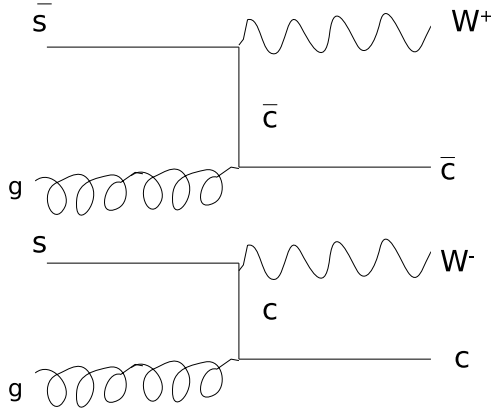
Figure 6.3. The Feynman diagrams for the production of $W + c$ events.

events agree with 0.5 within their uncertainties.

The one W+HF sample where we do see a charge asymmetry is in the $W + c$ events. The c quark is generated with a W^- and $\bar{c}$ is generated with a W^+, as shown in Figure 6.3, so a $W + c$ event will look like a N_+ type event. Remember we require the lepton+jets events to have 4 or more jets but $W + c$ events typically have only one jet. $W + c$ events can only pass our selection requirements if there are at least three additional gluon jets produced and the c jet is tagged as a b-jet. Therefore it is really hard for a $W + c$ event to enter our sample. Since the amount of $W + c$ background we expect in our sample is very small, we assumed a P_b of 0.5 ± 0.0 for $W + c$ events. Nevertheless we checked for the expected charge asymmetry in $W + c$ events requiring only one jet. Table 6.3 shows the number of N_+ and N_- type events and the corresponding P_b for the $W + c$ sample requiring only one jet in the event. The first row in the table shows the charge asymmetry where we have not done any parton level check. Once we require the c jet in the event to be matched to a c quark (second row), the charge asymmetry is slightly higher $\approx 62\%$ confirming the expected charge asymmetry. The result in the third row is obtained using only events where the c jet does not match to the c quark.

W+HF	# of tags	N_{total}	N_+	N_-	P_b
	2 tags	223	111	112	0.50 ± 0.03
$W + b\bar{b}$	1 tag	1329	658	671	0.50 ± 0.01
	Pretag	2127	1085	1042	0.51 ± 0.01
	2 tags	28	10	18	0.36 ± 0.09
$W + c\bar{c}$	1 tag	515	268	247	0.52 ± 0.02
	Pretag	2194	1134	1060	0.52 ± 0.01
	2 tags	22	10	12	0.45 ± 0.1
$W + c$	1 tag	438	224	214	0.51 ± 0.03
	Pretag	2369	1256	1113	0.53 ± 0.02

Table 6.2. The charge asymmetry (P_b) from MC for W+HF background events. The N_{total} column shows the total number of events left after applying the lepton+jets and top charge selection cuts on the MC sample. The N_+ and N_- columns show how N_{total} is shared between SM-like events (N_+) and exotic-like (N_-) events respectively.

	N_+	N_-	P_b
No parton level check	1072	672	0.61 ± 0.01
jet matched to c quark	1026	634	0.62 ± 0.01
jet not-matched to c quark	46	38	0.55 ± 0.05

Table 6.3. The charge asymmetry (P_b) for $W + c$ loose SecVtx tagged events requiring 1 jet in the event. P_b of 0.5 indicates no charge asymmetry. The first row in the table shows the result where we have not done any parton level check. Once we require the c jet in the event to be matched to a c quark (second row), the charge asymmetry is slightly higher. The result in the third row is obtained using only events where the c jet is not matched to a c quark.

6.3 QCD (non-W)

6.3.1 Production and Estimation

Because quarks have color charge, they can radiate gluons and this radiation can be detected as jets in the CDF detector. If one of these jets is misidentified as an isolated lepton, the QCD event can pass our lepton+jets sample selection. Due to the fact that these events do not include production of a W boson, they are also called non-W events (see Figure 6.4 as an example). There are several reasons that an object identified in the detector as a lepton is not actually a lepton coming from the decay of a W boson. The first possibility is that a jet can fake a lepton if part of the jet passes the isolation requirement. The second possibility of a non-W background is direct $b\bar{b}$
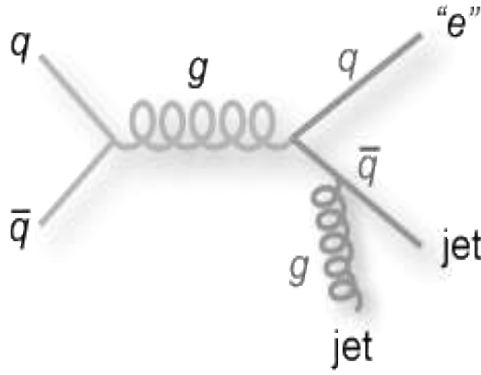
Figure 6.4. Feynman diagram of a multijet non-W event where the $q\bar{q}$ annihilation resulted in a gluon and then produced two light jets with an additional gluon jet from one of the light quarks. One of the light jets can be misidentified as a lepton and if there is additional gluon radiation the event can pass the lepton+jets event selection.

or $c\bar{c}$ production. In $b\bar{b}$ events, it is possible that a real lepton from the semileptonic decay of a B hadron can pass our isolated lepton requirements. A third possibility is a photon conversion that is misidentified as an electron or a high momentum pion that satisfies the muon criteria.
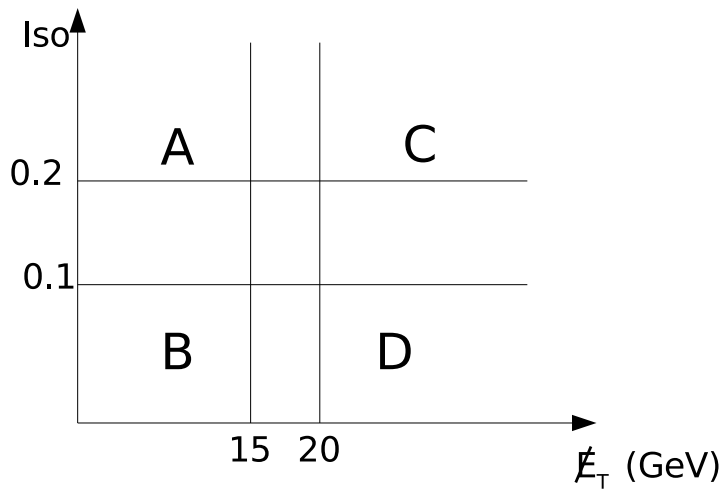
Figure 6.5. A drawing representing the 4 regions in isolation versus $\vec{\rlap{E}}_T$ that are used for the QCD background prediction.

One of the methods at CDF for estimating the QCD background in the b-tagged lepton+jets sample is called the Iso vs. $\vec{\rlap{E}}_T$ method where the isolation of the lepton

(Iso) is plotted against the missing transverse energy, $\vec{E}_T$. The plot is divided into 4 regions, named A, B, C and D as shown in Figure 6.5. The isolation and $\vec{E}_T$ cuts used for each region are given below:

A: $Iso > 0.2$, $\vec{E}_T < 15$ GeV

B: $Iso < 0.1$, $\vec{E}_T < 15$ GeV

C: $Iso > 0.2$, $\vec{E}_T > 20$ GeV

D: $Iso < 0.1$, $\vec{E}_T > 20$ GeV

Note that higher values of isolation do not indicate more isolated leptons. For example, $Iso = 0.1$ for electrons means the energy in a cone drawn around the electron cluster is 10% of the cluster energy (see Section 3.2.1 for details). Therefore leptons with $Iso < 0.1$ are much more isolated from the jet activity than leptons with $Iso > 0.2$. The large $\vec{E}_T$ and small isolation region, region D, is where top events are concentrated. The events in region A have large isolation values so they are events with non-isolated leptons and small $\vec{E}_T$. This region is enriched in QCD background and has only a small amount of W events. To get the number of non-W events in the signal region (region D), Equation 6.3 is applied to the b-tagged lepton+jets sample, where N_A, N_B, N_C, N_D are the number of events in regions A, B, C and D.

$$N_D = N_C \times \frac{N_B}{N_A} \qquad (6.3)$$

The total number of non-W background events in the signal region is estimated as the number of non-isolated lepton candidates in the high $\vec{E}_T$ region scaled by the ratio of isolated to non-isolated lepton candidates in the low $\vec{E}_T$ region, which is dominated by background. We expect 0.4 ± 0.3 QCD background events in the lepton+jets channel after the top charge selection requirements.

6.3.2 Charge Asymmetry

In order to estimate the charge asymmetry for non-W events, we used the Isolation versus $\vec{\rlap{E}}_T$ method as described above. Running over the lepton+jets top data sample we tried to calculate the charge asymmetry for each region. To prevent double counting of the same physical object, for example non-isolated electrons and muons, jets that were matched to a lepton within a cone of $\Delta R < 0.4$ were removed. We also applied an additional correction to the $\vec{\rlap{E}}_T$ for the presence of non-isolated muons as explained in Section 3.4. After all of the top charge selection requirements were applied, there were almost no events left in any of the regions so we could not calculate the purity.

Since we could not get a direct QCD purity result from the top data sample, we tried to calculate the purity using other data and MC samples. To better understand the details of the studies tried, let's review again the different sources of QCD backgrounds. QCD arise from

1. events where a non-isolated lepton passes isolation requirements
2. multijet events where a jet fakes a lepton
3. events where a photon conversion is misidentified as an electron or
events where a high momentum pion is misidentified as a muon.

Out of these three cases the type of events for which we expect to see a charge asymmetry towards Standard Model top are $b\bar{b}$ and $c\bar{c}$ events which are only a subset of the first case. Here we will describe the studies we did to measure the charge asymmetry due to the first and second cases.

For the first case, we choose $b\bar{b}$ events from the dijet data and calculate the purity for these QCD events. The dijet data include leptons that are coming from the decay

Iso vs. $\vec{\not{E}}_T$ region	N_+	N_-	P_b
A	117446	113567	0.508 ± 0.001
B	61612	55178	0.528 ± 0.001
C	4137	3662	0.530 ± 0.006

Table 6.4. Purity (P_b) of events from dijet data before b-tagging is applied on the jet that is back-to-back from the jet including the lepton.

Iso vs. $\vec{\not{E}}_T$ region	N_+	N_-	P_b
A	6448	5376	0.55 ± 0.01
B	1489	1282	0.54 ± 0.01
C	339	263	0.56 ± 0.02

Table 6.5. Purity (P_b) of events from dijet data after b-tagging is applied on the jet that is back-to-back from the jet including the lepton. We use the purity in region A for the QCD background.

of heavy flavor hadrons. For a $b\bar{b}$ event to pass our lepton+jets selection cuts, the lepton from the semileptonic decay of the B hadron must be chosen as the isolated lepton. Since the charge of a lepton from a semileptonic hadron decay is correlated to the flavor of the b and our Wb pairing method will most likely match this lepton with the b jet it is coming from, the event will look SM-like. Therefore we expect to see a tendency towards SM-like top events from $b\bar{b}$ background events. For the selection of $b\bar{b}$ events from the dijet data we required a b-tagged jet that is back-to-back in azimuth to the jet the lepton came from. Tables 6.4 and 6.5 show the purities found for each iso vs. $\vec{\not{E}}_T$ region for the events chosen before and after b-tagging the jet opposite the lepton jet. For the tagged case in the QCD-enriched region A, we measured a purity of 0.55 ± 0.01.

To check the purity for the second type of QCD events we used a dijet MC sample where a jet was assigned to be a fake electron if the jet passed the following requirements[9]:

Jet $E_T > 20$ GeV

Jet $\eta < 1.1$

---

[9]These selection cuts come from a fake lepton study done at CDF [38].

Number of tracks in jet≥ 4

$0.8 < E_{\mathrm{em}}/E_{\mathrm{total}} < 0.95$

where $E_{\mathrm{em}}/E_{\mathrm{total}}$ is the fraction of electron energy as measured in the electromagnetic calorimeter to the total energy of the jet. The charge of the fake lepton was taken from the curvature of the highest P_T track in the jet. We found a purity of 0.50 ± 0.01 which verified that there is no charge asymmetry for QCD events with a fake lepton.

As a result of the above studies, we decided to use the result of 0.55 ± 0.01 from the study done using data for the overall QCD background. This is however a conservative result since 0.55 was obtained from $b\bar{b}$ events which is the only charge asymmetric component of QCD events and will actually be diluted by other charge symmetric QCD events. We used the 0.55 for the purity but made the errors asymmetric using $+0.01$ and -0.05 to cover the charge symmetric result of 0.5.

6.4 Mistags (W+Light Flavor)

6.4.1 Production and Estimation

One other type of background to the lepton+jets sample are the W boson events produced in association with light quarks (W+Light Flavor) where a light quark jet is tagged as a b-jet. These jets that are tagged as b-jets even though they do not contain a true b quark are called mistagged jets. To calculate the background from mistags, we make use of the transverse decay length (L_{xy}) of the tracks coming from the secondary vertex. As mentioned in Section 3.5.3, L_{xy} is a signed distance and if a secondary vertex is reconstructed on the opposite side of the primary vertex from the jet direction it has a negative value (see the bottom diagram in Figure 6.6). In the negative L_{xy} case, the jet is much more likely to have come from a light quark than a b quark. By parameterizing the negative L_{xy} values as a function of the jet

148
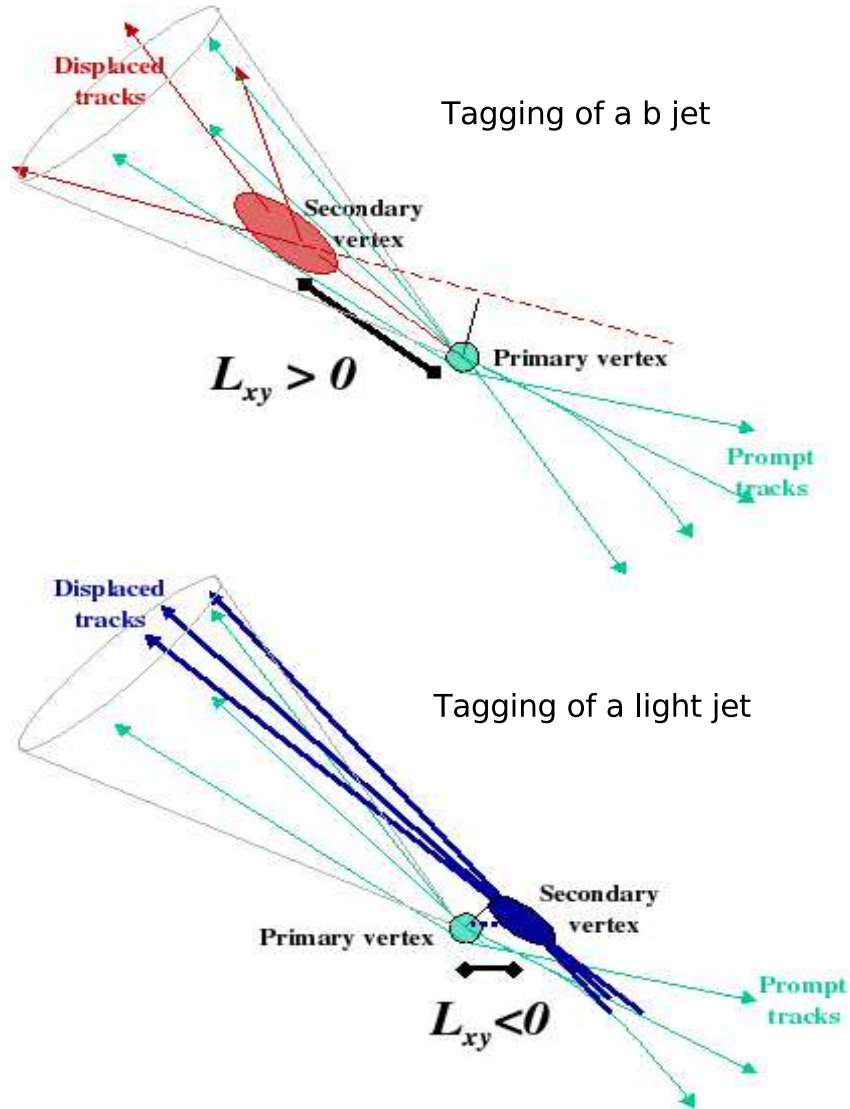
Figure 6.6. The transverse decay length (L_{xy}) of the tracks coming from the secondary vertex. If a secondary vertex is reconstructed on the opposite side of the primary vertex from the jet direction $L_{xy} < 0$ and the jet is much more likely to have come from a light quark than a b quark.

E_T, η and number of SecVtx tracks in the jet, the probability to mistag a jet can be given by a matrix known as the mistag matrix [29]. A jet that has at least two SecVtx tracks as described in Section 3.5.3 is called a "taggable jet" and the size of the mistag background is estimated by weighting each taggable jet with its mistag rate in the lepton+jets data sample before b-tagging. In the case of double tagged events using the loose SecVtx tagger, we expect 0.1 ± 0.1 events to be from mistags.

6.4.2 Charge Asymmetry

In order to check the charge asymmetry in the mistags, MC samples of $W+1$ or more light quarks (W+LF) were generated using MADGRAPH+PYTHIA with the W required to decay leptonically. Due to low statistics in the samples after the top charge analysis selection, the requirements on the number of b-tags was loosened. By checking the single and pretagged samples instead of the double tagged sample, we confirmed that there was no charge asymmetry introduced by mistags that would bias the top charge measurement towards the Standard Model or exotic model (See Table 6.6). So we use a P_b of 0.5 ± 0.0 for the mistag background.

# of tags	N_{total}	N_+	N_-	P_b
2 tags	7	2	5	0.3 ± 0.2
1 tag	350	192	158	0.55 ± 0.04
Pretag	3766	2025	1731	0.54 ± 0.02

Table 6.6. Jet charge purity (P_b) for the mistags background obtained from W+LF MC.

6.5 Diboson

6.5.1 Production and Estimation

The diboson background consists of WW, WZ and ZZ events. In a WW event, if one W decays to a lepton-neutrino pair and the other to two light quarks plus if there is at least one gluon jet from initial or final state radiation, the event can mimic

a lepton+jets $t\bar{t}$ event. Similarly, a ZZ event can pass our selection cuts if one of the Z bosons decays hadronically to $c\bar{c}$ or $b\bar{b}$ pair while the other Z decays to two leptons. The large $\vec{\not{E}}_T$ requirement can be satisfied if one of the leptons from the Z goes undetected giving the appearance of missing transverse energy. For the WZ background to mimic a lepton+jets $t\bar{t}$ event, the W boson must decay leptonically and the Z to a $c\bar{c}$ or $b\bar{b}$ pair. The ALPGEN event generator is used to predict the amount of diboson backgrounds in the sample. The ALPGEN events are then put through the HERWIG program for hadronization and parton showering. The estimation of the diboson background events is obtained by using Equation 6.4.

$$N_{\text{bckg}} = \sigma \cdot BR \cdot SF \cdot \left(\sum_{\text{lepton}} \int Ldt \cdot \epsilon_{\text{accept}} \cdot \epsilon_{\text{tag}} \right) \qquad (6.4)$$

where ϵ_{accept} is the acceptance efficiency of the process (WW, WZ or ZZ). The ϵ_{tag} is the b-tagging efficiency and $\int Ldt$ is the integrated luminosity. The sum is taken over all possible lepton types. Here σ and BR represent the theoretical cross-section and branching ratio for a given process (WW, WZ or ZZ) and SF is the b-tagging scale factor. Since the diboson cross-section is small, we expect only 0.04 ± 0.02 diboson events after the top charge selection.

6.5.2 Charge Asymmetry

There is no physical mechanism in diboson events that can cause a bias toward $+2/3$ or $-4/3$ events. The expected symmetric behavior was confirmed by measuring the charge asymmetry in diboson MC samples as shown in Table 6.7 and 0.5 was used for the diboson background purity.

# of tags	N_{total}	N_+	N_-	P_b
2 tags	4	2	2	0.5 ± 0.3
1 tag	94	40	54	0.43 ± 0.06
Pretag	570	303	267	0.53 ± 0.02

Table 6.7. Jet charge purity (P_b) for diboson background obtained from diboson ($WW + WZ + ZZ$) MC samples.
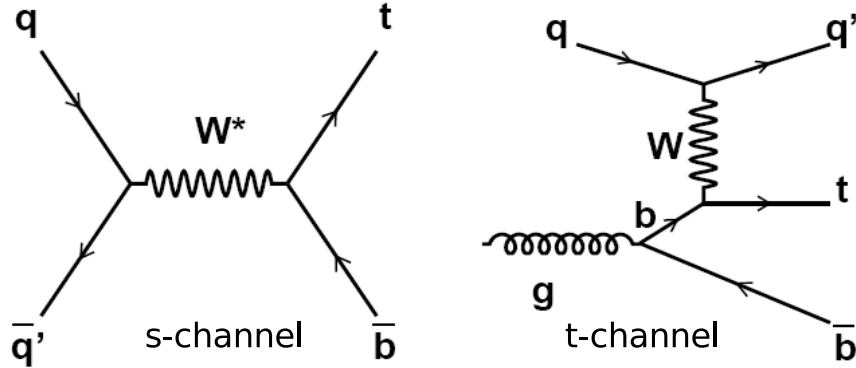
Figure 6.7. The left (right) picture is a Feynman diagram for s-channel (t-channel) single top production.

6.6 Single Top

6.6.1 Production and Estimation

At the Tevatron, in addition to $t\bar{t}$ pair production via the strong interaction, a top quark can also be produced singly through the electroweak interaction. A single top event can be produced in association with a b quark through $q\bar{q}$ annihilation (s-channel). It can also be produced through the W-gluon fusion process (t-channel) in which an initial gluon splits into a $b\bar{b}$ pair and a b quark interacts with a virtual W as shown in Figure 6.7. The single top background is just 1% of the total background to the loose b-tagged lepton+jets events and obtained by running on MadEvent MC [39] events generated for s-channel and t-channel seperately. The single top prediction was calculated in a similar manner to what was done for the diboson prediction. The expected number of single top events is 0.13 ± 0.01 after the top charge selection.

6.6.2 Charge Asymmetry

Even though single top events are considered background for $t\bar{t}$ events, they are not actually a background for the top charge measurement since we could measure the charge of the single top quark in the event. However, our event selection was optimized for $t\bar{t}$ events and by treating single top as a background we checked our expectation of seeing a charge asymmetry towards N_+. If the matching between the lepton and b-jet is correct, the event should contribute to N_+ because it was generated as SM top. We ran on MadEvent MC data and performed our event selection. When we count the number of events for which the b jet is confirmed to be coming from a top decay using the MC generator information, the relative number of N_+ events increases as expected. Even though this asymmetry is suppressed due to incorrect pairing assignments and the imperfectness of the jet charge calculation, we still observe a charge asymmetry of 0.58 ± 0.02 for the case where we count all events whether or not the pairing was correct. Once we make sure that our Wb pairing method picked the correct pair, the charge asymmetry goes up to 0.69 ± 0.03. The results are summarized in Table 6.8. We used a $P_b = 0.58 \pm 0.02$ for the single top background.

	P_b
no parton level check	0.58 ± 0.02
b jet coming from top	0.69 ± 0.03

Table 6.8. Purity for the b-tagged single top MC events.

6.7 Background Summary

We started with the backgrounds from the CDF top cross-section analysis [29] and then applied the top charge selection requirements to find the expectation for each background in this analysis as shown in Table 6.1. In addition we need the background purity to determine whether the background looks more like the Standard Model or

the exotic model. The measured P_b was 0.5 for most of the backgrounds indicating no charge asymmetry. Two places we observed a slight excess of SM-like events were from the QCD and single top backgrounds. Table 6.9 shows the purities obtained for all of the backgrounds.

Background	# of b-tags	P_b
W+HF		
$W + bb$	2 tags	0.50 ± 0.03
	1 tag	0.50 ± 0.01
	Pretag	0.51 ± 0.01
$W + cc$	2 tags	0.36 ± 0.09
	1tag	0.52 ± 0.02
	Pretag	0.52 ± 0.01
$W + c$	2 tags	0.45 ± 0.11
	1tag	0.51 ± 0.03
	Pretag	0.53 ± 0.02
Mistag		
	2 tags	0.29 ± 0.17
	1tag	0.55 ± 0.04
	Pretag	0.54 ± 0.02
Diboson		
	2 tags	0.50 ± 0.25
	1tag	0.43 ± 0.06
	Pretag	0.53 ± 0.02
QCD		$0.55^{+0.01}_{-0.05}$

Table 6.9. Background Purity P_b (Charge Asymmetry) for each background calculated from the equation $P_b = N_+/(N_+ + N_-)$. A purity of 0.5 implies charge symmetry.

CHAPTER 7

Systematic Uncertainties

> What men really want is not knowledge but certainty.
> Bertrand Russell (1872–1970), a British philosopher.

Apart from the statistical uncertainty introduced due to the finite statistics in our data sample, there are errors that can arise from uncertainties in the CDF detector's performance, errors in the calibration of calorimetry and tracking systems, as well as imperfections and assumptions in the analysis method or MC modeling. Systematic uncertainties are introduced to cover these errors. The uncertainty on the acceptance of signal and background events is included in the predictions obtained from the cross-section analysis [29]. For the top charge measurement, we have studied how systematic uncertainties affect the efficiencies (ϵ) and purities (P) of the Wb pairing and jet charge. The systematic uncertainties associated with the top charge measurement are listed below and will be explained in detail in the following subsections.

- Jet Energy Scale

- Initial/Final State Radiation

- Top Mass Uncertainty

- b-Tagging

- Parton Distribution Functions (PDF)

155

- Monte Carlo Modeling

7.1 Jet Energy Scale

The energy of jets measured using calorimeter towers does not correspond exactly to the energy of the initial parton that the jet came from so the jet energies must be scaled back to particle or parton level as described in Section 3.3.2. All of the various sources of uncertainty in the CDF jet energy scale listed below can cause a discrepancy between the modeling of jets in Monte Carlo and the actual jets observed in data.

- Relative Jet Energy Scale

- Raw Jet Energy Scale

- Absolute Jet Energy Scale

- Multiple Interactions

- Underlying Event

- Out-of-cone Corrections

In CDF there are different levels of jet energy scale (JES) corrections where for each level one or more of the sources of JES corrections are applied. For this thesis, we have used the level of JES corrections that corrects jets back to the particle level. Therefore multiple interactions, underlying event and out-of-cone corrections that correct jet energies from the particle level to parton level were not considered in this analysis, so no systematic uncertainty was taken for these.

As explained in Section 3.3.2, the relative JES corrections remove the η dependence of the calorimeter. The amount of uncertainty on the energy of a jet differs with different η ranges and the largest uncertainty ($\approx 7\%$) is assigned in the region with

156

$|\eta| > 2$. The amount of systematic uncertainty for jets in the region $0.2 < |\eta| < 0.6$ is shown (dashed green line) in Figure 7.1 as a function of jet P_T. The raw JES correction, which accounts for the non-linear response of the calorimeter, is around 5%. For the absolute JES, the jet energy measurement relies on the detector simulation and fragmentation model, so the uncertainty on the correction comes from calorimeter simulation errors and the uncertainty in the fragmentation modeling which increases with increasing P_T of the jets. The dashed blue line in Figure 7.1 shows how the amount of the systematic uncertainty changes with jet P_T for the absolute JES correction. The jet energy corrections simply scale a jet's four-vector by a factor, so its angle is unaffected but its energy and momentum will change.

In order to determine the jet energy scale uncertainty on the pairing and jet charge efficiencies and purities, we shift the jet energies up and down by the total uncertainty (or 1σ) which is calculated by summing the uncertainties from the various JES corrections. Then we recalculate the pairing and jet charge efficiencies and purities for these shifted samples. The results are shown in Table 7.1. We calculate the deviation (in percent) of each shifted sample from the unshifted sample and take the average of the two results as the systematic uncertainty. There is no significant error introduced by the JES on the jet charge efficiency and purity. The uncertainties on the pairing efficiency ($\epsilon_{\text{pairing}}$) and purity ($P_{\text{pairing}}$) are 3.1% and 1% respectively.

Jet energy	$\epsilon_{\text{pairing}}$ %	P_{pairing} %	ϵ_{JQ} %	P_{JQ} %
Nominal	56.3 ± 0.3	84.3 ± 0.3	98.0 ± 0.07	60.7 ± 0.3
$+1\sigma$	57.1 ± 0.3	85.0 ± 0.3	97.95 ± 0.07	60.7 ± 0.3
-1σ	53.6 ± 0.3	83.3 ± 0.3	98.03 ± 0.07	60.8 ± 0.3

Table 7.1. Pairing and jet charge efficiencies and purities obtained by shifting the jet energies up and down by 1σ.
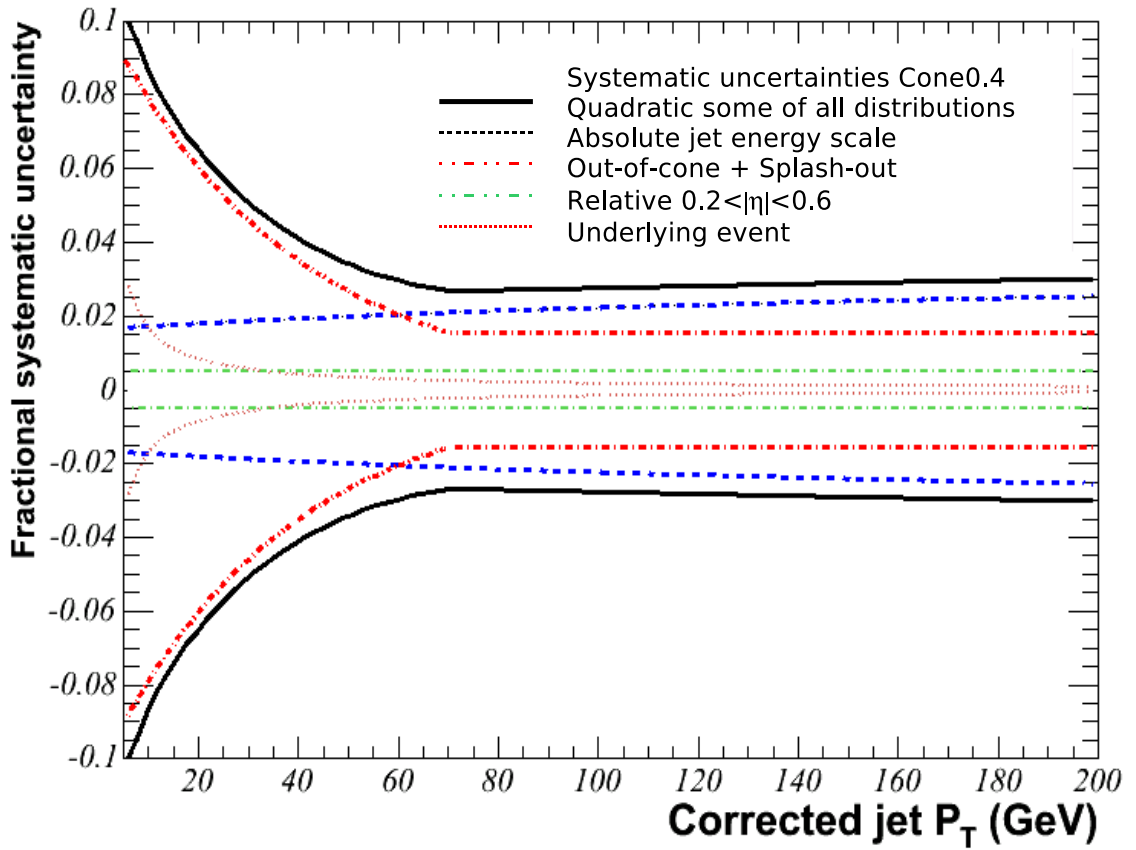
Figure 7.1. The Jet Energy Scale (JES) uncertainties in the central calorimeter ($0.2 < \eta < 0.6$) as a function of a jet's transverse momentum. The level of JES corrections we applied considers only the relative and absolute JES corrections.

7.2 Initial/Final State Radiation

For the top charge measurement, we assumed that the four highest E_T jets in the event were the jets associated with the partons from the top decay. However MC studies indicate that sometimes at least one of the four highest E_T jets does not come from the $t\bar{t}$ decay. In such cases the jet is usually from initial or final state radiation. The initial state radiation (ISR) uncertainty is assigned due to the possibility of picking a gluon or photon jet that was radiated before the $p\bar{p}$ interaction as one of the objects from the top decay. Similarly, gluons or photons radiated from final state particles, called final state radiation (FSR), can enter our selection. Because a gluon jet is totally uncorrelated with the $t\bar{t}$ decay products, ISR/FSR has the potential to degrade the pairing and jet charge purities. We checked how the efficiencies and purities changed if we used a top MC where the amount of ISR (FSR) was changed by varying the QCD scale parameter Λ_{QCD}. We used two different MC samples for ISR (FSR) where one of the MC samples included less ISR (FSR) and the other included more ISR (FSR) compared to the default MC sample (see Table 7.2). The Λ_{QCD} was doubled for the more ISR/FSR and halved for the less ISR/FSR samples. We calculated the percent difference between each shifted sample and the default sample and cite one error (for the ISR/FSR combined) calculated as the sum in quadrature of the largest shifts for ISR and FSR. The ISR/FSR gives an uncertainty of around 1.4% on the jet charge purity but has no significant effect on the jet charge efficiency. The uncertainties on the pairing efficiency and purity are 2.3% and 1.2% respectively.

ISR/FSR	$\epsilon_{\text{pairing}}$ %	P_{pairing} %	ϵ_{JQ} %	P_{JQ} %
Nominal	56.3 ± 0.3	84.3 ± 0.3	98.0 ± 0.1	60.7 ± 0.3
ISR less	55.2 ± 0.5	83.9 ± 0.6	97.9 ± 0.1	59.9 ± 0.6
ISR more	54.7 ± 0.5	85.1 ± 0.5	97.8 ± 0.2	60.8 ± 0.6
FSR less	55.6 ± 0.5	83.7 ± 0.6	98.1 ± 0.2	60.5 ± 0.6
FSR more	56.6 ± 0.6	84.3 ± 0.6	97.8 ± 0.2	60.6 ± 0.6

Table 7.2. Pairing and jet charge efficiencies and purities obtained from top MC samples that contain more or less ISR (FSR) with respect to the default sample.

7.3 Top Mass Uncertainty

The mass of the top quark is used as a constraint in the reconstruction of $t\bar{t}$ events in the kinematic fitter. Our nominal purity and efficiency values were obtained with a top mass constraint of 175 GeV/c^2 on a MC input sample of equal top mass. We looked at the efficiency and purity variations when a different top mass was used as an input to the fitter. We have run on MC samples where the generated top mass was 170 or 180 GeV/c^2, but used a constraint of 175 GeV/c^2 in the fitter. The results are given in Table 7.3. We calculated the difference in purity between the value obtained from the 170 GeV/c^2 sample and the default sample. We did the same with the 180 GeV/c^2 sample and took the average of the two results as the top mass systematic which is 2.4% for the pairing purity (P_{pairing}) and 0% for the jet charge purity (P_{JQ}). The systematic uncertainties on the efficiencies were calculated in the same way and resulted in 1% for the pairing efficiency ($\epsilon_{\mathrm{pairing}}$) and 0% for the jet charge efficiency (ϵ_{JQ}). Considering the latest top mass result [40] and its uncertainty of 170.9 ± 1.8 GeV/c^2, we conclude that 2.4% and 1% for the pairing purity and efficiency are actually conservative results since a 5 GeV difference was used between the top mass value of the default MC sample and the other MC samples chosen to investigate the top mass uncertainty.

Top mass	$\epsilon_{\mathrm{pairing}}$ %	P_{pairing} %	ϵ_{JQ} %	P_{JQ} %
175 Gev/c^2	56.3 ± 0.3	84.3 ± 0.3	98.0 ± 0.1	60.7 ± 0.3
170 Gev/c^2	54.7 ± 0.4	81.0 ± 0.5	97.9 ± 0.1	60.8 ± 0.5
180 Gev/c^2	55.5 ± 0.4	85.0 ± 0.4	98.0 ± 0.1	60.7 ± 0.5

Table 7.3. Pairing and jet charge efficiencies and purities obtained from top MC samples where the generated top mass was changed to 170 and 180 GeV/c^2 but the kinematic fitter was still constrained to 175 GeV/c^2.

7.4 *b*-tagging

In our $t\bar{t}$ MC sample, 25% of the events are double *b*-tagged. However, this percentage is sensitive to the amount of initial and final state radiation. Hard gluon bremsstrahlung either in the initial or final state can produce a jet which can be mistaken for the *b* quark jet. These are called mistagged jets. As explained in Section 3.5.3 the mistag rate is actually a matrix and generally is smaller in MC than that observed in data. This affects the number of events where the assigned *b*-jets are not actually from *b* quarks.

For finding the uncertainty introduced due to errors in *b*-tagging, we applied the mistag matrix on non-heavy flavor jets and assigned those jets as tagged if their mistag probability was greater than a given random probability. This way we increased the light flavor content in the MC. However to get the same composition as in data, we need to decrease the heavy flavor content at the same time. For heavy flavor jets, we made use of the ratio of the *b*-tagging efficiency in MC over the data which is known as the *b*-tagging scale factor. If the jet was tagged we threw a random number that we compared with the *b*-tagging scale factor of 95%. If the random number is smaller than 0.95, we kept the jet as tagged but if it is larger, we declared it as untagged. By applying the mistag matrix to non-heavy flavor jets and the *b*-tag scale factor to heavy flavor jets in our default $t\bar{t}$ MC, we obtained a new sample. We calculated the non-*b* fraction (f_{nonb}) which is the number of double-tagged events where one or more of the tagged jets was not matched to a *b* quark, in both the newly obtained MC sample and the default MC sample. By dividing the f_{nonb} from the new MC by the f_{nonb} from the default MC, we obtained a non-*b* scale factor (SF_{nonb}) of 1.05 ± 0.05. We assigned 100% error on the SF_{nonb} which accounts for a 20% increase on the mistag probability. How the SF_{nonb} is used for the top charge measurement will be covered in Chapter 9 and is not used as a systematic uncertainty here.

7.5 Parton Distribution Functions (PDFs)

As mentioned earlier, quark confinement requires outgoing partons to combine with other quarks in order to form color-neutral hadrons thus making it impossible to observe bare quarks and gluons. However it is possible to construct parton distribution functions that represent the probability density to find partons in a hadron with a certain longitudinal momentum fraction (x) of the proton's momentum and momentum scale (Q^2) which is the square of the momenta involved in the event. These functions are extremely difficult to calculate theoretically but are instead constrained by particle physics experiments. Several sets of parton distribution functions (PDFs) are available such as CTEQ5L [15] and MRST [41]. Our default MC uses the PDF set CTEQ5L. Since different PDFs can affect event kinematics, a set of CTEQ and MRST PDFs are compared to the nominal PDF set and the effect on our efficiencies and purities is taken as a systematic uncertainty. Instead of generating a different set of MC for each PDF set, we reweight one MC sample. To do this, each event in the default MC sample is reweighted by the relative probability of this event in a different PDF set given the event's Q^2, that is the momentum fractions of the interacting partons from the proton and anti-proton. We calculated the efficiencies and purities for each of the PDF sets, compared each to the default CTEQ5L result and added the differences in quadrature. As a result we assign a 1.1% and 0.3% systematic uncertainty for the pairing efficiency and purity respectively. The calculated jet charge efficiency and purity using different PDFs were the same within errors as in the nominal (CTEQ5L) case so no error was assigned.

7.6 Monte Carlo Modeling

We assign an uncertainty to account for different models in different MC generators. For example different parton showering models are used in the PYTHIA and HERWIG

MCs that may affect the efficiency and performance of the analysis method. We compared the pairing and jet charge efficiencies and pairing purity obtained from HERWIG $t\bar{t}$ MC with the results from our default PYTHIA MC sample (see Table 7.4).

MC sample	$\epsilon_{\text{pairing}}$ %	P_{pairing} %	ϵ_{JQ} %
PYTHIA	56.3 ± 0.3	84.3 ± 0.3	98.0 ± 0.1
HERWIG	56.1 ± 0.5	84.6 ± 0.5	97.9 ± 0.1

Table 7.4. Pairing and jet charge efficiencies and pairing purity obtained from PYTHIA (default sample) and HERWIG MC.

We take the percent difference in the pairing efficiency between HERWIG and PYTHIA as the systematic uncertainty on the $\epsilon_{\text{pairing}}$ which is 0.8%. Similarly the percent difference in the pairing purity P_{pairing} is 0.3%. Because we calibrate the jet charge purity in data, no systematic due to the MC generator is assigned for P_{JQ}. The systematics that arise from the calibration of the jet charge purity using data were addressed in Chapter 5.

7.7 Systematics Summary

Table 7.5 summarizes the sources and size of the systematic uncertainties for the top charge measurement in the lepton+jets channel. Out of all the sources of systematic uncertainties we considered, the largest systematic on the pairing efficiency was found to be from the jet energy scale and the next largest from ISR/FSR. We did not assign any systematic uncertainty for the jet charge efficiency since the result was always the same as the nominal within errors. Regarding the pairing purity the largest uncertainty is due to shifting the top mass while the largest uncertainty for the jet charge purity is from ISR/FSR. The combined uncertainties on the efficiencies and purities are calculated by adding the individual uncertainties in quadrature assuming they are uncorrelated. The total uncertainty on the pairing efficiency and purity are

Systematics (in %)	$\epsilon_{\text{pairing}}$	P_{pairing}	ϵ_{JQ}	P_{JQ}
ISR/FSR	2.3	1.2	0	1.4
MC modeling	0.8	0.3	0	-
JES	3.1	1.0	0	0
PDF	1.1	0.3	0	0
Top Mass	1.0	2.4	0	0
total	4.2	2.9	0	1.4

Table 7.5. Summary of the systematic uncertainties in % where ϵ_{JQ} ($\epsilon_{\text{pairing}}$) and P_{JQ} (P_{pairing}) are the efficiency and purity for jet charge (Wb pairing). Because the JQ is calibrated on data we do not assign a systematic uncertainty on P_{JQ} due to different MC modeling effects.

found to be 4.2% and 2.9% respectively, while the total uncertainty on the jet charge purity is 1.4%. How these uncertainties are used to get the uncertainty on the top charge result will be explained in Chapter 9.

CHAPTER 8

Studies in the Dilepton Channel

> What we observe is not nature itself, but nature exposed
> to our method of questioning.
>
> Werner Heisenberg (1901–1976), a German physicist.

8.1   Introduction

As we mentioned earlier, the final CDF top charge measurement is based on the combined result from the lepton+jets and dilepton channels. This chapter briefly describes the studies done in the dilepton channel. As in the lepton+jets channel, the dilepton selection starts with the standard high P_T lepton samples. Because the dilepton events have two leptons and more missing transverse energy from the two neutrinos in the event, the selection requirements are different than the lepton+jets event selection. In summary the dilepton selection consists of two leptons with $E_T > 20$ GeV, two or more jets with $E_T > 15$ GeV, $\vec{\displaystyle{\not}E}_T > 25$ GeV, $H_T > 200$ GeV (where $H_T = P_{T,\text{lepton}} + E_{T,\text{jet}} + \vec{\displaystyle{\not}E}_T$) and a treatment to remove Z boson events. The two highest E_T jets in the event are assumed to be the two b-jets where one of them is tagged using the tight SecVtx algorithm. Details of the dilepton $t\bar{t}$ selection can be found in reference [42].

8.2 Wb **Pairing**

In the dilepton channel, there are 2 b-jets (b_1, b_2) and 2 leptons (l_1, l_2), hence two possibilities for matching the b-jet with the correct lepton (l_1b_1 and l_2b_2 or l_1b_2 and l_2b_1). We calculate the invariant mass of the lepton and b-jet (M_{lb}) for all four lb pairs and make use of the fact that the incorrect pairing is likely to have a large M_{lb}^2 value where M_{lb}^2 is defined as in Equation 8.1.

$$M_{lb}^2 = (E_l + E_b)^2 - (\vec{p}_l + \vec{p}_b)^2 \tag{8.1}$$

As can be seen in Figure 8.1, the M_{lb}^2 distribution for incorrect pairings has a long tail at large M_{lb}^2 while the M_{lb}^2 distribution for correct pairings is populated in the low M_{lb}^2 region. For example, if the M_{lb}^2 value of the l_1b_1 combination is the largest among the four M_{lb}^2 values, then we know that l_1b_1 is most likely an incorrect pairing. So we choose the other combination l_1b_2 which also implies l_2 should be matched to b_1 as the correct combination.

$M_{lb,\mathrm{max}}^2$ cut	$\epsilon_{\mathrm{pairing}}$	P_{pairing}	D	ϵD^2
no cut	1	0.686 ± 0.004	0.4	0.1
5,000	0.990 ± 0.001	0.688 ± 0.004	0.4	0.1
10,000	0.859 ± 0.003	0.722 ± 0.004	0.4	0.2
15,000	0.627 ± 0.004	0.812 ± 0.004	0.6	0.2
22,000	0.369 ± 0.004	0.959 ± 0.003	1.0	0.3
30,000	0.226 ± 0.003	0.994 ± 0.001	1.0	0.2

Table 8.1. List of various $M_{lb,\mathrm{max}}^2$ cuts and their performance for dilepton events with one jet tagged with the tight SecVtx algorithm. Higher performance is indicated by a high ϵD^2 where ϵ is the pairing efficiency ($\epsilon_{\mathrm{pairing}}$) and D ($D = 2 \times P_{\mathrm{pairing}} - 1$) is the dilution.

What about the case where all 4 M_{lb}^2 values are close to each other? Picking the correct lb pair is harder in this case compared to a case where the correct and incorrect pairings have very different M_{lb}^2 values. The distribution for the maximum M_{lb}^2 ($M_{lb,\mathrm{max}}^2$) value of the four invariant mass values as shown in Figure 8.2 can then help us distinguish between the correct and incorrect pairs. As expected $M_{lb,\mathrm{max}}^2$ have
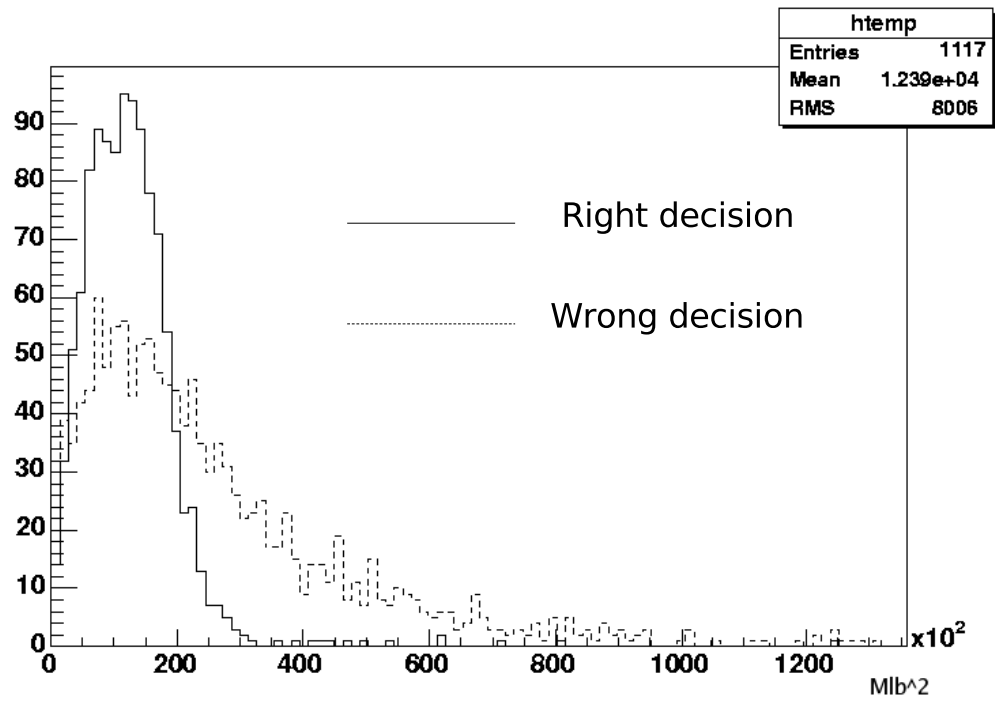
Figure 8.1. M_{lb}^2 distribution for the correct b-lepton pairings (solid) and incorrect b-lepton pairings (dashed).
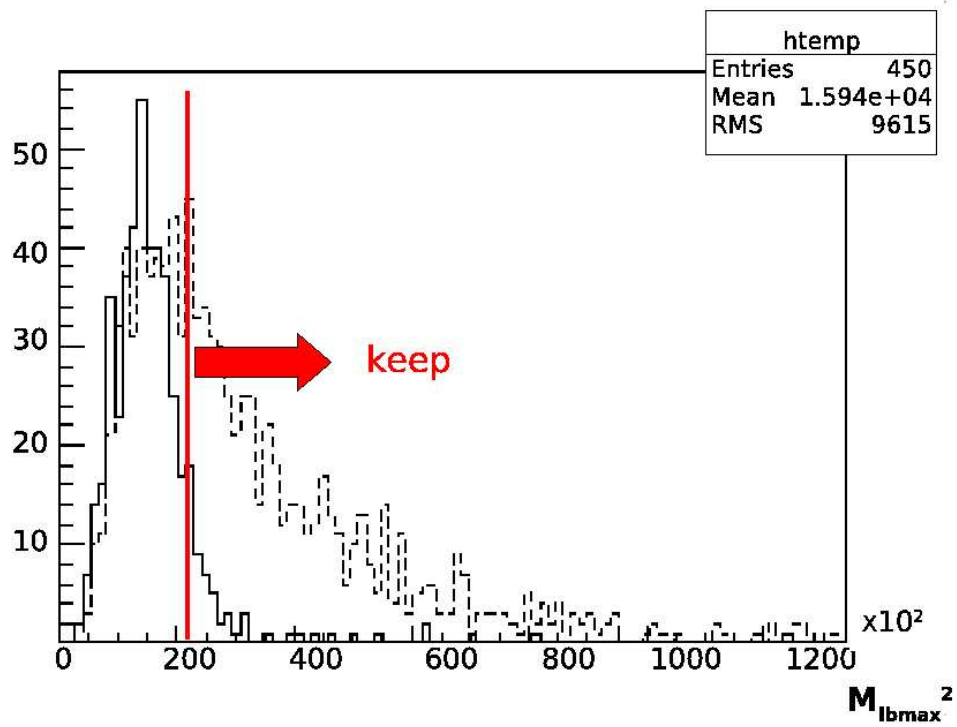
Figure 8.2. $M^2_{lb,\mathrm{max}}$ distribution for which we made the right decision (dashed) and for the wrong decision (solid).

small values when we made the wrong decision and greater values when we made the correct decision. An improvement can be achieved by rejecting events that have $M^2_{lb,\text{max}}$ below a certain value. By removing events with $M^2_{lb,\text{max}} < 22,000$ GeV$^2/c^4$, we keep most of the correctly paired events and get rid of most of the incorrectly paired events. Table 8.1 shows the performance of using different $M^2_{lb,\text{max}}$ cuts. For the events with $M^2_{lb,\text{max}} < 22,000$ GeV$^2/c^4$, we obtained a pairing efficiency of 37% and purity of 96%. The pairing efficiency ($\epsilon_{\text{pairing}}$) is the number of events remaining after the $M^2_{lb,\text{max}}$ cut over the total number of dilepton candidates containing at least one tight SecVtx b-tag. The pairing purity (P_{pairing}) is the ratio of the number of correctly paired events containing at least 1 tight b-tag where the 2 jets are indeed matched to b quarks over the total number of events with $M^2_{lb,\text{max}} > 22,000$ GeV$^2/c^4$.

Sample	$\epsilon_{\text{pairing}}$	P_{pairing}	ϵD^2
1 tight b-tag	0.369 ± 0.004	0.959 ± 0.003	0.311
Pretag	0.374 ± 0.003	0.960 ± 0.002	0.317

Table 8.2. Comparison of pretagged dilepton events with events where one of the b-jets is tagged by the tight SecVtx tagger. Higher performance is indicated by a high ϵD^2 where ϵ is the pairing efficiency, $\epsilon_{\text{pairing}}$, and D ($D = 2 \times P_{\text{pairing}} - 1$) is the dilution. The loss of efficiency due to b-tagging is not included in the tagged case.

As part of the optimization studies, the effect of b-tagging was also investigated. Table 8.2 shows the pairing purity and efficiency for single-tagged and pretagged (before b-tagging) dilepton events. At first glance it looks like b-tagging does not effect the performance since the ϵD^2 values in Table 8.2 are very close. However directly comparing the ϵD^2 values in the table is not the right thing to do. For the correct comparison between the tagged and pretagged samples, we need to consider the effect of b-tagging on the pairing purity and also on the amount of background in the sample. In the dilepton channel after the top charge cuts the background fraction goes from 44% to 0% once one of the b-jets is tagged. Therefore even though the pairing purity for the pretag is found to be 96% in the MC, it is diluted by the

background for which we have a purity of 0.5. In addition, to calculate ϵD^2 we must also consider the b-tagging efficiency. Combining the effect of the $M_{lb,\max}^2$ cut with the b-tagging efficiency, we obtained an ϵD^2 of 0.185 for the single tight SecVtx tagged case. When no b-tagging is used the ϵD^2 is 0.099, which is significantly lower, so we decided to use the single tight SecVtx tagged events for the dilepton channel.

8.3 b Flavor Tagging

For flavor tagging the b-jets we used the same momentum weighted jet charge method as described for lepton+jets in Section 4.4.2.2. The same selections as listed in Section 4.5 were used. The results for the efficiency and purity of the jet charge algorithm with and without b-tagging are given in Table 8.3.

# of b tags	ϵ_{JQ}	P_{JQ}
0	0.767 ± 0.002	0.598 ± 0.003
1	0.863 ± 0.002	0.61 ± 0.003

Table 8.3. Purity (P_{JQ}) of the momentum weighted jet charge along with the efficiency (ϵ_{JQ}) using the dilepton events from $t\bar{t}$ MC.

8.4 Backgrounds

The expected number of events for the backgrounds in the dilepton channel were already calculated for the CDF top cross-section measurement [42] and are given in Table 8.4 for 955 pb^{-1} of data. Remember that for the top charge measurement we also need to know if any of the backgrounds favor one hypothesis (Standard Model or exotic model) over the other.

As we did in the lepton+jets channel, for each dilepton background we measured the background purity (P_b) defined as the number of events that look like Standard Model over the total number of events (see Equation 6.2). A purity of 0.5 implies the

background	DIL σ prediction	N_b
Drell-Yan	$0.38^{+0.76}_{-0.38}$	$0 + 0.38$
Fakes	$0.81^{+1.62}_{-0.81}$	$0 + 0.81$
Diboson	0.0 ± 0.1	$0 + 0.1$
Total	$1.19^{+1.79}_{-0.9}$	$0 + 0.9$

Table 8.4. Table of expected background events in the dilepton (DIL) channel for 955 pb^{-1} of data. The second column gives the prediction for each background taken from the CDF dilepton cross-section (σ) measurement where one of the b-jets is tight SecVtx tagged. After the top charge analysis cuts, the expected number of background events (N_b) is 0.

background is SM like 50% of the time and exotic model like 50% of the time. There are three backgrounds in the dilepton channel which are described below:

- Drell-Yan

  One of the dominant background processes in the dilepton channel are Drell-Yan events where a Z boson decays to two leptons. We used ALPGEN+HERWIG samples that include $Z, \gamma \to l\bar{l}$ and within the poor statistics obtained after cuts, we calculated a jet charge purity of 0.47 ± 0.03.

- Fakes

  These are the type of events where one or more of the jets are misidentified as leptons in the detector, called fakes. We ran the top charge dilepton analysis code on $W+3$ jets MC samples and obtained a purity of 0.49 ± 0.08, meaning no charge asymmetry was seen within its uncertainty.

- Diboson (WW, WZ)

  WW and WZ events can pass the dilepton selection cuts if both bosons decay leptonically and the leptons are chosen as the two leptons in the dilepton top event. When the dilepton top charge selection cuts were applied to WZ MC events, no events were left. In order to gain more events, we dropped the requirement of removing Z boson events and obtained a purity of 0.50 ± 0.02.

Using a WW MC sample, we found a purity of 0.48 ± 0.02. These results are consistent with 0.5 confirming that there is no charge asymmetry.

Table 8.5 summarizes the charge asymmetry study on the dilepton background events. No charge asymmetry is found.

Background	P_{JQ}	ϵ_{total}
Drell-Yan	0.47 ± 0.03	0.64 ± 0.04
Fakes	0.49 ± 0.08	0.41 ± 0.05
WW	0.48 ± 0.02	0.27 ± 0.01
WZ	0.50 ± 0.02	0.15 ± 0.01

Table 8.5. Jet charge purity (P_{JQ}) for dilepton backgrounds and the efficiency ($\epsilon_{\text{total}} = \epsilon_{\text{pairing}} \times \epsilon_{JQ}$) for accepting each background.

8.5 Systematics

The sources of systematic uncertainties in the dilepton channel are the same (except for one) as in the lepton+jets channel and are calculated in the same way. The error on the event predictions obtained from the dilepton cross-section analysis [42] include an uncertainty due to acceptance effects. Apart from the effect of systematic uncertainties on the dilepton acceptance, we studied how the systematic uncertainties affected the Wb pairing and jet charge efficiencies and purities.

There is one additional source of uncertainty in the dilepton channel compared to the lepton+jets channel. This is due to the polarization of the W boson which refers to the direction of the spin of the W boson with respect to its direction of motion. The angle (θ^*) between the lepton in the W rest frame and the W direction in the top rest frame is used for the W polarization studies. The relation of θ^* with different W polarization states is depicted in Figure 8.3 where the bold blue arrows represent the spins of fermions and the bold red arrow represents the spin of the W boson. The angle θ^* is a function of the invariant mass of the ℓ and b (M_{lb}) and since we used M_{lb}^2 to pick the correct lb pair, we need to assign an uncertainty due to different W
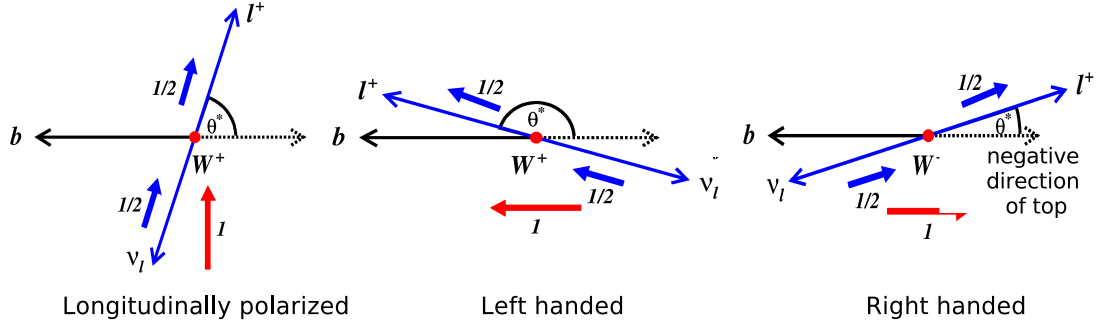
Figure 8.3. θ^* for different W polarization states.

Systematics (in %)	$\epsilon_{\text{pairing}}$	P_{pairing}	ϵ_{JQ}	P_{JQ}
ISR/FSR	4.6	1.0	1.2	1.9
MC generator	0	0	1.0	-
JES	4.2	0.8	0.4	0
PDF	4.1	0.3	0	0
top mass	7	2	0	0
W polarization	1.5	0	0	0
total	**10.3**	**2.4**	**1.6**	**1.9**

Table 8.6. Summary of systematics uncertainties (in %) in Dilepton channel.

polarizations. In the Standard Model, the W is expected to be longitudinally polarized 70% of the time and left-handed 30% of the time. We varied the fraction of the longitudinally polarized Ws from 70% down to 65% and also up to 75% in the $t\bar{t}$ MC and checked the effect on the pairing purity and efficiency. Varying the longitudinal polarization fraction by 5% caused a 1.5% increase in the pairing efficiency and did not affect the purity.

The summary of the systematic uncertainties in the dilepton channel is given in Table 8.6 and corresponds to the systematics discussed in Chapter 7. The total systematic uncertainties on $\epsilon_{\text{pairing}}$, P_{pairing}, ϵ_{JQ} and P_{JQ} are 10.3%, 2.4%, 1.6% and 1.9% respectively. How these are used in the final top charge result is explained in Chapter 9.

CHAPTER 9

Signal and Background Estimates

> Knowledge is to comprehend knowledge
> Comprehension is to know yourself
> if you don't know yourself
> What is the point of your studies?
> Yunus Emre (1238-1320), a Turkish Poet.

9.1 Expected Number of Events

This chapter describes how we calculate the expected number of signal (N_s) and background events (N_b) that will be used for the top charge result as well as how we calculate the total signal purity (P_s) and total background purity (P_b).

Using 955 pb^{-1} of data for the dilepton channel and 695 pb^{-1} of data for the lepton+jets channel, we calculated the number of signal and background events we expect for the top charge analysis. These are calculated by multiplying the predicted number of events in each channel obtained from the σ analyses [29, 42] by the total efficiency (ϵ_{total}) which is the product of the pairing and jet charge efficiencies. The expected number of events for each background was already presented in Chapter 6 for the lepton+jets channel and Chapter 8 for the dilepton channel. Here we summarize them again along with the expected number of signal events in Table 9.1. We expect $35.4 \pm 0.4(\text{stat})\pm 7.9(\text{sys})$ lepton+jets events and $10.3 \pm 0.4(\text{stat})\pm 1.7(\text{sys})$ dilepton

background	prediction	$\epsilon_{\text{total}} = \epsilon_{\text{pairing}} \cdot \epsilon_{JQ}$	N_{s} or N_{b}
Lepton+Jets (695 pb^{-1})			
W+HF	5.3 ± 1.2	0.15 ± 0.01	0.8 ± 0.2
QCD fakes	1.8 ± 1.0	0.2 ± 0.1	0.4 ± 0.3
Diboson	0.4 ± 0.01	0.11 ± 0.05	0.04 ± 0.02
Mistag	1.32 ± 1.19	0.08 ± 0.03	0.1 ± 0.1
Single Top	0.77 ± 0.06	0.16 ± 0.01	0.13 ± 0.01
Total	9.6 ± 2.0	-	1.4 ± 0.3
Signal	$64.9^{\pm 0.6(\text{stat})}_{\pm 14.3(\text{sys})}$	$0.545^{\pm 0.002(\text{stat})}_{\pm 0.023(\text{sys})}$	$35.35^{\pm 0.37(\text{stat})}_{\pm 7.92(\text{sys})}$
Dilepton (955 pb^{-1})			
Drell-Yan	$0.38^{+0.76}_{-0.38}$	-	$0 + 0.38$
Fakes	$0.81^{+1.62}_{-0.81}$	-	$0 + 0.81$
Diboson	0.0 ± 0.1	-	$0 + 0.1$
Total	$1.19^{+1.79}_{-0.9}$	-	$0 + 0.9$
Signal	$32.4^{\pm 1.3(\text{stat})}_{\pm 4.0(\text{sys})}$	$0.318^{\pm 0.004(\text{stat})}_{\pm 0.033(\text{sys})}$	$10.30^{\pm 0.43(\text{stat})}_{\pm 1.66(\text{sys})}$
Dilepton and Lepton+Jets Combined			
Total Background			$1.40^{+0.96}_{-0.34}$
Total Signal			$45.7^{\pm 0.6(\text{stat})}_{\pm 8.1(\text{sys})}$

Table 9.1. Table of expected signal and background events. The second column shows the predicted number of events that are taken from the cross-section measurements. The lepton+jets predictions include the efficiency of the lepton+jets event selection with both b-jets tagged with the loose SecVtx algorithm. The dilepton predictions include the efficiency of the dilepton event selection with one of the b-jets tagged with the tight SecVtx algorithm.

events, for a total of 45.7 ± 0.6(stat)± 8.1(sys) events for the top charge measurement. In addition we expect to have $1.40 + 0.96 - 0.34$ background events, all of which are coming from the lepton+jets channel since the background is reduced to zero in the dilepton channel after requiring one b-tag. The error on the background prediction is statistical only. Since the amount of background is so small and does not affect the top charge result we did not calculate the systematic uncertainties on it. The second column in Table 9.1 shows the predictions from the σ analysis for the number of background and signal events. The third column is the efficiency for accepting background events due to the top charge analysis cuts. The numbers in the fourth column show the expected number of events for the backgrounds (N_{b}) and signal (N_{s}) obtained by multiplying the second and third columns.

9.2 Calculating Signal Purity

The calculation of the signal purity (P_s) is not as straightforward as calculating the number of signal events. The starting point is to multiply the pairing and jet charge purities ($P_\text{pairing} \cdot P_{JQ}$) to get a combined purity. However, this product gives the fraction of events with the correct top charge only if the methods used for Wb pairing and b-jet flavor tagging are both correct for the event. If one of the methods fails we get an incorrect top charge. However if both methods fail, we actually get the correct top charge. For example, assume we have a Standard Model $t\bar{t}$ event, so $t \to W^+b$ and $\bar{t} \to W^-\bar{b}$. If our pairing method incorrectly matches W^+ to the $\bar{b}$ but then the flavor tagging method assigns the wrong flavor, declaring the $\bar{b}$ a b instead, we actually get the same result ($+2/3$) as expected for the Standard Model case. So considering the case where a correct result is obtained, one can write P_s as the combination of two pieces, the first for when both methods (pairing and jet charge) are correct and the second for when both methods are incorrect, as shown in Equation 9.1.

$$P_s = P_\text{pairing} \cdot P_{JQ} + (1 - P_\text{pairing}) \cdot (1 - P_{JQ}) \tag{9.1}$$

Yet Equation 9.1 is neither complete nor correct. Even though we used b-tags in the events, we still have events where the selected b-jets are not actually coming from b quarks. There are only a few of these events but even so we must consider their effect on the purity. Defining f_nonb as the fraction of events where one or more of the jets tagged as b's are not actually matched to b quarks, P_s can be expressed as

$$P_s = f_\text{nonb} \cdot 0.5 + (1 - f_\text{nonb}) \cdot (P_\text{pairing} \cdot P_{JQ} + (1 - P_\text{pairing}) \cdot (1 - P_{JQ})) \tag{9.2}$$

Note that only the events for which the jets are real b-jets contribute to the P_s in Equation 9.1. The non-b jets are incorrect pairings and have no sign correlation with the associated lepton, so they contribute to P_s with a purity of 0.5 as shown in the

176

	Dilepton	Lepton+Jets
f_{nonb}	0.075 ± 0.001	0.055 ± 0.001
SF_{nonb}	1.05 ± 0.05	1.05 ± 0.05
P_{pairing}	$0.959 \pm 0.003(\text{stat}) \pm 0.013(\text{sys})$	$0.844 \pm 0.003(\text{stat}) \pm 0.024(\text{sys})$
P_{JQ}	$0.603 \pm 0.005(\text{stat}) \pm 0.011(\text{sys})$	$0.608 \pm 0.003(\text{stat}) \pm 0.009(\text{sys})$
SF_{JQ}	$1.03 \pm 0.02(\text{stat}) \pm 0.04(\text{sys})$	$1.03 \pm 0.02(\text{stat}) \pm 0.04(\text{sys})$

Table 9.2. All of the ingredients, fraction of pairs with no b-jets (f_{nonb}), correction due to the mistag rate difference between MC and data (SF_{nonb}), purity of pairing (P_{pairing}), purity of jet charge (P_{JQ}) and its scale factor (SF_{JQ}), to correct the P_{JQ} obtained from MC for the dilepton and lepton+jets channels.

first term of Equation 9.2. The second term is from Equation 9.1 but multiplied by the fraction of real b events $(1 - f_{\text{nonb}})$.

Equation 9.2 now has all the pieces but is still not accurate. Remember that we do not rely on the MC to give us the correct b and non-b fractions. In Chapter 7, we discussed the uncertainty introduced due to b-tagging and calculated a non-b scale factor, SF_{nonb}. To get the correct fraction of non-b events (f_{nonb}) we must multiply it by the SF_{nonb} as in Equation 9.3.

$$P_s = f_{\text{nonb}} \cdot SF_{\text{nonb}} \cdot 0.5 + (1 - f_{\text{nonb}} \cdot SF_{\text{nonb}}) \cdot (P_{\text{pairing}} \cdot P_{JQ} + (1 - P_{\text{pairing}}) \cdot (1 - P_{JQ})) \quad (9.3)$$

In addition we do not rely on the jet charge purity from MC but instead did a calibration study in data as described in Chapter 5. This resulted in a scale factor for the jet charge method of $SF_{JQ} = 1.03 \pm 0.02(\text{stat}) \pm 0.04(\text{sys})$. We must also correct the jet charge purity (P_{JQ}) by the scale factor SF_{JQ} which causes Equation 9.3 to become:

$$P_s = f_{\text{nonb}} \cdot SF_{\text{nonb}} \cdot 0.5 + (1 - f_{\text{nonb}} \cdot SF_{\text{nonb}}) \cdot (P_{\text{pairing}} \cdot P_{JQ} \cdot SF_{JQ} + (1 - P_{\text{pairing}}) \cdot (1 - P_{JQ} \cdot SF_{JQ}))$$

$$(9.4)$$

.

Putting all the ingredients from Table 9.2 into Equation 9.4, we measure the signal purity (P_s) to be $0.58 \pm 0.01(\text{stat}) \pm 0.02(\text{sys})$ for the lepton+jets channel and

177

Channel	$P_{\rm s}$	N_+	N_-
Lepton+jets	$0.58^{\pm0.01(\text{stat})}_{\pm0.02(\text{sys})}$	$20.6^{\pm0.4(\text{stat})}_{\pm4.7(\text{sys})}$	$14.8^{\pm0.3(\text{stat})}_{\pm3.4(\text{sys})}$
Dilepton	$0.60^{\pm0.01(\text{stat})}_{\pm0.02(\text{sys})}$	$6.2^{\pm0.3(\text{stat})}_{\pm1.0(\text{sys})}$	$4.1^{\pm0.2(\text{stat})}_{\pm0.7(\text{sys})}$
Total	$0.59^{\pm0.01(\text{stat})}_{\pm0.02(\text{sys})}$	$26.8^{\pm0.5(\text{stat})}_{\pm4.8(\text{sys})}$	$18.9^{\pm0.4(\text{stat})}_{\pm3.4(\text{sys})}$

Table 9.3. Signal purity for each channel separately and both channels combined (total) and the number of SM like (N_+) and exotic model like (N_-) events.

$0.60 \pm 0.01(\text{stat}) \pm 0.02(\text{sys})$ for the dilepton channel. Combining both channels and taking into account the number of events from each channel, the total signal purity is $P_s = 0.59 \pm 0.01(\text{stat}) \pm 0.02(\text{sys})$. Knowing the signal purity and expected number of signal events we can further calculate how many Standard Model like events (N_+) and how many exotic quark like events (N_-) are expected. We expect a total of $N_+ = 26.8 \pm 0.5(\text{stat}) \pm 4.8(\text{sys})$ and $N_- = 18.9 \pm 0.4(\text{stat}) \pm 3.4(\text{sys})$ events. These results are shown in Table 9.3.

9.3 Calculating Background Purity

We have already presented the purity for each background in the lepton+jets channel in Chapter 6. For the dilepton channel we have no background left after the top charge selection cuts and none of the backgrounds showed a charge asymmetry towards SM or exotic model events, so we use a total purity of 0.5 ± 0.0 for the dilepton backgrounds. In this section we will describe how we combine the background purities for the lepton+jets channel.

The total background purity ($P_{\rm b,total}$) can be calculated by summing the N_+ values for each background and dividing the result by the total number of events as shown in Equation 9.5 where the sum is taken over all backgrounds.

$$P_{\rm b,total} = \frac{\sum_{\rm i}(N_+)_{\rm i}}{\sum_{\rm i}(N_+)_{\rm i} + \sum_{\rm i}(N_-)_{\rm i}} \tag{9.5}$$

We can not directly use Equation 9.5 because there are two asymmetric backgrounds

background	P_{b}	N^+	N^-
Lepton+Jets (695 pb^{-1})			
W+HF	0.5 ± 0.0	0.38 ± 0.09	0.38 ± 0.09
QCD fakes	$0.55^{+0.01}_{-0.05}$	$0.2^{+0.1}_{-0.1}$	$0.1^{+0.1}_{-0.1}$
Diboson	0.5 ± 0.0	0.02 ± 0.01	0.02 ± 0.01
Mistag	0.5 ± 0.0	0.06 ± 0.05	0.06 ± 0.05
Single Top	0.58 ± 0.02	0.07 ± 0.01	0.05 ± 0.01
Total	$0.52^{+0.01}_{-0.01}$	0.73 ± 0.18	0.68 ± 0.16
Dilepton (955 pb^{-1})			
Drell-Yan	0.5 ± 0.0	$0 + 0.4$	$0 + 0.4$
Fakes	0.5 ± 0.0	$0 + 0.8$	$0 + 0.8$
Diboson	0.5 ± 0.0	$0 + 0.1$	$0 + 0.1$
Total	0.5 ± 0.0	$0 + 0.9$	$0 + 0.9$
Total Background	$0.52 \pm 0.01(\mathrm{stat})$	$0.7^{+0.9}_{-0.2}$	$0.7^{+0.9}_{-0.2}$

Table 9.4. Background purity for each channel separately and both channels combined and the number of SM like (N_+) and exotic model like (N_-) events expected.

that have a purity higher than 0.5. These are the QCD and single top backgrounds as can be seen in Table 9.4. For the backgrounds which we did not expect any charge asymmetry and the measured P_{b} agreed with 0.5 within its uncertainty, we use a P_{b} of 0.5 ± 0.0 when calculating $P_{\mathrm{b,total}}$. In order to find $P_{b,total}$ we need an equation that combines the asymmetric and symmetric parts of the backgrounds correctly. Let N_1 and N_2 represent the amount of the two asymmetric backgrounds (QCD, single top) and N_3 the total amount of the symmetric backgrounds ((W+HF) + Diboson + Mistag). $P_{\mathrm{b,total}}$ can then be expressed as in Equation 9.6.

$$P_{\mathrm{b,total}} = \frac{N_1 \cdot f_1 + N_2 \cdot f_2 + N_3 \cdot f_3}{N_1 + N_2 + N_3} \tag{9.6}$$

where f_1, f_2 and f_3 are the fractions of Standard Model like events for QCD, single top and the symmetric backgrounds respectively. Using Equation 9.6 we measure the total background purity ($P_{\mathrm{b,total}}$) as $0.52 \pm 0.01(\mathrm{stat})$.

N_s	$91.3 \pm 1.1(\text{stat})\pm16.2(\text{sys})$
N_b	$2.8 \pm 1.9(\text{stat})$
P_s	$0.59 \pm 0.01(\text{stat})\pm0.02(\text{sys})$
$P_{b,\text{total}}$	$0.52 \pm 0.01(\text{stat})$

Table 9.5. Expected number of background and signal pairs together with the corresponding purities.Since the amount of background is so small and does not affect the top charge result the effect of systemmatic uncertainties were not checked on backgrounds. The uncertainty on N_b and P_b are statistical only.

9.4   Summary

Table 9.5 shows the total signal and background estimates for the top charge measurement. Note that the estimates in Table 9.5 are twice of what was shown in Table 9.1. This is because we have two top charge measurements per $t\bar{t}$ event, one from the t decay side and another from the $\bar{t}$ decay side. The calculation of the combined signal purity (P_s) was more complicated. While obtaining P_s we considered that if the pairing is wrong and if the jet charge is also wrong we still get the same answer as having everything correct. There is also a small probability that the b-jets were misidentified, in which case they will have a random charge correlation with the lepton. The jet charge purity was also corrected by a scale factor obtained from data. The calculation of the total background purity ($P_{b,\text{total}}$) was based on combining the charge symmetric and asymmetric parts of the background correctly where 0.5 ± 0.0 was used for the purity of the charge symmetric backgrounds. The results for P_s and $P_{b,\text{total}}$ are summarized in Table 9.5. We expect $91.3\pm1.1(\text{stat})\pm16.2(\text{sys})$ signal events and $2.8\pm1.9(\text{stat})$ background events. We expect to classify the signal events (as SM or exotic model like) correctly $59\pm1(\text{stat})\pm2(\text{sys})\%$ of the time. In addition we expect the background events to look like SM top events $52\pm1\%$ of the time.

CHAPTER 10

Statistical Treatment

> The null hypothesis is never proved or established, but
> is possibly disproved, in the course of experimentation.
> Every experiment may be said to exist only to give the facts
> a chance of disproving the null hypothesis.
>
> R. A. Fisher (1890–1962), an English statistician.

For the measurement of the top quark's charge, we count the number of Wb pairs that support the Standard Model (SM) and those that support the exotic model (XM). A positive lepton combined with a b-jet or a negative lepton combined with a $\bar{b}$-jet are assigned as a SM-like pair, whereas a positive lepton combined with a $\bar{b}$-jet or a negative lepton combined with a b-jet are assigned as a XM-like pair. The total number of pairs that are SM-like is defined as N_+ while the total number of pairs that support the XM is defined as N_-. Once we obtain N_+ and N_- from the data, we can compare the measurement with the Standard Model expectation and quantify the degree of evidence in favor of the SM over the XM. In this pursuit, we would like to use a parameter of interest that has quite different probability density functions between the two hypotheses. We have chosen the fraction (f_+) of events following the SM as our parameter of interest and it is expressed as in Equation 10.1.

$$f_+ = \frac{N_+}{N_+ + N_-} \tag{10.1}$$

In our MC, f_+ should be equal to 1 since the MC was generated according to the SM. However, due to the limitations in our methods for pairing and b-flavor tagging this is not always the case, but we still expect f_+ to be close to 1 if the result is consistent with the SM given the performance of our methods (refer to Section 4.2 for more details).

If we had an infinite amount of statistics our result would be the measured value of f_+. However with finite statistics we can just find the best estimation for f_+ and compare it with the SM expectation. In order to do this, the Particle Data Book [43] advocates two approaches: the Bayesian and frequentist. We have exploited both approaches for this result and they are described below.

10.1 Profile Likelihood

A common method used in high energy physics to find the most likely value of a parameter of interest is a maximum likelihood approach [44, 45] that selects the hypothesis (H) which maximizes the conditional probability of a set of observations X assuming H is true $P(X|H)$. The maximum likelihood is calculated based on the distribution of the parameter of interest, in our case f_+, as shown in Equation 10.2

$$L(f_+|X) = \prod_{i=1}^{n} P(X_i|f_+) \tag{10.2}$$

where we have n independent observations $X = (X_1, X_2, .., X_n)$. The top charge measurement, like most other measurements, involves not only a parameter of interest but also nuisance parameters that are incompletely known and add to the uncertainty of the parameter of interest. In our case, we have 4 nuisance parameters:

N_s: Number of signal events

P_s: Purity of the signal

N_b: Number of background events

P_b: Purity of the background

With the nuisance parameters, $\theta = (N_s, P_s, N_b, P_b)$, the likelihood function can be written as in Equation 10.3 where the maximum occurs when the unknown nuisance parameters are replaced by their maximum likelihood estimates.

$$L(f_+, \theta | X) = \prod_{i=1}^{n} P(X_i | f_+, \theta) \tag{10.3}$$

If the likelihood function is calculated at specific values of $f_+ = f_{+0}$ and then divided by the maximum likelihood where f_+ is not restricted to a specific value we get a likelihood ratio (Equation 10.4). The maximum likelihood estimates are substituted for both f_+ and the nuisance parameters for the denominator. In the context of nuisance parameters the function λ is also called the "profile likelihood" [46].

$$\lambda(f_{+0} | X) = \frac{max(L(f_{+0}, \theta | X); \theta)}{max(L(f_+, \theta | X); f_+, \theta)} \tag{10.4}$$

Note that the denominator is just a number obtained by fitting f_+ along with all the nuisance parameters. The advantage of a profile likelihood is that all the nuisance parameters are eliminated from the likelihood function and the likelihood is just a function of f_+. Further taking $-2Ln\lambda$ corresponds to a χ^2 distribution [46]. Larger λs correspond to smaller χ^2s. So the minimum point of the $-2Ln\lambda$ distribution with respect to f_+ (see Figure 10.1 as an example) gives the best value of f_+ for which the nuisance parameters have their best estimates. For more details on profile likelihood see reference [46].

10.1.1   Nuisance Parameters

As mentioned earlier, we have 4 nuisance parameters: the number of signal events (N_s), the purity of the signal (P_s), the number of background events (N_b) and the purity of the background (P_b). We have already given the expected number for each

of these nuisance parameters in Table 9.5. Here I will remind you how each of these parameters is calculated and show the relation of the nuisance parameters with our parameter of interest (f_+) which is the fraction of SM-like Wb pairs.

The number of expected signal events (N_s) is the number of events left after the Wb pairing (PR) and flavor tagging (FT) criteria are applied. So N_s includes the efficiency of the PR and FT methods in it. N_b is the amount of all the backgrounds added together and it also has the effect of the efficiency of the PR and FT methods included in it. P_s is the signal purity calculated as described in Section 9.2. P_b is the probability that all the backgrounds look like the Standard Model, so it is a combination of all the individual purities from the individual backgrounds. This is all summarized in the following table.

N_s	expected number of signal events after PR and FT
N_b	expected number of total background events after PR and FT
P_s	expected purity of the PR and FT methods on the signal
P_b	expected charge asymmetry for the background

The f_+ shown in Equation 10.1 can now be written as a function of these nuisance parameters. This is illustrated in the following equations where $< N_+ >$ and $< N_- >$ are the means of the Poisson distributed N_+ and N_-.

$$< N_+ > \ = \ N^+ f_+ + N^- (1 - f_+) + N_b^+ \tag{10.5}$$

$$< N_- > \ = \ N^- f_+ + N^+ (1 - f_+) + N_b^- \tag{10.6}$$

$$\text{where :} \tag{10.7}$$

$$N^+ \ = \ P_s N_s \tag{10.8}$$

$$N^- \ = \ (1 - P_s) N_s \tag{10.9}$$

$$N_b^+ \ = \ P_b N_b \tag{10.10}$$

$$N_b^- \ = \ (1 - P_b) N_b \tag{10.11}$$

10.1.2 Likelihood Expression

Our likelihood expression has five terms:

$$L = L_{\rm s} \cdot L_{\rm b} \cdot L_{\sigma_s} \cdot L_{P_{\rm s}} \cdot L_{P_{\rm b}} \tag{10.12}$$

A signal part ($L_{\rm s}$) which is Poisson distributed, the background ($L_{\rm b}$) which is Gaussian distributed, a term related to the uncertainty on the number of signal events (L_{σ_s}) which is also Gaussian distributed, plus purities for the signal ($L_{P_{\rm s}}$) and background ($L_{P_{\rm b}}$) which are Gaussian distributed. The signal term ($L_{\rm s}$) can be expressed as:

$$L_{\rm s} = L_{{\rm s}_+} \cdot L_{{\rm s}_-} = \frac{<N_+>^{x^+} e^{(-<N_+>)}}{x^+!} \cdot \frac{<N_->^{x^-} e^{(-<N_->)}}{x^-!} \tag{10.13}$$

where x_+ and x_- are the number of events following the $+2/3$ and $-4/3$ hypotheses in the data, respectively.

The second term in Equation 10.12 deals with the background component and can be written as:

$$L_{\rm b} = \frac{1}{\sigma_{N_{\rm b}}\sqrt{2\pi}} e^{-\frac{(y_{\rm b}-N_{\rm b})^2}{2\sigma_{N_{\rm b}}^2}} \tag{10.14}$$

where $N_{\rm b}$ and $\sigma_{N_{\rm b}}$ are the number of background and its uncertainty from the background prediction studies. $y_{\rm b}$ is chosen from a Gaussian distribution whose mean is $N_{\rm b}$. The third term of Equation 10.12 considers the uncertainty on the signal:

$$L_{\sigma_s} = \frac{1}{\sigma_{N_{\rm s}}\sqrt{2\pi}} e^{-\frac{(y_{\rm s}-N_{\rm s})^2}{2\sigma_{N_{\rm s}}^2}} \tag{10.15}$$

where $N_{\rm s}$ and $\sigma_{N_{\rm s}}$ are the expected numbers for the number of signal and its uncertainty respectively. $y_{\rm s}$ is chosen from a Gaussian distribution whose mean is $N_{\rm s}$.

The fourth and fifth parts in Equation 10.12 are the the purity terms for signal and background respectively:

$$L_{P_{\mathrm{s}}} = \frac{1}{\sigma_{P_{\mathrm{s}}}\sqrt{2\pi}} e^{-\frac{(z_{P_{\mathrm{s}}}-P_{\mathrm{s}})^2}{2\sigma_{P_{\mathrm{s}}}^2}} \qquad (10.16)$$

$$L_{P_{\mathrm{b}}} = \frac{1}{\sigma_{P_{\mathrm{b}}}\sqrt{2\pi}} e^{-\frac{(z_{P_{\mathrm{b}}}-P_{\mathrm{b}})^2}{2\sigma_{P_{\mathrm{b}}}^2}} \qquad (10.17)$$

where P_{s} (P_{b}) and $\sigma_{P_{\mathrm{s}}}$ ($\sigma_{P_{\mathrm{b}}}$) are the measured signal (background) purity and the error on it. $z_{P_{\mathrm{s}}}$ ($z_{P_{\mathrm{b}}}$) is chosen from a Gaussian distribution whose mean is P_{s} (P_{b}).

10.1.3 Calculation of the Profile Likelihood

To solve for f_+, we must find the minimum of the total likelihood expression given in Equation 10.12 which is the product of all the likelihood expressions from Equation 10.13 to 10.19. To do this, we take the partial derivatives of the total likelihood expression with respect to each nuisance parameter and set that equation equal to zero acquiring a system of nonlinear equations.

$$\frac{\partial L}{\partial N_{\mathrm{s}}} = 0 \qquad (10.18)$$

$$\frac{\partial L}{\partial P_{\mathrm{s}}} = 0 \qquad (10.19)$$

$$\frac{\partial L}{\partial N_{\mathrm{b}}} = 0 \qquad (10.20)$$

$$\frac{\partial L}{\partial P_{\mathrm{b}}} = 0 \qquad (10.21)$$

Since the computations are difficult to perform by hand, appropriate software has been implemented within the MINUIT package[31], which is a widely used parameter fitting program in high energy physics.

As mentioned earlier, the likelihood curve (the numerator of Equation 10.4) is calculated by scanning through different values of f_+ and at each value we let MI-
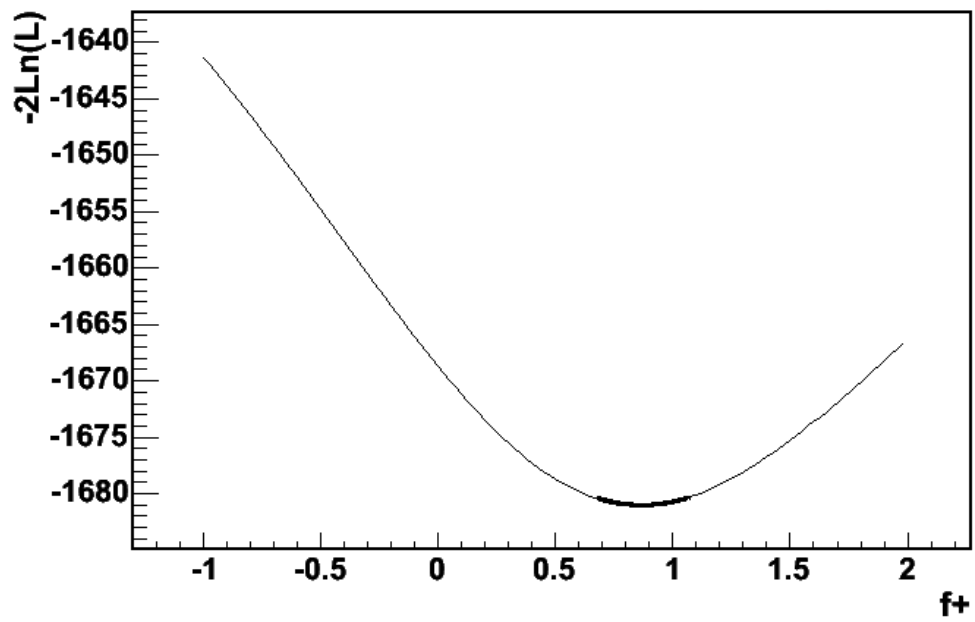
Figure 10.1. The $-2Ln\lambda$ curve obtained from one pseudo-experiment assuming the SM is true.

NUIT minimize the likelihood for all the nuisance parameters. Figure 10.1 shows the $-2Ln\lambda$ curve assuming the SM is true obtained from one pseudo-experiment. A pseudo-experiment is constructed by randomly generating N_+, N_- and the nuisance parameters given the expected values obtained for each from MC studies. In other words, we generate fake data based on our predictions. Below is the list of variables and how we simulate them in pseudo-experiments:

x^+: draw a random number from a Poisson distribution whose mean is N^+.

x^-: draw a random number from a Poisson distribution whose mean is N^-.

y_b: draw a random number from a Gaussian distribution whose mean and σ are the expected number of background events (N_b) and its error (σ_{N_b}).

y_s: draw a random number from a Gaussian distribution whose mean and σ are the expected number of signal events (N_s) and its error (σ_{N_s})

z_{P_s}: draw a random number from a Gaussian distribution whose mean and σ are the expected signal purity (P_s) and its error (σ_{P_s}).

z_{P_b}: draw a random number from a Gaussian distribution whose mean and σ are the expected background purity (P_b) and its error (σ_{P_b}).

If we repeat the above procedure 100,000 times and pick the minimum of the likelihood curve as f_+ and make a distribution of these values, we get the red curve in Figure 10.2. This is the f_+ curve assuming the SM is true and it peaks at 1 as expected. The black curve is generated in the same way but assuming the XM is true ($f_+ = 0$).

10.2 Extracting a Limit

Once we have obtained the N_+ and N_- values from our data and used them to draw the $-2Ln\lambda$ curve, the minimum of the curve is the maximum likelihood estimator
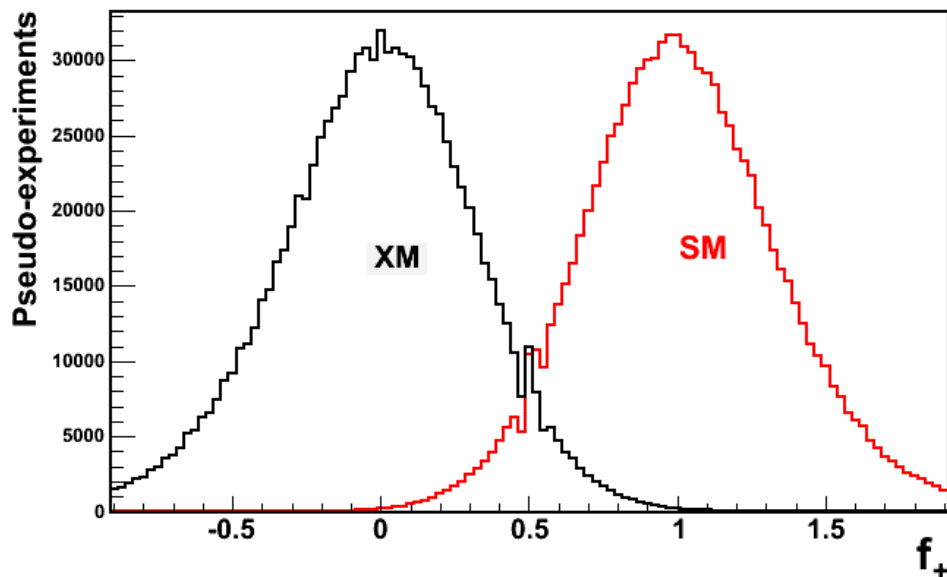
Figure 10.2. f_+ distributions for the Standard Model (SM) and exotic model (XM) hypothesis obtained from 100,000 pseudo-experiments.

for f_+. If we had a large amount of data, this would be our result. However given our small data sample, we instead followed a hypothesis testing procedure which is a common frequentist method. The hypothesis test [44, 45] is a statistical test that rejects or accepts a null hypothesis (H_0) given a data sample with two kinds of events corresponding to hypotheses H_0 or H_1.

10.2.1 Hypothesis Test

The theory of hypothesis tests is based on the frequency interpretation of probability and allows the rejection or non-rejection of the null hypothesis based on the probability of mistakenly inferring that the data support the other hypothesis more. Here we will briefly describe the theory of hypothesis tests.

One starts by choosing a test statistic (X) that is a variable which provides some discrimination between the two hypothesis H_0 and H_1. In a hypothesis test, before

making a measurement of X, one should identify a critical region or rejection region (R) that is a set of values of the test statistic for which the null hypothesis is rejected in a hypothesis test. The rejection region includes the values in an area which is calculated as:

$$\alpha = \int_{-\infty}^{X_{\text{cut}}} f(X|H_0)dX \tag{10.22}$$

where $f(X|H_0)$ is the probability density of X given H_0 and X_{cut} is the value of X at the decision boundary that separates the rejection region from the rest of the region (see Figure 10.3 as an example).

Given the data point corresponding to the observed value of X (X_{data}), the decision can be made depending on whether the X_{data} lies within the rejection region R ($X_{\text{data}} \in R$) or outside R ($X_{\text{data}} \notin R$). This gives us four possible outcomes as shown in Table 10.1. If X_{data} falls in R then H_1 is true (case A). If it does not fall within R then H_0 is true (case C). But there is still some probability to fall within R and be wrong meaning H_0 is the true hypothesis(case B). There is also some probability to be outside of R and be wrong meaning H_1 is the true hypothesis (case D).

Case	position of X_{data}	chosen hypothesis
A	$X_{\text{data}} \in R$	H_1
B	$X_{\text{data}} \in R$	H_0
C	$X_{\text{data}} \notin R$	H_0
D	$X_{\text{data}} \notin R$	H_1

Table 10.1. The four different possible outcomes from a hypothesis test.

		DECISION	
		Reject H_0	Don't reject H_0
	H_0	Type 1 Error (B)	Right Decision (C)
TRUTH	H_1	Right Decision (A)	Type II Error (D)

Table 10.2. Different decisions in a hypothesis test and the corresponding case from Table10.1 in the parenthesis.

If the decision is A or C, we have chosen the correct hypothesis. If it is B or D, we have the wrong conclusion. In the case of a wrong conclusion: case B is called a
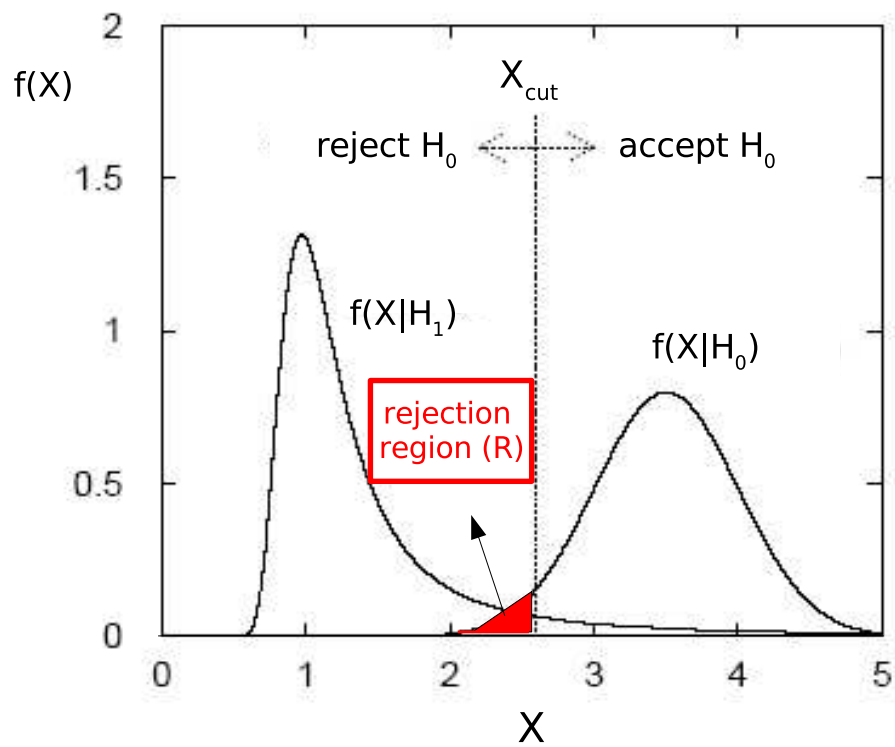
Figure 10.3. An example with the parameter of interest X for two hypotheses, H_0 and H_1. The filled area under H_0 is the chosen rejection region. If the observed X falls in this region H_0 is rejected otherwise H_0 is accepted.

"mistake of the first kind", while case D is called a "mistake of the second kind" (see Table 10.2). The probability to reject H_0 if H_0 is true (case B or an error of the first kind) is given by Equation 10.22. The probability to accept H_0 if H_1 is true (Case D or an error of the second kind) is calculated as:

$$\beta = \int_{X_{\text{cut}}}^{\infty} f(X|H_1)dX \tag{10.23}$$

where $f(X|H_1)$ is the probability density of X given H_1. In statistical terms, α is called the significance of the test, while $1 - \beta$ is called the power of test.
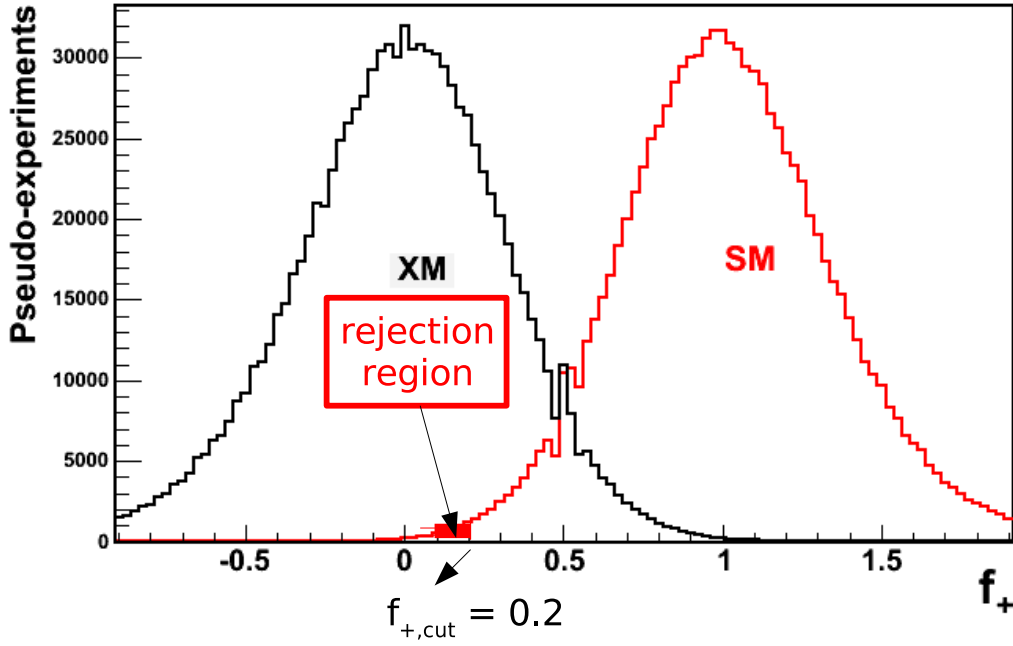

Figure 10.4. f_+ distributions obtained from pseudo-experiments assuming the Standard Model (SM) or exotic model (XM). The filled red area is the a priori 1% for α that corresponds to a f_+ of 0.2.

10.2.2 The Null Hypothesis and Test Statistic for the Top Charge Analysis:

In the top charge analysis, we need to choose the Standard Model (SM) or exotic model (XM) as our null hypothesis. The null hypothesis should be chosen as the one for which errors of the first kind are more important than errors of the second kind. The reason for this is that α (the error of the first kind) is chosen by the experimenter and then the rejection region is based on α. Considering the experimental support in favor of the SM, incorrectly rejecting the SM is more important than incorrectly accepting it. Therefore, we take the SM as our null hypothesis.

Our test statistic is the parameter of interest (f_+) that was introduced at the beginning of the chapter. If we perform pseudo-experiments using the profile likelihood and our nuisance parameters, we get a f_+ distribution peaking at 1 for the SM and a distribution peaking at 0 for the XM as shown in Figure 10.4.

10.2.3 Choice of α for the Top Charge Analysis and the Calculation of $1 - \beta$

As we have already mentioned, α is an a priori value chosen by the experimenter and once α is determined, β is also determined. For this analysis, we chose $\alpha = 0.01$, meaning the probability of incorrectly rejecting the SM is 1%. This implies that the rejection region covers 1% of the area under the SM distribution (see Figure 10.4). For an $\alpha = 0.01$, the corresponding boundary of the rejection region is 0.2.

To find the relationship between α and β we performed pseudo-experiments. We scanned the XM curve in Figure 10.4 by throwing random f_+ values. We calculated the area under the SM curve as given in Equation 10.24 for each f_+ value. In other words, we calculated the following integral

$$y = \int_{-\infty}^{f_{+\text{PE}}} f(X|SM)dX \tag{10.24}$$

where $f_{+\mathrm{PE}}$ represents the f_+ value chosen for a specific pseudo-experiment. The resulting distribution is shown in Figure 10.5. The dashed line in the figure is located at our a priori chosen α value of 0.01 and the area to the left of it corresponds to $1 - \beta$. We can see now how the $1 - \beta$ increases as we choose (shift the dashed line to) higher α values. The area to the left of $\alpha = 0.01$ is 81% of the whole area and is the level of rejecting the XM if SM is true.
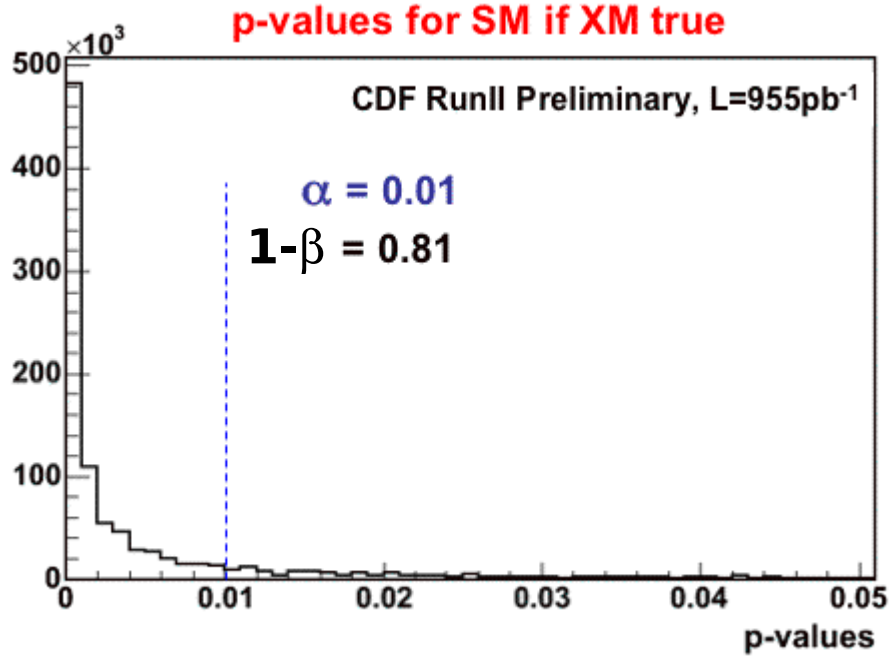
Figure 10.5. The p-value distribution for the SM assuming the XM is true. The dashed blue line represents the a priori α value of 1% (probability of incorrectly rejecting the SM if SM is true) and its corresponding $1 - \beta$ value or Power of Test (probability of rejecting the SM if XM is true).

10.2.4 p-value

In high energy physics, we often report a p-value instead of α as the significance of the test. The p-value is defined as:

$$\mathrm{p-value} = \int_{-\infty}^{f_{+\mathrm{data}}} f(X|H_0)dX \tag{10.25}$$

194

The p-value definition is equivalent to that for α, but the integral is evaluated at the measured f_+ value ($f_{+\text{data}}$) rather than at an a priori chosen value. The p-value can be interpreted as the probability of getting a result as extreme or more extreme than the one we observed if the proposed null hypothesis is correct. A small p-value provides evidence against the null hypothesis. If the p-value is less than α, we reject the null hypothesis; otherwise, we accept the null hypothesis.

10.2.5 Interpretation of the Possible outcomes

When we perform our top charge analysis on data, we will calculate a p-value. As explained above, if the p-value is greater than $\alpha = 1\%$, meaning outside of the rejection region, the measurement will be therefore consistent with the SM (null-hypothesis). The significance level of the test would be the α value, i.e. 1%, while the power of the test or the probability that the test will reject a false null hypothesis is given by $1 - \beta$. In our case, it corresponds to the probability to reject the SM if the XM is true. This probability is calculated as 81% in Section 10.2.3. So if our p-value is greater than α, we will reject the XM with an 81% CL, but if the p-value is smaller than α, we will reject the SM with 99% confidence.

10.3 Bayesian Treatment

Besides the frequentist approach, we also tried a Bayesian treatment which reflects the degree of belief for a given hypothesis, while the frequentist approach tells more about the confidence of accepting or rejecting the hypothesis. Since there is no direct analogy to the p-value from the frequentist treatment in the Bayesian framework, we can not directly compare results obtained from the different treatments. In the Bayesian structure instead of generating pseudo-experiments based on our expectations, we require prior probabilities for the SM and XM. Even though we believe more strongly in the SM due to the great amount of experimental evidence we give the SM and XM

equal probabilities to be true. In other words, $f_+ = 1$ or 0 are equally probable. The Bayesian result will tell us how much our prior beliefs change in the light of observed data (a posteriori). Below is the Bayes' theorem written considering SM and XM as the two hypotheses.

$$\frac{P(SM|X)}{P(XM|X)} = \frac{P(X|SM) * P(SM)}{P(X|XM) * P(XM)} \tag{10.26}$$

where the individual terms are described below:

$P(SM|X)$: posterior probability of the Standard model ($f_+ = 1$)

$P(XM|X)$: posterior probability of the exotic model ($f_+ = 0$)

$P(X|SM)$: likelihood at $f_+ = 1$

$P(X|XM)$: likelihood at $f_+ = 0$

$P(SM)$: prior probability of the Standard model

$P(XM)$: prior probability of the exotic model

2Ln(BF)	Strength of Evidence
0-2	Not worth more than a bare mention
2-6	Positive
6-10	Strong
> 10	Very Strong

Table 10.3. Standard scale for the Bayes factor [47].

From Equation 10.28 we define the Bayes factor (BF) as the ratio of posterior odds divided by the ratio of prior odds.

$$BF = \frac{P(X|SM)}{P(X|XM)} \tag{10.27}$$

Assuming P(SM) and P(XM) in Equation 10.28 to be 1 for both hypotheses and given no systematic uncertainties, the Bayes factor reduces to a simple likelihood ratio. In our case, we just evaluate the likelihood at $X = 1$ and $X = 0$ and take the ratio.

To include systematics we integrate separately the numerator and denominator over the nuisance parameters. The BF is commonly quoted as 2Ln(BF) and the measured value compared to the standard BF scale given in Table 10.3. The higher the value the stronger the degree of evidence for the Standard Model.

10.4   Summary

In this chapter two different statistical approaches that will be used for quoting our top charge result were presented. The first approach is frequentist for which we used a profile likelihood along with a hypothesis test. From data we will measure the fraction of SM-like events (f_+) and report a p-value. We chose the probability of incorrectly rejecting the SM if the SM is true to be $\alpha = 1\%$. We also calculated the probability of rejecting the SM if the XM is true as 81%. Accepting or rejecting the SM will depend on our p-value being larger or smaller than the chosen $\alpha = 1\%$. The second approach is Bayesian. For the Bayesian result, we will calculate the Bayes factor (BF) and quote 2Ln(BF).

CHAPTER 11

Summary and Results

> Experimental confirmation of a prediction is a
> measurement, but an experiment disproving a prediction is
> a discovery.
>
> Enrico Fermi (1901-1954), an Italian physicist.

We have completed the first measurement of the top quark's charge from CDF
using both the lepton+jets channel and the dilepton channel. We used CDF data
with an integrated luminosity of around 1 fb^{-1} for the measurement. In the Standard
Model (SM) the top quark is expected to have a charge of 2/3 but an alternative
theory (XM) has been proposed with the charge being 4/3. To determine the top's
charge we needed three components: the charge of the W boson, the flavor of the b
quark and the correct pairing between the W boson and b quark. We use the charge
of the lepton (electron or muon) from the leptonically decaying W to determine the
W's charge. The flavor of the b-jet is assigned using a momentum weighted jet charge
algorithm. In the lepton+jets channel ($t\bar{t} \rightarrow l\nu\bar{b}bjj$) we kinematically reconstructed
the events using a χ^2 fitting technique to get the correct Wb pairing. For the dilepton
channel $t\bar{t} \rightarrow l\nu l\nu\bar{b}b$ we used the invariant mass of the lepton plus b-jet.

After knowing the correct Wb pairing and the flavor of the b-jets, we count and
assign each Wb pair to be Standard Model like (SM-like) or exotic model like (XM-
like). Using 695 pb^{-1} of data for the lepton+jets channel and 955 pb^{-1} of data for

	channel cuts	ℓb pairing cuts	jet charge cuts	SM	XM
Lepton+Jets	91 (75)	48 (51)	94 pairs (94)	53	41
Dilepton	31 (33)	10 (11)	16 pairs (17)	9	7
Total	122 (108)	58 (62)	110 pairs (111)	62	48

Table 11.1. The observed number of Standard Model like (SM) and Exotic Model like (XM) Wb pairs for each channel along with the expectations shown in parenthesis. The second column shows the number of events after the lepton+jets/dilepton event selection cuts. In the third column are the number of events left after pairing. The fourth column gives the number of Wb pairs for which the b-jet charge was calculated using the momentum weighted jet charge algorithm. Finally the last two columns show the number of SM and XM-like pairs in our data.

the dilepton channel, a total of 110 Wb pairs were found. Out of the 110 Wb pairs, 62 had a SM signature with $t \rightarrow W^+b$ or $t \rightarrow W^-\bar{b}$ resulting in a charge of 2/3 or -2/3. The remaining 48 Wb pairs exhibited the XM signature resulting in a charge of 4/3 or -4/3. Table 11.1 shows the final data yields after each analysis cut. The details of the lepton+jets and dilepton data yields can be found in Appendix B. We used the number of SM-like events and XM-like events to measure various statistical results that will be the topic of the next section. Figure 11.1 shows the data (points) for the product of the W-charge and the associated b-jet charge for the lepton+jets events compared to distributions obtained from MC for signal and backgrounds. Note that W-charge$\times b$-jet-charge is negative for SM-like events and positive for XM-like events. A similar distribution for the dilepton events is shown in Figure 11.2. The comparison of different kinematic distributions such as χ^2, number of tracks and the P_T of the tracks between the data and MC can be seen in Appendix C.

11.1 Results

We decided to report our results using both a frequentist and a Bayesian approach. For the frequentist method we used a profile likelihood method to measure the fraction (f_+) of events that have a SM-like signature. Using 62 SM-like and 48 XM-like Wb pairs as inputs for N_+ and N_- to the profile likelihood (λ) we fit for the fraction
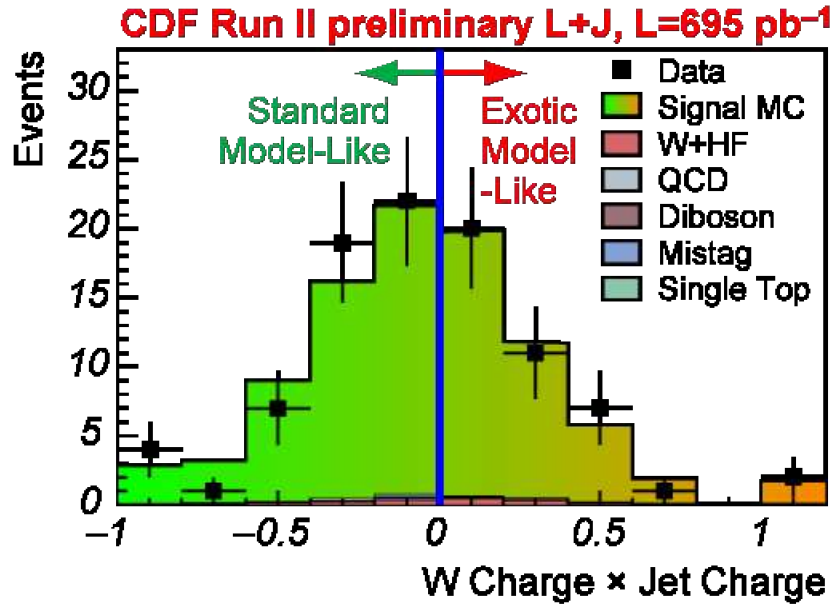
Figure 11.1. The W-charge×b-jet-charge points showing the number of SM-like and XM-like data events (black squares) on top of the distributions obtained from signal and background MC samples for the lepton+jets channel in 695 pb^{-1}.
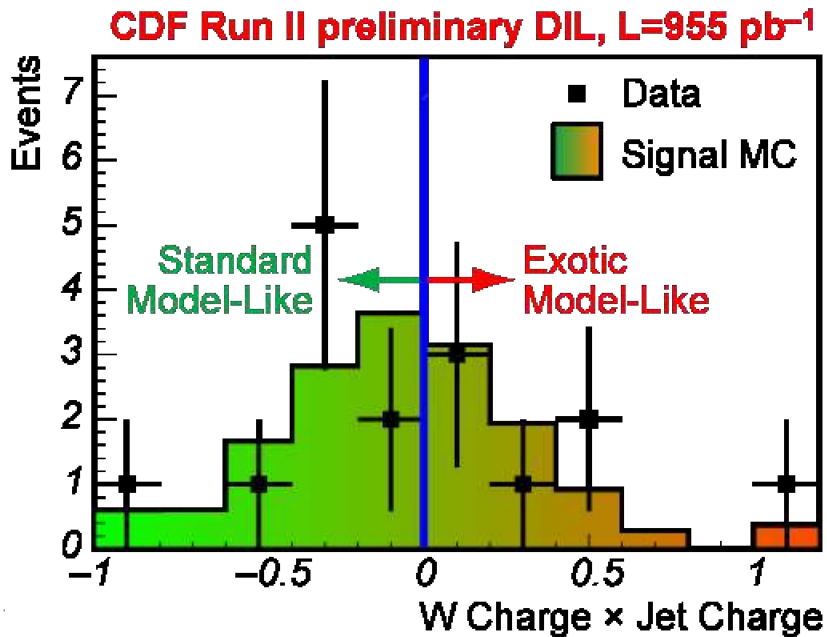

Figure 11.2. The W-charge×b-jet-charge distribution showing the number of SM-like and XM-like events in data (black squares) on top of the expected distributions obtained from signal MC sample for the dilepton channel in 955 pb^{-1}. There are no background events for dilepton channel therefore no distribution for backgrounds.

(f_+) of SM-like events and obtained 0.88 as the best fit f_+ value. Figure 11.3 shows the $-2Ln\lambda$ curve. For our f_+ value of 0.88, we integrated the area under the SM distribution and obtained a p-value of 0.35. Since our p-value is greater than our a priori value of $\alpha = 0.01$ (the probability of incorrectly rejecting SM), we exclude the exotic quark hypothesis at an 81% confidence level as explained in Section 10.1.2. We also computed the Bayes factor as described in the previous chapter. For our Bayesian approach, we calculated a Bayes Factor (BF) of 2.Ln(BF)=8.54 which translates into the conclusion that the data strongly favors the Standard Model over the exotic model.
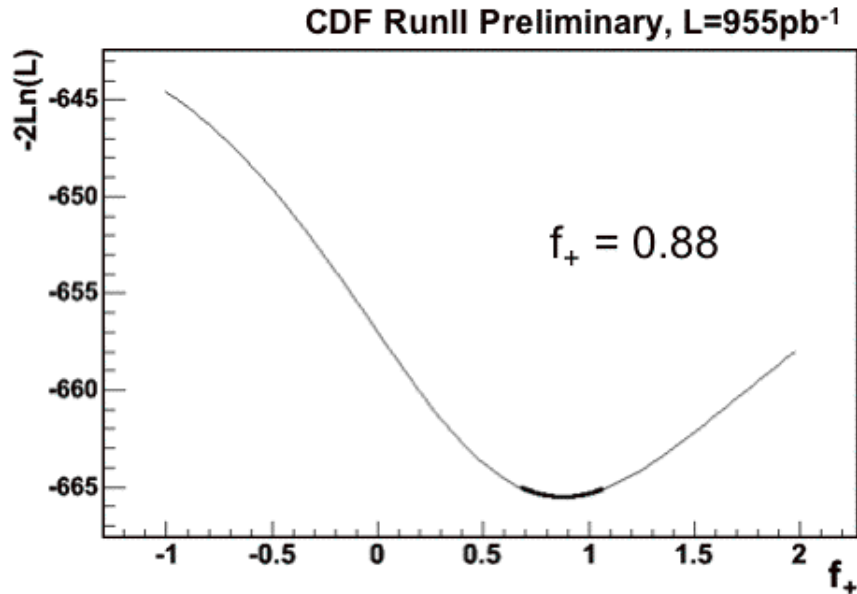

Figure 11.3. $-2Ln\lambda$ curve for the pairs observed. The best f_+ found is 0.88.

This was the first measurement of the top charge at CDF. The measurement was done with close to 1 fb^{-1} of data using the top decay products. We classified the data events as SM-like or exotic-like depending on the charge of the b-jet and the associated W boson. The data agree with the Standard Model expectation of a charge of $2/3$ and excludes the exotic model case at an 81% confidence level.
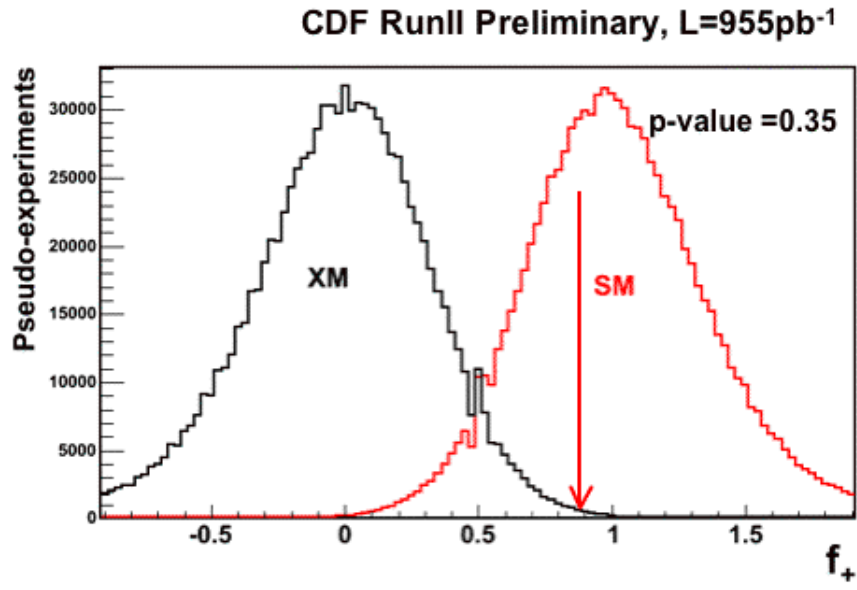
Figure 11.4. f_+ distributions for SM and XM. The red arrow indicates the observed f_+ value of 0.88. The area to the left of 0.88 under the SM curve is 0.35.

APPENDIX A

Charge Bias

A bias was found for the jet charge algorithm between b and $\bar{b}$-jets in MC. The purity of the jet charge (JQ) algorithm obtained using only $\bar{b}$-jets was different from the purity obtained using only b-jets. To determine where the bias was coming from, we first looked at different track types. There are two types of tracks at CDF, offline and secondary vertex. Secondary vertex tracks are chosen from offline tracks with tighter cuts on the track momentum, silicon information and impact parameter. We have checked if a similar asymmetry exists when offline tracks are used instead of secondary vertex tracks. The discrepancy was worse with offline tracks. Using offline tracks, the JQ purity is 0.616 ± 0.005 and 0.643 ± 0.005 for b and $\bar{b}$-jets respectively. The comparison for the jet charge distributions between b-jets and $\bar{b}$-jets using offline tracks (top plot) or using SecVtx tracks (bottom plot) can be seen in Figure A.1. It can be seen that the $\bar{b}$-jet charge distribution is shifted more to the positive side than the b-jet charge distribution is shifted to the negative side. The reason for this bias was suspected to be an excess of positive charges in an event. Figure A.2 shows the x and y coordinates of the primary particles coming from the interaction point that are associated with negative (left plot) or positive (right plot) tracks. An excess of positively charged particles is seen, notice there are more dots in the right-hand plot. This is believed to be the result of interactions between particles from the collision and the detector's material which produces more proton (+) tracks than anti-proton

(-) tracks. Looking at the P_T distribution of the positive and negative particles in Figure A.3, we can also see the excess of positive particles at low P_T. One other possible source of charge bias is from errors made when assigning a charge to the tracks at the reconstruction level. This has been confirmed by comparing the number of positively and negatively charged particles at MC generator level with the number found at the detector reconstruction level. These studies are all done with the top MC sample. We confirmed that the inconsistency between b and $\bar{b}$ purity observed in the MC also exists in data by looking at dijet data. The results are shown in Table A.1. Refer to Chapter 5 for the explanation of how the jet charge purity is measured in data.

	P_b	$P_{\bar{b}}$
Offline tracks	0.530 ± 0.006	0.565 ± 0.006
SecVtx tracks	0.547 ± 0.006	0.544 ± 0.006

Table A.1. The jet charge purity P_b ($P_{\bar{b}}$) for b ($\bar{b}$) jet candidates calculated with offline tracks or with secondary vertex tracks using dijet data. SecVtx tracks have less bias than the offline tracks.
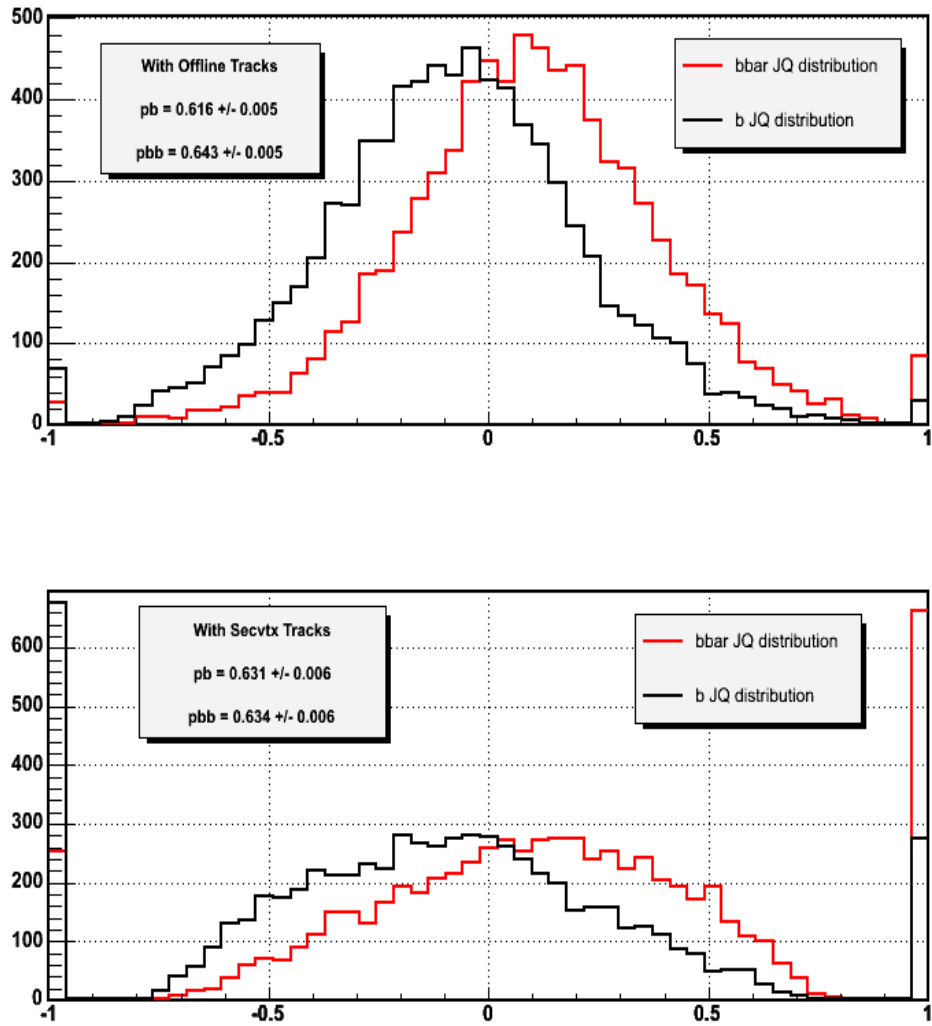
Figure A.1. The jet charge distribution for b (black line) and $\bar{b}$ (red line) jets using offline tracks (top plot) or using secondary vertex tracks (bottom plot).
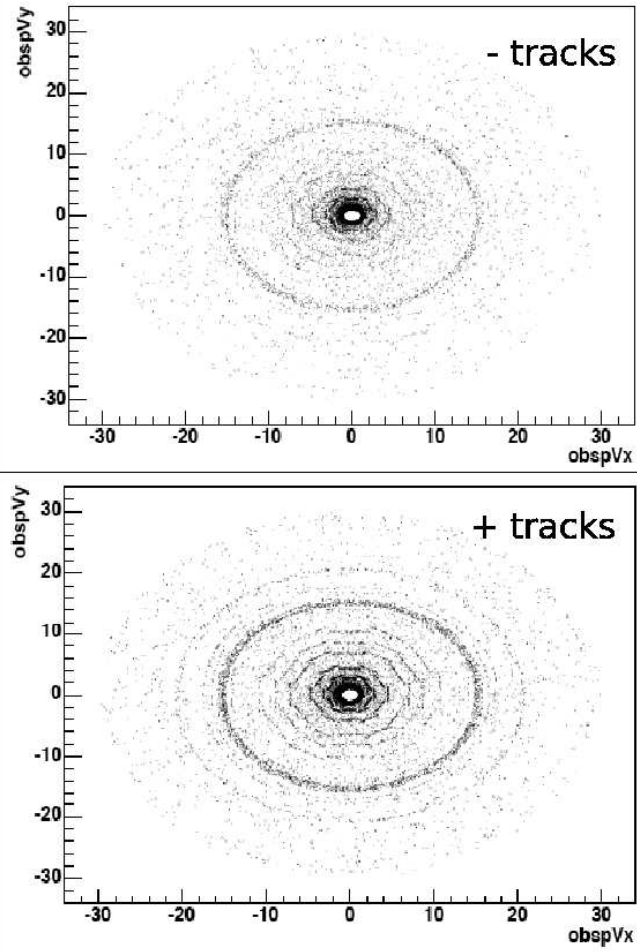
Figure A.2. Vertex of the positive (right) and negative (left) tracks on the $x - y$ plane. V_x (V_y) shows the x (y) position of the vertex.
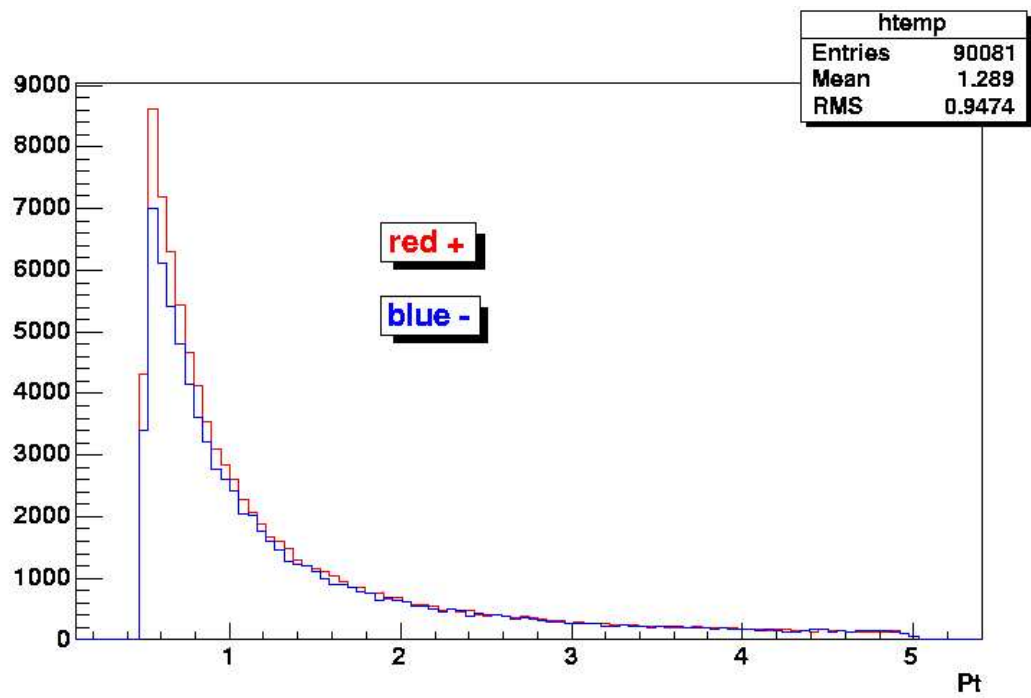
Figure A.3. P_T distribution for the positive (red line) and negative (blue line) particles that come from the interaction point.

APPENDIX B

Data Yield Details

The number of events in the lepton+jets channel after each analysis cut are given in Table B.1. The events are separated into three categories depending on the detector subsystem the lepton is observed in: CEM electron, CMUP muon or CMX muon. For example, the 727 $t\bar{t}$ events with a CEM electron were reduced to 53 events after tagging both of the b-jets. After the χ^2 cut, the number was reduced to 29 events of which the leptonic b-jet charge was defined for all, but the hadronic b-jet charge was defined for only 28.

	L+J cuts	$>= 2tag$	χ^2 cut	leptonic b defined	hadronic b defined
CEM	727	53	29	29	28
CMUP	310	28	13	13	12
CMX	91	10	6	6	6

Table B.1. The number of lepton+jets (L+J) events after the lepton+jets selection (column 2), requiring a double b-tag (column 3) and χ^2 cut(column 4). The last two columns show the number of measurements for the leptonic side and hadronic side respectively.

Remember that we have two top charge measurements for each event, one from the leptonic side and one from the hadronic side, as long as the jet charge algorithm can be applied on the tagged jets. A breakdown of how the events are divided between the SM and XM cases are presented in Table B.2 and Table B.3 for the leptonic and hadronic sides respectively. The second column in these tables displays W charge (WQ) and b-jet charge (JQ) signs. For example a "+-" indicates the charge of the W

L + J:leptonic side	WQ JQ	electron	muon	total
SM	−+	10	5	15
SM	+−	9	4	13
XM	−−	4	5	9
XM	++	6	5	11

Table B.2. The results from the leptonic b-jet side of the lepton+jets events showing the SM-like or XM-like Wb pairs depending on lepton type (electron or muon). The positive/negative signs in the second column indicates the signs of the charge of the W boson and the associated b-jet. For example a "+-" implies W boson charge is + and the b-jet's charge is -.

L + J:hadronic side	WQ JQ	electron	muon	total
SM	−+	9	1	10
SM	+−	9	6	15
XM	−−	6	9	15
XM	++	4	2	6

Table B.3. The results from the hadronic b-jet side of the lepton+jets events showing the SM-like or XM-like Wb pairs depending on lepton type (electron or muon). The positive/negative signs in the second column indicates the signs of the charge of the W boson and the associated b-jet. For example a "+-" implies W boson charge is + and the b-jet's charge is -.

boson was positive while the charge for the associated SecVtx tagged jet was found to be negative. Notice that the product of WQ×JQ is always negative in the case of the SM and is always positive in the case of the XM. The third and fourth columns show how the events are distributed based on lepton type, electron or muon. The dilepton channel results for the W charge and b-jet charge (JQ) are shown in Table B.4.

| Dilepton	WQ JQ	1st b jet	2nd b jet
SM	−+	1	3
SM	+−	4	1
XM	−−	3	2
XM	++	2	0

Table B.4. Results from the dilepton events showing the W charge (WQ) and b-jet Charge (JQ). The 1st b-jet column is the highest E_T b-jet. The 2nd b-jet column is the other b-jet in the event.

APPENDIX C

Data MC Comparison Plots

The distributions for various kinematic variables in data are compared to the expectations obtained from MC samples for signal and background. All of the MC plots are normalized to the data area.
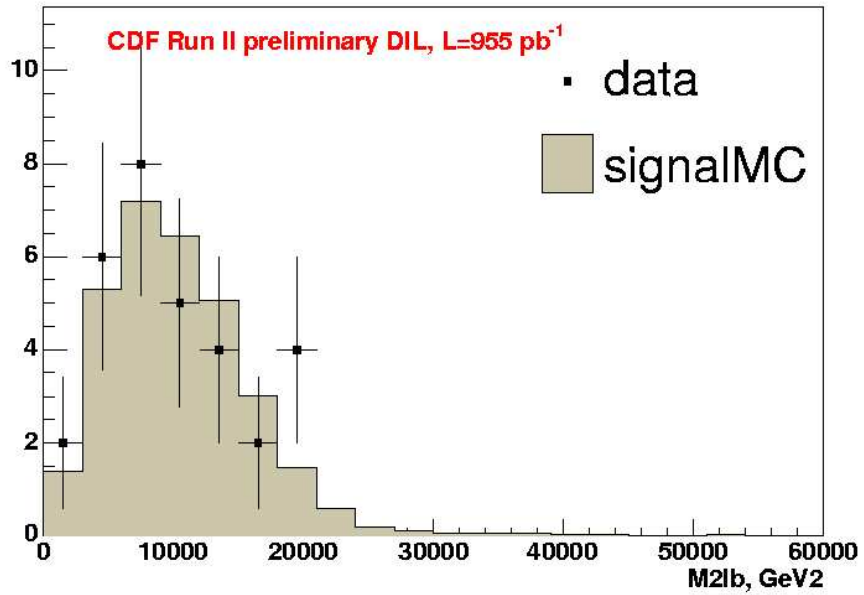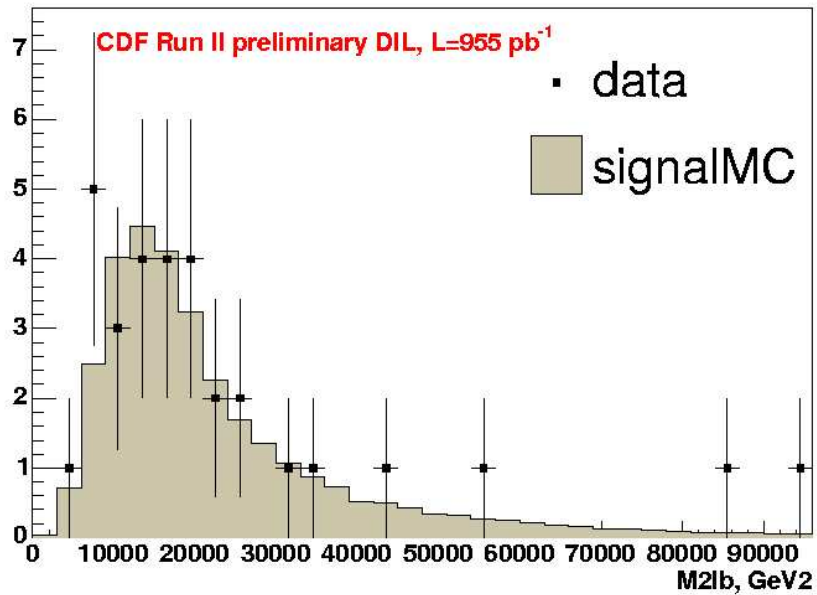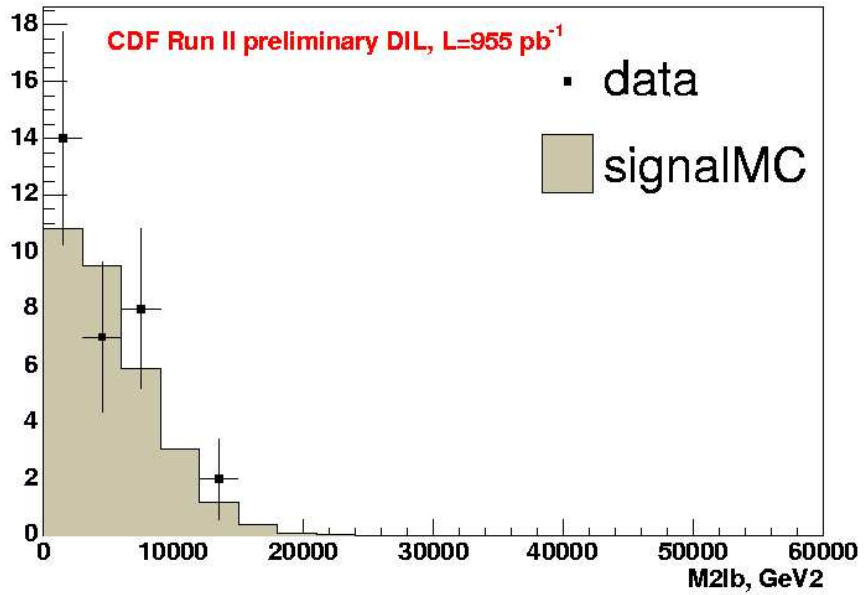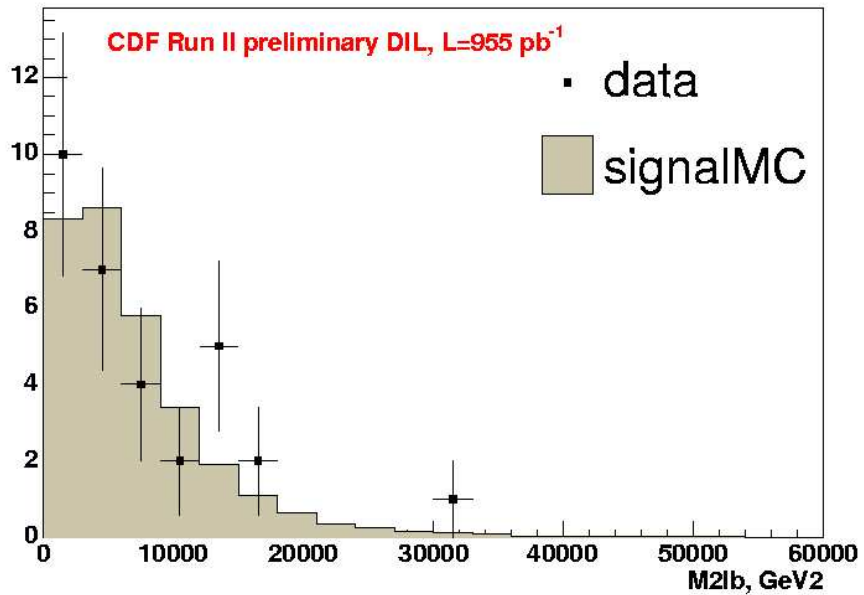
Figure C.1. At the top is the M_{lb}^2 distribution of the Wb pair that has the larger M_{lb}^2 value among the two correct Wb pairs. At the bottom is the M_{lb}^2 distribution of the Wb pair that has larger M_{lb}^2 value among the two wrong Wb pairs. The distributions include the events after the dilepton event selection requirements.
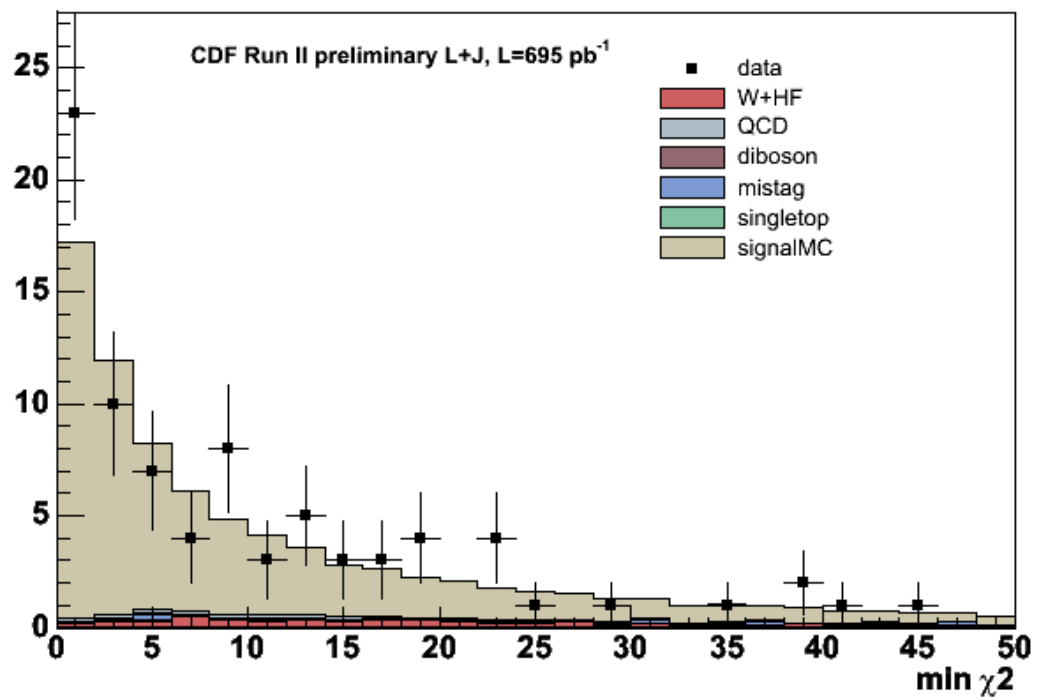
Figure C.2. At the top is the M_{lb}^2 distribution of the Wb pair that has the smaller M_{lb}^2 value among the two correct Wb pairs. At the bottom is the M_{lb}^2 distribution of the Wb pair that has the smaller M_{lb}^2 value among the two wrong Wb pairs. The distributions include the events after the dilepton event selection requirements.

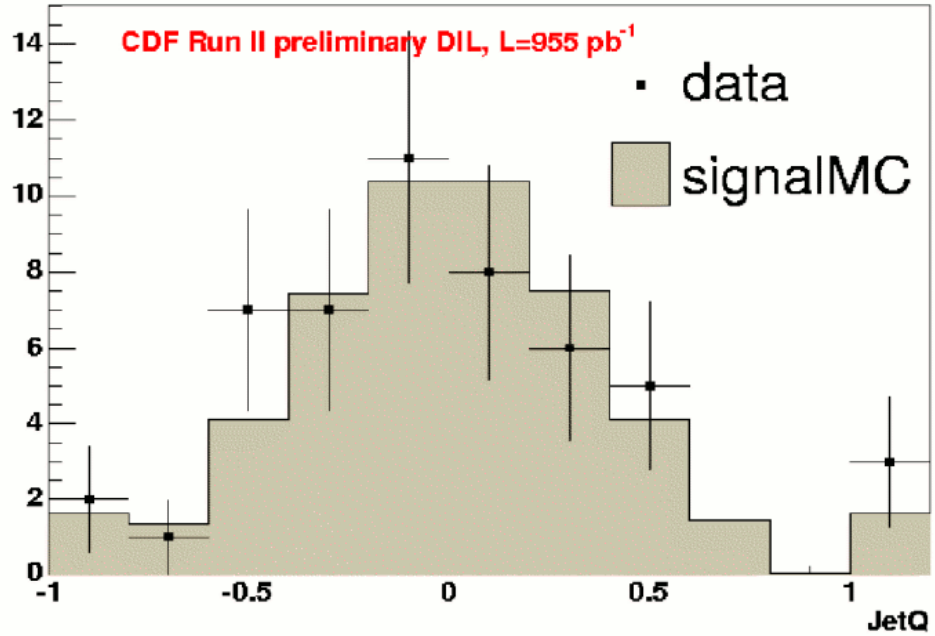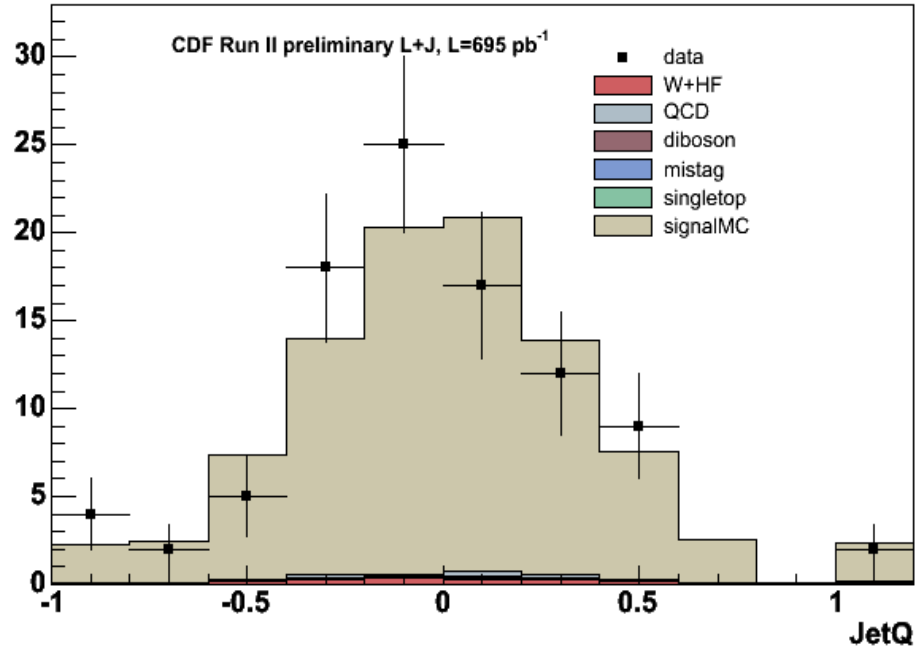**Figure C.3.** The χ^2 distribution after lepton+jets cuts and tagging the two b jets.

Figure C.4. The jet charge of *b*-jets that are used for the final top charge measurement for the lepton+jets (left) and dilepton (right) channels.
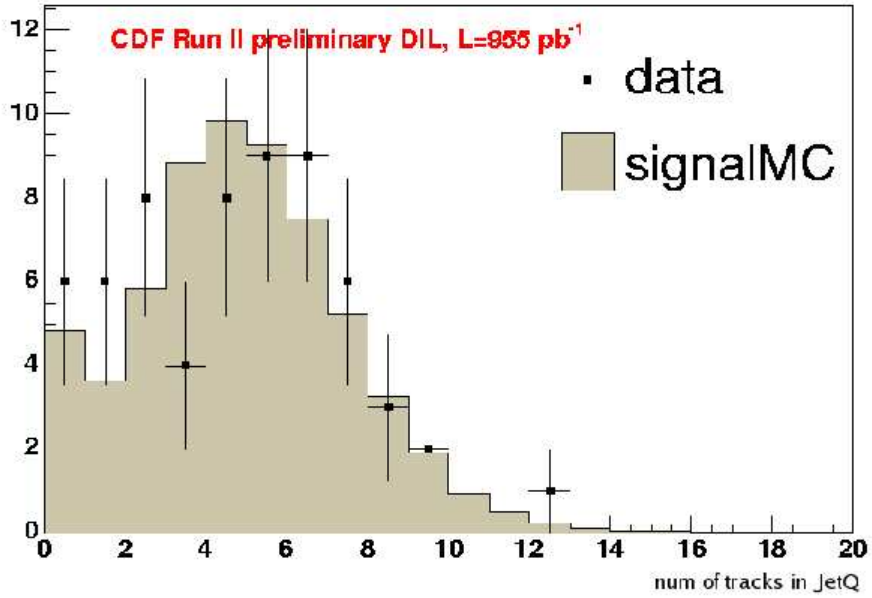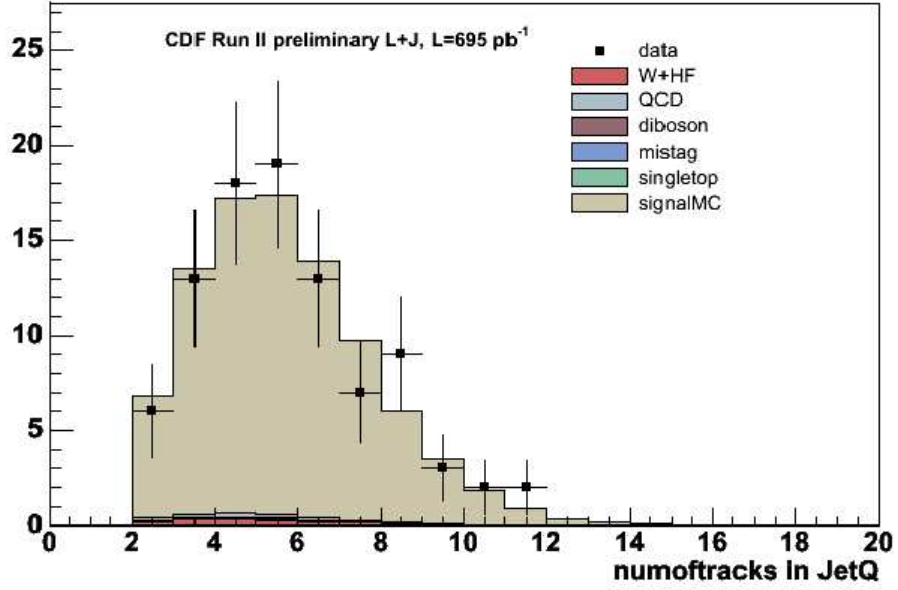
Figure C.5. The distribution for the number of tracks used in the jet charge algorithm for the lepton+jets (left) and dilepton (right) channels.
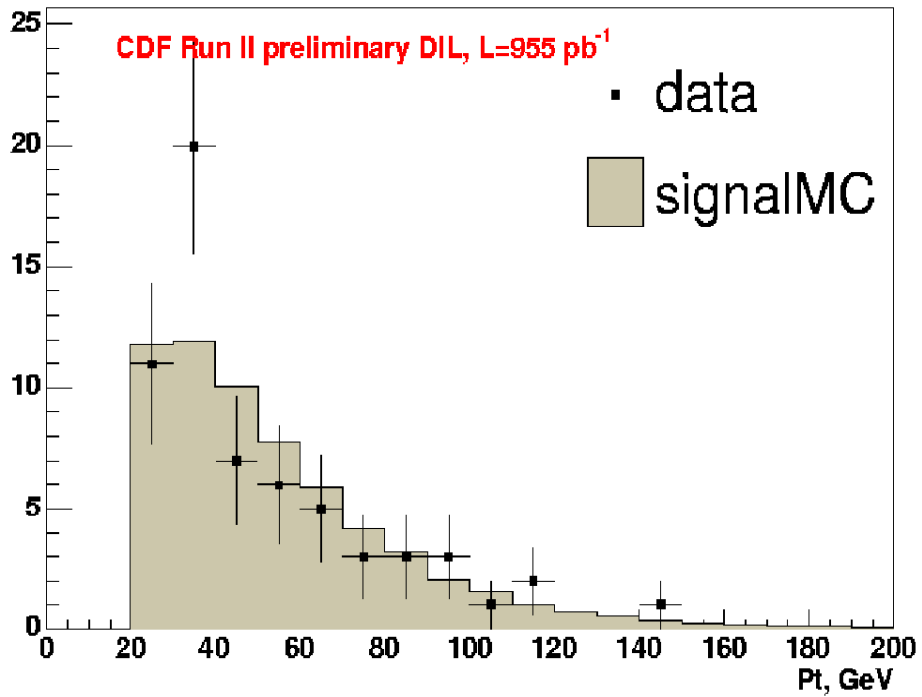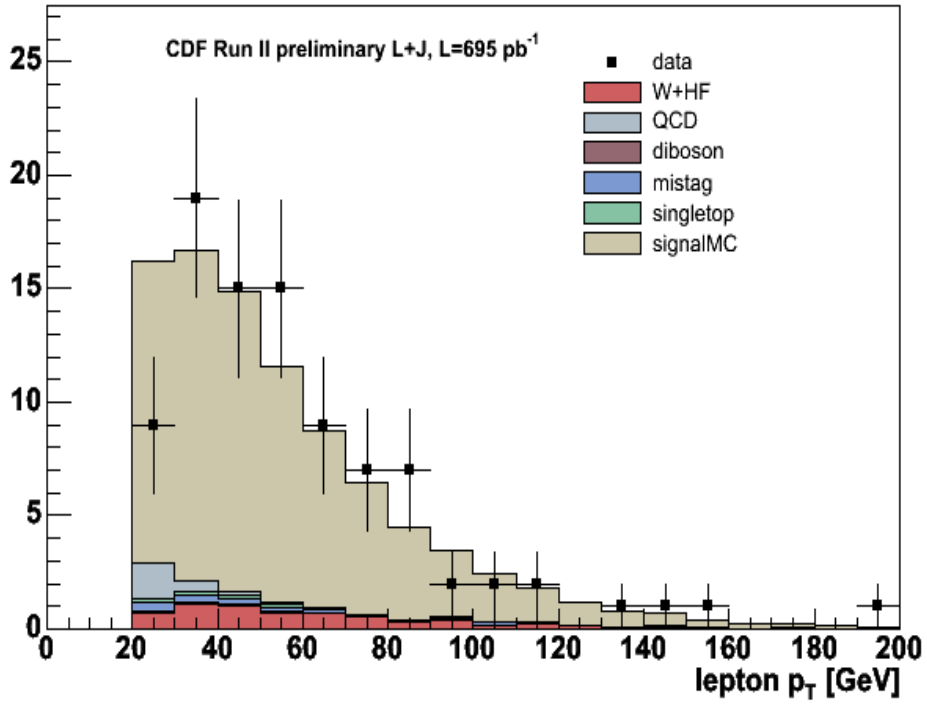
Figure C.6. Lepton P_T distributions obtained after the lepton+jets (left) and dilepton (right) selection cuts and b-tagging.
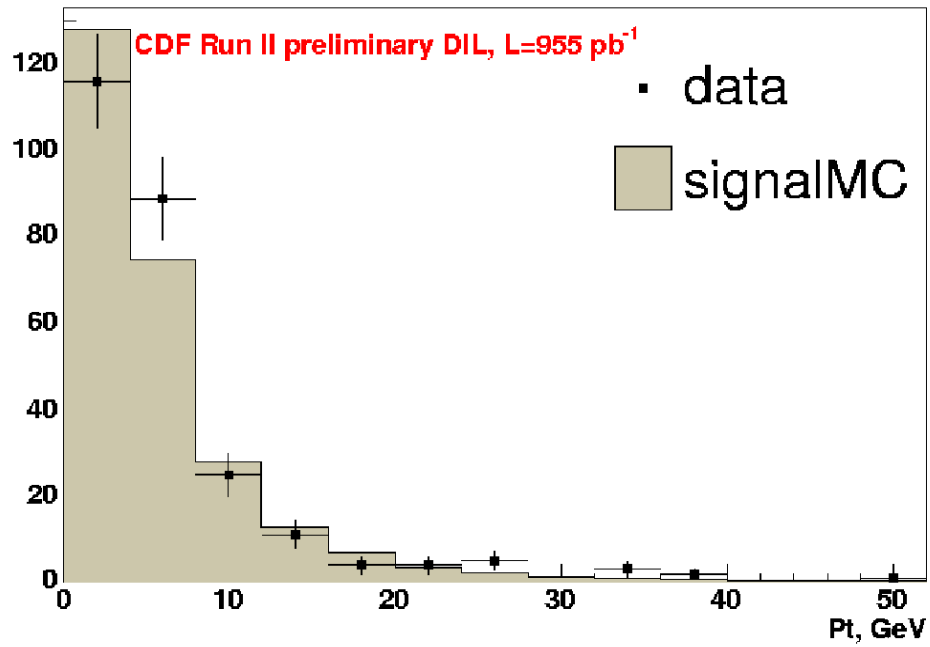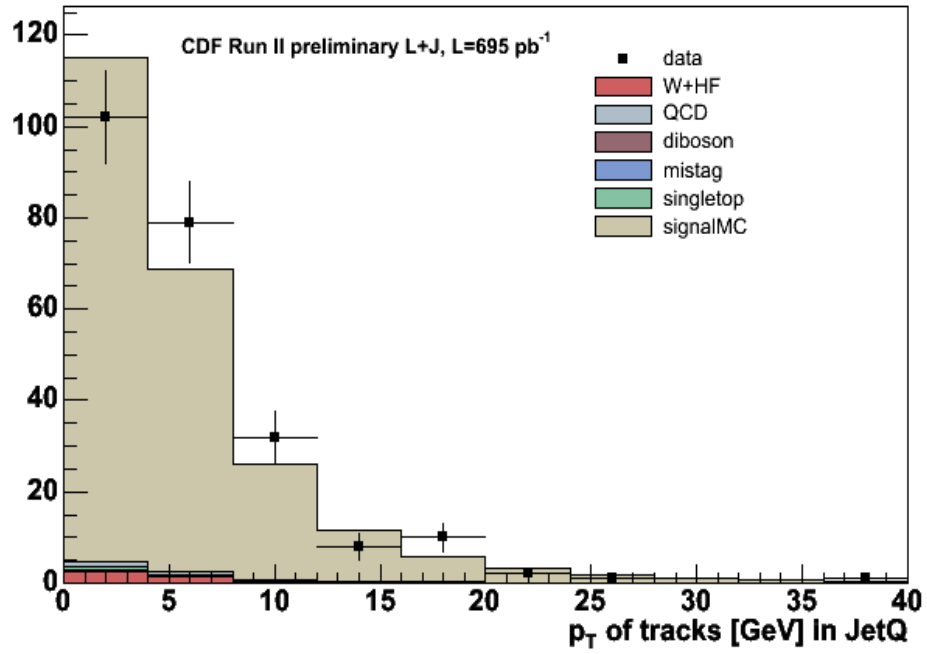
Figure C.7. The distribution for the P_T of tracks used in the jet charge algorithm for the lepton+jets (left) and dilepton (right) channels.

BIBLIOGRAPHY

[1] Francis Halzen and Alan D. Martin, "Quarks and Leptons: An Introductory Course in Modern Particle Physics ", JohnWiley & Sons (1984).

[2] Donald H. Perkins, "Introduction to High Energy Physics", Addison-Wesley Longman (1982).

[3] F. Abe *et al.* (The CDF Collaboration), Phys. Rev. **D 51**, 4623 (1995).

[4] F. Abe *et al.* (The CDF Collaboration), Phys. Rev. Lett. **74**, 2626 (1995).

[5] S. Abachi *et al.* (The D0 Collaboration), Phys. Rev. Lett. **74**, 2632 (1995).

[6] U. Baur, M. Buice, L.H. Orr, "Direct Measurement of the Top Quark Charge at Hadron Collider", Phys. Rev. **D 64**, 094019 (2001).

[7] D. Chang, W. Chang, E. Ma, "Alternative Interpretation of the Tevatron Top Events", Phys. Rev. **D 59**, 091503 (1999).

[8] http://www-bdnew.fnal.gov/operations/rookie_books/rbooks.html

[9] http://www-bdnew.fnal.gov/tevatron/

[10] F. Abe *et al.* (The CDF Collaboration), Nucl. Instr. Meth. **A 271**, 378 (1988).

[11] D. Acosta *et al.* (The CDF Collaboration), Phys. Rev. D. **71**, 052003 (2005).

[12] R. Blair *et al.* (The CDF Collaboration), "The CDF-II detector: Technical Design Report", FERMILB-PUB-96/390-E (1996).

[13] A. Still *et al.* (The CDF Collaboration), Nucl. Instr. Meth. **A 447**, 1 (2000).

[14] T. Sjostrand *et al.*, Comput. Phys. Commun. **135**, 238 (2001).

[15] J. Pumplin *et al.*, J. High Energy Phys. **0207**, 012 (2002).

[16] D. J. Lange, Nucl. Instr. Meth. **A 462**, 152 (2001).

[17] G. Marchesini *et al.*, Comput. Phys. Commun. **67**, 465 (1992); G. Corcella *et al.*, J. High Energy Phys. **0101**, 010 (2001).

[18] R. Brun and F. Carminati, CERN Programming Library Long Writeup, W5013 (1993).

[19] T. Affolder *et al.* (The CDF Collaboration), Nucl. Instr. Meth. A **526**, 249 (2004).

[20] G. Grindhammer, M. Rudowicz, and S. Peters, Nucl. Instr. Meth. **A 290**, 469 (1990).

[21] T. Affolder *et al.* (The CDF Collaboration), Nucl. Instr. Meth. **A 447**, 1 (2000).

[22] F. Abe *et al.* (The CDF Collaboration), Phys. Rev. **D 45**, 1448 (1992).

[23] A. Bhatti *et al.* (The CDF Collaboration), Nucl. Instr. Meth. **A 566**, 375 (2006).

[24] T. Sjostrand and M. van Zijl, Phys. Rev. **D 36**, 2019 (1987)

[25] R. Field, "Herwig, Jimmy and PYTHIA Tune A", talk given at TeV4LHC at Fermilab (2004).

[26] T. Affolder *et al.* (The CDF Collaboration), Phys. Rev. **D 64**, 032002 (2001).

[27] A. Abulencia *et al.* (The CDF Collaboration), Phys. Rev. **D 74**, 072006 (2006).

[28] D. Acosta *et al.* (The CDF Collaboration), Phys. Rev. **D 71**, 052003 (2005).

[29] D. Acosta *et al.* (The CDF Collaboration), Phys. Rev. **D 52**, 052003 (2005).

[30] Trevor Vickey, Ph.D Thesis, "Measurement of W Boson Polarization in Top Quark Decay", University of Illinois, FERMILAB-Thesis-2004-49.

[31] F. James and M. Roos, MINUIT: "A System for Function Minimization and Analysis of the Parameter Errors and Correlations", Comput. Phys. Commun. **10**, 343 (1975).

[32] A. Abulencia *et al.*. (The CDF Collaboration), Phys. Rev. **D 73**, 032003 (2006).

[33] http://neuralnets.web.cern.ch/NeuralNets/nnwInHep.html

[34] Ford Garberson, Joe Incandela, Chris Neu, "SecVTx b-Tag Efficiency Measurement Using Muon Transverse Momentum for 1.2 fb^{-1} Analyses", CDF Note 8640.

[35] M. Soderberg, D. Gerdes, C. Neu,"Measurement of the W+bb b-jet Cross-Section Using 695 pb^{-1}", http://www-cdf.fnal.gov/physics/new/top/2006/xs_wbb/

[36] Simon Sabik, Pierre Savard, "Track reconstruction efficiency in jets", CDF Note 6894.

[37] M.L. Mangano et al., J. High Energy Phys., 0307, 001, 2003.

[38] This study was done by Dr.Un-ki Yang using CDF data. There is no internal CDF Note or publication currently.

[39] T. Stelzer and W. F. Long, Comput. Phys. Commun. 81, 337, 1994.

[40] http://www-cdf.fnal.gov/physics/new/top/top.html

[41] A.D. Martin, R.G. Roberta, W.J. Stirling and R.S. Thorne, Eur. Phys. **J.C 4**, 463 (1998).

[42] D. Acosta *et al.* (The CDF Collaboration),Phys. Rev. Lett. **93**, 14 (2004).

[43] http://pdg.lbl.gov/

[44] Glen Cowan, "Statistical Data Analysis", Clarendon Press, Oxford, 1998.

[45] Siegmund Brandt, "Data analysis : statistical and computational methods for scientists and engineers", Springer, New York, 1999.

[46] W. A. Rolke, A. M. Lopez, J. Conrad, "Limits and Confidence Intervals in the Presence of Nuisance Parameters", Nucl. Instr. Meth., **A 551**, 493, 0403059 (2005).

[47] Robert E. Kass and Adrian E. Raftery, J. Amer. Statist. Assoc. **90**, 773 (1995).