A Simultaneous Measurement of the b-tagging Efficiency Scale Factor and the $t\bar{t}$ Production Cross Section at the Collider Detector at Fermilab

Nazim Hussain

Master of Science

Department of Physics

McGill University, Montréal

July 2011

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

DEDICATION

To those who dare to dream.

ACKNOWLEDGEMENTS

Many thanks to Dr. John Freeman who initially suggested that it might be beneficial to perform this old analysis using new data and new tricks. His guidance and description of the overall larger picture helped to keep this analysis focused.

This thesis could not have been completed without the patience, support, code and humour of Dr. Tom Schwarz. Much of the knowledge I gained regarding the realities of an experimental top quark physics analysis was acquired under his tutelage. For that I owe much thanks.

I tip my hat to the occupants of Room 345. Their assistance over the years has been invaluable, whether they realize it or not.

Finally, I am deeply grateful to my supervisor, Prof. Andreas Warburton. His exemplary supervision over the years has provided many reasons to be thankful, but most of all, I am indebted to him for his willingness to give me a chance, at a time when no one else would.

ABSTRACT

The ability to compare results between Monte Carlo and data is imperative in modern experimental high-energy physics analyses. The b-tagging efficiency Scale Factor (SF) allows for an accurate comparison of b quark identification in data samples and Monte Carlo. This thesis presents a simultaneous measurement of the SF for the SecVtx algorithm and the $t\bar{t}$ production cross section using 5.6 fb^{-1} of $p\bar{p}$ collision data at $\sqrt{s} = 1.96$ TeV collected by the Collider Detector at Fermilab (CDF) experiment. The $t\bar{t}$ cross section was measured to be 7.26 ± 0.47 pb, consistent with prior CDF analyses. The tight SF value was measured to be 0.925 ± 0.032 and the loose SF value was measured at 0.967 ± 0.033. These are the most precise SF SecVtx measurements to be performed at CDF to date.

ABRÉGÉ

Il est essentiel, chez les expériences modernes de physique des hautes énergies, de pouvoir comparer les résultats obtenus des cacluls Monte Carlo aux résultats provenant de données expérimentales. Le facteur d'échelle de l'efficacité du « b-tagging » (FE) permet de faire une comparaison légitime entre l'identification des b quarks dans des échantillons de données et des échantillons de Monte Carlo. Cette thèse présente une mesure simultanée du FE (pour l'algorithme SecVtx) et de la section efficace de la production de $t\bar{t}$, utilisant 5.6 fb^{-1} de données de collision $p\bar{p}$ à $\sqrt{s} = 1.96$ TeV. Ces données furent accumulées par l'expérience « Collider Detector at Fermilab » (CDF). La section efficace de la production de $t\bar{t}$ fut établie à 7.26 ± 0.47 pb, une valeur en accord avec les mesures antécédentes du collaboration CDF. La valeur du FE serré fut établie à 0.925 ± 0.032 et 0.967 ± 0.033 pour celle du SF lâche. Celles-ci représentent, à date, par rapport à la précision, les meilleures mesures du FE de SecVtx faites à l'expérience CDF.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
Introduction

1.1 Introduction

The universe abounds with mystery. As such, human beings continuously seek to answer questions regarding the world around them. Some of the dominant questions surround the composition of matter. What is matter made of? Can the objects that we interact with be broken down to more fundamental constituents? While it was once theorized that the protons and neutrons in the nuclei of atoms were indivisible particles, this is now known not to be the case. These nucleons are in fact composed of quarks and gluons, just two of the rich variety of fundamental particles that are predicted by the Standard Model of particle physics (SM).

In order to study subatomic phenomena, we observe high energy particle collisions and examine the new particles formed from such interactions as well as their decay products. Naturally, cosmic rays interacting with terrestrial objects provide such a source of collisions. However, in order to have precise control over the energies involved in the collision as well as to closely monitor the results, we must use human inventions.

The Tevatron, located at Fermilab in Batavia, IL, USA, is the world's highest energy matter-antimatter collider. It is a ring of over 6 kilometres in circumference where protons and their antimatter counterpart, antiprotons, are accelerated to over 99.99% the speed of light and collided at a centre-of-mass energy ($\sqrt{s}$) of 1.96 TeV. To study these collisions, the Tevatron is home to two general purpose particle detectors, the oldest being the Collider Detector at Fermilab (CDF) experiment. Although the source of many exciting

discoveries during its 26 year operational history, perhaps the most significant was the discovery of the top quark in 1995. Since then, a variety of top quark properties have been measured at CDF.

One such property is the top-antitop production cross section ($\sigma_{t\bar{t}}$), an effective measure of the probability of the production of the top quark and its antimatter partner, the antitop quark, at the given collision energy. The top quark (t quark), the most massive of the quarks, commonly decays into the bottom quark (b quark), the second-most massive quark. Consequently, many measurements of the $t\bar{t}$ cross section rely on the identification of these b quarks via so-called b-tagging techniques. Such techniques are not perfect and it is quite possible that they overestimate or underestimate the actual quantity of b quarks. The b-tagging efficiency is exactly such a measure of the methods' effectiveness. Commonly, in experimental particle physics analyses, researchers must compare the data collected from the detectors to theoretical predictions obtained from Monte Carlo calculations (MC). However, the b-tagging efficiency, as measured in data samples, may differ from that measured in MC. This means that in order to make adequate comparisons between the two, the appropriate scaling must be known. The b-tagging efficiency scale factor (SF) is such a measure; it is the ratio between the b-tagging efficiency as measured in data versus the b-tagging efficiency measured in Monte Carlo.

Typically, for most analyses that use b-tagging, the SF is an input value into a large analysis mechanism. However, like all experimentally obtained quantities, the SF measurement has an associated uncertainty. Consequently, any analysis that uses the SF has this uncertainty incorporated as part of its overall systematic uncertainty. At CDF, researchers are currently engaged in an attempt to discover the last SM particle that has yet to be observed: the Higgs boson. If this particle exists, it is believed that very few signal events

will be seen by the detector. Thus, it is of utmost importance to try to keep the systematic uncertainties of the analysis as low as possible; if they are too large, it may not be possible to say with certitude that a discovery has been made. As a result, any endeavour to reduce the systematic uncertainties in a Higgs boson analysis may have tremendous positive repercussions. Therefore, reducing the uncertainty of the SF measurement is important for these Higgs boson searches.

At CDF, the typically used SF value comes from a measurement where the b-tagging is performed on low-momentum particles. However, for Higgs boson searches, as well as many other modern analyses, the b-tagging is performed at higher momentum values. Through extrapolation, one can determine the SF to be used in these higher regimes. However, it is conceivable that by performing the SF measurement initially in a high momentum sample, a more precise measurement can be made. In theory, such increased SF precision would lead to less uncertainty for any analysis that uses the SF as an input.

This thesis presents a simultaneous measurement of the b-tagging efficiency scale factor and the $t\bar{t}$ production cross section at CDF, performed using 5.6 fb^{-1} of proton-antiproton collision data at a centre-of-mass energy of 1.96 TeV. The underlying methods used are not novel; they have been used at CDF before to perform the measurement in 2006 and 2008. This thesis, however, takes advantage of having access to five times as much data as the last measurement as well as a different approach to calculating the contributions of background processes. Furthermore, the measurement is also performed independently for two different time periods of the detector's lifetime and ascertains whether a change in SF has occurred or not. Due to the simultaneous nature of the SF measurement, this thesis will also report a new measurement of the $t\bar{t}$ cross section at CDF. Although the final quoted value of SF in this

thesis is applicable only to measurements made at CDF, the underlying process of calculating the SF can be performed at any experiment where the $t\bar{t}$ cross section is measured.

Chapter 2
The Standard Model

Modern particle physics is described by a theory known as the Standard Model. Developed in the 1960s and 1970s, the SM has been remarkably accurate in describing the vast majority of experimentally observed particle physics interactions. In fact, while there are indications that the SM is not the complete description of particle physics in our universe, many theorists believe that it is most likely a subset of a larger, broader theory, rather than being wrong altogether [1]. In order to fully understand the significance of the SF and $\sigma_{t\bar{t}}$, we must first examine the fundamentals of the SM. We begin by introducing the most common particles of the SM, the quarks, leptons and force carriers. We will then briefly discuss the Higgs boson, the most experimentally elusive of the SM particles. Given that the SF relates intrinsically to b-tagging and that the $t\bar{t}$ cross section is measured simultaneously with SF in this analysis, we will discuss in greater depth the bottom and top quarks.

It should be noted that, unless otherwise stated, natural units are used in this thesis, where $\hbar = c = 1$.

2.1 Quarks, Leptons and Force Carriers

The SM is a collection of quantum field theories that describe the most fundamental of matter interactions [1]. The majority of SM particles are known as fermions, a collection of spin-1/2, indivisible objects. These fermions are further separated into collections of quarks and leptons.

All of the quarks are charged particles and they can experience three of the fundamental forces: the weak nuclear force, the strong nuclear force and the electromagnetic force. They experience the force of gravity; however, it

should be noted that the SM does not include gravity as one of its forces. As such, it is an incomplete description of the universe. But, at the femtometre scale of particle interactions, gravity is several orders of magnitude weaker than the other forces and can effectively be ignored. Thus, we refrain from further discussion of gravity in this thesis.

Quarks are further subdivided into three generations, illustrated in Table 2.1 [2]. An individual quark can be further classified as being up-type or down-type, depending on its electric charge. The up, charm and top quarks are up-type quarks with an electric charge of +2/3 and the down, strange and bottom quarks with a charge of -1/3 are of the down-type variety. In addition to carrying electrical charge, quarks also possess colour charge. Each quark may be one of three colours (red, green, blue) or their corresponding anticolour (anti-red, anti-green, anti-blue). The principle of colour confinement postulates that observed states have 0 net colour charge [3]. This means that bare quarks are generally not seen in nature; instead, we observe composite quark bodies known as hadrons. There are two types of hadrons: baryons, which are composed primarily of three quarks, each of a different colour, and mesons, which are primarily composed of a colour-anticolour pair of quarks [4].

Like the quarks, the leptons are grouped in three generations; however, they only experience the electromagnetic and weak nuclear forces. These particles may be classified as either charged leptons or uncharged neutrinos. Although it has been demonstrated experimentally that neutrinos have mass, for the sake of simplicity, in this thesis, they will be assumed massless as per their original depiction in the SM [1]. This is a reasonable assumption given that the latest estimates on neutrino masses are several orders of magnitude below the lightest of the charged lepton masses [2].

	Flavour	Mass (MeV)	Charge (e)
Quarks	*First Generation:*		
	up (u)	1.7-3.3	+2/3
	down (d)	4.1-5.8	-1/3
	Second Generation:		
	charm (c)	1270	+2/3
	strange (s)	101	-1/3
	Third Generation:		
	top (t)	172000	+2/3
	bottom (b)	4190	-1/3
Leptons	*First Generation:*		
	electron (e)	0.511	-1
	electron neutrino (ν_e)	assumed massless	0
	Second Generation:		
	muon (μ)	106	-1
	muon neutrino (ν_μ)	assumed massless	0
	Third Generation:		
	tau (τ)	1777	-1
	tau neutrino (ν_τ)	assumed massless	0

Table 2.1: Standard Model quarks and leptons. Although neutrinos have been experimentally verified as being massive particles, this thesis assumes that they are massless.

The fermions interact with each other via the gauge bosons; these are spin-1 particles that are carriers of the respective forces. The electromagnetic force is mediated by the massless photon. The strong nuclear force, on the other hand, is mediated by the gluon, a massless particle that carries colour charge. Given that there are three different colours and each gluon carries two possible colours, there are a total of eight different types of gluons in the SM. Finally, the weak force is mediated by the massive W and Z bosons. The W is found in both positively and negatively electrically charged varieties and the Z is available as an uncharged particle [4]. The fundamental quarks and leptons of the SM are illustrated in Table 2.1 and the force carriers in Table 2.2 [2].

It should be noted that the SM predicts that all of its fundamental particles have a corresponding antiparticle. Although some neutral particles are

Name	Force	Mass (GeV)	Charge (e)
photon (γ)	Electromagnetic	0	0
Z^0	Weak Nuclear	91.2	0
$W^{\pm}$	Weak Nuclear	80.4	± 1
gluon (g)	Strong Nuclear	0	0

Table 2.2: Standard Model force carriers.

their own antiparticle (like the photon and Z boson [1]), most are distinctly different. In particular, the antiparticles of charged particles share the same mass and spin, but possess the opposite electrical charge. Antiparticles are denoted by an overbar on top of the symbol for their corresponding particle.

Although the particles mentioned above are fundamental, this does not mean that they are all stable. In fact, the majority have lifetimes of order microseconds or less before they decay into something more stable [2].

2.2 The Higgs Boson

The Higgs boson (H^0) is the only particle predicted by the SM that has yet to be observed. Consequently, at the Fermilab Tevatron (and the Large Hadron Collider at CERN), many researchers are actively engaged in the search for the Higgs. In the SM, the electromagnetic and weak nuclear forces can be unified to create the electroweak interaction, described by the $SU(2) \times U(1)$ symmetry group. It is theorized that this symmetry can be spontaneously broken by the so-called Higgs mechanism, involving a scalar Higgs field. This process leads to the acquisition of mass by the SM particles. However, if the theory is correct, we expect to observe a spin-0 boson, the Higgs boson, the quantum of the Higgs field [5].

Although the SM predicts the existence of the Higgs boson, it does not predict its mass. The LEP experiment at CERN and Fermilab's CDF and DØ experiments have excluded the SM Higgs boson from having a mass below 114 GeV at the 95% confidence level. Furthermore, CDF and DØ have also

excluded higher masses in the 158-175 GeV range at the 95% confidence level. These exclusion limits are illustrated graphically in Figure 2.1 [6]. In addition, precision electroweak measurements indirectly place limits on the Higgs boson mass being less than 186 GeV [2]. Thus, experimentally, many searches for SM Higgs particles are focused in the 114-158 GeV range. These masses are lower than that of the top quark. If this mass range is correct, it would mean that the Higgs boson can be produced at the current generation of particle colliders.

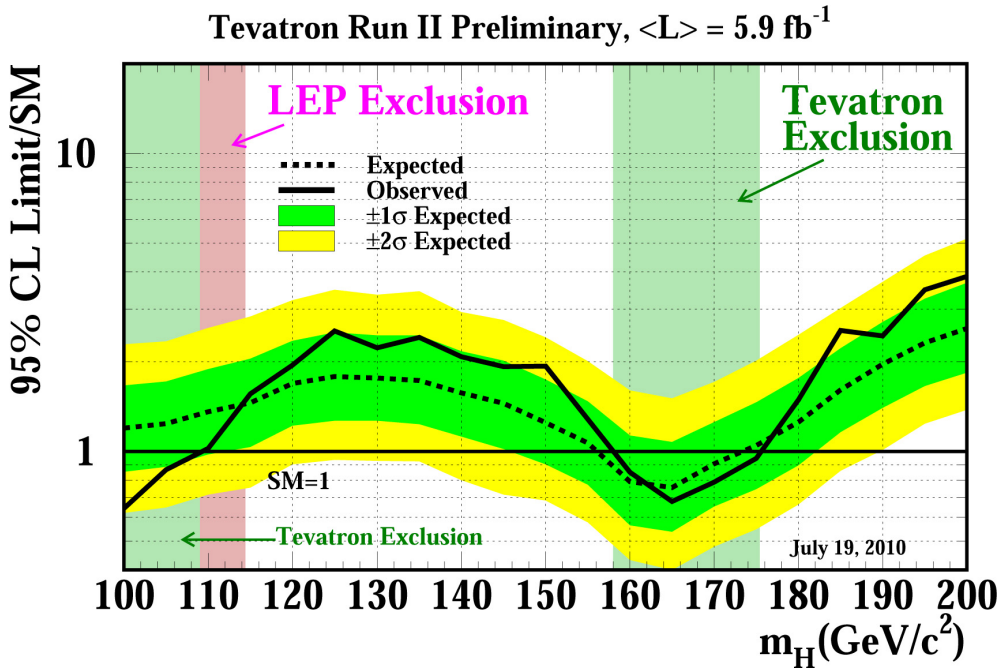


Figure 2.1: Higgs exclusion limits from direct searches. The y-axis shows the ratio of the SM Higgs cross section (the notion of a cross section is further elucidated in Section 5.2). Note, as well, that masses greater than 186 GeV are also excluded (indirectly) from precision electroweak measurements.

The Higgs is believed to couple to fermions with strengths in proportion to their mass [5]. As such, some of the Higgs decays deemed most likely to be seen at particle colliders include interactions with the b and the t quarks (the most massive of the quarks). A sample Higgs decay involving heavy quarks is illustrated in Figure 2.2 [7].

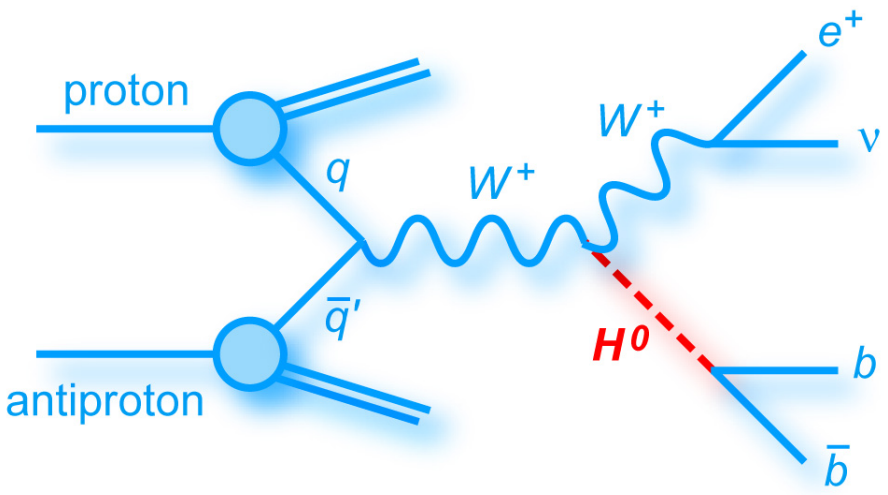


Figure 2.2: Higgs radiated from a W boson and decaying into a $b\bar{b}$ pair. This is believed to be one of the more promising decay channels for Higgs searches at Fermilab.

2.3 The Bottom Quark

The b quark is the down-type quark of the third generation. It is the second-most massive of the quarks with a bare mass equivalent to roughly five times that of the proton. Although the b quark can decay via W boson-mediated weak processes to a c or u quark, these processes are highly suppressed. As a result, the b quark is relatively long lived with a lifetime of order 10^{-12} seconds [3]. As we shall see later, this long lifetime makes b quarks the most identifiable quark within particle detectors. Two of the most common b-decays are illustrated in Figure 2.3 [3]; a weak decay into leptons and a weak decay into quarks. In addition, because of the relatively high mass of the b quark in comparison to its decay products, it is not uncommon for the decay products to be produced at large angles relative to the initial trajectory of the parent b quark. This characteristic leads to easier identification of b quarks.

Of course, due to colour confinement, the bare b quark is never observed within particle detectors. Rather, what is noted, are the myriad b-hadrons.

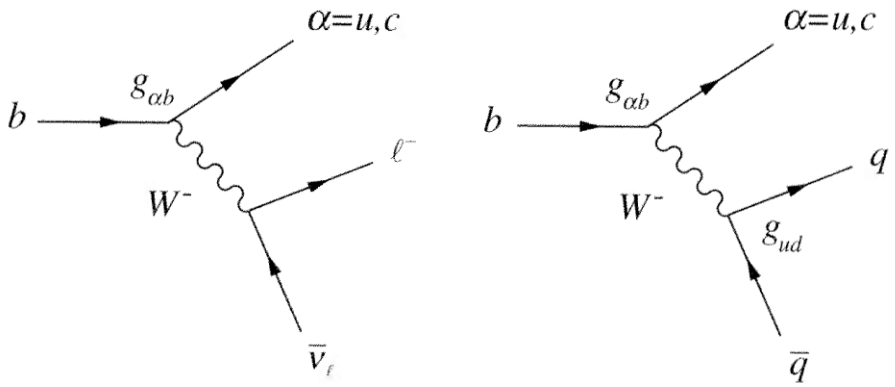


Figure 2.3: Common b quark decays. The diagram on the left shows the b quark decaying into a u or a c and a pair of leptons, while the diagram on the right shows a decay into u or a c and a pair of quarks. This pair may be either $d\bar{u}$, $d\bar{c}$, $s\bar{u}$ or $s\bar{c}$.

However, because these hadrons contain at least one unstable quark, they will decay as well.

2.4 The Top Quark

The t quark is the up-type quark of the third generation and the most massive of all of the quarks. It is roughly 40 times more massive than the b quark. To put this mass in perspective, the rest mass of the top quark is comparable to that of the nucleus of a gold atom. Unlike the b quark, the t quark has a very short lifetime of the order 10^{-25} seconds. Given that the strong interaction operates on longer timescales, t quarks do not hadronize prior to decay [8]. Rather, upon creation, they mainly decay to b quarks. Thus, the identification of b-decays is critical if one hopes to study t quarks.

At the Tevatron, most of the t quarks produced are in $t\bar{t}$ pairs as opposed to being produced singly. Furthermore, roughly 85% of the $t\bar{t}$ pairs are produced from the annihilation of a quark and an antiquark. The other 15% is mainly from gluon-gluon fusion [9]. These processes are illustrated in Figure 2.4 and Figure 2.5, respectively [7].

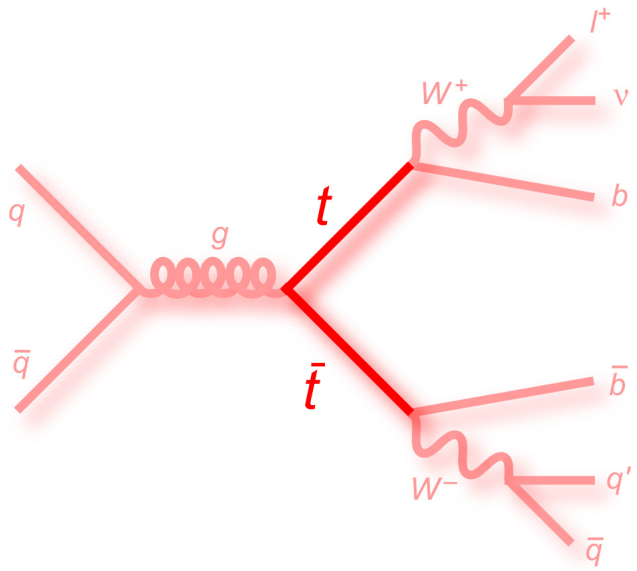


Figure 2.4: $t\bar{t}$ production from $q\bar{q}$ annihilation. The b quarks subsequently decay by the same processes described in the previous section.

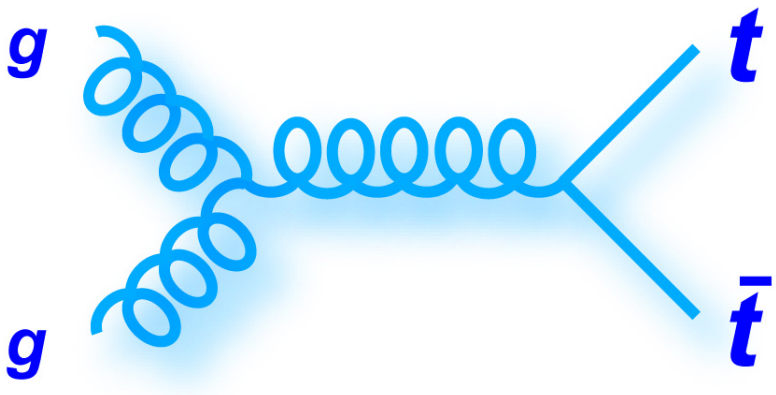


Figure 2.5: $t\bar{t}$ production from gluon-gluon fusion. The top quarks decay the same way as shown in Figure 2.4.

The study of the t quark is a rich area of research at modern hadron colliders. This is partially due to the hope that the top quark will provide some fundamental insight into the SM. For example, there is no explanation in the SM as to why the t quark is as massive as it is, in particular in comparison to the other quarks. Nor is there any justification for having three generations of quarks. Perhaps a better understanding of the top quark may help address these questions. Furthermore, given that the t quark is expected to couple strongly to the Higgs boson, the understanding of top quark behaviour is beneficial to Higgs searches.

Chapter 3
Experimental Infrastructure

In order to produce the top quark in a laboratory environment, we must collide particles. However, due to its high mass, the particles need to be collided at tremendous energies. The Tevatron accelerator at Fermilab was the first collider to produce the top quark and, until the recent startup of the CERN Large Hadron Collider (LHC), the only place to study t quark physics. However, to observe $t\bar{t}$ interactions, simply colliding particles at high energy is not sufficient; there must be a detector to observe and record the results of the collisions. At the Tevatron, there are two such detector experiments: DØ and the Collider Detector at Fermilab (CDF). Since the results of this thesis are based on data collected from CDF, it is natural that we introduce the Tevatron and CDF as the experimental apparatus.

3.1 The Tevatron

Located at Fermilab in Batavia, Illinois, the Tevatron is the second most powerful particle collider in the world. It is a synchrotron of over 6 km in circumference where protons and antiprotons are accelerated and then collided at $\sqrt{s} = 1.96$ TeV ($p\bar{p}$ collisions). Although the Tevatron has been in existence since 1983, the Fermilab accelerator complex has undergone many enhancements and modifications since that time [10]. In the interest of relevance, we shall focus on the period after 2001, the so-called Run II period of accelerator (and detector) operations, when the data used in this thesis were gathered.

The complete chain of accelerators at Fermilab is illustrated in Figure 3.1 [11]. We will briefly discuss the procedure of beam production and acceleration through this system. The first step in the acceleration process occurs at the

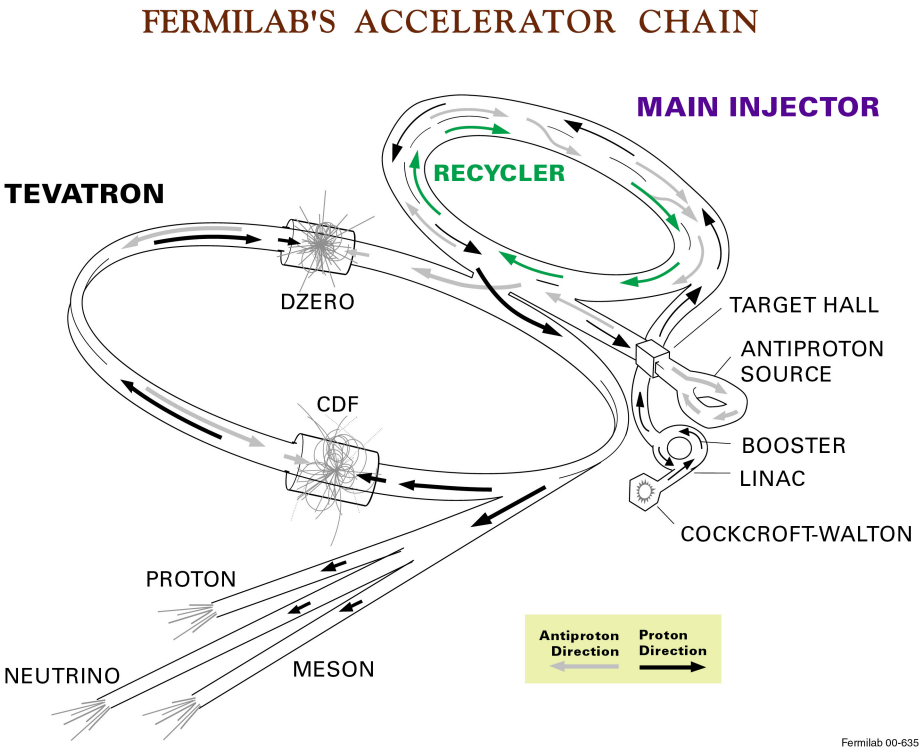


Figure 3.1: Diagram of Fermilab's accelerators (not to scale).

Cockcroft-Walton Preaccelerator (Preacc). A source in the Preacc converts hydrogen gas to ionized hydrogren gas (H^-). The Preacc then accelerates the H^- to 750 keV through the use of a highly charged surface held at a 750 kV potential [12]. The H^- then enters the Linear Accelerator (Linac) where it is accelerated to 400 MeV using radio frequency (RF) cavities. It is also focused into a beam through the use of quadrupole magnets within the RF cavities. From the Linac, the H^- beam enters the Booster. Here, the beam is passed through a thin layer of carbon, stripping it of its electrons [13]. What remains is a pure beam of protons (H^+). This beam is then accelerated to 8 GeV via the Booster's synchrotron. The latter is a ring of 75 meters in radius equipped with 19 RF cavities to perform the acceleration. The beam then moves into another synchrotron, the Main Injector (MI). This device is elliptical in shape and has a circumference of approximately 3200 m [12, 14]. Through the use

of 18 RF cavities, the proton beam is accelerated to 150 GeV and via the use of kicker magnets, as the machine's name implies, it is injected into the main Tevatron ring. Through the use of 8 RF cavities, the Tevatron accelerates the proton beam from 150 GeV to 980 GeV. It must be noted that in order to reduce the energy cost required for the steering of the beam through 6 km of tunnel, the Tevatron uses superconducting magnets, cooled to a temperature of 4K through the use of liquid helium [15]. Only the Tevatron uses such magnets; the other accelerators at Fermilab do not have cryogenic cooling systems [12].

As described earlier, the Tevatron is a $p\bar{p}$ collider. The process described above, however, only mentions the creation of a 980 GeV proton beam. Antiproton creation starts with the MI sending a 120 GeV proton beam to a fixed target of Inconel (a nickel-iron alloy) [16]. Many particles are produced from this interaction, including antiprotons. Using magnets to select particles of the desired charge and momentum, 8 GeV antiprotons are collected from this spray of particles and directed towards the Debuncher. The latter is a rounded triangular synchrotron with a mean radius of 90 m. Using stochastic cooling processes, the Debuncher creates a stable beam of 8 GeV antiprotons [12]. From here, the beam is transferred to the Accumulator, a storage ring located in the same tunnel as the Debuncher. Further stochastic cooling is applied to the antiproton beam. The beam then travels to the Recycler, an antiproton storage ring housed in the same tunnel as the MI. The Recycler performs more stochastic cooling on the beam as well as electron cooling. Finally, the 8 GeV antiproton beam is directed into the MI where it is accelerated to 150 GeV and injected in the Tevatron. Within the Tevatron, the antiproton beam is accelerated to its final energy of 980 GeV.

The proton and antiproton beams travel in opposite directions within the Tevatron and are collided at $\sqrt{s} = 1.96$ TeV at two locations along the ring: the collision halls of the DØ and CDF experiments. Upon collision, a multitude of particles is created, many of which are detected by the experiments. An important concept at particle colliders, like the Tevatron, is that of instantaneous luminosity. The luminosity is the number of collisions per unit time per unit area [4]. At the Tevatron, the area in question is the cross-sectional area of the $p\bar{p}$ interaction. In particle physics, these areas are typically expressed in barns (b), where 1 b = 10^{-28} m^2. A commonly discussed quantity is the integrated luminosity ($\mathcal{L}$), the integral of the instantaneous luminosity with respect to time (generally expressed in units of pb^{-1} or fb^{-1}). $\mathcal{L}$ is directly proportional to the total number of collisions that have occurred. As such, when expressing the amount of data delivered by a collider (or acquired by a detector experiment), we generally use $\mathcal{L}$. As of 2011, each of the Tevatron general purpose detector experiments have collected data samples with an integrated luminosity of order fb^{-1}.

3.2 The Collider Detector at Fermilab Experiment (CDF)

CDF, like the other detector on the Tevatron ring, DØ, is a general purpose particle detector. The data collected by the experiment can be used for a wide array of particle physics analyses ranging from basic SM measurements [17] to searches for extra dimensions [18]. Many of the particles formed from the initial $p\bar{p}$ collision from the Tevatron (the so-called "hard interaction") are stopped by the detector. Through an examination of the particles' interactions with the detector components, these stopped particles may be identified. Consequently, the particles produced from the hard interaction and their subsequent decays and interactions with each other can be reconstructed.

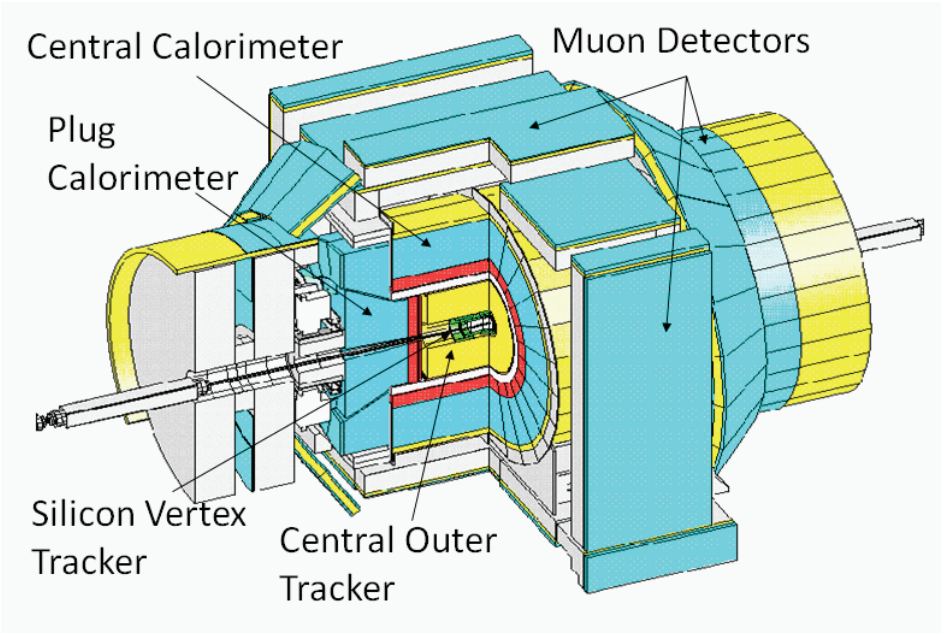


Figure 3.2: Schematic of CDF Detector.

3.2.1 Coordinate Systems and Transverse Planes

The CDF detector is illustrated in Figure 3.2 [19]. Due to its cylindrical symmetry, a cylindrical coordinate system is adopted when discussing the detector [20]. The z-axis is taken to be along the beam line, with θ chosen as the polar angle relative to the beam line, and ϕ, the azimuthal angle about the same axis. The trajectory of a particle within the detector can thus be described by θ, ϕ and z_0, the intersection point of the particle trajectory with the z-axis. Rather than reference the angle θ, it is often convenient to work with the pseudorapidity η, defined as $\eta = -\ln(\tan(\frac{\theta}{2}))$.

Within the detector, protons and antiproton bunches are collided at a rate of 2.5 MHz. Due to this high rate and limitations in computer speed and storage, not all events are recorded. Considering that protons and antiprotons move along the z-axis, after the collision, most particles continue moving along the z direction with slight deviation. Events that are more interesting from a

physics perspective are those where the resulting particles from the hard interaction move in a plane perpendicular to the beam line (the transverse plane). As a consequence, many measurements are made with regards to p_T and E_T, the particle's momentum and energy in the transverse plane, respectively.

3.2.2 Tracking System

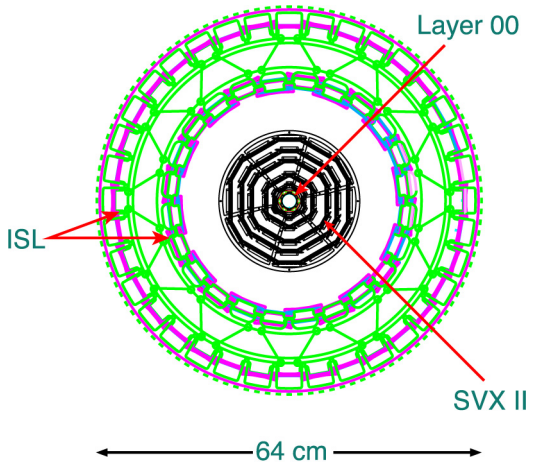


Figure 3.3: End view of the CDF silicon detector system.

Closest to the beam pipe lies the Silicon Vertex Tracker (SVX), a multi-layer silicon ionization detector [21]. It offers the best tracking resolution in the detector and is used for the identification of particle trajectories and decay vertices. The SVX is composed of three separate sub-detectors: Layer 00, SVX II and the Intermediate Silicon Layers (ISL). Layer 00 is a 300 μm thick layer of silicon sensors. It is cylindrical in shape with a radius of 1.3 cm and is the closest point of the detector to the beam crossing interaction point. SVX II is composed of five layers of silicon located at radii of 2.4 to 10.7 cm from the centre of the detector. It is separated into three barrels, each of which is further separated in 12 wedges in ϕ [20]. The SVX II provides coverage of roughly $|z| = 45$ cm on both sides of the interaction point. Finally, the ISL provides two additional layers of silicon detectors at radii of 20.2 and 29.1 from the interaction point. The entire silicon system provides high resolution

tracking in the range $|\eta| < 2$. To give a better indication of the location of SVX subcomponents, the entire SVX system of CDF is illustrated in Figure 3.3 [21].

Surrounding the SVX is the Central Outer Tracker (COT), an open-cell drift chamber filled with argon and ethane. The COT provides coverage in the regions of $|z| < 155$ cm and $|\eta| \lesssim 1$ at a radial distance of $40 < r < 137$ cm from the interaction point [21]. There are eight concentric superlayers (the positions of which are indicated in Figure 3.4 [21]), each consisting of 12 layers of sense wires. As charged particles pass through the argon-ethane mixture, the produced ionization is measured by the sense wires. The sense wires in the superlayers alternate between a configuration where all of the wires are in the purely z direction and another where the wires are at a 2° angle relative to the beam line. At each end of the wires sits an aluminum plate that holds both the sense wires as well as sheets held at a potential to generate a 1.9 kV/cm electric field [20, 21]. By determining which of the sense wires registered a signal, a particle's trajectory through the COT (track) can be determined. This COT track can be traced back to the SVX as well, so as to provide detailed tracking information near the beam line.

Both the SVX and the COT lie inside a 5 m long superconducting solenoid [20]. This device creates a magnetic field of 1.4 T, so that based on a track's curvature and direction, the particle's charge and momentum can be determined. The SVX and the COT are collectively referred to as the *tracking system* of the detector. Since these two detectors are ionization based, uncharged particles are not detected until they exit the solenoid and enter the calorimeters.

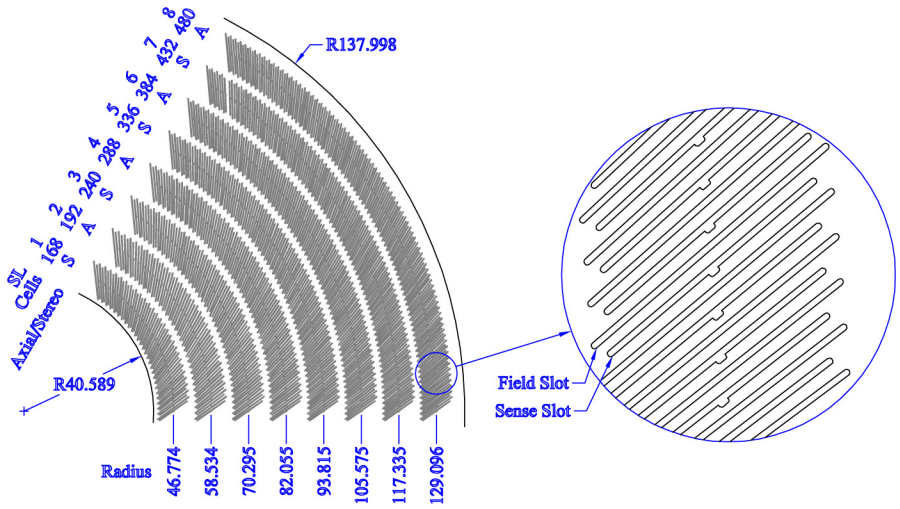


Figure 3.4: Cross-sectional view of the COT. Take note of the radial distance from the z-axis of the layers (distances in cm).

3.2.3 Calorimetry

The CDF calorimetry system is scintillator based with separate electromagnetic (EM) and hadronic sections and is used to measure particle energy deposition [21]. The EM portions comprise polystyrene scintillator sandwiched between sheets of lead, while the hadronic calorimeter, designed to stop more massive hadrons, has its scintillator sheets sandwiched between steel [20]. The actual detection elements that read the scintillation are known as "towers" and they extend through both the EM and hadronic sections [22].

The lead and steel in the calorimeters serve the purpose of not only slowing down the particles so that they deposit their energy in the calorimeters, but also to cause *showers*. When an incoming particle interacts with the metal in the calorimeters, it may produce other particles that will further interact with each other as well as the lead or steel. This cascade of particles is known as a shower and helps to disperse the energy deposited in the detector.

There are two sections to the CDF calorimetry system, each covering different regions in η. The central barrel calorimeter covers $|\eta| < 1$ and the plug

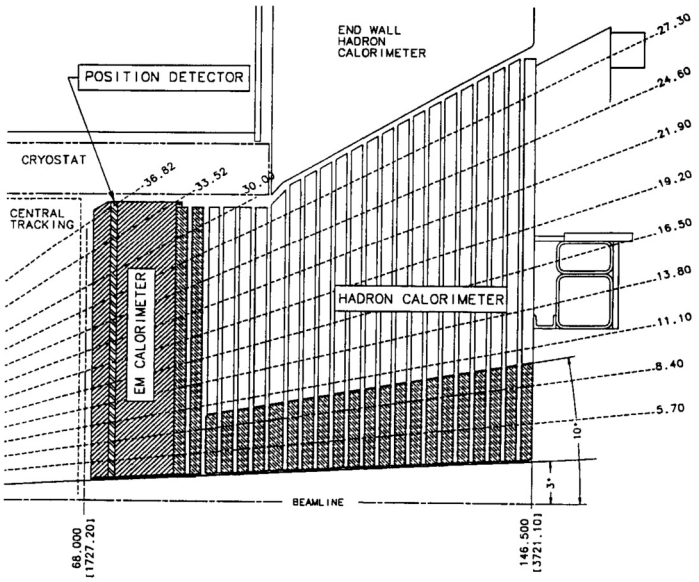


Figure 3.5: Cross section of the upper section of the plug calorimeter. Notice the division between the EM and hadronic portions.

calorimeters cover $1.1 < |\eta| < 3.64$ [20]. For the barrel calorimeter, the scintillator plates are parallel to the beam line, whereas in the plug calorimeters, the plates are in the transverse direction. A cross section of the plug calorimeters is provided in Figure 3.5 [22].

3.2.4 Muon System

The outermost portion of CDF is the Muon Detector. This combination of wire drift chambers and scintillators is located behind a thick layer of steel absorber. In order to penetrate this absorber in the far ends of the detector, a p_T of greater than 3.0 GeV is required [20]. Any charged particle that can get this far into the detector but still be stopped is most likely a muon. A schematic representing various types of particles interacting with the different detector components is seen in Figure 3.6 [23].

3.2.5 Trigger System

Due to the large number of collisions that occur within CDF and the limitations in computer processing and storage, it is not possible to record

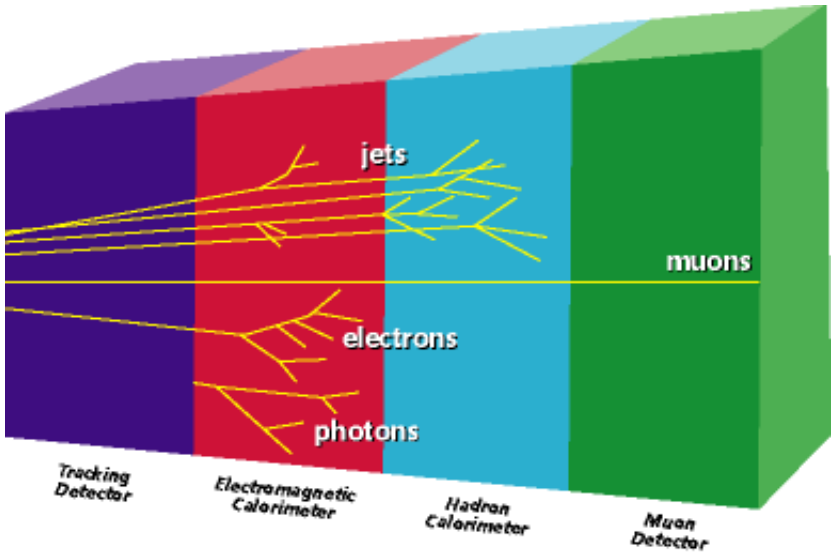


Figure 3.6: Particle penetration into different detector components. Note that the uncharged photon is not detected by CDF until it reaches the EM calorimeter.

all of the events. The triggering system reduces the raw event rate in the detector of 2.5 MHz to 75 Hz of data recording [20]. The system has three stages; the first being purely hardware based, the second being a combination of hardware and software and the third being a farm of processors performing event reconstruction [21].

The level 1 trigger is used to select events based upon the information obtained by the tracking, calorimetry and muon systems. There are three parallel hardware processing streams that are used to make decisions. The first identifies objects of interest within the calorimeters based upon the energy deposit patterns read by the calorimeter towers. In addition to the deposition geometry, further decisions to keep the event can be made upon a rough estimate of the candidate object's E_T by assuming a collision vertex of $z = 0$. The second stream selects muon candidates via p_T thresholds and timing of signals between different portions of the muon chambers [20, 21]. The third stream is the eXtremely Fast Tracker (XFT) which uses information from specific COT superlayers to construct tracks and further extrapolate them to their position

in the calorimetry and muon systems. Coincidences between XFT data and the calorimeter object and muon candidates identified from the other two streams are used to select the events to pass on to the level 2 trigger. Of the initial 2.5 MHz of events, the level 1 trigger selects 20 kHz to pass to the level 2 trigger.

The level 2 trigger (L2) uses a more sophisticated algorithm for the reconstruction of electromagnetic objects within the calorimeter compared to that used in level 1. This allows for greater energy resolution of the calorimeter objects [21]. The level 2 trigger also uses SVX information to search for events with displaced tracks. The calorimeter objects and tracking information are then combined with the data from the level 1 trigger and the level 2 trigger software selects events of potential interest based on E_T, p_T, geometry and tracking considerations. The maximum L2 acceptance rate is 300 Hz [20]. These selected events are then passed to the level 3 trigger.

The level 3 trigger attempts to fully reconstruct the event and determine the trajectories of the objects produced from the $p\bar{p}$ collision as they travel through the various detector subcomponents. Selection is made based on a variety of E_T and p_T thresholds as well as tracking characteristics. Ultimately, the level 3 trigger selects the 75 Hz of events that are recorded for further event reconstruction and processing [20].

It should be noted that the original CDF trigger system has undergone upgrades and improvements since the initial Run II data taking of the CDF detector in 2002. Most notable was the L2 tracking upgrade introduced in 2008. This upgrade gave the L2 trigger an additional stream of data from the Level 1 XFT. In essence, the L2 trigger was given greater track resolution information with regards to ϕ and θ [24]. Consequently, the level 2 trigger was able to improve its three dimensional reconstruction of tracks within the detector.

Once the events selected from the level 3 trigger are written to disk, the data are not ready for use in physics analyses. They must first undergo a reconstruction process of greater thoroughness than that performed by the level 3 trigger. Once that reconstruction is complete, we can then proceed in selecting the events that may represent the $t\bar{t}$ signal of interest. Now that we understand the principal detector components, we can discuss how the detector subsystem signals are used for particle identification and event reconstruction.

Chapter 4
Event Reconstruction

Although much information is provided by the raw signals obtained from the detector, a single SVX track or energy deposit in a calorimeter is not sufficient to indicate that a top quark event has occurred. Instead, we must piece together evidence from several detector subsystems in order to obtain a candidate signal event. This chapter discusses the various constituents of the detector signature relevant for $t\bar{t}$ identification at CDF. We conclude the chapter by describing the lepton+jets event selection signal that is used as the primary identifier for top quark events within this thesis.

4.1 Collision Vertices

Collisions from the Tevatron are not simple affairs where one proton collides head-on with an antiproton. The width of the beams at the interaction point is approximately 30 μm. This is too large to allow a one-to-one interaction interaction between a single proton and antiproton. Instead, the beams are composed of bunches of protons and antiprotons and at the interaction point, these bunches are crossed within the detector [15]. An illustration of the crossing of beams and the collision of bunches is seen in Figure 4.1 [25]. Since there is space within the bunches, most of the particles from one bunch do not interact with those from the other during crossing. Statistically, however, at each bunch crossing, one proton will interact with one antiproton. The location of this interaction is known as the vertex of the hard interaction. If multiple particles in the bunch have collided, there will be multiple vertices.

Given that the SVX is the closest component of detector to the colliding beams, it is instrumental in identifying vertices. Recall that the SVX is an

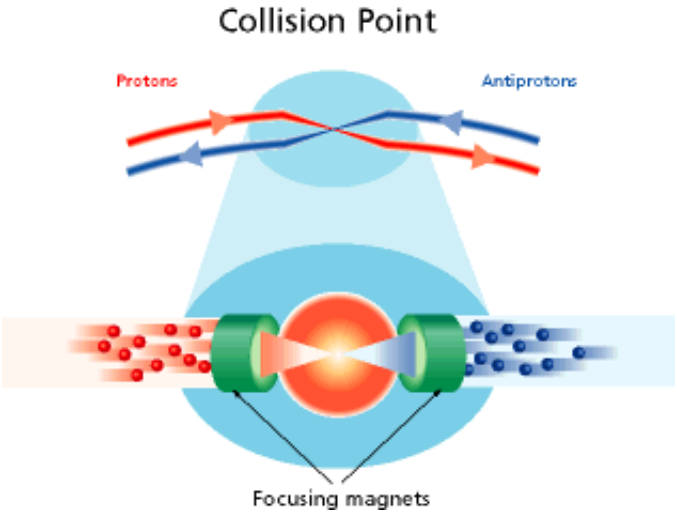


Figure 4.1: An artistic representation of beam crossings at CDF. Notice that the particles in each beam are collected in bunches. The collisions of individual particles from these bunches lead to vertices observed in the SVX.

ionization detector, so as charged particles interact with it, they leave tracks corresponding to their trajectories. A vertex corresponds to the intersection of such tracks. Since there may be multiple vertices per bunch crossing, the vertex with the highest p_T value is labeled as the primary vertex of the interaction while all other vertices are referred to as secondary vertices.

Not all of the secondary vertices represent hard interactions, especially if they are located off of the beam line; they may in fact relate to the decay of particles formed as a result of the hard interaction. Some of these secondary vertices actually correspond to b quarks. At CDF, since many analyses rely upon the identification of such quarks, certain algorithms, such as SecVtx, have been developed to identify (or tag) b quarks. SecVtx will be discussed in greater detail in Section 4.4.

4.2 Charged Leptons

Although the electron, muon and tauon (and their antiparticles) are all charged leptons, in this thesis, we shall restrict the use of the term to electrons and muons (and their antiparticles). In general, the identification of electrons

is performed in the EM calorimeter. During interaction with the detector, a high momentum electron will leave isolated energy deposits in small contiguous groups of towers (known as a "cluster") [26]. The signature used to identify the electron is a signal seen in a sequence of clusters in the central EM calorimeter in the pseudorapidity range of $|\eta| < 1.1$. The spatial distribution of the trajectory within the calorimeter can also be traced back to tracks within the COT and the SVX in order to determine the electron's path closer to the beam line.

Muons are detected by the presence of an ionization signal within the muon drift chambers on the outside of the detector. Any charged particle that is recorded by this detector subsystem is most likely a muon. The CDF muon system is set up as a series of four-layer stacks of drift chambers. The reconstructed muon is seen as a "stub", a line segment through one of the four-layer stacks [26]. The movement of the muon can be traced back to the energy depositions in the calorimeter and the tracking system to reconstruct the complete trajectory of the particle.

Muon and electron candidates can be discerned by the CDF trigger system. Specific triggers can be set to select events that show the presence of a certain number of charged leptons or events whose charged leptons satisfy certain p_T threshold requirements.

4.3 Jets and Jet Energy Corrections

When a particle undergoes decay or hadronization within CDF, due to conservation of momentum, the daughter particles typically travel in approximately the same direction as the parent. A jet is a roughly collinear collection of particles that results from the decay of a quark or gluon (both of which are partons) [27]. Due to the principle of conservation of energy, the energy of the jet's constituent particles should equal the energy of the parent parton.

A cartoon depiction of a jet formed after a collision is presented in Figure 4.2 [28].

A jet is a collection of objects that is treated as a single entity. Depending on the particular analysis at hand, this may make calculations easier to perform because instead of keeping track of multiple properties of different particles, one needs only to account for the characteristics of a single object. As a result, we commonly discuss such jet properties as jet E_T or jet p_T. Jets can be crudely reconstructed in the triggering system and their presence can be used in the decision to keep an event or not. Spatially, a jet is defined through the creation of a jet cone. The vertex of the cone is situated at the location of the parent parton. The centroid corresponds to a position in η-ϕ space with coordinates of (η^{jet}, ϕ^{jet}). Similarly, the location of the other extremity of the cone is given by $(\eta^{tower}, \phi^{tower})$, the location of the tower within the calorimeter where the furthest of the jet's particles have been detected. The jet axis is defined by the line that goes from the jet vertex to the base of the cone. Finally, the size of the jet is given by the radius R, defined as $R = \sqrt{(\eta^{tower} - \eta^{jet})^2 + (\phi^{tower} - \phi^{jet})^2}$. R is a dimensionless quantity, with R values of 0.4, 0.7 and 1.0 in use at CDF [29].

Given this definition, it is not surprising that there are errors introduced when calculating the jet energy. For example, one can easily imagine that, within the jet cone, there may be particles that are not due to the decay of the parent parton but come from other events within the detector. Thus, the raw jet energy that is measured by the calorimeters needs to undergo a correction before it can be used in calculations in physics analyses. At CDF a variety of calculations are applied, which correct for the following issues [29]:

- Calorimeters having a non-homogeneous response in η-space.
- Different calorimeter responses to jets of different momentum.

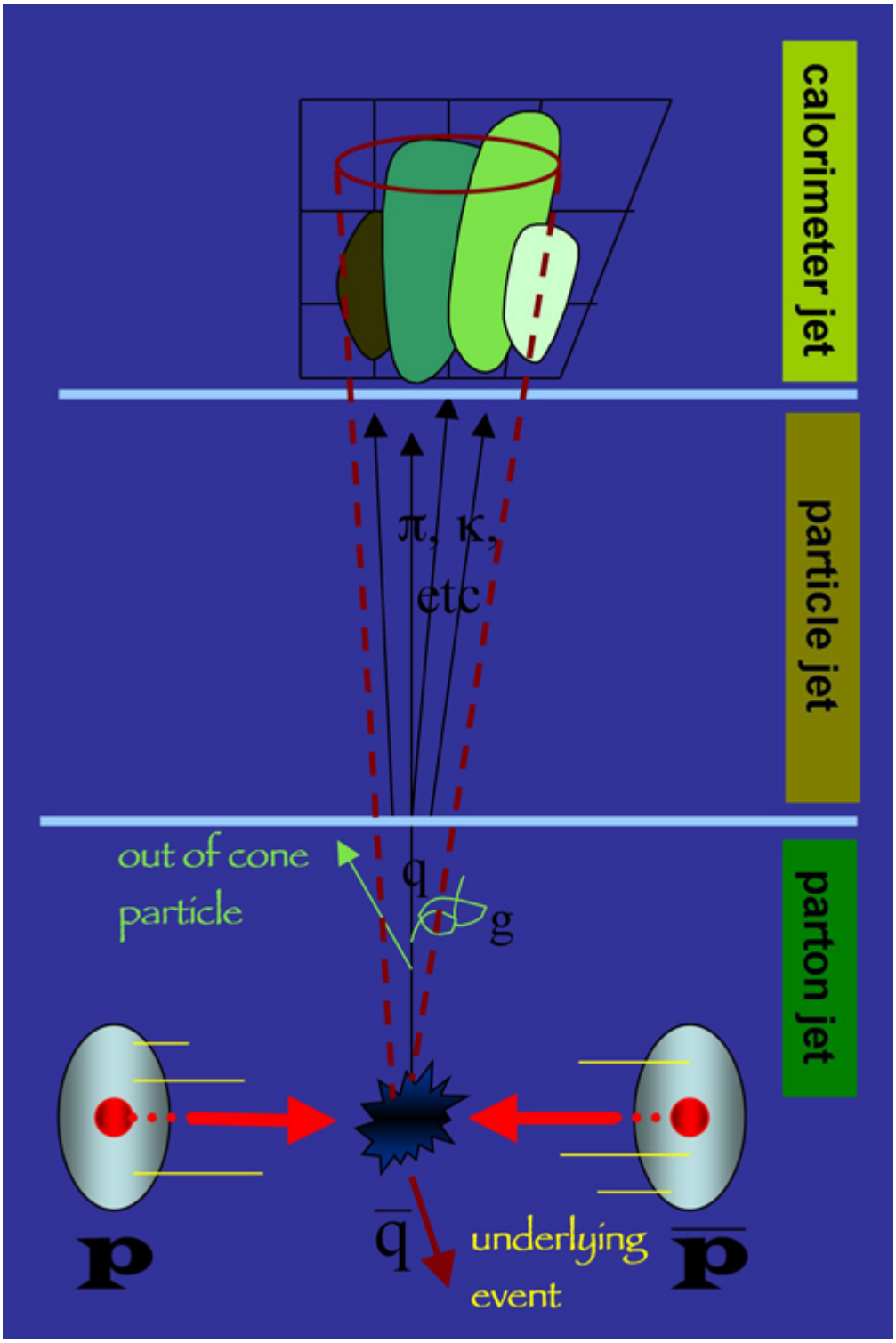


Figure 4.2: Cartoon depiction of a jet. The resulting quark created from the hard interaction (denoted as "underlying event" in the diagram) decays and forms other particles that travel in a roughly linear direction. The shape of the jet cone has been traced, extending from the quark vertex to the energy deposited in the calorimeter.

- Detector-independent parton radiation events.
- Particles related to the jet's parent parton that may lie outside of the jet cone.
- Multiple hard interactions when the Tevatron beams are crossed; particles unrelated to the parent parton may be found inside the jet cone.

When we refer to jet energy in this thesis, unless it is otherwise specified, we mean the corrected jet energy, after it has been subjected to the above corrections.

4.4 SecVtx and *b*-tagging

There are multiple algorithms in use at CDF for the identification of b quarks (a process known as *b-tagging*). The analysis in this thesis, specifically, uses the SecVtx algorithm that was originally developed at CDF for use in the top quark discovery searches [26]. Since then, SecVtx has undergone continuous improvement and remains the most commonly used b-tagger in top quark analyses.

SecVtx exploits the fact that b quarks are relatively long-lived particles. Primarily, it looks at the displacement of secondary event vertices relative to the primary vertex corresponding to the hard interaction [26]. The exact transverse position of the primary vertex is determined by looking at the tracks located within ± 1 cm in the z direction of the best guess of the vertex's location. The impact parameter, d_0, is a measure of the distance from the beam line to a track's closest approach in the transverse plane. It carries an associated uncertainty, σ_{d_0}, which includes both the uncertainty on the track's position as well as the beam line's position. The location of the vertex is first fitted by using all of the tracks in the above specified z-range and where $|d_0/\sigma_{d_0}| < 3$. Any track which contributes $\chi^2 > 10$ to the fit (or if χ^2/(degrees of freedom)> 5, the track which has the largest χ^2 contribution) is ignored

and the fit is performed again using the remaining tracks. This recursive fitting of the vertex position is repeated until no tracks with an individual χ^2 contribution greater than 10 remains; if all of the tracks are eliminated, the beam line profile is used to estimate the primary vertex position. The uncertainty of such fits for the primary vertex position is of order 10 μm [26, 30].

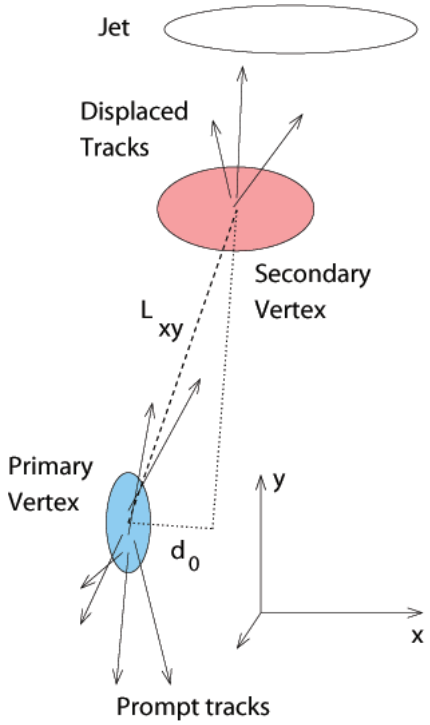


Figure 4.3: A diagram representing the displacement of the secondary vertex with regards to the primary. The presence of displaced tracks (tracks whose intersection clearly does not seem to be located at the primary vertex) is used to calculate the location of the secondary vertex.

In order to identify secondary vertices, SecVtx looks for the presence of jets. Tracks within the jet cone are identified and are evaluated against a selection of criteria relating to their p_T, the number and quality of silicon tracks and the χ^2/(degrees of freedom) of the final track fit [26]. A jet is deemed "taggable" only if it has at least two tracks that satisfy the criteria. The algorithm then uses these selected tracks to identify the secondary vertex.

Initially, it attempts to reconstruct the location of the secondary vertex by using three tracks, each with a $p_T > 0.5$ GeV (and at least one with a $p_T > 1$ GeV) and $|d_0/\sigma_{d_0}| > 2.5$. If reconstruction is unsuccessful, or the jet only has two usable tracks, an attempt is performed using two tracks with a minimum p_T of 1 GeV (and at least one with a $p_T > 1.5$ GeV) and $|d_0/\sigma_{d_0}| > 3.5$. Figure 4.3 [31] presents a depiction of the displaced tracks used to reconstruct the secondary vertex.

Once a secondary vertex is identified in relation to a jet, the SecVtx algorithm calculates its two-dimensional decay length, L_{2D}. This value corresponds to the projection onto the jet axis of the vector pointing from the primary vertex to the secondary vertex. L_{2D} is assigned a sign relative to the jet direction: it is positive if the absolute difference in ϕ between the jet axis and the secondary vertex vector is less than 90° and negative if it is greater than 90°. Jets that correspond to the decay of B hadrons (hadrons containing a b quark) are expected to have large, positive values of L_{2D} due to the long lifetime of the b quark [26]. In order to reduce the number of mistagged jets from this procedure, a secondary vertex is only deemed to be a "good vertex" corresponding to a heavy quark if the quality ratio $L_{2D}/\sigma_{L_{2D}} > 3$, where $\sigma_{L_{2D}}$ is the total estimated error on L_{2D}. As expected, this error includes contributions from the error in determining the locations of the primary and secondary vertices. Any given jet will have at most one good vertex. If it does, that jet is said to have been b-tagged and we consider it to have a b quark present.

It should be noted that b-tagging can be defined as either "tight" or "loose" depending on the quality of the impact parameters. The method described above corresponds to the tight tagging scenario where, if only two usable jet tracks are present, $|d_0/\sigma_{d_0}| > 3.5$. For loose tagging, this quality ratio is reduced slightly to $|d_0/\sigma_{d_0}| > 3.0$ [30]. Loose tagging, which permits

an increased number of falsely tagged jets, does allow for the identification of a greater number of b-tagging candidates and may be useful for analyses that are statistically limited.

4.5 Missing Transverse Energy

Neutrinos are undetectable by CDF. They interact very weakly with matter and, despite the lead and steel shielding in the calorimetry and muon systems, neutrinos pass through undetected. Certain particle decay processes involving the W boson lead to the emission of neutrinos, thus it is important to have a means of indirectly identifying their presence.

The beams of protons and antiprotons that collide within the detector have momentum vectors almost purely within the z direction. The transverse momentum and transverse energy of the beam particles are negligible. Thus, due to conservation of energy and momentum, we expect that for every jet or track with transverse energy, there must be another jet or track positioned in such a way so that the sum of momentum in the transverse plane is 0. If no such track is found, the event is said to contain *missing transverse energy* ($\not\!\!E_T$). This $\not\!\!E_T$ is taken either as the signal of a neutrino or a loss of one or more particles through uninstrumented regions of the detector. $\not\!\!E_T$ can be identified by the trigger system and used in the decision of whether to record an event or not.

Typically, neutrinos are produced in association with the corresponding charged lepton from the same generation [2]. It should be noted that when $\not\!\!E_T$ is calculated with relation to electron production, since the electron is stopped by the detector's calorimetry system, the energy deposited in the calorimeter is used in the $\not\!\!E_T$ calculation. On the other hand, when a muon is involved in the event, the $\not\!\!E_T$ calculation must take into account not only the ionization signals

recorded by the muon chamber, but also the significantly smaller amount of energy deposited by the muon in the calorimeters.

4.6 Event Selection: Lepton + Jets

Using the methods described in this chapter, the particle decays and interactions that occur after the hard interaction can be reconstructed. Through the identification of vertices, jets, leptons and b quarks, one can combine this information to create specific search signatures. Since we are primarily concerned with $t\bar{t}$ events, in this thesis we select events that have a signature that most likely corresponds to the decay of a top quark. In Figure 4.4 [7], we see a possible $t\bar{t}$ pair decay; one top quark decays into a b quark, a charged lepton and a neutrino (leptonic W decay) and the other decays into a b quark as well as two other quarks (hadronic W decay). Such an event can be identified within the detector via the presence of a charged lepton, $\not\!\!E_T$ and jets (due to the hadronization of the quarks). This is known as the *lepton + jets* signature and is the signal that is used in this analysis. There are other possible $t\bar{t}$ topologies that may occur at CDF, such as the case where both of the virtual W bosons decay leptonically (the so-called di-lepton channel, characterized by charged leptons and $\not\!\!E_T$). However, events with these alternate signals are produced with a smaller signal-to-background ratio than lepton + jets events [32]. Thus, for the sake of simplicity, these other signal channels are not used in this thesis.

All lepton + jets events are not necessarily top quark events. We perform a selection of events that match certain criteria to increase the chances of looking at events that actually deal with $t\bar{t}$ production. Candidate events are selected via the high p_T lepton triggers as well as the $\not\!\!E_T$ + Jets trigger. The event must have one charged lepton with $p_T > 20$ GeV, $\not\!\!E_T$ of greater than 20 GeV and at least 3 jets in the $|\eta| < 2.0$ region of the detector with a

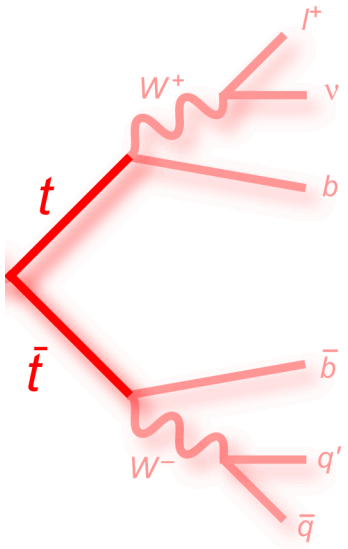


Figure 4.4: Decay signature of the $t\bar{t}$ lepton + jets selection. The presence of a charged lepton, $\not\!\!E_T$ and three or more jets characterizes the selection.

corrected $E_T > 20$ GeV. These thresholds are chosen taking into consideration the mass of the top quark and the knowledge that $t\bar{t}$ pairs at CDF are produced effectively at rest. A further constraint is placed that at least one of the jets has been b-tagged by SecVtx. Also, an additional cut is placed that the scalar sum of the p_T of the lepton and jets (H_T) be at least 250 GeV. Several background processes, such as W boson decays that are not top quark related, carry the lepton + jets signature and this final cut is intended to eliminate a portion of the contributions. From the 5.6 fb^{-1} of data in the original data set, approximately 5000 events are ultimately selected using these selection criteria.

Now that we have seen how events are reconstructed and particles are identified from the CDF data, we can focus our discussion on how these particles are used to perform the measurement of interest of this thesis.

Chapter 5
b-Tagging Efficiency Scale Factor and $t\bar{t}$ Production Cross Section

The primary result presented in this thesis is a simultaneous measurement of the b-tagging efficiency scale factor (SF) and the $t\bar{t}$ cross section at CDF. It is imperative that these two concepts be clearly elucidated before any discussion of the results. This chapter presents a description of the previous methods used to calculate SF at CDF and the motivation to perform a simultaneous measurement of SF with the $t\bar{t}$ cross section. Afterwards, a description of the method used to calculate the $t\bar{t}$ cross section will be presented with a focus on the estimation of backgrounds.

5.1 b-Tagging Efficiency Scale Factor

In experimental particle physics, Monte Carlo calculations play an important role in the analysis process. Given that most particle physics interactions are probabilistic in nature with several outcomes of varying degrees of probability, Monte Carlo calculations are an important way to model theoretical predictions as well as backgrounds. When b-tagging is performed using SecVtx (or any b-tagging algorithm), it is possible that the tagging behaves differently in data samples and Monte Carlo samples. Such differences are potentially problematic if one wishes to make comparisons between data and Monte Carlo for a quantity involving b-tagging.

In the b-tagging process, as described in Section 4.4, a jet is identified as having a b quark parent or not. However, this tagging is independent of reality; a tagged jet may in fact contain no b quark while an untagged jet may indeed possess one. The ability of the b-tagging algorithm to correctly identify the b quarks is quantified by its efficiency, ϵ. The efficiency is expressed as

a decimal value normalized to 1, where a value less than one means fewer b quarks are identified than are actually present, while greater than one means that the algorithm overestimates the number of b quarks. For SecVtx, the efficiency for tagging in data samples is approximately 0.4 [33].

5.1.1 Theoretical Description of the SF

The SF is the ratio $\epsilon_{data}/\epsilon_{MC}$, the ratio of the b-tagging efficiency as measured in data compared to the measurement in Monte Carlo. As its name implies, the scale factor allows for the proper scaling between b-tagging in data and Monte Carlo, meaning that adequate comparisons between the two can be made. Because the SF is an experimentally measured value, it carries an uncertainty. For analyses that use b-tagging, this uncertainty constitutes part of the overall systematic uncertainty for the analysis. It is evident, then, that any attempt to reduce the uncertainty on the SF will reduce the systematics associated with any analysis that uses b-tagging.

One of the current investigations that could greatly benefit from reduced systematics is the search for the Higgs boson. For example, in a recent analysis searching for a light mass Higgs boson of 115 GeV, at CDF, the expected number of signal events for certain decay channels involving a b quark is approximately 2. This value stands in stark contrast to the expected number of background events bearing the same signature, which is over 200 [34]. If one wishes to make a conclusive statement regarding the observation of a signal, the uncertainty on the number of background events must be reduced to a level less than the number of signal events. Thus, any reduction of the systematic uncertainties could provide assistance for such an analysis. A small signal compared to larger backgrounds is a common trait of all low mass Higgs boson searches at CDF.

Historically, at CDF, the SF has been measured using two distinct methods: one of which involves performing the measurement in a sample rich in electrons, while the other uses a muon-rich sample. Initially, the results obtained from these two different methods were averaged to obtain the official CDF SF value. However, since 2008, the Electron Method has been used exclusively as the sole method of SF measurement. We shall now discuss both of these methods in further detail.

5.1.2 The Electron Method

The currently established method of measuring the SF at CDF is the Electron Method. It is thus named because it uses a sample rich in electrons. Events are selected that contain two jets with $E_T > 5$ GeV that lie "back-to-back", meaning that they are separated by $\Delta\phi > 2$. One of these jets, named the "electron jet", must contain an electron with $E_T > 9$ GeV within a jet cone of $R = 0.4$. This electron may originate from two primary sources: from a photon that interacted with matter in the detector and created an electron-positron pair (photon conversion) or from the decay of heavy flavour hadrons (hadrons containing a b or a c quark) [35, 36]. The jet without the electron is known as the "away jet" and it may or may not be b-tagged by SecVtx.

In simple terms, this method exploits the fact that the electron jets with an electron from photon conversion are topologically similar to those from heavy flavour events. However, since the electrons from the former case can be readily identified, the observation of which electron jets are tagged by SecVtx gives a good estimate of the algorithm's efficiency.

The detailed calculation involves the following variables [35]:

- C, the number of events where the electron comes from a photon conversion.

- N, the number of events where the electron does not come from a photon conversion.
- N^+, N^-, the number of positively (contains a b quark) and negatively (contains no b quark) tagged electron jets. Similarly, C^+, C^- refer to positively and negatively tagged photon conversion events.
- N_+, N_-, the number of positively and negatively tagged away jets. Similarly, C_+, C_- refer to the positively and negatively tagged away jets
- α, an empirically determined mistag asymmetry correction term. α multiplied by the amount of negatively tagged jets gives the amount of positively tagged jets that were mistagged.
- F^{HF}, the fraction of electron jets that are due to heavy flavour events.
- F^{HF}_{a-tag}, the fraction of away jets that are due to heavy flavour events.
- ϵ, the tagging efficiency. ϵ_{data} and ϵ_{MC} are the respective tagging efficiencies in data and Monte Carlo.

The first step consists of separating the sample into a subsample where there are no tagged away jets. In such a sample, all b-tags are from the electron jet and the efficiency, ϵ_{1-tag}, can be calculated rather easily. We can then get F^{HF} via the following calculation [35]:

$$F^{HF} = \frac{N^+ - \alpha N^-}{N\epsilon_{1-tag}}. \tag{5.1}$$

Using this result, we can then calculate the heavy flavour fraction in the away jets using the following:

$$F^{HF}_{a-tag} = 1 - \frac{\frac{C_+}{N_+} - \frac{C^+ - \alpha C^-}{N^+ - \alpha N^-}}{\frac{C}{N} - \frac{C^+ - \alpha C^-}{N^+ - \alpha N^-}}(1 - F^{HF}). \tag{5.2}$$

Finally, we can use F^{HF}_{a-tag} to solve for the efficiency:

$$\epsilon = \frac{N^+_+ - \alpha N^-_+}{N_+} \cdot \frac{1}{F^{HF}_{a-tag}}. \tag{5.3}$$

The above procedure can be performed for data to obtain ϵ_{data}. The tagging efficiency in Monte Carlo, ϵ_{MC}, is significantly easier to calculate; one needs only to compare the known quantity of b quarks from the simulation to the amount that are actually tagged by SecVtx. The ratio of the two efficiencies constitutes the SF. Also, we will obtain different values of SF depending upon whether we are using a tight or loose SecVtx selection. The last time the SF calculation was performed at CDF using the Electron Method was in 2010. The results are summarized in the table below [35]:

SecVtx Tightness	SF	Uncertainty
Tight	0.959	$\pm$0.054
Loose	0.981	$\pm$0.069

Table 5.1: SF values obtained from the Electron Method in 2010. The uncertainties quoted include both systematic and statistical contributions.

5.1.3 The Muon Method

An alternative method of measuring the SF that was employed at CDF is the Muon Method. Similar to the Electron Method, the approach consists of using a sample of jets that have a charged lepton to independently calculate the quantity of b quarks present and comparing this result to the value obtained from SecVtx b-tagging. As the name implies, the Muon Method uses a sample of jets that have a muon candidate. Specifically, the sample must contain a jet of E_T greater than 9 GeV with a muon with a track of $p_T > 9$ GeV (the "muon jet"). The muon jet must have an R value of 0.4 and the muon must lie in the region $|\eta| < 0.6$. In the sample, the muon jet must be accompanied by an "away jet" of $E_T > 15$ GeV. This jet must lie in $|\eta| < 1.5$ and have the further requirement that $|\phi_{away-jet} - \phi_{\mu-jet}| > 2.0$. In addition, the away jet must have at least two good SecVtx tracks using the loose tagging requirements [37].

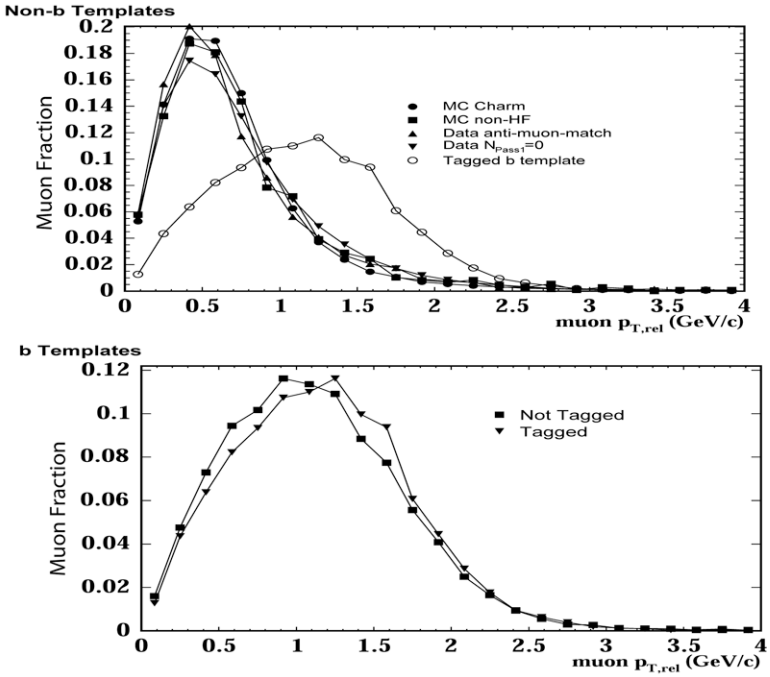


Figure 5.1: Template fitting for the Muon Method. By comparing the contributions to the $p_{T,rel}$ spectrum from both the b templates and non-b templates, one can calculate the number of b quarks within the sample. Notice how the $p_{T,rel}$ distributions are dramatically different in shape for the non b contributions.

One of the signatures of a muon from a B hadron decay is that it has a high transverse momentum relative to the jet axis ($p_{T,rel}$). Decays from hadrons without a b quark produce muons with less relative transverse momentum [37]. Thus, an examination of the $p_{T,rel}$ spectrum of the muons relative to their associated jets can determine the quantity of b quarks within a sample. The first step in the Muon Method consists of separating the samples based upon whether the muon jet is b-tagged or not. In each of these subgroups, the $p_{T,rel}$ spectrum (the distribution of $p_{T,rel}$ amongst the muons in the sample) has two distinct contributions: those from muons related to B hadron decay and those that are not. Using Monte Carlo calculations, one can calculate the $p_{T,rel}$ distributions from b-related processes (b templates) and non-b-related processes (non-b templates). By fitting the data's $p_{T,rel}$ distribution to the

Monte Carlo templates (an example of such fits is presented in Figure 5.1 [37]), one can calculate the number of b quarks in the tagged and untagged samples. ϵ_{data} is then found by dividing the number of b jets from the tagged subsample to the sum of the b jets in the tagged and untagged subsamples [37].

The calculation of ϵ_{MC} for the muon sample is rather straightforward; similar to the Electron Method case, the amount of b quarks generated in the Monte Carlo sample is known and this can be easily compared to the number obtained from running SecVtx on the sample. The SF is then calculated by dividing the two efficiency values. The last time the Muon Method was accepted for use in CDF analyses was in 2007. The results are presented in the table below [37].

SecVtx Tightness	SF	Uncertainty
Tight	0.932	$\pm$0.051
Loose	0.944	$\pm$0.053

Table 5.2: SF values obtained from the Muon Method in 2007. The uncertainties quoted include both systematic and statistical contributions.

5.1.4 Method Discrepancies

As of February 2010, the complete data set at CDF had an integrated luminosity of 5.6 fb^{-1}. This data set is separated into 29 data taking periods (ranging from period 0 to period 28), each generally covering a span of several months. For analyses using data up to the end of period 17 (the period ending the 16th of April, 2008), the official SF value was obtained by combining the results of the Electron Method and the Muon Method [38]. Prior to the end of period 17, the two methods were used in tandem and provided results that were consistent with each other within error. However, after period 17, a rift began to form between the two different methods.

The difference between the results is particularly dramatic if we measure the scale factor exclusively for periods 18-28. The results from the latest scale factor calculations for this period are presented in Table 5.3 [39, 40].

SF Type	Value	Statistical Uncertainty
Electron Method Tight	0.895	±0.016
Muon Method Tight	0.789	±0.012
Electron Method Loose	0.944	±0.015
Muon Method Loose	0.826	±0.012

Table 5.3: SF values calculated in 2010 for data periods 18-28, grouped by tightness. Notice how the values calculated from the Electron Method do not agree with those obtained from the Muon Method.

If we compare these results to those seen in Table 5.1 (whose results come from data for periods 0-28), we quickly note that the scale factor as measured in periods 18-28 has decreased compared to the results spanning periods 0-28. The fact that a change was noticed in periods 18-28 is not completely surprising given that were was a substantial upgrade to the CDF triggering system after period 17, allowing for better track reconstruction (as discussed in Section 3.2.5). What is perhaps more surprising is the large difference between the SF results obtained from the Muon Method compared to the Electron Method for both the tight and the loose tagging cases. We expect the results of both methods to agree with each other because the SecVtx b-tagging efficiency SF is a property of the tagging algorithm, independent of the methods used to measure it. It should be mentioned that the entries in Table 5.3 do not include systematic uncertainties. However, their inclusion would probably not cause the Electron Method to agree with the Muon Method. If we assume these systematics to be ±0.07 (a rounding of the largest total uncertainty in Table 5.1), the Electron and Muon Method are still not in mutual accord.

This particular discrepancy has never been explained to satisfaction. Although the official SecVtx SF used by CDF used to be made based on a combination of results from the Electron and Muon Methods, today, the results of the Electron Method are used exclusively as the official scale factor value. However, there is no concrete evidence that this is the correct choice; it is possible that the actual SF lies closer to the value predicted from the Muon Method, or radically different from both methods.

The measurement performed in this thesis allows for the determination of the scale factor independently from the results obtained by the muon and Electron Methods. It should allow us to confirm whether a drop in scale factor occurred after period 17 and whether the choice of using exclusively the SF results obtained from the Electron Method is appropriate. Furthermore, it may also provide lower uncertainty than the Electron Method.

5.2 The Top-Antitop Production Cross Section

The majority of top quarks that are produced at the Tevatron are produced in $t\bar{t}$ pairs. A measure of the probability of such a process occurring is given by the production cross section (σ). We shall now discuss the cross section in slightly more quantitative terms, as well as the method used in this thesis to measure it.

5.2.1 Theoretical Description of the $t\bar{t}$ Cross Section

Let us consider, for example, the following interaction where two particles (generically named 1 and 2) collide [41]:

$$1 + 2 \rightarrow Products. \tag{5.4}$$

The cross section is given by

$$\sigma = \frac{Transition\ Rate}{Incident\ Particle\ Flux}. \tag{5.5}$$

In the above expression, the transition rate is the number of times that the interaction occurs within a unit time. The incident particle flux is the number of 1 and 2 particles per unit area per unit time. Consequently, the cross section has units of area. In particle physics, as discussed in Section 3.1, the common unit of area is the barn (b), where 1 barn = 10^{-28} m^2.

Since the products of Equation 5.4 may be formed at a variety of angles and have different angular distributions, it is important to consider the differential cross section, $d\sigma/d\Omega$. This is simply the number of particles that scatter over a particular solid angle Ω. An integration over all solid angles should give the total cross section σ.

In a purely general quantum mechanical form, for an interaction of the form

$$1 + 2 \rightarrow 3 + 4 + ... + n, \tag{5.6}$$

$d\sigma$ is given by [4]:

$$d\sigma = \frac{|\mathcal{M}|^2 S}{4 \cdot \sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2)^2}} \left[\prod_{i=3}^{n} \left(\frac{d^3\mathbf{p}_i}{(2\pi)^3 2E_i}\right)\right] \times (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4 - ... - p_n). \tag{5.7}$$

In the above expression,

- S is a statistical counting factor of value $1/k!$ for each group of k identical particles
- p_i is the 4-momentum (a 4-vector containing energy and three directions of momentum) of the ith particle
- $\mathbf{p}_i$ is the momentum of the ith particle
- E_i is the energy of the ith particle
- $\delta^4(x)$ is an energy-conserving Dirac delta function

- $\mathcal{M}$ is the quantum mechanical matrix element, which can be calculated given the value of certain Standard Model constants and particle masses.

In the case where there are only two particles in the final state (like in $t\bar{t}$ production), in the center-of-momentum frame of reference, Equation 5.7 simplifies to

$$\frac{d\sigma}{d\Omega} = \left(\frac{1}{64\pi^2}\right) \frac{|\mathcal{M}|^2 S}{(E_1 + E_2)^2} \frac{|\mathbf{p}_f|}{|\mathbf{p}_i|}, \tag{5.8}$$

where E_1 and E_2 correspond to the energy of the incoming particles and $\mathbf{p}_i$ and $\mathbf{p}_f$ correspond to the initial and final momenta of either of the outgoing particles. The integration required to calculate σ is actually quite complicated to perform for collisions at hadron colliders. Particles 1 and 2 are not the protons and antiprotons of the hard interaction, but actually the quarks and the gluons within the hadrons. The momentum of these partons is modeled by Parton Distribution Functions (PDFs), which are necessary to include in the cross section calculation [1].

To make matters more complicated, it is not only the processes described in Figures 2.4 and 2.5 that produce $t\bar{t}$ pairs. These are the leading order diagrams representing $t\bar{t}$ production, but there are infinitely more processes whose diagrams have the same beginning and final particles but different interactions in between. While such diagrams contribute to the cross section, the general rule is that the more complex the diagram, the smaller its contribution. Therefore, it is not uncommon to see the calculation performed only at leading order, or at next-to-leading order (NLO), meaning the leading order diagram as well as the next most significant contribution.

The latest NLO theoretical calculation of the $t\bar{t}$ production cross section for the Tevatron center-of-mass energy of 1.96 TeV (assuming a top quark mass of 175 GeV) [42] is

$$\sigma_{t\bar{t}} = 6.7^{+0.7}_{-0.9}\ \text{pb}. \tag{5.9}$$

5.2.2 Prior $\sigma_{t\bar{t}}$ Measurements at CDF

It is interesting to compare the theoretical NLO calculation of the $t\bar{t}$ cross section to the results that have been obtained experimentally. The best measurement to date is a combination of four CDF analyses performed using 2.9 to 4.6 fb^{-1} of $p\bar{p}$ collision data at a center-of-mass energy of 1.96 TeV. The analysis in question combines the four following results [43]:

- Measurement of $\sigma_{t\bar{t}}$ using lepton + jets events where the signal events have been distinguished from background events through the use of an artificial neural network (a computer algorithm that can be trained to identify patterns).
- Measurement of $\sigma_{t\bar{t}}$ using lepton + jets events and SecVtx b-tagging.
- Measurement of $\sigma_{t\bar{t}}$ using dilepton events (where the top and antitop quarks each decay into a lighter quark and a W boson and both of these W bosons decay into a charged lepton and a neutrino).
- Measurement of $\sigma_{t\bar{t}}$ using all hadronic events (where the top and antitop quarks each decay into a lighter quark and a W boson and both of the W bosons decay into quarks).

Assuming a mass of the top quark of 172.5 GeV, the $t\bar{t}$ cross sections of the four constituent analyses are presented in Figure 5.2 [43].

The combination of the four measurements, including the uncertainties due to statistical and systematic considerations, as well as the uncertainty on the measurement of the integrated luminosity acquired by the detector, is

$$\sigma_{t\bar{t}} = 7.50 \pm 0.31(\text{stat}) \ \pm 0.34(\text{syst}) \ \pm 0.15(\text{lumi}) \text{ pb}. \tag{5.10}$$

By combining the uncertainties in quadrature, the $t\bar{t}$ production cross section value from CDF is

$$\sigma_{t\bar{t}} = 7.50 \pm 0.48 \text{ pb}. \tag{5.11}$$

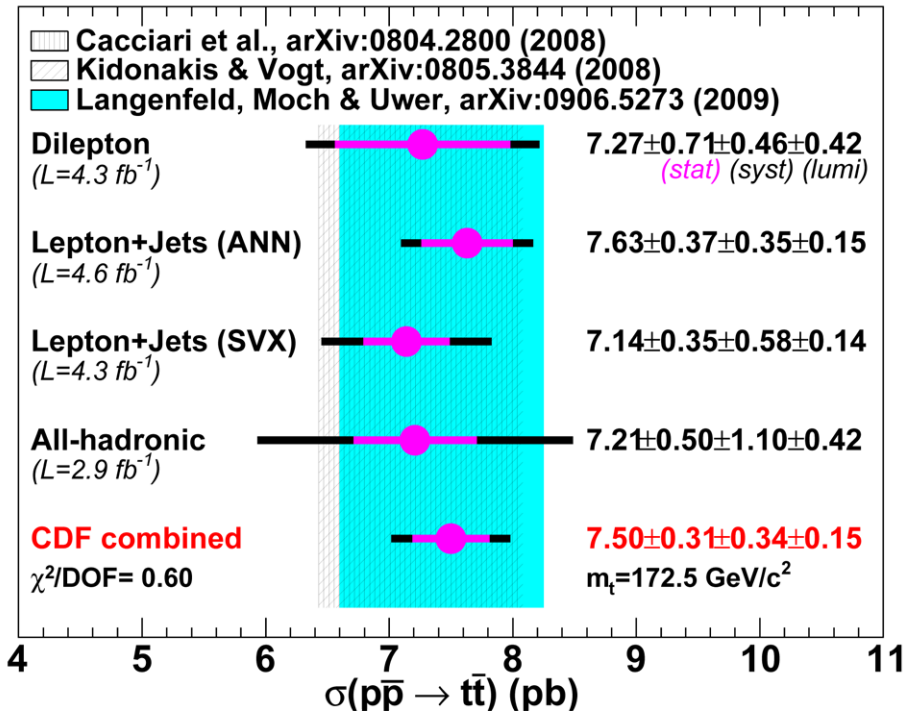


Figure 5.2: $t\bar{t}$ production cross section measurements from four distinct CDF analyses. The value in red corresponds to the combined $\sigma_{t\bar{t}}$ value.

This value agrees within uncertainty with the theoretical prediction presented in Section 5.2.1. However, the central value of the measurement, uncertainties notwithstanding, is higher than the theoretical NLO prediction and does not fall within the latter's uncertainty bounds. A portion of these differences can be explained through the use of different top quark mass assumptions. It may also be that the NLO prediction does not reasonably depict the measured processes and that diagrams of higher order have a significant effect.

5.2.3 Experimentally Measuring the Cross Section

Using the lepton + jets signal and the SecVtx algorithm for b-tagging, one can go about experimentally measuring $\sigma_{t\bar{t}}$ at CDF. Conceptually, the calculation involved in the measurement is quite straightforward. The cross section can be calculated from [26]:

$$\sigma_{t\bar{t}} = \frac{N_{obs} - N_{bkg}}{A \cdot \epsilon_{data} \cdot \mathcal{L}}. \tag{5.12}$$

In this expression, N_{obs} corresponds to the number of observed candidate events that pass the selection criteria, while N_{bkg} is the number of non $t\bar{t}$ events that pass the selection criteria. Hence, the numerator represents the total number of $t\bar{t}$ events within the data sample. In the denominator, ϵ_{data} is the b-tagging efficiency in data and $\mathcal{L}$ is the total integrated luminosity of the data sample. Finally, A is the acceptance. This value is defined as the fraction of produced $t\bar{t}$ events that satisfy all of the selection criteria. The acceptance is particularly important because not all of the $t\bar{t}$ events produced by the Tevatron will possess the lepton + jets signature; without the appropriate acceptance value, the measurement would not be of the total production cross section. Furthermore, A also includes the geometric acceptance of the detector. There are gaps between subsystems and uninstrumented regions, thus not all signal events are recorded by the detector. The acceptance value used is calculated through the use of Monte Carlo and a software-based geometry description of CDF [26].

If we assume that the Monte Carlo calculations are reasonably well understood and that the data's integrated luminosity is well known, we see that it is the calculation of N_{bkg} that causes the most difficulty in our determination of $\sigma_{t\bar{t}}$. The method used in this thesis is to construct a likelihood based on the data, $t\bar{t}$ cross section and the predicted background contributions for that cross section. The measured value of $\sigma_{t\bar{t}}$ and its associated statistical uncertainty are extracted from this likelihood. This particular procedure of calculating $\sigma_{t\bar{t}}$ is not novel, it has been performed at CDF before, albeit using a smaller data sample [44].

5.2.4 Background Estimation via Method 2

In order to estimate the background contributions to top quark production in the lepton + jets channel, at CDF one commonly uses the process known

as *Method 2 for You*, or simply "Method 2." This is a multi-step procedure that calculates the effects of various background processes of varying degrees of importance. For most steps in the process, backgrounds are calculated for both b-tagged events ("tag" label in the equations below) and untagged events ("pretag" label).

Several electroweak processes contribute a portion of the overall background. Most notably the production and subsequent decay of two W bosons, two Z bosons or a W and a Z boson (WW, ZZ, and WZ processes, respectively) all produce a charged lepton, neutrino and jets. Using the theoretical cross section value for each of these processes ($\sigma_{p\bar{p}\to X}$, where X represents the different electroweak processes), the number of predicted background events for a given integrated luminosity $\mathcal{L}$ is

$$N_{p\bar{p}\to X}^{pretag} = \sigma_{p\bar{p}\to X} \cdot A \cdot \mathcal{L} \tag{5.13}$$

$$N_{p\bar{p}\to X}^{tag} = \sigma_{p\bar{p}\to X} \cdot A \cdot \epsilon_{MC} \cdot \mathcal{L}. \tag{5.14}$$

In these expressions, A is the acceptance and ϵ_{MC} is the b-tagging efficiency in Monte Carlo. This calculation can be performed for each of the different electroweak processes mentioned above and summed to give the expected number of background events due to these processes. This total electroweak background contribution is illustrated in the equation below:

$$N_{ewk}^{tag,\,pretag} = \sum_{X} N_{p\bar{p}\to X}^{tag,\,pretag}. \tag{5.15}$$

A portion of the background contributions arises from events that do not involve the electroweak bosons. These events are the so-called quantum chromodynamic (QCD) background events. In order to estimate their contribution, Method 2 compares a $\not\!\!E_T$ distribution from a MC representing QCD processes to MC representing signal events. In a similar vein to the $p_{T,rel}$ distributions

discussed during the Muon Method of SF measurement, these $\not\!\!E_T$ distributions are histograms that relate the proportion of events to different values of $\not\!\!E_T$. By comparing the QCD and signal distributions, one can estimate the fraction of events that are due to QCD background (F_{QCD}). The number of QCD background events is thus estimated as [44]:

$$N_{QCD}^{pretag} = F_{QCD}^{pretag} \cdot N^{pretag} \tag{5.16}$$

$$N_{QCD}^{tag} = F_{QCD} \cdot N^{tag}. \tag{5.17}$$

All backgrounds that are not due to purely electroweak or QCD processes are referred to as "W + jets" backgrounds. For pretag events, the number of W + jets background events is arrived at by using the QCD fraction to calculate the number of non-QCD events that pass the selection. The electroweak background and the estimate of $t\bar{t}$ signal events from Monte Carlo is then subtracted. This process is illustrated in the equation below:

$$N_{W+jets}^{pretag} = N_{pretag} \cdot \left(1 - F_{QCD}^{pretag}\right) - N_{ewk}^{pretag} - N_{t\bar{t}}^{pretag}. \tag{5.18}$$

For the tagged events, special care must be taken to account for processes that will yield a W boson in addition to a b or a c quark. Using Monte Carlo calculations, the fraction of $Wb\bar{b}$, $Wc\bar{c}$ and Wc events is calculated (f_{HF}). However, this value must be modified by a correction factor K, which compensates for differences in data and Monte Carlo. Finally, a portion of the W + jets events is actually due to top quark production, but in the so-called "single top" channels. These are channels that produce one top quark instead of a $t\bar{t}$ pair. The amount of predicted single top events must be subtracted out alongside the electroweak and signal contributions. The total number of

tagged W + jets is given by:

$$N^{tag}_{W+jets} = (N_{pretag}(1 - F_{QCD}) - N_{ewk} - N_{single\,top} - N_{t\bar{t}}) \cdot f_{HK} \cdot K \cdot \epsilon. \quad (5.19)$$

One final background contribution that is considered by Method 2 is the effect of mistagging. Sometimes, a secondary vertex is identified by the SecVtx algorithm due to the crossing of poor quality tracks. However, such a vertex is not due to a b quark, despite being identified by the algorithm as such. The negative tag rate, the frequency at which mistags occur, is measured using high statistics data samples and is parametrized by the five jet variables of E_T, number of good SVX tracks, sum of all jet E_T in the event, jet η and jet ϕ [44]. Using the negative tag rate, the probability of a given light quark jet being mistagged as a b-jet can be determined. Ultimately, this probability can be applied to the individual SecVtx jets of a data sample to determine the expected number of mistagged jets (N_-). The number of mistagged background jets is then calculated as follows:

$$N^{tag}_{mistagged} \quad (5.20)$$
$$= \frac{N_-}{N^{pretag}} \cdot \left(N^{pretag} - N^{pretag}_{t\bar{t}} - N^{pretag}_{QCD} - N^{pretag}_{W+jets} - N^{pretag}_{ewk} - N^{pretag}_{single\,top}\right).$$

The Method 2 process, when provided adequate Monte Carlo samples, calculates Equations 5.14, 5.17, 5.19 and 5.20 and sums them up to obtain N_{bkg} in Equation 5.12. Method 2 also calculates the associated uncertainty with N_{bkg}, since many of the inputs to the calculation involve estimations from Monte Carlo processes that inherently have an associated error.

5.2.5 The Scale Factor as a Cross Section Input

Equation 5.12, used to experimentally calculate the $t\bar{t}$ production cross section, can be rewritten as follows [45]:

$$\sigma_{t\bar{t}} = \frac{N_{obs} - N_{bkg}}{A \cdot (SF) \cdot \epsilon_{MC} \cdot \mathcal{L}}, \tag{5.21}$$

where SF is the b-tagging efficiency scale factor. By using this version of the expression, the SF is used as an input to calculate the cross section. Obviously, this would imply that any error in the value of SF used would negatively impact the $\sigma_{t\bar{t}}$ measurement. In the latest CDF analysis measuring the $t\bar{t}$ cross section with b-tagging, the systematic uncertainty due to the SF was estimated at approximately 5% [44].

5.2.6 Log Likelihood Function to Determine the Cross Section

Although it might appear straightforward to calculate $\sigma_{t\bar{t}}$ using Equation 5.21 once the background calculation is performed, the reality is somewhat more subtle. Some of the values used in the Monte Carlo calculations needed for the background determination rely upon an estimate of $\sigma_{t\bar{t}}$ as an input to the simulation. The use of Equation 5.21 thus appears like a circular argument: $\sigma_{t\bar{t}}$ is used as an input to calculate the number of background events which is then used to calculate $\sigma_{t\bar{t}}$. Actually, one might consider this entire exercise of ultimately measuring the SF somewhat circular. The tagging of b quarks is used to measure $\sigma_{t\bar{t}}$ and then this value is used to provide b-tagging scale factor information. In order to avoid this circularity, the cross section is not calculated by using Equation 5.21, but rather through the use of a Poisson log likelihood function [44]:

$$-2 \cdot \ln L = -2 \cdot (N_{obs} \cdot \ln (D \cdot \sigma_{t\bar{t}} + N_{bkg}) - \ln (N_{obs}!) - (D \cdot \sigma_{t\bar{t}} + N_{bkg})). \tag{5.22}$$

In this expression, L represents the likelihood and D is equal to the denominator of Equation 5.21.

Generically, a likelihood function $L(\theta)$ provides a quantitative measure of the likelihood of a certain unknown parameter θ given a set of input data [2]. It is common to work with the normalization $-2 \cdot \ln L(\theta)$ because the most likely value of the parameter θ corresponds to the minimum of the function. Furthermore, when $-2 \cdot \ln L(\theta)$ is plotted, the one standard deviation error bounds of the most likely value are given by the intersection of the function and a constant line located one unit above the minimum [46].

In Equation 5.22, our unknown parameter is $\sigma_{t\bar{t}}$. Thus, we can determine the most likely cross section value by calculating the right hand side of Equation 5.22 for multiple cross section values and plotting $-2 \cdot \ln L$ versus $\sigma_{t\bar{t}}$. The minimum of the plot corresponds to the most likely cross section value and hence, a $\sigma_{t\bar{t}}$ measurement.

Note that this most likely cross section value is for one particular value of SF. This is because Equation 5.22 uses D as an input, and D is the denominator of Equation 5.21, which includes the SF. It thus stands to reason that the minimization of the log likelihood function in Equation 5.22 (using different $\sigma_{t\bar{t}}$ inputs) can be performed using a variety of different SF values. So, we can calculate the most likely $\sigma_{t\bar{t}}$ value for several different SF values. By examining the value of $-2 \cdot \ln L$ associated with each of these $\sigma_{t\bar{t}}$-SF pairs, we can further ascertain which pair is the most likely.

5.2.7 1-Tag and 2-Tag Cross Section Calculations

One of the requirements in the event selection is that the event have at least one SecVtx tag. We can choose to further divide our tagged data sample into two categories: 1-tag, where all selected events have only one b-tag, and 2-tag, where all of the events have 2 or more b-tags. It is possible then to

calculate the $t\bar{t}$ cross section independently using only 1-tag events and using only 2-tag events. With regards to Equation 5.21, the acceptance value A, as well as N_{obs} and N_{bkg}, will be different for both situations. However, since the $t\bar{t}$ production cross section is a physical observable that is independent of the selection used, one expects that both the 1-tag and 2-tag samples should yield the same value for the cross section.

We can exploit this knowledge to measure $\sigma_{t\bar{t}}$. The process to determine $\sigma_{t\bar{t}}$-SF pairs described in Section 5.2.6 can be performed for a variety of SF values for both 1-tag and 2-tag samples. We can plot these points on a $\sigma_{t\bar{t}}$ versus SF plot, taking care to join the points from the 1-tag sample together and separately join the points from the 2-tag samples, making two distinct curves. Since the 1-tag and 2-tag samples should have the same $t\bar{t}$ cross section value, we expect there to be an intersection at the most likely value of $\sigma_{t\bar{t}}$. This intersection represents another $\sigma_{t\bar{t}}$ measurement.

Up to now, we have discussed how the SF is experimentally measured at CDF, as well as the process of measuring the $t\bar{t}$ production section using a lepton + jets event selection, SecVtx b-tagging, and a log likelihood function. We have also alluded to the notion of varying the SF while performing the $\sigma_{t\bar{t}}$ measurement. In Chapter 6, we shall further quantify this notion by presenting the methods used to simultaneously measure the SF and $t\bar{t}$ production cross section.

Chapter 6
Simultaneously Measuring the SF and $t\bar{t}$ Cross Section

In the previous chapter, we saw that the SF could be used as an input to calculating the $t\bar{t}$ cross section. Using a particular SF value, we can create a log likelihood function to estimate the actual value of the cross section parameter. By varying both the SF and $\sigma_{t\bar{t}}$, we can simultaneously determine what the most likely value is for both of these unknowns. This chapter presents the primary method used to perform the calculation: the Likelihood Surface Method, in which a multidimensional likelihood function is constructed and minimized to determine the SF and $\sigma_{t\bar{t}}$. As a cross check, an independent calculation is performed by creating $\sigma_{t\bar{t}}$ versus SF plots for 1-tag events and 2-tag events separately. Using the knowledge that the actual cross section should be identical regardless of the tagging restriction, the intersection of such curves gives another simultaneous measure of SF and $\sigma_{t\bar{t}}$. Finally, the chapter concludes with a discussion of the prior analyses that have performed such a simultaneous measurement.

Although this thesis uses data from CDF to perform this measurement, these methods can be applied elsewhere, at other experiments. As long as the $t\bar{t}$ cross section can be measured using b-tagging techniques, the SF can be measured simultaneously using the methods elucidated in this chapter.

6.1 The Likelihood Surface Method

The primary method used to calculate the SF and $\sigma_{t\bar{t}}$ has been named the Likelihood Surface Method (LSM). It involves measuring the $t\bar{t}$ cross section from CDF data using the event selection, methodologies and background estimation via Method 2 discussed in Chapters 4 and 5. Recall that Method 2

requires the input of an estimated $t\bar{t}$ cross section to perform the background estimate. So, if we assume a single SF value, we can calculate the log likelihood for a variety of cross section values. Specifically, in the LSM, we calculate the log likelihood for $t\bar{t}$ cross sections between 5 and 10 pb in increments of 0.1 pb for a single SF value using 1-tag events that pass our lepton + jets selection criteria. We can then perform these steps again and calculate the log likelihoods of $t\bar{t}$ cross sections using 2-tag events. By this point, for a particular SF value, we have 1-tag and 2-tag log likelihoods for cross sections in the 5 to 10 pb range. We can then multiply the 1-tag and 2-tag log likelihood information to arrive at a combined log likelihood for each of the cross section values.

At this point in the process, we have a list of cross sections and associated log likelihood values, but they are all for a single value of SF. The next step consists of repeating the above procedure but for SF values between 0.80 and 1.10 in increments of 0.005. After doing this, we have a collection of ordered triplets: ($\sigma_{t\bar{t}}$, SF, log likelihood) for SF in the range of 0.8 to 1.1 and $\sigma_{t\bar{t}}$ in the range of 5 to 10 pb. We can plot these points, using $\sigma_{t\bar{t}}$ on the x-axis, SF on the y-axis and log likelihood along z. As such, these coordinates represent discrete points on a three dimensional log likelihood surface. The minimum of the surface corresponds to the most likely SF and $\sigma_{t\bar{t}}$ value within the selected range.

6.1.1 Surface Fitting to an Elliptical Paraboloid

In order to minimize our log likelihood surface, we must first fit the points to a function. For two dimensional log likelihood functions, it is common to fit a parabola near the location of the minimum and analytically calculate the minimum using the parameters of the fit. In three dimensions, the analogue of the parabola is the elliptical paraboloid. This choice does not seem all that surprising if we examine the shape of the surface, illustrated in Figure 6.1.

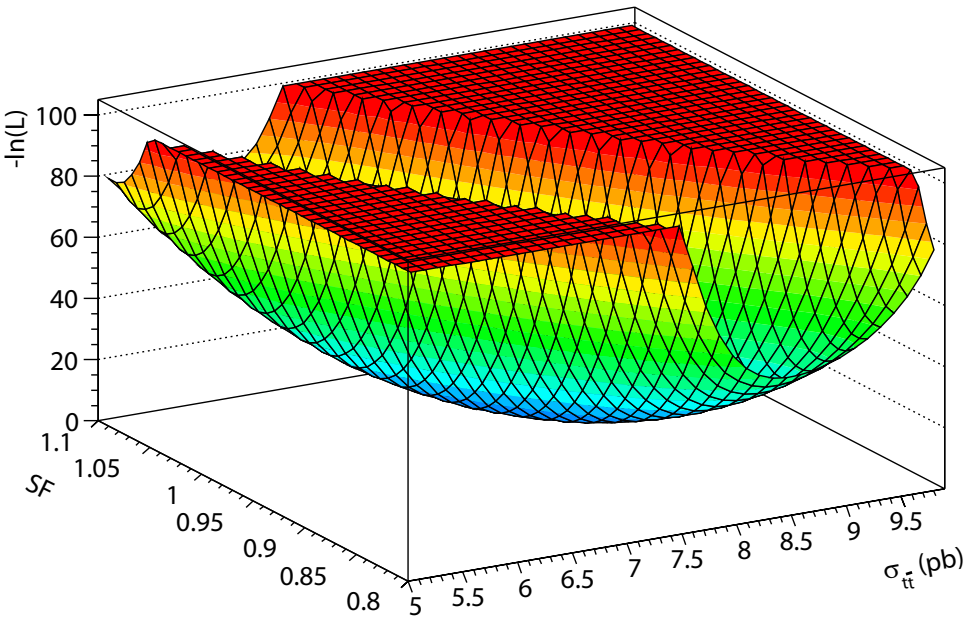


Figure 6.1: An example of a log likelihood surface. This particular plot comes from tight tagging data up to period 28. The log likelihood range has been restricted so as to emphasize the apparently elliptical paraboloid shape near the minimum.

The points are fitted to a generic elliptical paraboloid of the form

$$z = ax^2 + by^2 + cxy + dx + fy + g, \tag{6.1}$$

where x and y are the independent variables ($\sigma_{t\bar{t}}$ and SF, in our case) and the other letters are parameters of the fit. The fit uses a recursive algorithm applied over the range of $0.8 < \text{SF} < 1.1$ and $6.25 < \sigma_{t\bar{t}} < 8.25$ pb in order to find the values of the parameters. The equation of the fitted paraboloid is then used to find the minimum of the function. The coordinates of the minimum correspond to the most likely value of SF and the $t\bar{t}$ cross section. An example of the location of the minimum of the surface is demonstrated in Figure 6.2, a contour plot of the log likelihood surface.

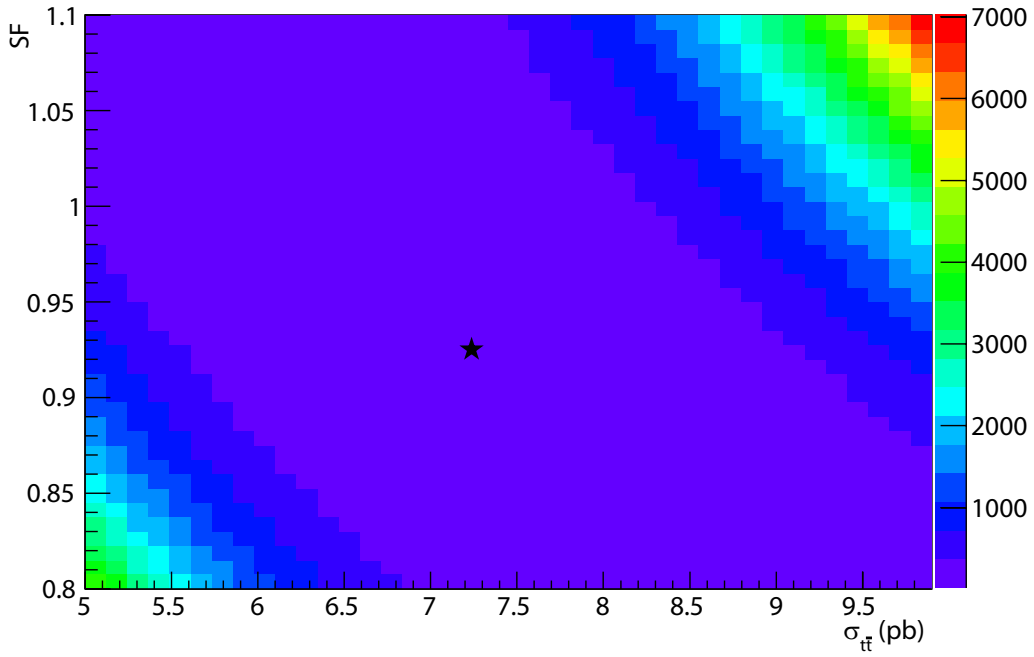


Figure 6.2: The location of the minimum (indicated by a star) is shown on a projected contour plot of a log likelihood surface. The surface presented here is the same one shown in Figure 6.1, though now we show the entirety of the log likelihood range.

By varying both the $t\bar{t}$ cross section in our background calculations and the SF, the LSM allows for the simultaneous calculation of the SF and $\sigma_{t\bar{t}}$ at CDF.

6.1.2 Statistical Uncertainty on the Elliptical Paraboloid

For a three dimensional likelihood surface, the 1 standard deviation (1σ) uncertainty bound is given by the two dimensional shape formed from the intersection of a plane with the likelihood surface. We explained in Section 5.2.5 that for a $-2 \cdot \ln L$ parabola, the 1σ error bound is given by the curve's intersection with a line 1 unit above the minimum. In 3 dimensions, due to the extra degree of freedom, the intersection height is different. For a 3D surface, the 1σ error contour of a $-2 \cdot \ln L$ surface is given by the intersection of the surface and a plane located 2.30 units above the surface minimum (or 1.15 units if we opt to use a $-\ln L$ surface instead of $-2 \cdot \ln L$ [46]). In the case of an elliptical paraboloid, this error bound is in the shape of an ellipse. An example of such a shape is presented in Figure 6.3.

In order to extract the individual 1σ uncertainty on the cross section and SF, we use the log likelihood surface fit parameters. If we assume that the x and y coordinates of the surface minimum are given by (x_{min}, y_{min}), we can calculate the z coordinate of the 1σ contour (which we call z_{ell}). We can calculate z_{ell} by invoking Equation 6.1, evaluated at (x_{min}, y_{min}), and using the values of the parameters obtained from the fitting process:

$$z_{ell} = ax_{min}^2 + by_{min}^2 + cx_{min}y_{min} + dx_{min} + fy_{min} + g + 1.15. \tag{6.2}$$

The statistical errors on the SF and cross section are given by

$$\sigma_{t\bar{t}\,err} = \sqrt{x_{min}^2 + \frac{f^2 + 4bz_{ell}}{4ab - c^2}} \tag{6.3}$$

$$SF_{err} = \sqrt{y_{min}^2 + \frac{d^2 + 4az_{ell}}{4ab - c^2}}. \tag{6.4}$$

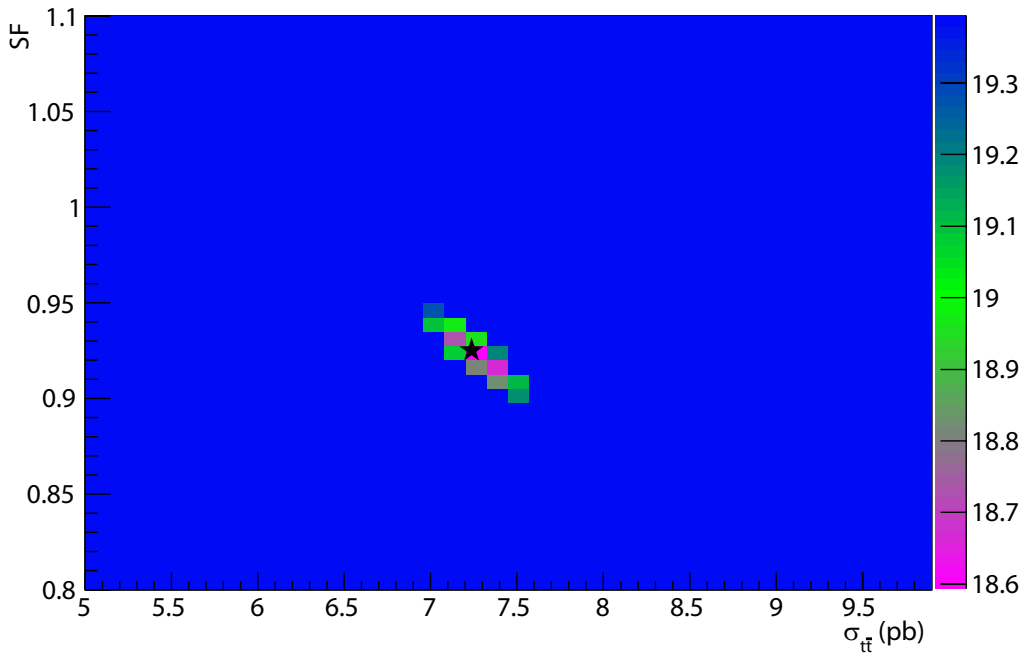


Figure 6.3: 1σ uncertainty contour of the log likelihood surface minimum, for the same surface illustrated in Figure 6.2. The minimum, once again, is indicated by a star.

6.2 The Intersection Method

The LSM described in Section 6.1 provides a means to simultaneously calculate the SF and $t\bar{t}$ cross section as well as the statistical uncertainty associated with these values. As a means of verifying the central value obtained from the LSM, an alternate method, deemed the Intersection Method, is used. The Intersection Method starts the same way as the LSM through the calculation of cross section log likelihood values in the domain of 5 to 10 pb in intervals of 0.1 pb for a given value of SF using 1-tag data events. These log likelihoods are plotted and a parabola is fitted in the region of 6.25 to 8.25 pb, as seen in Figure 6.4.

The minimum of the parabola is calculated and the corresponding cross section value associated with the minimum is taken to be the most likely cross section in the 5 to 10 pb range. Ignoring the actual log likelihood value, this

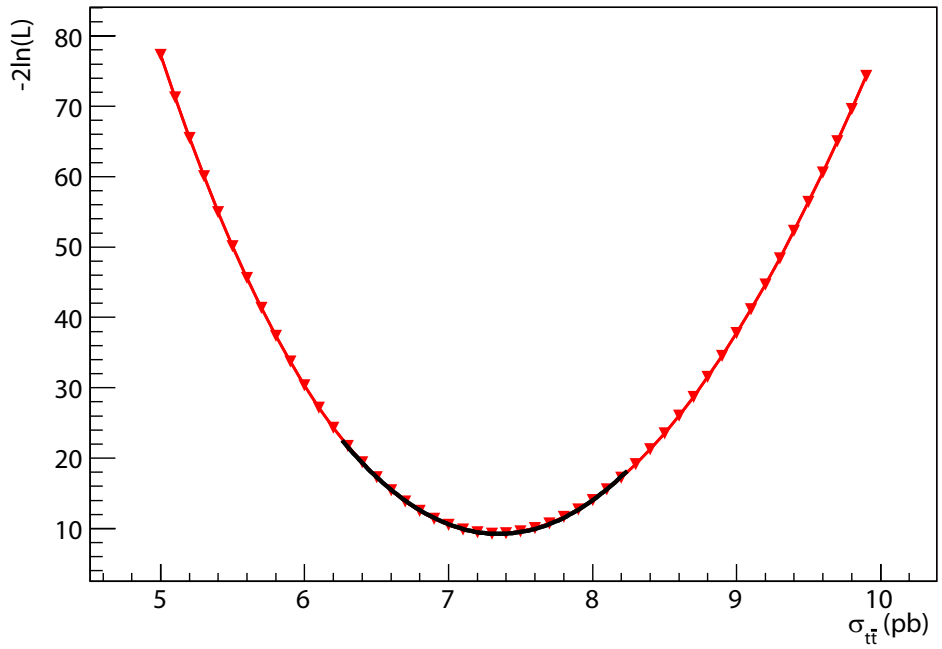


Figure 6.4: Log likelihoods of different cross section values for a particular value of SF (in this case, SF=0.91 in 1-tag tight tagging up to period 28). A parabola (as seen in black) is fit to the data and minimized so as to obtain the most likely cross section value. The $\sigma_{t\bar{t}}$ value at the minimum is taken as the $\sigma_{t\bar{t}}$ value for this particular value of SF.

method allows us to assign to a given value of SF the most likely $\sigma_{t\bar{t}}$ value. We can repeat this same process for SF values between 0.80 and 1.10 in increments of 0.005. At the end of this procedure, we have an estimate of the most likely cross section value for all of these SF inputs. We can then plot these values with SF on the x-axis and $\sigma_{t\bar{t}}$ on the y-axis; an example of this is provided in Figure 6.5.

This axis selection may appear somewhat counterintuitive considering the selection used in the LSM. However, for a two dimensional plot, it makes sense because Equation 5.21 indicates that the cross section calculation uses SF as an input. This choice also satisfies the conventions of error bar placement in two-dimensional plots. Generally speaking, if only one of the two variables has an associated error, that variable is represented along the y-axis. Because the SF for the points on the plot are not extrapolated but come from the strict

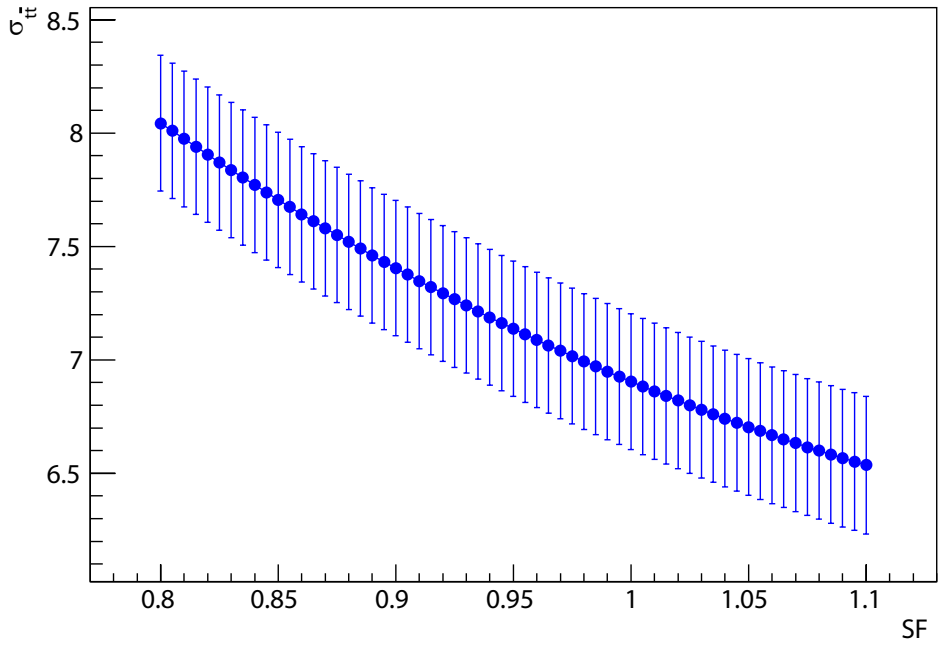


Figure 6.5: Plot of $\sigma_{t\bar{t}}$ versus SF for tight, 1-tag data up to period 28. The correlated error bars are 1σ uncertainty from the minimization of the log likelihood functions (an example of which was provided in Figure 6.4).

variation of the SF parameter, the scale factor is taken to have no error. Thus, it is the cross section that is plotted on the y-axis. Finally, this axis selection is consistent with other analyses using similar methods [47, 45]. These analyses are further discussed in Section 6.4.

The procedure to create such a plot of $\sigma_{t\bar{t}}$ versus SF can be repeated for the 2-tag data events as well. Although the plots based on 1-tag and 2-tag data may show differences, we expect them to intersect. Since the $t\bar{t}$ production cross section is physically independent of the event selection criteria, there should be a point where it is the same for both the 1-tag and 2-tag plots (this is illustrated in Figure 6.6). The coordinates of the intersection correspond to the most likely $\sigma_{t\bar{t}}$ value and the most likely SF value. Hence, the Intersection Method allows for another simultaneous measurement of the b-tagging efficiency scale factor and the $t\bar{t}$ production cross section.

We should note that we expect an anti-correlation between the SF and $\sigma_{t\bar{t}}$. Because the SF is in the denominator of Equation 5.21, it is expected that an increase in SF should lead to a decrease in $\sigma_{t\bar{t}}$ and that a decrease in SF will lead to an increase in $\sigma_{t\bar{t}}$.

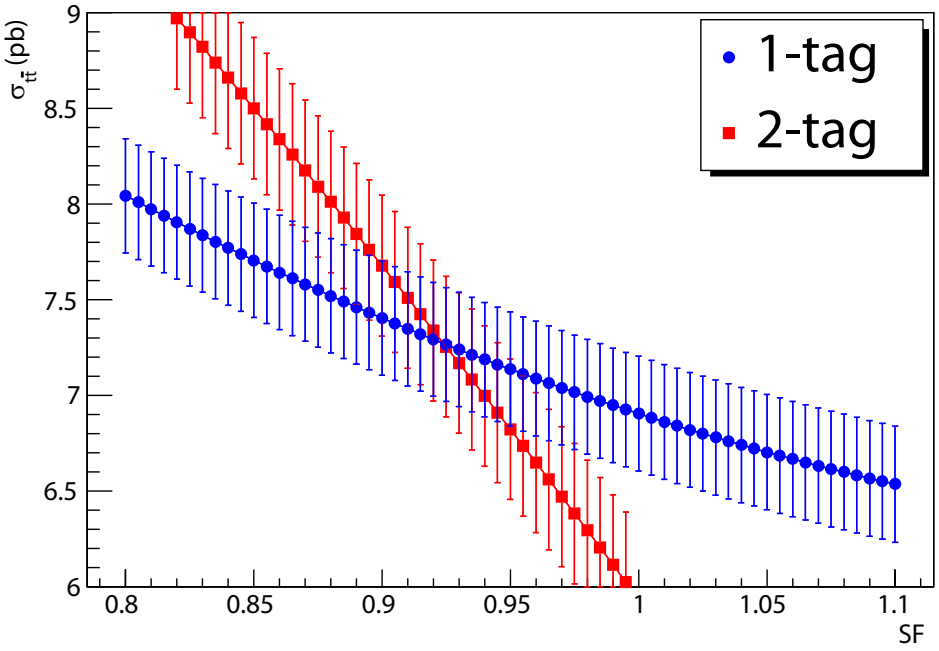


Figure 6.6: Intersection of 1-tag and 2-tag $\sigma_{t\bar{t}}$ versus SF plots, for tight tagging data up to period 28. The coordinates of the intersection represent the simultaneous measurement of the SF and the $t\bar{t}$ production cross section.

6.2.1 Determining the Location of the Intersection

In order to determine the location of the intersection, both the 1-tag and 2-tag plots are fitted to second order polynomials along the domain of $0.8 < \text{SF} < 1.1$. The point of intersection is determined by simultaneously solving the system of equations using the equations of the fits.

The Intersection Method is not used as the primary means of calculating the SF and $\sigma_{t\bar{t}}$ because it is subject to a somewhat higher level of abstraction (compared to the LSM) between the calculation of different cross section log likelihoods for a given value of SF and the final quoted results. In the log

likelihood surface, there are 51 points associated with each SF value and the surface fit is applied to this large number of points. In the Intersection Method, however, a single most likely $\sigma_{t\bar{t}}$ value is calculated for each SF value through the use of a parabolic fit and minimization (for both the 1-tag and 2-tag events) and then this collection of ordered pairs, $(\mathrm{SF}, \sigma_{t\bar{t}})$ information, is fit to another polynomial. In other words, a fit is performed for each verified SF value, and then a fit is performed again to obtain the equation of the $\sigma_{t\bar{t}}$ versus SF plot for the domain of $0.8 < \mathrm{SF} < 1.1$. This process is then repeated for the 2-tag case. With the LSM, there is only one fit performed and it is done over a larger set of points.

6.2.2 Statistical Uncertainty in the Intersection Method

It was previously mentioned that the $t\bar{t}$ cross section values in the $\sigma_{t\bar{t}}$ versus SF plots have an associated uncertainty. This is the 1σ error bound of the minimum of the $-2 \cdot \ln L$ versus $\sigma_{t\bar{t}}$ plot (seen in Figure 6.4) used to determine the most likely cross section for a particular SF value. After the minimization is performed, the intersection of the parabola with a horizontal line located 1 unit above the minimum is used to determine the 1σ errors. The $\sigma_{t\bar{t}}$ values corresponding to those bounds are used as the error bars for each point in the $\sigma_{t\bar{t}}$ versus SF plots.

Just as a second order polynomial is fitted to the central value of the points in the $\sigma_{t\bar{t}}$ versus SF plots, a parabola is also fitted to the upper and lower bounds of the collection of points as well. This means that rather than having an intersection point where the 1-tag and 2-tag lines cross, we actually have an intersection *area* where the two bands cross. An example of such a four-cornered intersection area is illustrated in Figure 6.7. Using the equations of the fits, the coordinates of two of the corners of the shape are determined (the two that are furthest from the intersection point in both the x and y

directions). The difference between the x coordinates of the corners and the x coordinate of the intersection point is taken as the error in the SF, while the difference between the y coordinates of the corners and the y coordinate of the intersection is used as the error in the $t\bar{t}$ cross section.

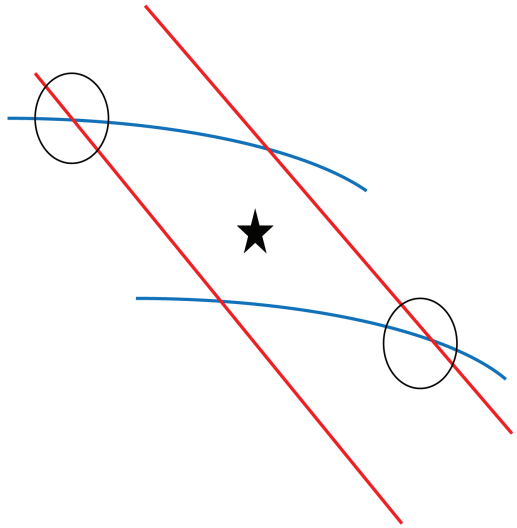


Figure 6.7: Schematic example of the intersection area. The red and blue lines represent the fits to the upper and lower error bars of the 1-tag and 2-tag curves in the $\sigma_{t\bar{t}}$ versus SF plots. The intersection point is indicated by the star and the two circles are the locations of the two corners used to calculate the error on the intersection coordinates.

6.3 Systematic Uncertainties

Up to this point, we have only discussed the statistical uncertainties in the LSM and the Intersection Method. There are two primary sources of systematic uncertainty that must also be accounted for in this analysis. The first is the mistagging rate of the SecVtx algorithm. Mistagging was introduced in Section 5.2.4 as an effect that must be compensated for in the calculation of the number of background events. However, the misidentification of secondary vertices is a significant enough effect that it must be treated as a systematic

uncertainty as well: an excess of identified b-jets will lead to a higher $t\bar{t}$ production cross section value, whereas a deficit will lead to a lower measured value. The SecVtx uncertainty is estimated at approximately $\pm 20\%$ [26]. To calculate the effect of this systematic, for any data sample used to simultaneously measure the SF and $\sigma_{t\bar{t}}$ value, the measurement must be performed again using the appropriate higher and lower mistagging multiplier (1.2 and 0.8, respectively, for a 20% uncertainty).

The other source of systematic uncertainty comes from the $Wb\bar{b}$, $Wc\bar{c}$, and Wbc kfactors. Most of the Monte Carlo calculations that are used in the Method 2 background calculations are performed at leading order. However, since many theoretical predictions of physical processes are performed at next-to-leading order (such as the theoretical estimates of the $t\bar{t}$ cross section), it is important to have an approximate scaling to next-to-leading order for processes that significantly impact the calculation of backgrounds. With regards to b-tagging, processes that generate a W boson that subsequently decays into either a b quark or a c quark are particularly relevant because they represent the production of heavier quarks. The kfactor is the approximate multiplier to go from leading order to next-to-leading order effects for these processes. Method 2 uses a kfactor of 1.5, but this value assumes an error of ± 0.3 [48]. Consequently, to evaluate the effect of this systematic error, all measurements of SF and $\sigma_{t\bar{t}}$ are performed using kfactors of 1.2 and 1.8, in addition to the standard 1.5.

6.4 Prior Simultaneous Measurements

The idea of simultaneously measuring the b-tagging efficiency scale factor with the $t\bar{t}$ cross section is not unique to this thesis. This measurement has been performed twice before at CDF: once in 2006 using 695 pb^{-1} of data [47] and again in 2008 using 1.12 fb^{-1} [45]. Both of these analyses construct

and minimize multidimensional log likelihood surfaces as well as perform a measurement using the Intersection Method. The primary differences in the analysis presented in this thesis are the use of a larger data sample and the $t\bar{t}$ event background estimation via Method 2. The older analyses determined background contributions through the use of Monte Carlo modeling of the specific background processes. As a consequence, their systematic uncertainty analysis involved a series of pseudo experiments assuming a Gaussian distribution of the Monte Carlo errors. Such a background estimation procedure, though thorough, is quite time consuming. We present the SF and $\sigma_{t\bar{t}}$ results obtained from both of these prior analyses in Table 6.1 [45, 47]. The uncertainties for the SF in the original documentation of these analyses are not broken down into their systematic and statistical contributions, so we present their combined uncertainties. Although the $\sigma_{t\bar{t}}$ values are cited with their respective systematic and statistical uncertainties, for the sake of consistency with the SF values, we present the combined values. Furthermore, for the sake of brevity, the $\sigma_{t\bar{t}}$ values shown in Table 6.1 are actually the mean of the $t\bar{t}$ cross sections calculated for the tight and loose tagging cases.

Analysis Year	Loose SF	Tight SF	$\sigma_{t\bar{t}}$ (pb)
2006	0.97 ± 0.10	0.98 ± 0.10	8.6 ± 0.9
2008	0.98 ± 0.07	0.99 ± 0.08	8.1 ± 0.6

Table 6.1: Results from prior simultaneous SF and $\sigma_{t\bar{t}}$ measurements.

One might argue that the SF and $\sigma_{t\bar{t}}$ values obtained from these analyses are largely irrelevant today; the SF may have changed with the modified detector and the latest $\sigma_{t\bar{t}}$ measurements have improved experimental techniques with greater precision. In addition, the Monte Carlo used for background estimation has undergone modifications. While those are legitimate criticisms, let us simply say that both the SF and $\sigma_{t\bar{t}}$ values measured by these analyses

agreed with the currently accepted SF and experimentally measured $\sigma_{t\bar{t}}$ values at CDF of the time.

The use of Method 2 represents a modernization of the prior analyses and an opportunity to leverage the enhanced experimental methods that have been devised at CDF since the 2008 analysis. In addition, performing the analysis on a data set five times larger than that available in 2008 should yield a result with lower statistical uncertainties. Finally, the 2006 and 2008 analyses were both performed using data where the Electron Method and the Muon Method had SF values that agreed with each other. The analysis in this thesis represents the first time that a method, other than the Muon or Electron Methods, has been used to measure the SF at CDF when those traditional methods gave differing results.

The SF result measured simultaneously with the $t\bar{t}$ cross section in this thesis represents an opportunity to ascertain whether the CDF collaboration made the correct decision in choosing to rely on the Electron Method rather than the Muon Method for its official SF value. Furthermore, given the genuinely different approach to measuring SF using the methods described in this chapter (compared to the Muon and Electron Methods described in Chapter 5), it is possible that the simultaneous measurement provides smaller uncertainties than the official SF value currently in use.

Chapter 7
Results and Discussion of the Scale Factor and $t\bar{t}$ Production Cross Section

In this chapter, we present the results from performing a simultaneous measurement of the b-tagging efficiency scale factor and the $t\bar{t}$ production cross section. All of these results were obtained using CDF lepton + jets data (and *not* the electron-rich and muon-rich samples used in the Electron and Muon Methods, respectively) and the procedures detailed in Chapters 5 and 6. The results are presented initially for the tight tagging scenario and then followed by the results for the loose tagging case. The chapter concludes by comparing our obtained SF and $\sigma_{t\bar{t}}$ results with prior measurements.

7.1 Tight Tagging

We shall first discuss the tight tagging results using data from three different time intervals. Firstly, we examine the data up to period 28. This data has an integrated luminosity of 5.6 fb^{-1} and represents the entirety of the CDF data set up to February 2010. We then look at the data from periods 0-17 (a total of 2.7 fb^{-1}), when both the Electron Method and Muon Methods were used for calculating the SF, and compare these results to those from periods 18-28 (a total of 2.9 fb^{-1}), where only the Electron Method was used, to determine if there has been any change.

7.1.1 Periods 0-28

Using the LSM, we can determine simultaneously the SF and $\sigma_{t\bar{t}}$ value using data up to period 28. The log likelihood surface up to the one sigma statistical error contour is presented in Figure 7.1. This is not the entirety of the surface, rather the z axis' $(-\ln L)$ contibutions have been truncated to

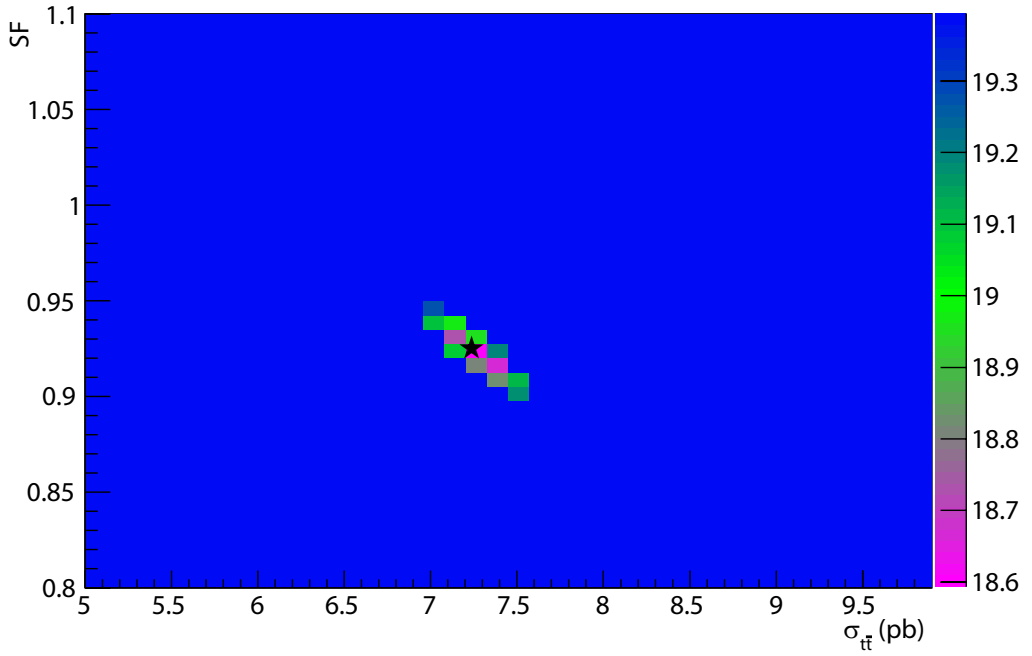


Figure 7.1: Log likelihood surface for tight tagging data up to period 28, truncated at the 1σ uncertainty bound. The minimum of the likelihood surface is indicated by a star.

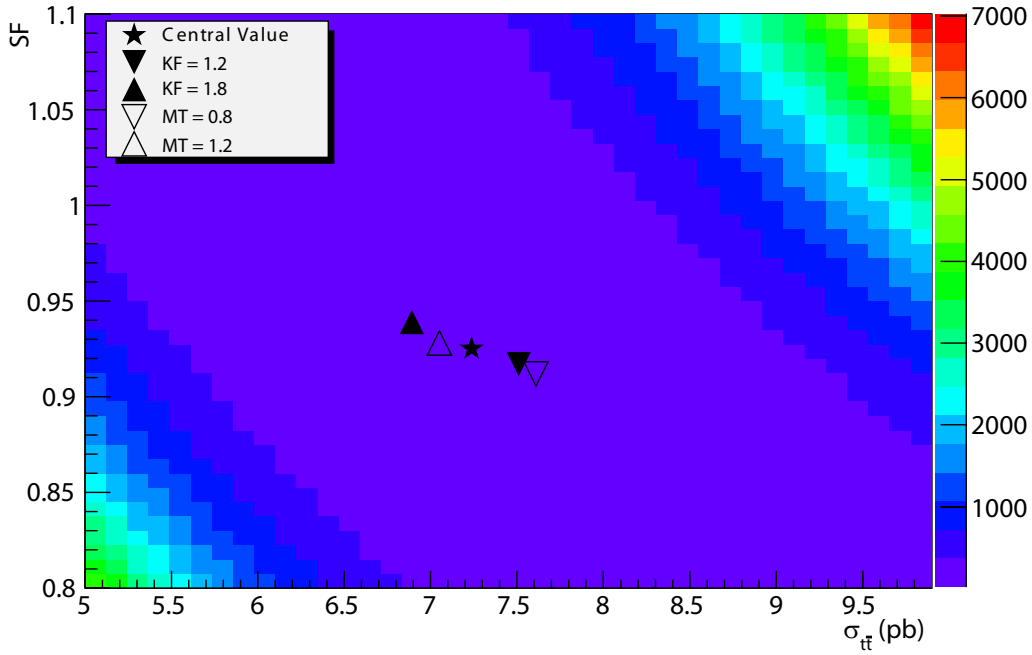


Figure 7.2: Complete log likelihood surface for tight tagging data up to period 28. The likelihood contours correspond to the surface whose minimum is indicated by a star. As a comparison, the minima of the surfaces corresponding to kfactor (KF) and mistagging (MT) systematics are indicated as well.

the level of the 1σ error bound only. The coordinates of the minimum of the surface indicate that SF=0.925±0.028 and $\sigma_{t\bar{t}}$= 7.24±0.31 pb. We restrict a more thorough comparison of these values to prior analyses to Section 7.3, but for now, let us simply comment that these values appear reasonable. The SF agrees within error with the currently accepted tight tagging SF determined from the Electron Method in 2010 (as seen in Table 5.1) and the $t\bar{t}$ production cross section agrees within error with the theoretical prediction mentioned in Section 5.2.1. However, the uncertainties that we have quoted thus far for our measured values are incomplete; they only embody the statistical errors from the log likelihood surface.

To ascertain the impact of the systematic uncertainties, we can perform the same analysis two additional times using the kfactors (KF) of 1.2 and 1.8 instead of the standard 1.5. We can also perform the analysis an additional two times using the mistagging rates (MT) of 0.8 and 1.2 rather than the standard 1.0. Recall that these different kfactors and mistagging rates for the systematic uncertainty calculations were introduced in Section 6.3. In this thesis, we shall use the term "central value" to indicate a measured value without its associated uncertainties. The complete log likelihood surface for the central value case (KF=1.5 and MT=1.0) is presented in Figure 7.2. In this figure, the minimum of the surface is marked with a star. In order to judge the effect of the systematic uncertainties, on this same surface contour plot, we have also indicated the position of the minima of the log likelihood surfaces constructed using the different KF and MT systematic uncertainty values.

The shifts of the systematic log likelihood surface minima with regards to the central value are not completely surprising. We expect that a higher kfactor should lead to a decrease in the measured value of $\sigma_{t\bar{t}}$. $Wb\bar{b}$, $Wc\bar{c}$

and Wbc production play a role in the estimate of the number of W + jets background events, as discussed in Section 5.2.4. A larger kfactor means more W + jets background events in our event selection, thus a lower number of signal events and consequently a smaller measured $t\bar{t}$ production cross section. Similarly, we expect that a higher mistagging value should lead to a lower measured $\sigma_{t\bar{t}}$. A greater amount of mistags means that a smaller number of genuine b quarks have been identified and thus a smaller value of $\sigma_{t\bar{t}}$ will be measured. The systematic uncertainties are collected in Table 7.1.

Systematic Type	Positive Contribution	Negative Contribution
SF Kfactor	+0.014	-0.008
$\sigma_{t\bar{t}}$ Kfactor (pb)	+0.27	-0.35
SF Mistagging	+0.003	-0.013
$\sigma_{t\bar{t}}$ Mistagging (pb)	+0.38	-0.19

Table 7.1: Systematic uncertainties for the LSM for tight tagging data up to period 28.

If we combine the systematic uncertainty contributions in quadrature, we obtain the following results for tightly tagged SecVtx data up to period 28, calculated via the Likelihood Surface Method:

$$\mathrm{SF} = 0.925 \pm 0.028(\mathrm{stat})\ ^{+0.014}_{-0.016}(\mathrm{syst}) \tag{7.1}$$

$$\sigma_{t\bar{t}} = 7.24 \pm 0.31(\mathrm{stat})\ ^{+0.46}_{-0.39}(\mathrm{syst})\ \mathrm{pb}. \tag{7.2}$$

If we conservatively assume that the systematic uncertainties are symmetric and equal to the absolute value of the largest of the systematic uncertainty bounds, we can combine the uncertainties to arrive at

$$\mathrm{SF} = 0.925 \pm 0.032 \tag{7.3}$$

$$\sigma_{t\bar{t}} = 7.24 \pm 0.55\ \mathrm{pb}. \tag{7.4}$$

Let us now compare these values of SF and $\sigma_{t\bar{t}}$ to those obtained for the same set of data using the Intersection Method rather than the LSM. In Figure 7.3, we present the $\sigma_{t\bar{t}}$ versus SF plots for both the 1-tag and 2-tag selections of tightly tagged data up to period 28. Somewhat analogous to Figure 7.1, this plot shows only the statistical uncertainties of the lines, indicated by colour bands. Recall that in such a plot, the measured SF and $\sigma_{t\bar{t}}$ value come from the coordinates of the intersection of the 1-tag and the 2-tag lines. Using the area of intersection of the two colour bands, we see that measured values are SF=$0.924^{+0.054}_{-0.061}$ and $\sigma_{t\bar{t}} = 7.27^{+0.65}_{-0.58}$ pb. Although the uncertainty bounds are larger compared to those from the LSM, the results from both of these methods agree with each other. In fact, the statistical uncertainties notwithstanding, the coordinates corresponding to the minimum of the LSM and the intersection of the 1-tag and 2-tag plots differ by less than 1%.

The intersection plot in Figure 7.3 shows some interesting features regarding the tagging of the data. We notice immediately that the statistical uncertainty bounds on the 2-tag curve are less than those on the 1-tag curve. This is expected because, after the event selection criteria have been applied, there are significantly more 2-tag events than 1-tag events. It should be noted that the relative 2-tag to 1-tag event yields are governed by a variety of factors and not solely the tagging efficiency. A prior analysis using SecVtx b-tagging determined that there was an order of magnitude difference between the number of 2-tag events compared to 1-tag [44]. Another aspect noted in these plots is that the value of $\sigma_{t\bar{t}}$ decreases as the scale factor increases. This behaviour is expected due to the anti-correlation between SF and $\sigma_{t\bar{t}}$ discussed in Section 6.2.

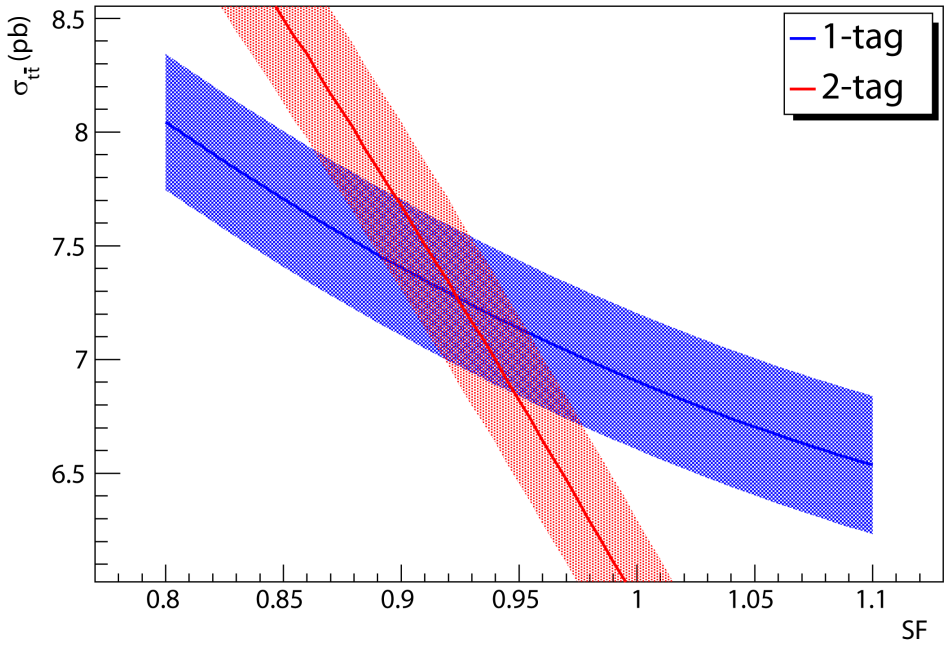


Figure 7.3: $\sigma_{t\bar{t}}$ versus SF plot for tight tagging data up to period 28. The central values are the solid lines and the bands represent the statistical uncertainties. The area of intersection is where the colour bands overlap.

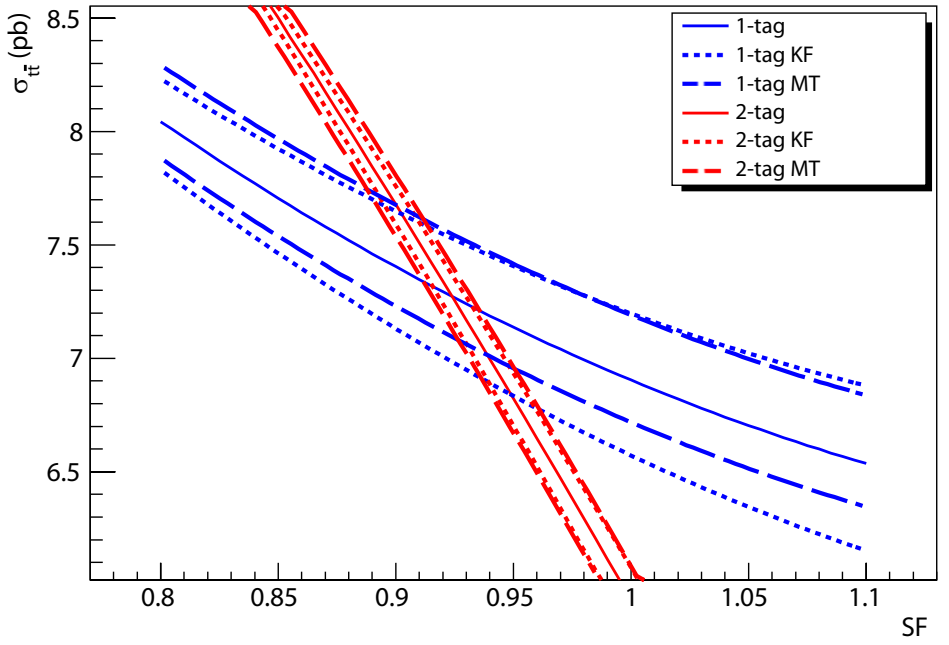


Figure 7.4: Systematic uncertainties on the $\sigma_{t\bar{t}}$ versus SF intersection plot for tight tagging up to period 28. The central values of the 1-tag and 2-tag lines (without statistical uncertainty considerations) are plotted with solid lines and the kfactor (KF) and mistagging (MT) systematic uncertainties are plotted using dashed lines.

The systematic uncertainties of the Intersection Method are illustrated in Figure 7.4. For the sake of visual clarity, the statistical uncertainty bands have been suppressed on the standard 1-tag and 2-tag curves (the non-systematic curves). The mistagging systematic uncertainties are represented by the longer-dashed lines while the kfactor systematics are the shorter-dashed lines. As was seen in the LSM results, the KF=1.2 and MT=0.8 systematics have greater $\sigma_{t\bar{t}}$ values than the central values, while the KF=1.8 and MT=1.2 systematics are lower. The systematic uncertainty contributions from the Intersection Method are summarized in Table 7.2 below.

Systematic Type	Positive Contribution	Negative Contribution
SF Kfactor	+0.035	-0.029
$\sigma_{t\bar{t}}$ Kfactor (pb)	+0.40	-0.49
SF Mistagging	+0.027	-0.036
$\sigma_{t\bar{t}}$ Mistagging (pb)	+0.47	-0.32

Table 7.2: Systematic uncertainties for the Intersection Method for tight tagging data up to period 28.

By combining the systematic uncertainties, we arrive at the following values for the SF and $\sigma_{t\bar{t}}$, as calculated by the Intersection Method for tight tagging up to period 28:

$$\mathrm{SF} = 0.924^{+0.054}_{-0.061}(\mathrm{stat})\ ^{+0.044}_{-0.046}(\mathrm{syst}) \tag{7.5}$$

$$\sigma_{t\bar{t}} = 7.24^{+0.65}_{-0.58}(\mathrm{stat})\ ^{+0.62}_{-0.58}(\mathrm{syst})\ \mathrm{pb}. \tag{7.6}$$

If we assume the uncertainties are symmetric, using the larger of the two uncertainty bounds for both the statistical and the systematic error contributions, the combined uncertainty values are

$$\mathrm{SF} = 0.924 \pm 0.076 \tag{7.7}$$

$$\sigma_{t\bar{t}} = 7.24 \pm 0.90\ \mathrm{pb}. \tag{7.8}$$

Both the SF and $\sigma_{t\bar{t}}$ values calculated from the Intersection Method are in excellent agreement with the values obtained from the LSM. In fact, the combined uncertainty bounds of the LSM expressed in Equations 7.3 and 7.4 fit entirely within the bounds described in Equations 7.7 and 7.8. This agreement gives us confidence in the SF and $\sigma_{t\bar{t}}$ values obtained using the LSM.

7.1.2 Periods 0-17

Let us now take a look at the SF and $t\bar{t}$ cross section values calculated from tight tagging data in periods 0-17, a subset of the data from periods 0-28. In order to determine whether the SF changed after period 17 when the CDF trigger changes were introduced and the Electron Methods and Muon Methods started to give differing SF measurements, it is important to look at the data from periods 0-17 independently from those of periods 18-28.

Our discussion will follow the same methods used to examine the data from periods 0-28, discussed in Section 7.1.1. We begin by looking at the LSM 1σ uncertainty contour for tight tagging data up to period 17, as seen in Figure 7.5. The minimum of the surface has coordinates of SF=0.963 ± 0.038(stat) and $\sigma_{t\bar{t}}$=7.54 ± 0.44(stat) pb. Immediately, we notice that the central value appears higher than in periods 0-28 for both the SF and the $t\bar{t}$ cross section. A more definitive comparison can be made by observing the systematic uncertainties from the kfactor and mistagging, as illustrated in Figure 7.6. Visually, the contours appear very similar to those seen in Figure 7.2 for the tight tagging data up to period 28; however, the whole surface seems to have shifted up and to the left (as indicated by the central value with larger SF and $\sigma_{t\bar{t}}$ coordinates). Table 7.3 collects the systematic uncertainties.

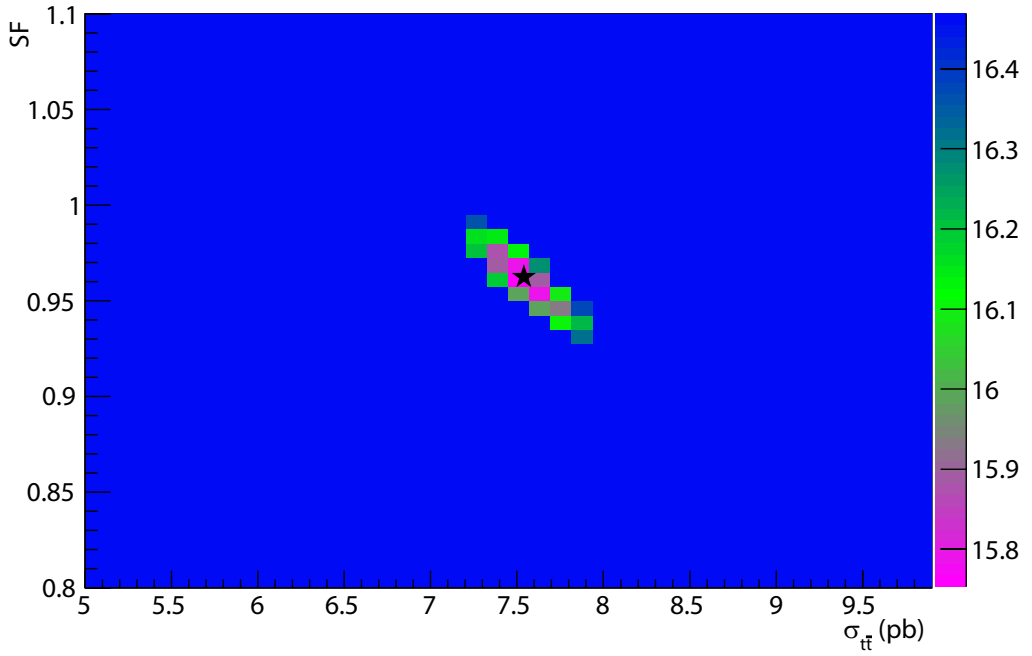


Figure 7.5: Log likelihood surface for tight tagging data up to period 17, truncated at the 1σ uncertainty bound.

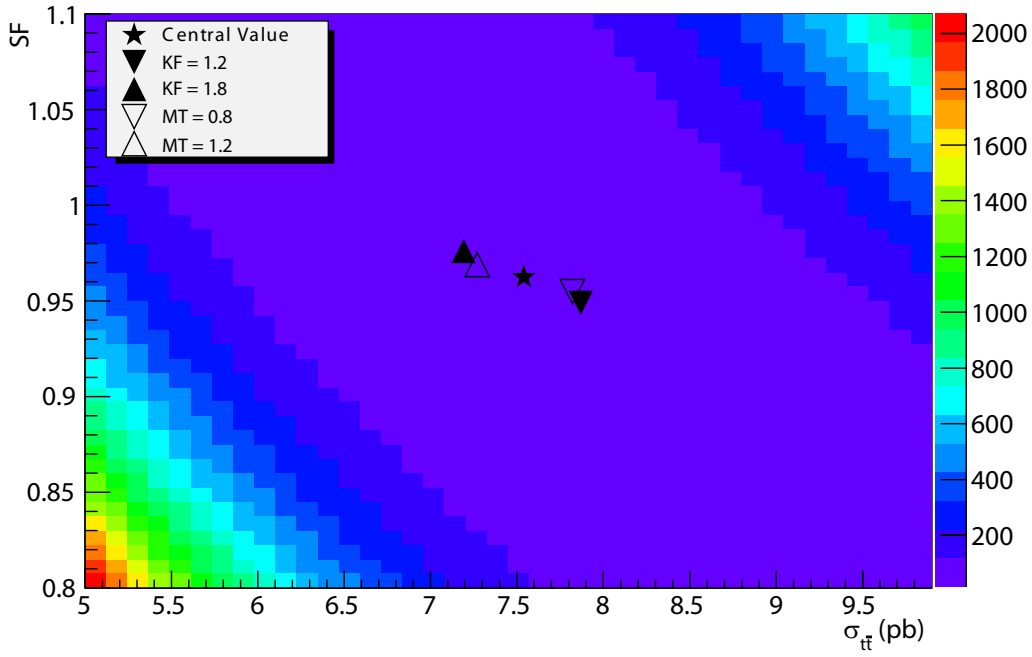


Figure 7.6: Complete log likelihood surface for tight tagging data up to period 17. The likelihood contours correspond to the surface whose minimum is indicated by a star (central value). The minima of the surfaces corresponding to kfactor and mistagging systematics are indicated as well.

Systematic Type	Positive Contribution	Negative Contribution
SF Kfactor	+0.014	-0.014
$\sigma_{t\bar{t}}$ Kfactor (pb)	+0.33	-0.35
SF Mistagging	+0.006	-0.008
$\sigma_{t\bar{t}}$ Mistagging (pb)	+0.28	-0.27

Table 7.3: Systematic uncertainties for the LSM for tight tagging up to period 17.

If we combine the systematic uncertainties, the measured values are

$$\text{SF} = 0.963 \pm 0.038(\text{stat}) \, ^{+0.015}_{-0.016}(\text{syst}) \tag{7.9}$$

$$\sigma_{t\bar{t}} = 7.54 \pm 0.44(\text{stat}) \, ^{+0.43}_{-0.44}(\text{syst}) \text{ pb}. \tag{7.10}$$

If we combine the statistical and systematic uncertainties using the conservative methods of Section 7.1.1, our final measured values for tight tagging up to period 17 using the LSM are

$$\text{SF} = 0.963 \pm 0.041 \tag{7.11}$$

$$\sigma_{t\bar{t}} = 7.54 \pm 0.62 \text{ pb}. \tag{7.12}$$

The higher statistical uncertainties compared to the LSM results for periods 0-28 are to be expected given that there is over double the integrated luminosity of the period 0-17 data in the period 0-28 data. Since the systematic uncertainties agree within 5% with those seen in the LSM for periods 0-28, it is the larger statistical uncertainty that primarily contributes to the combined uncertainties being larger for periods 0-17 versus the 0-28 data set. Both the SF and $\sigma_{t\bar{t}}$ values agree within error with the values presented in Equations 7.3 and 7.4. Because it is a physical observable, we expect that the $t\bar{t}$ production cross section measurement be unchanged in the smaller data set, and our results are indeed consistent with that. However, the central value of the SF is higher than the period 0-28 case. Furthermore, based upon its

uncertainty bounds, there is the possibility that the actual SF in periods 0-17 is higher than the SF in periods 0-28.

We can attempt to corroborate these SF and $\sigma_{t\bar{t}}$ values for tight tagging data up to period 17 by looking at the results from the Intersection Method. The intersection of the 1-tag and 2-tag $\sigma_{t\bar{t}}$ versus SF plots with statistical uncertainties is illustrated in Figure 7.7 and the systematic uncertainties are presented in Figure 7.8. The systematics are collected in Table 7.4. The measured values of SF and $\sigma_{t\bar{t}}$ are thus

$$\mathrm{SF} = 0.961^{+0.076}_{-0.093}(\mathrm{stat})\ ^{+0.050}_{-0.045}(\mathrm{syst}) \tag{7.13}$$

$$\sigma_{t\bar{t}} = 7.56^{+0.93}_{-0.75}(\mathrm{stat})\ ^{+0.55}_{-0.62}(\mathrm{syst})\ \mathrm{pb}. \tag{7.14}$$

Systematic Type	Positive Contribution	Negative Contribution
SF Kfactor	+0.037	-0.031
$\sigma_{t\bar{t}}$ Kfactor (pb)	+0.40	-0.48
SF Mistagging	+0.033	-0.032
$\sigma_{t\bar{t}}$ Mistagging (pb)	+0.38	-0.40

Table 7.4: Systematic uncertainties for the Intersection Method for tight tagging up to period 17.

We note that the statistical uncertainties are quite large compared to the tightly tagged Intersection Method results for data up to period 28. The fact that both the 1-tag and 2-tag curves have larger statistical errors than their counterparts in Section 7.1.1 means that the area of intersection of the curves is significantly larger. If we conservatively combine the statistical and systematic uncertainties, the measured values for tightly tagged data up to period 17 are

$$\mathrm{SF} = 0.96 \pm 0.11 \tag{7.15}$$

$$\sigma_{t\bar{t}} = 7.6 \pm 1.1\ \mathrm{pb}. \tag{7.16}$$

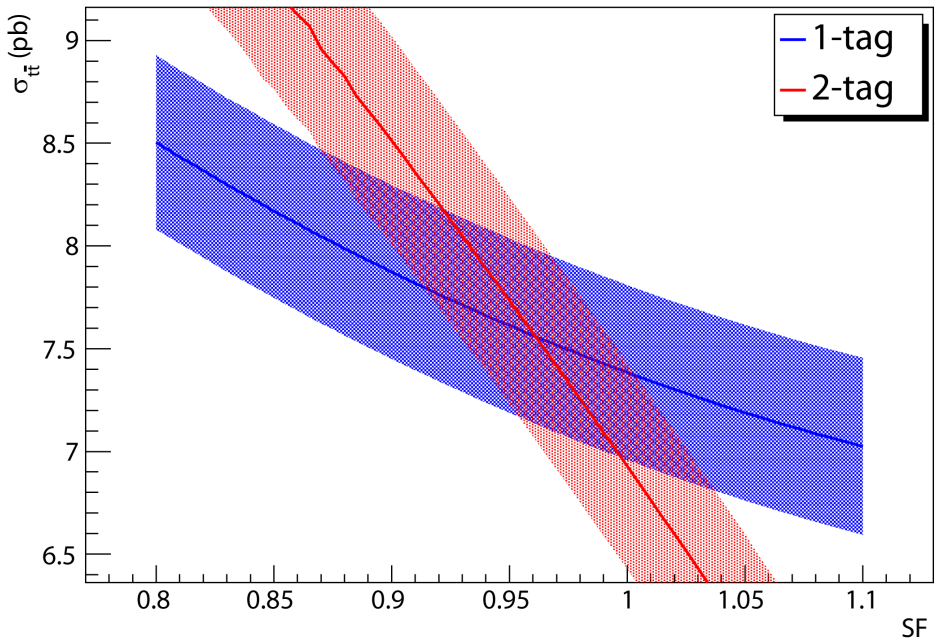


Figure 7.7: $\sigma_{t\bar{t}}$ versus SF intersection plot for tight tagging data up to period 17. The bands represent the statistical uncertainties only.

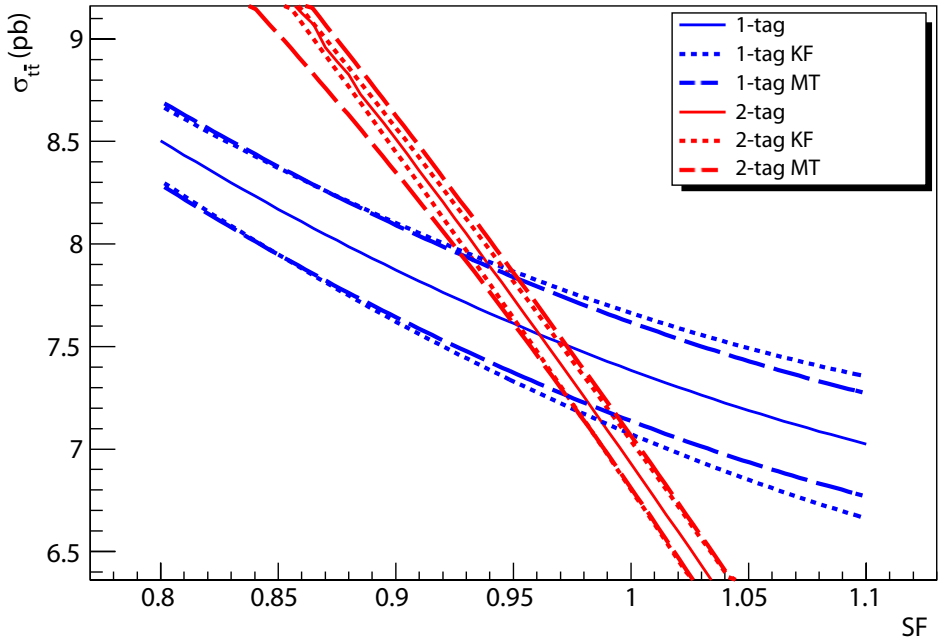


Figure 7.8: Systematic uncertainties on the $\sigma_{t\bar{t}}$ versus SF intersection plot for tight tagging up to period 17. The central values of the 1-tag and 2-tag lines (without statistical uncertainty considerations) are plotted with solid lines and the kfactor (KF) and mistagging (MT) systematic uncertainties are plotted using dashed lines.

Although the uncertainty bounds are large, the values obtained from the Intersection Method (in particular the central values) are in agreement with the results obtained from the LSM.

7.1.3 Periods 18-28

Let us now focus our attention on the SF and $\sigma_{t\bar{t}}$ values calculated for tight tagging data in periods 18-28. The SF results from the Muon Method diverged from those obtained by the Electron Method using data from these periods. Given that the simultaneous measurement of the scale factor with the $t\bar{t}$ cross section represents a markedly different way of measuring SF (compared to the Electron and Muon Methods), it is imperative to perform our measurement using data from periods 18-28. The result may be used to confirm or reject the SF measurements of the other methods.

The 1σ log likelihood surface is presented in Figure 7.9 and the full log likelihood surface with the systematic uncertainties is depicted in Figure 7.10. Compared to the data from periods 0-17, it is immediately apparent that the central value of the minimum of the surface has a lower SF. To fully quantify these results, let us look at the systematic uncertainties, presented in Table 7.5.

Systematic Type	Positive Contribution	Negative Contribution
SF Kfactor	+0.011	-0.010
$\sigma_{t\bar{t}}$ Kfactor (pb)	+0.29	-0.33
SF Mistagging	+0.002	-0.004
$\sigma_{t\bar{t}}$ Mistagging (pb)	+0.22	-0.23

Table 7.5: Systematic uncertainties for the LSM for tight tagging in periods 18-28.

The values of the SF and $\sigma_{t\bar{t}}$ measured by the LSM for tight tagging periods 18-28 are

$$\text{SF} = 0.861 \pm 0.038(\text{stat}) \, ^{+0.011}_{-0.011}(\text{syst}) \tag{7.17}$$

$$\sigma_{t\bar{t}} = 7.24 \pm 0.44(\text{stat}) \, ^{+0.36}_{-0.40}(\text{syst}) \text{ pb}. \tag{7.18}$$

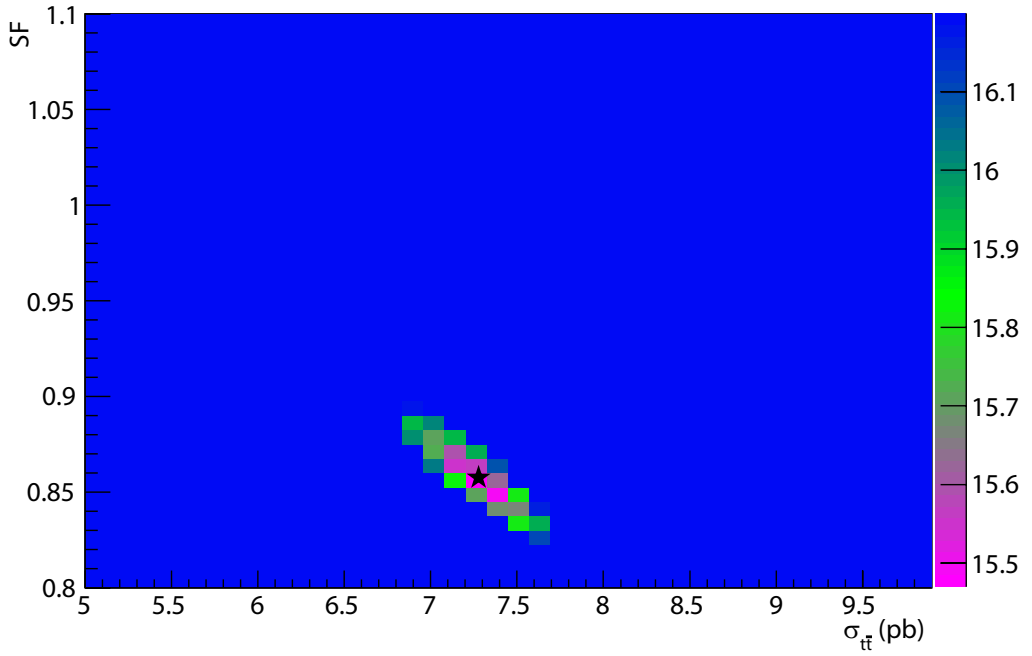


Figure 7.9: Log likelihood surface for tight tagging data for periods 18-28, truncated at the 1σ uncertainty bound.

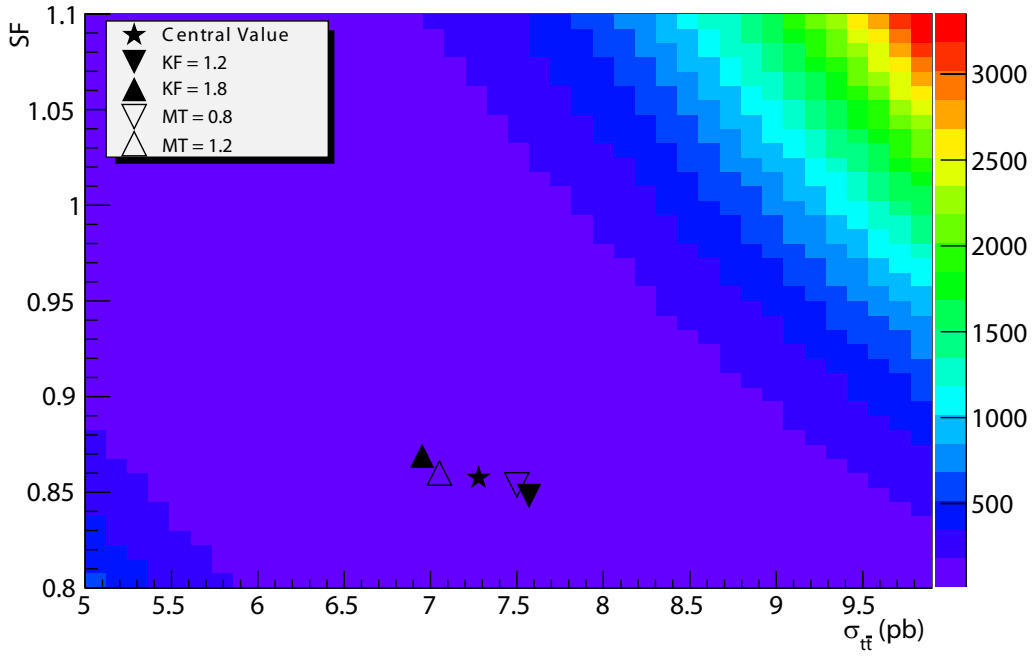


Figure 7.10: Complete log likelihood surface for tight tagging data in periods 18-28. The likelihood contours correspond to the surface whose minimum is indicated by a star (central value). The minima of the surfaces corresponding to kfactor and mistagging systematics are indicated as well.

If we conservatively combine the errors, we get values of

$$\text{SF} = 0.861 \pm 0.040 \tag{7.19}$$

$$\sigma_{t\bar{t}} = 7.24 \pm 0.59 \text{ pb}. \tag{7.20}$$

Although the $t\bar{t}$ production cross section result is consistent with the LSM values for periods 0-28 and 0-17, the SF is markedly different. It does not agree within uncertainty with the LSM value for periods 0-17; the central value is lower by over 10%. Also interesting to note, when this SF value is compared to the Electron and Muon Method values in Table 5.3, it agrees with the Electron Method value, but not the Muon Method one.

Given these interesting results, it is worth trying to replicate them using the Intersection Method. The 1-tag and 2-tag $\sigma_{t\bar{t}}$ versus SF plots with their statistical uncertainties are presented in Figure 7.11. It should be noted that the statistical uncertainties for the intersection will be somewhat underestimated because the area of intersection seems to extend below SF=0.8, the lowest value of SF used to perform the calculations. The systematic uncertainties from kfactor variations and mistagging rates are illustrated in Figure 7.12. These systematics are fully described in Table 7.6.

Systematic Type	Positive Contribution	Negative Contribution
SF Kfactor	+0.032	-0.027
$\sigma_{t\bar{t}}$ Kfactor (pb)	+0.40	-0.48
SF Mistagging	+0.031	-0.030
$\sigma_{t\bar{t}}$ Mistagging (pb)	+0.40	-0.42

Table 7.6: Systematic uncertainties for the Intersection Method for tight tagging in periods 18-28.

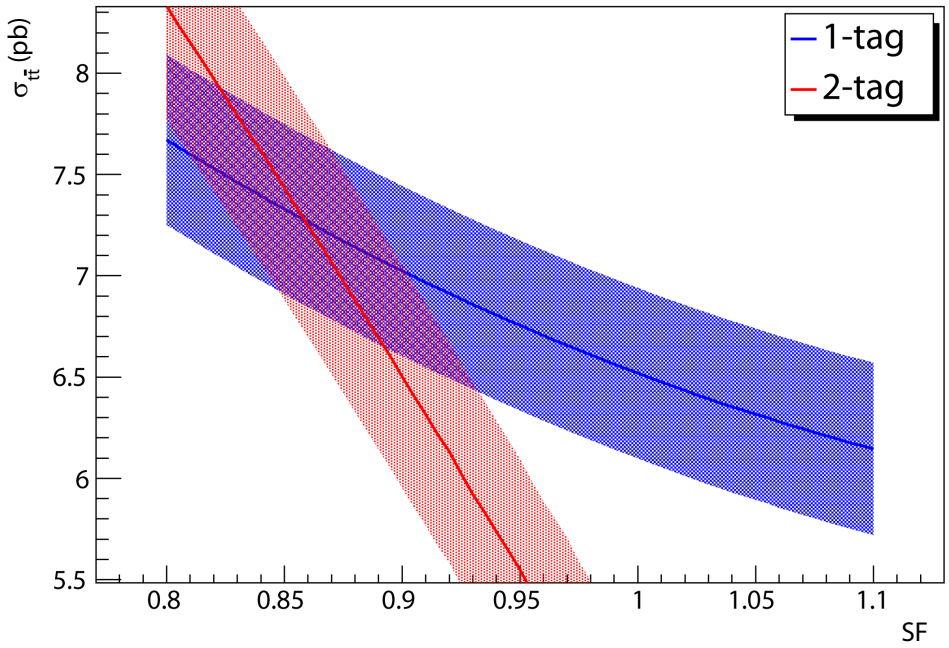


Figure 7.11: $\sigma_{t\bar{t}}$ versus SF intersection plot for tight tagging data in periods 18-28. The bands represent the statistical uncertainties only.

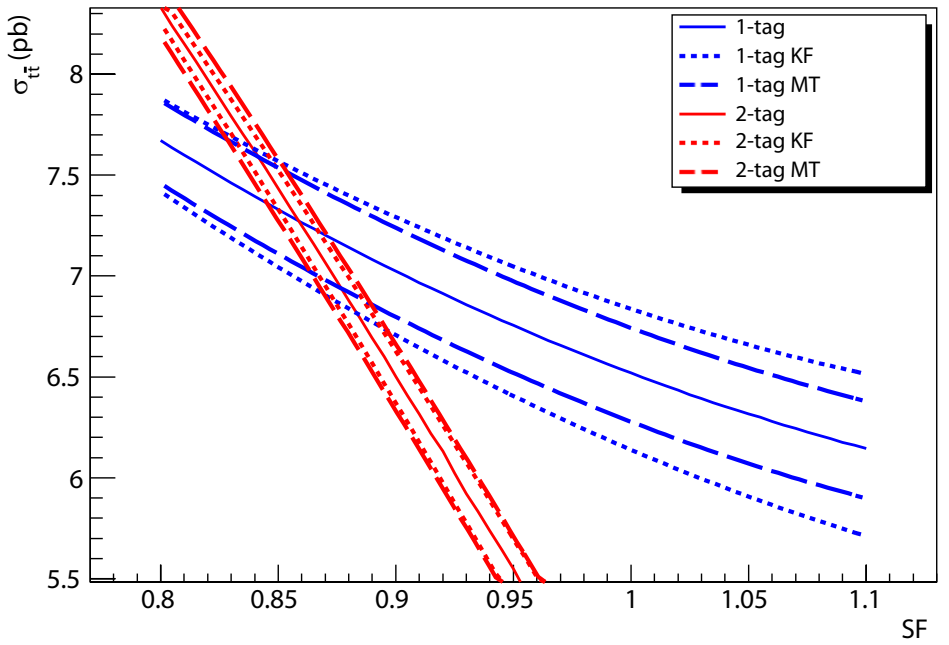


Figure 7.12: Systematic uncertainties on the $\sigma_{t\bar{t}}$ versus SF intersection plot for tight tagging data in periods 18-28. The central values of the 1-tag and 2-tag lines (without statistical uncertainty considerations) are plotted with solid lines and the kfactor (KF) and mistagging (MT) systematic uncertainties are plotted using dashed lines.

The measured SF and $\sigma_{t\bar{t}}$ from the Intersection Method for tight tagging data in periods 18-28 are

$$\mathrm{SF} = 0.859^{+0.081}_{-0.085}(\mathrm{stat})\ ^{+0.046}_{-0.040}(\mathrm{syst}) \tag{7.21}$$

$$\sigma_{t\bar{t}} = 7.27^{+0.81}_{-0.85}(\mathrm{stat})\ ^{+0.57}_{-0.64}(\mathrm{syst})\ \mathrm{pb}. \tag{7.22}$$

If we conservatively combine the errors, we arrive at

$$\mathrm{SF} = 0.859 \pm 0.087 \tag{7.23}$$

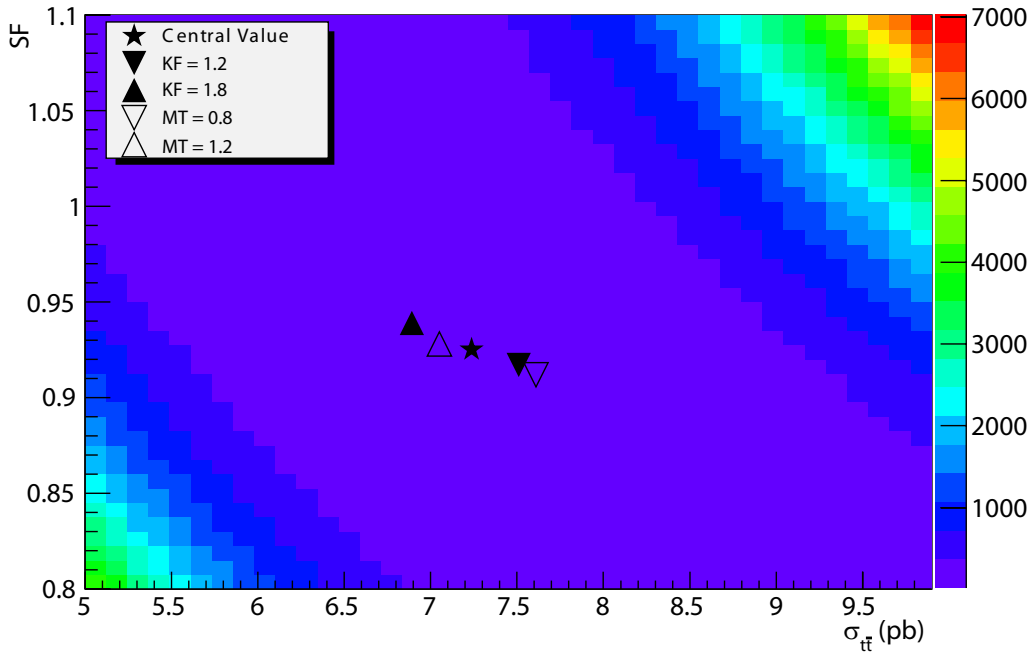
$$\sigma_{t\bar{t}} = 7.3 \pm 1.1\ \mathrm{pb}. \tag{7.24}$$

Both of these central values agree with the results from the LSM for the data from this period. We are therefore confident in the results obtained via the Likelihood Surface Method.

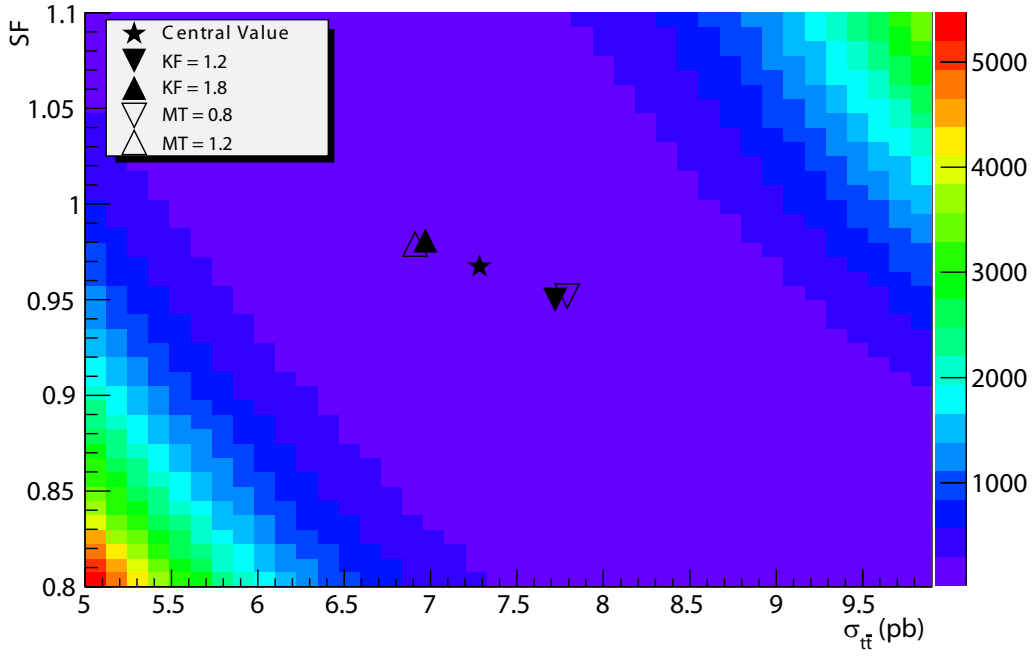
7.2 Loose Tagging

Both the Electron and Muon Methods calculate the scale factor for tight and loose tagging and arrive at a different value for each. Since the track quality restrictions are relaxed for the SecVtx loose tagging case, a different number of events (compared to tight tagging) are selected from the total yield of events within the data. Generally, the loose tagging SF results are higher than the tight tagging results.

The SF and $\sigma_{t\bar{t}}$ values for loose tagging were determined using the same methods as the tight tagging case. We start by finding the minimum of a log likelihood surface and double-check the central value using the Intersection Method. To illustrate the similarities between the surfaces for the tight tagging case and the loose tagging case, Figure 7.13 depicts two such surfaces side-by-side; Figure 7.13(a) is a contour diagram for tight tagging up to period 28 (including the systematic uncertainties), while Figure 7.13(b) is the contour diagram from the LSM for loose tagging up to period 28. The major apparent

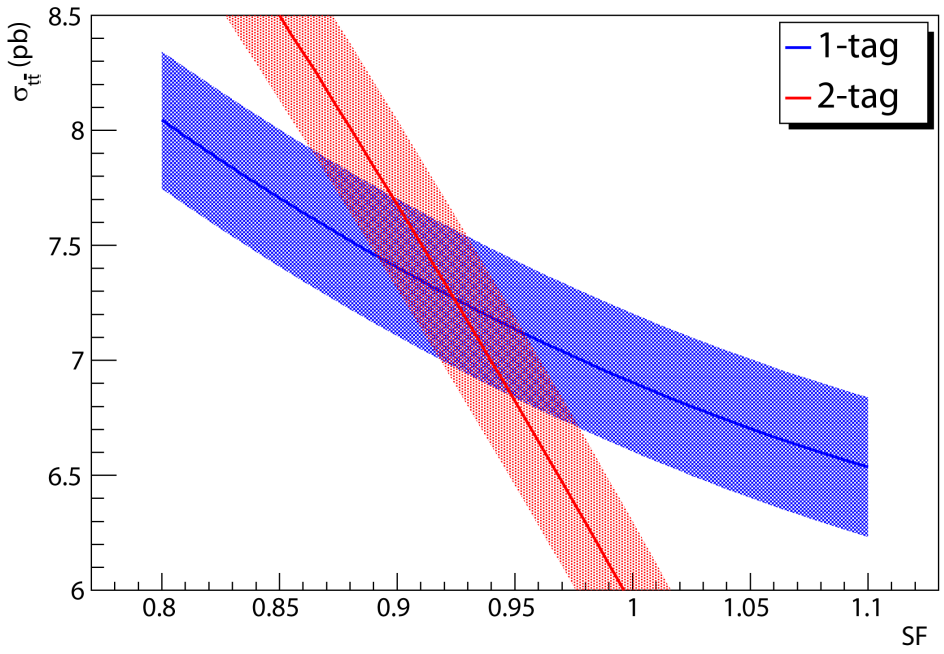


(a) Log likelihood surface and systematics for tight tagging data up to period 28.

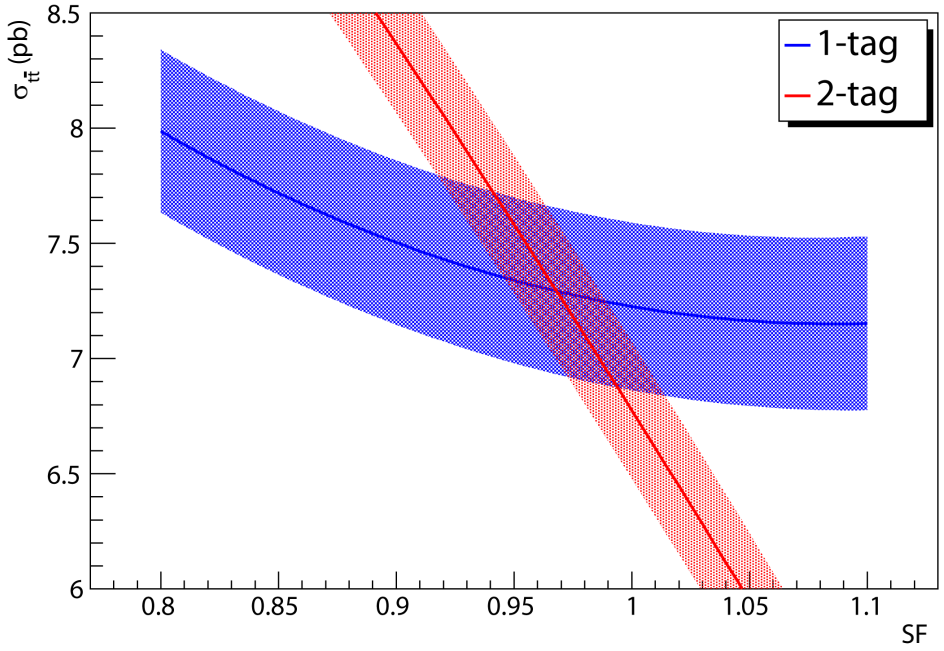


(b) Log Likelihood surface and systematics for loose tagging data up to period 28.

Figure 7.13: A comparison of log likelihood surfaces between tight and loosely tagged data up to period 28.



(a) Intersection Method statistical uncertainty plot for tight tagging data up to period 28.



(b) Intersection Method statistical uncertainty plot for loose tagging data up to period 28.

Figure 7.14: A comparison of Intersection Method plots between tight and loosely tagged data up to period 28.

differences are that the central value has a noticeably higher SF value and that the systematic uncertainties appear larger for the loose tagging data.

We can also examine the similarities between the intersections of the 1-tag and 2-tag $\sigma_{t\bar{t}}$ versus SF plots. The curves for tight tagging data up to period 28 are presented in Figure 7.14(a), while the loose tagging data for the same periods is presented in Figure 7.14(b). Perhaps the most obvious difference is that the uncertainty bounds are noticeably larger for the 1-tag $\sigma_{t\bar{t}}$ versus SF plot in the loose tagging case and that the curvature is more pronounced. However, we still obtain an area of intersection similarly shaped to those seen in the tight tagging scenarios.

For the sake of brevity and to avoid repetition of similar-looking graphics, for the loose tagging results, we shall omit the presentation of the LSM and Intersection Method plots. Instead, we present the SF and $\sigma_{t\bar{t}}$ results from both methods in Tables 7.7 to 7.10.

Once again, like in the tight tagging case, the results obtained from the LSM agree within uncertainty with those from the Intersection Method. The central values all agree within 1% between the two methods. Thus, moving forward, since there is less uncertainty in the LSM results, those are the ones that shall be quoted as the final values.

Like the tight tagging results, the loose tagging results show that there is a higher SF value in the period 0-17 data versus the 18-28 data. The SF values for the two periods do not agree within uncertainty. With regards to the $t\bar{t}$ cross section, all values from the three data taking periods agree with each other.

Data Periods	SF	Stat. Uncertainty	Pos. Syst. Uncertainty	Neg. Syst. Uncertainty	Combined Uncertainty
0-28	0.967	±0.024	+0.018	-0.023	±0.033
0-17	1.003	±0.034	+0.025	-0.021	±0.042
18-28	0.918	±0.034	+0.018	-0.016	±0.038

Table 7.7: SF results for loose tagging using the LSM.

Data Periods	SF	Pos. Stat. Uncert.	Neg. Stat. Uncert.	Pos. Syst. Uncert.	Neg. Syst. Uncert.	Total Uncert.
0-28	0.968	+0.046	-0.051	+0.054	-0.060	±0.079
0-17	1.004	+0.065	-0.074	+0.063	-0.058	±0.097
18-28	0.916	+0.065	-0.076	+0.056	-0.053	±0.094

Table 7.8: SF results for loose tagging using the Intersection Method.

Data Periods	$\sigma_{t\bar{t}}$ (pb)	Stat. Uncertainty	Pos. Syst. Uncertainty	Neg. Syst. Uncertainty	Combined Uncertainty
0-28	7.28	±0.30	+0.67	-0.49	±0.73
0-17	7.34	±0.42	+0.58	-0.65	±0.78
18-28	7.37	±0.43	+0.53	-0.56	±0.71

Table 7.9: $\sigma_{t\bar{t}}$ results for loose tagging using the LSM.

Data Periods	$\sigma_{t\bar{t}}$ (pb)	Pos. Stat. Uncert.	Neg. Stat. Uncert.	Pos. Syst. Uncert.	Neg. Syst. Uncert.	Total Uncert.
0-28	7.29	+0.51	-0.45	+0.76	-0.57	±0.92
0-17	7.4	+0.7	-0.6	+0.7	-0.7	±1.0
18-28	7.4	+0.8	-0.7	+0.6	-0.7	±1.1

Table 7.10: $\sigma_{t\bar{t}}$ results for loose tagging using the Intersection Method.

7.3 Further Discussion Regarding the SecVtx b-tagging Efficiency Scale Factor

Up until now, the LSM results for the b-tagging efficiency scale factor have been irregularly distributed throughout Sections 7.1 and 7.2. For convenience and to facilitate discussion, we collect all of the final SF results in Table 7.11. A similar table, collecting all of the measured $\sigma_{t\bar{t}}$ values associated to these scale factors, is presented in Table 7.12 in Section 7.4.

Tagging	Data Period	SF	Stat. Uncert.	Pos. Syst. Uncert.	Neg. Syst. Uncert.	Combined Uncert.
Tight	0-28	0.925	±0.028	+0.014	-0.016	±0.032
	0-17	0.963	±0.038	+0.015	-0.016	±0.041
	18-28	0.861	±0.038	+0.011	-0.011	±0.040
Loose	0-28	0.967	±0.024	+0.018	-0.023	±0.033
	0-17	1.003	±0.034	+0.025	-0.021	±0.042
	18-28	0.918	±0.034	+0.018	-0.016	±0.038

Table 7.11: Summary table of the measured SF results. All results are from the LSM exclusively.

One of the first things to be noted about the SF data is that the uncertainty on the SF measurement is statistically dominated. For each of the scale factors indicated in Table 7.11, the systematic uncertainties are less than the statistical ones. As expected, the statistical uncertainties are smaller for the period 0-28 data, because there is more data in those periods as compared to periods 0-17 and 18-28 (which are simply subsets of periods 0-28). This statistical dominance of the uncertainties is a positive feature with regard to experiment longevity; it means that, as CDF acquires more data, subsequent measures of the SF will have lower, overall uncertainties.

With regard to the central SF values, the trend noticed in both the loose and tight tagging scenarios is that the SF is higher in periods 0-17 than it is for periods 18-28. If we plot the values from Table 7.11 and the corresponding $\sigma_{t\bar{t}}$ values that were measured for the same tagging type and data period, we

see that the period 0-17 SF does not agree with the period 18-28 results (as illustrated in Figures 7.15 and 7.16). Such an observation would seem to imply that the SF was higher in the earlier data taking periods of the experiment than in the later periods. If we look purely at the central SF values, the SF has decreased by 12% in the tight tagging case and 9% in the loose case when comparing the older periods' SF to the more recent periods' SF. It makes sense then that the period 0-28 SF measurements lie somewhere in between the period 0-17 and 18-28 measurements; the overall SF for the entirety of the data taking has undergone an averaging out of the higher value in the earlier periods and the lower value in the later period.

Within uncertainty, the results from the period 0-28 loose and tight tagging agree with the standard CDF SF values used in current analyses (as indicated in Table 5.1). If we look at the central values, we see that they both agree with the Table 5.1 results within less than 3% (less than 1% for the tight tagging value). Such correspondence provides another degree of confidence in the results obtained in this thesis. We also note that the central values for SF noted in Tables 5.1, 5.2 and 5.3 are consistently higher for loose tagging compared to tight tagging. Our measured results show this same phenomenon in each of the different data periods analyzed.

If we focus our attention on the period 18-28 SF results, we can compare them to the results obtained by the Electron Method and Muon Method for the same periods (as presented in Table 5.3). Within uncertainty, our measured results for tight and loose tagging for these periods agree with the values obtained from the Electron Method. The values calculated using the Muon Method differ by 9% for the tight tagging case and 11% for the loose tagging case compared to our measure of SF from the LSM. Our SF values for period 18-28 do not agree with those from the Muon Method. Our results thus support

CDF's decision to rely exclusively on the Electron Method of measuring SF for data periods after 17. Amongst the three possible choices of using the Electron Method results only, the Muon Method results only or combining the two results to obtain the official value for use in analyses (as was done originally in periods 0-17), our results support the argument that using the Electron Method exclusively was the correct decision to take.

It should be noted that the tight and loose tagging SF values for periods 0-28 calculated through simultaneous measurement with the $t\bar{t}$ production cross section have smaller uncertainties than the values obtained from previous simultaneous measurements (as seen in Table 6.1). This is to be expected given the larger data set. However, it should be noted that the SF results from this thesis are more precise than the official CDF values currently in use. The combined uncertainty on the tight value of ± 0.032 is a 41% improvement over the currently used value of ± 0.054. In addition, the loose value combined uncertainty of ± 0.033 is a 52% improvement over the current analogous CDF uncertainty of ± 0.069. These lower uncertainty values provide a compelling argument for the use of the simultaneous SF and $\sigma_{t\bar{t}}$ measurement for determining the SF at CDF. A smaller overall uncertainty on the SF means a smaller systematic uncertainty on any analysis that uses SecVtx b-tagging. As mentioned in Section 5.1.1, such reduced systematics are particularly beneficial in Higgs boson searches or other searches where a b-tagging is performed and a small quantity of signal events is expected.

While the results confirm that there is a change in SF in the earlier data taking periods compared to the later ones, there is no thorough explanation as to *why* the change occurred. Both the CDF detector's hardware and software have undergone continuous improvements (in addition to physical wear on the hardware) since the installation of the experiment and the initial period 0 data

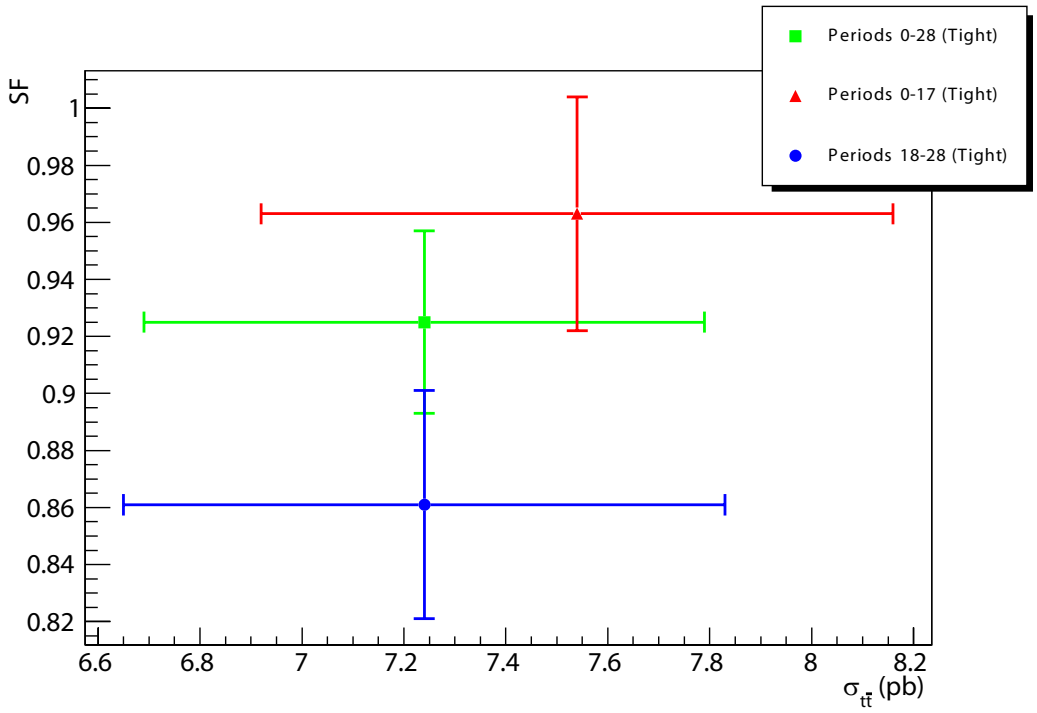


Figure 7.15: Final results for tight tagging. These 3 measurements were made using the LSM. Notice how the SF for periods 0-17 does not agree with the SF from periods 18-28. The three values of $\sigma_{t\bar{t}}$ do agree, though.

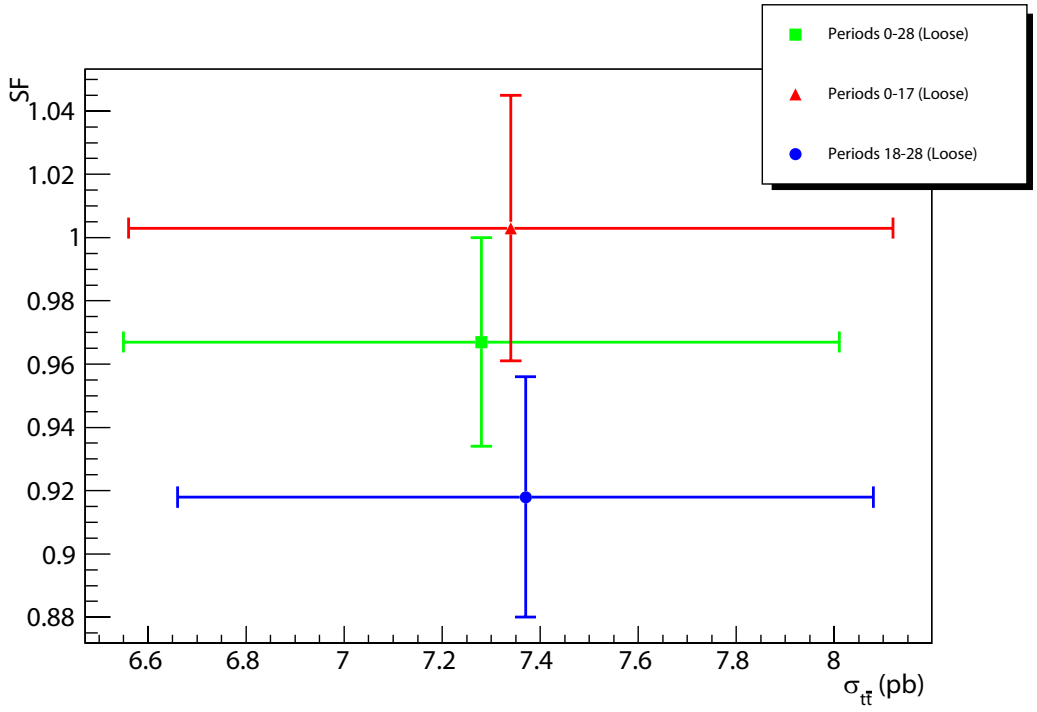


Figure 7.16: Final results for loose tagging. These 3 measurements were made using the LSM. As for the tight tagging case, the SF for periods 0-17 does not agree with the SF from periods 18-28. The three values of $\sigma_{t\bar{t}}$ do agree, though.

taking sessions. Although it is known that a trigger upgrade was performed after period 17, the results presented in this thesis alone are not sufficient to state that this was the cause of the noted SF change between the earlier and later data periods. An interesting extension to this analysis would be to perform the simultaneous measurement of the SF and the $t\bar{t}$ production cross section independently in several subsections of the overall CDF data sets. A suggestion would be to split the complete data set into 4 (or perhaps 8) smaller subsections and perform the measurement in each. This would allow a greater opportunity to observe the evolution of the SF over the lifetime of the detector. Such an analysis would indicate whether there is a stark change after period 17 or whether the change was more gradual. An abrupt change after period 17 would provide greater evidence that the trigger upgrade is the primary reason that the SF is different in the later data periods.

7.4 Further Discussion Regarding the Top-Antitop Production Cross Section

Let us now take a more comprehensive look at the $t\bar{t}$ cross section that has been measured in this thesis. For the sake of convenience, we collect the results of all of the different $\sigma_{t\bar{t}}$ values obtained using the LSM in Table 7.12.

Tagging	Data Period	$\sigma_{t\bar{t}}$ (pb)	Stat. Uncert.	Pos. Syst. Uncert.	Neg. Syst. Uncert.	Combined Uncert.
Tight	0-28	7.24	$\pm$0.31	+0.46	-0.39	$\pm$0.55
	0-17	7.54	$\pm$0.44	+0.43	-0.44	$\pm$0.62
	18-28	7.24	$\pm$0.44	+0.36	-0.40	$\pm$0.59
Loose	0-28	7.28	$\pm$0.30	+0.67	-0.49	$\pm$0.73
	0-17	7.34	$\pm$0.42	+0.58	-0.65	$\pm$0.78
	18-28	7.37	$\pm$0.43	+0.53	-0.56	$\pm$0.71

Table 7.12: Summary table of the measured $\sigma_{t\bar{t}}$ results. All results are from the LSM exclusively.

We notice that neither the systematic uncertainty nor the statistical uncertainty dominates the overall uncertainty. In some of the cases, the systematic

uncertainties are larger than the statistical ones, while, for others, both types of uncertainties are comparable. What is evident, though, is that the systematic uncertainties are larger in the loose tagging cases when compared to tight tagging measurements made over the same data-taking periods. Consequently, the overall uncertainties are larger for the loose tagging measurements in comparison to their tight tagging counterparts.

Within uncertainty, all of these results agree with each other. This behaviour is expected because the $\sigma_{t\bar{t}}$ production cross section for $p\bar{p}$ collisions at a specific center-of-mass energy is a physical observable that is independent of the amount of data collected or the criteria used for b-tagging. To arrive at a single $\sigma_{t\bar{t}}$ value from this analysis, two different approaches are possible:

(i) Perform a weighted average of the four results from periods 0-17 and periods 18-28.

(ii) Perform the arithmetic mean of the two period 0-28 results.

We have not listed taking the arithmetic mean of the six values in Table 7.12 as an option because the results from periods 0-17 and 18-28 are not independent of the period 0-28 results. The calculation of the average $\sigma_{t\bar{t}}$ using both listed methods is presented in Table 7.13.

Average Method	$\sigma_{t\bar{t}}$ (pb)	Stat. Uncertainty	Pos. Syst. Uncertainty	Neg. Syst. Uncertainty	Combined Uncertainty
(i)	7.37	$\pm$0.22	+0.24	-0.26	0.34
(ii)	7.26	$\pm$0.22	+0.41	-0.31	0.47

Table 7.13: $\sigma_{t\bar{t}}$ calculated by different averaging methods.

We see that both methods produce results that agree with each other, but the weighted average of the smaller data samples (Average Method (i)) has a smaller combined uncertainty. Let us err on the side of conservatism and choose the result from Average Method (ii) as our final $\sigma_{t\bar{t}}$ result; the method

does not rely on knowing the relative proportions of data in periods 0-17 versus 18-28, and the uncertainties are larger. Thus, the ultimate $\sigma_{t\bar{t}}$ value from this analysis is

$$\sigma_{t\bar{t}} = 7.26 \pm 0.22(\text{stat}) \, ^{+0.41}_{-0.31}(\text{syst}) \text{ pb}. \tag{7.25}$$

If we combine the statistical and systematic uncertainties, we arrive at

$$\sigma_{t\bar{t}} = 7.26 \pm 0.47 \text{ pb}. \tag{7.26}$$

As expected, this value is more precise than the $\sigma_{t\bar{t}}$ values calculated in the previous simultaneous measurement analyses (as depicted in Table 6.1). It is more interesting, perhaps, to compare our $t\bar{t}$ cross section value to some of its more modern counterparts.

It was previously mentioned in Section 5.2 that the latest theoretical prediction for the $t\bar{t}$ production cross section at CDF is $6.7^{+0.7}_{-0.9}$ pb. The value that we have calculated in Equation 7.26 is consistent, within error, with this theoretical prediction.

Furthermore, the value of 7.26±0.47 pb calculated in this thesis agrees with the combined CDF experimental result cited in Equation 5.11. Additionally, the central values differ by only 3%. The uncertainty of the $\sigma_{t\bar{t}}$ value obtained from the simultaneous measurement of the SF and $t\bar{t}$ cross section is somewhat smaller than the CDF combined result. However, this does not mean that the simultaneous approach has better precision. The combination $\sigma_{t\bar{t}}$ measurement includes a more thorough analysis of systematic uncertainties in measuring the $t\bar{t}$ production cross section as well as the contributions due to uncertainties regarding the integrated luminosity [43].

The excellent agreement between the $\sigma_{t\bar{t}}$ value measured in this thesis and the CDF combined result provides extra confidence in the experimental and

analytical methods used in this thesis and the validity of the measured values for the b-tagging efficiency scale factor.

Chapter 8
Conclusions

8.1 Conclusions

Many physics analyses at CDF require the identification of b quarks and use Monte Carlo calculations for the determination of background effects as well as theoretical predictions. Thus, a proper understanding of the b-tagging efficiency scale factor is imperative. This thesis reports a simultaneous measurement of the scale factor and the $t\bar{t}$ production cross section using 5.6 fb^{-1} of $p\bar{p}$ collision data at $\sqrt{s} = 1.96$ TeV, collected during operational periods 0 to 28 of the CDF detector. For tightly tagged SecVtx b-jets, the scale factor was measured to be SF$= 0.925 \pm 0.028(\text{stat})^{+0.014}_{-0.016}(\text{syst})$ and for loosely tagged jets, SF$= 0.967 \pm 0.024(\text{stat})^{+0.018}_{-0.023}(\text{syst})$. These values agree with the currently accepted SF values at CDF; however, they represent a 41% and 52% improvement, respectively, with regard to precision.

This thesis also reports on a measurement of the SF for tight and loose tagging in data periods 0-17 and 18-28, independently. The results from periods 18-28 agree within uncertainty with the SF results calculated using the Electron Method for SF determination and do not agree with the SF values calculated using the Muon Method for the same periods. We therefore conclude that the CDF collaboration made the correct choice in deciding to rely solely on the Electron Method SF results for data periods after 17. Furthermore, this thesis has also confirmed that the SF has decreased at CDF when comparing period 0-17 data versus period 18-28 data. The SF results for the earlier periods do not agree with those from the later periods. The central

values show a decrease of 12% in the tight tagging case and 9% for the loose tagging case.

With regard to the $t\bar{t}$ production cross section at CDF, the measurement reported in this thesis is $7.26 \pm 0.22(\text{stat})^{+0.41}_{-0.31}(\text{syst})$ pb. This value agrees both with the NLO theoretical predictions for the $t\bar{t}$ production cross section at Tevatron energies as well as the latest CDF $\sigma_{t\bar{t}}$ result arrived at by combining the results from several different signal channels.

The SF and $\sigma_{t\bar{t}}$ values reported in this thesis were measured simultaneously through the minimization of a multidimensional log likelihood surface. These results were verified by examining the intersection of $\sigma_{t\bar{t}}$ versus SF plots (constructed using likelihood information) for selected data events with 1 b-tag and 2 or more b-tags.

Since the uncertainty in measuring the SF is smaller using the simultaneous measurement method used in this thesis compared to the Electron Method currently used at CDF, we suggest that the collaboration adopt this simultaneous measurement method to determine the SF in its future data sets. Any analysis that uses SecVtx b-tagging can benefit from the reduced uncertainties provided by this method. Or, if one wishes to err on the side of caution, perhaps the official CDF SF value can come from a combination of the Electron Method with the $\sigma_{t\bar{t}}$ simultaneous method, somewhat analogous to how the official SF used to be a combination of the measured value from the Electron Method and the Muon Method.

The fact that the SF and $\sigma_{t\bar{t}}$ measured using the methods described in this thesis agree with the currently accepted experimental values demonstrates the feasibility of these methods. Although the SF was measured specifically for the SecVtx b-tagging algorithm at CDF, the methods used are generic enough

that they can be applied to other forms of b-tagging and at any other collider experiments where $t\bar{t}$ pairs are produced.

REFERENCES

[1] C. Burgess and G. Moore, *The Standard Model: A Primer* (Cambridge University Press, Cambridge, 2007).

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

[3] B. Martin and G. Shaw, *Particle Physics* (Wiley, Chichester, 2008), 3rd ed.

[4] D. J. Griffiths, *Introduction to Elementary Particles* (Wiley, Hoboken, 1987).

[5] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics* (Wiley, Toronto, 1984).

[6] T. Aaltonen *et al.* (CDF and DØ Collaborations), *Combined CDF and DØ Upper Limits on Standard Model Higgs Boson Production with up to 6.7 fb^{-1} of Data*, Preliminary Results prepared for the ICHEP 2010 Conference (2010).

[7] A. Heinson, *Feynman Diagrams for Top Physics Talks and Notes*, `http://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html` (2010).

[8] A. Quadt, Eur. Phys. J. C **48**, 835 (2006).

[9] T. Aaltonen *et al.* (The CDF Collaboration), Phys. Rev. D **81**, 012001 (2010).

[10] J. Peoples, *The Tevatron Collider: A Thirty Year Campaign*, Presented at the Fermilab Colloquium, March 3rd (2010).

[11] *Fermilab's Accelerator Chain*, `http://www-visualmedia.fnal.gov/VMS_Site/gallery/stillphotos/2000/0600/00-0635D.hr.jpg` (2000).

[12] C. Gattuso *et al.*, *Fermilab Accelerator Concepts Rookie Book v3.6*, `http://www-bdnew.fnal.gov/operations/rookie_books/Concepts_v3.6.pdf` (2010).

[13] B. Worthel *et al.*, *Fermilab Booster Rookie Book v4.1*, `http://www-bdnew.fnal.gov/operations/rookie_books/Booster_V4.1.pdf` (2009).

[14] T. Asher and S. Baginski, *Fermilab Main Injector Rookie Book v1.1*, http://www-bdnew.fnal.gov/operations/rookie_books/Main_Injector_v1.1.pdf (2010).

[15] D. Crawford, *Fermilab Tevatron Rookie Book v2.3*, http://www-bdnew.fnal.gov/operations/rookie_books/Tevatron_v2.3.pdf (2009).

[16] J. Morgan *et al.*, *Fermilab Antiproton Source Rookie Book v2.1*, http://www-bdnew.fnal.gov/operations/rookie_books/Pbar_v2.1.pdf (2010).

[17] T. Aaltonen *et al.* (The CDF Collaboration), Phys. Rev. D **82**, 031103 (2010).

[18] T. Aaltonen *et al.* (The CDF Collaboration), Phys. Rev. Lett. **101**, 181602 (2008).

[19] Modified from http://www-cdf.fnal.gov/physics/exotic/r2a/20050929.HWW/Detector3DV.gif (2005).

[20] A. Abulencia *et al.* (The CDF Collaboration), J. Phys. G **34**, 2457 (2007).

[21] T. Aaltonen *et al.* (The CDF Collaboration), Phys. Rev. D **77**, 112001 (2008).

[22] R. Blair *et al.* (The CDF II Collaboration), *The CDF II Detector Technical Design Report*, Public CDF Publication FERMILAB-PUB-96-390-E (1996).

[23] http://www-cdf.fnal.gov/events/pic/part.gif (2005).

[24] A. Lister, *The CDF L2 Track Trigger Upgrade*, Proceedings of the 34th International Conference on High Energy Physics (2008).

[25] http://ed.fnal.gov/projects/exhibits/searching/graphics/16_collision_point.gif.

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

[27] D. Green, *High p_T Physics at Hadron Colliders* (Cambridge University Press, Cambridge, 2005).

[28] http://www-cdf.fnal.gov/internal/physics/top/jets/jet.jpg.

[29] A. Bhatti *et al.*, Nucl. Instr. and Meth. A **566**, 375 (2006).

[30] D. Acosta *et al.* (The CDF Collaboration), *Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using Lepton +*

Jets Events with Secondary Vertex b-Tagging, Unpublished Internal Note CDF/PUB/TOP/PUBLIC/7801 (2005).

[31] S. Grinstein, `http://www-cdf.fnal.gov/internal/physics/top/RunIIBtag/bTag.html`.

[32] D. Acosta *et al.* (The CDF Collaboration), Phys. Rev. Lett. **95**, 102002 (2005).

[33] Y. Chen *et al.*, *Report from the b-tagging task force*, Unpublished Internal Note CDF/DOC/SEC_VTX/CDFR/10069 (2010).

[34] T. Aaltonen *et al.* (The CDF Collaboration), Phys. Rev. Lett. **104**, 141801 (2010).

[35] V. Giakoumopoulou and J. Freeman, *Summer 2010 SecVtx Scale Factors Calculated Using the Electron Method through Period 28*, Unpublished Internal Note CDF/PHYS/SEC_VTX/CDFR/10178 (2010).

[36] S. Grinstein, J. Guimarães da Costa, and D. Sherman, *The SecVtx Electron-Method Scale Factor for Winter 2007*, Unpublished Internal Note CDF/DOC/SEC_VTX/CDFR/8265 (2007).

[37] F. Garberson, J. Incandela, and C. Neu, *SecVtx b-Tag Efficiency Measurement Using Muon Transverse Momentum for 1.2 fb^{-1} Analyses*, Unpublished Internal Note CDF/DOC/SEC_VTX/CDFR/8640 (2007).

[38] F. Garberson *et al.*, *Combination of the SecVtx 1.2 fb^{-1} b-Tagging Scale Factors*, Unpublished Internal Note CDF/DOC/SEC_VTX/CDFR/8666 (2007).

[39] V. Giakoumopoulou, *Electron-based SecVtx SFs up to period 28*, Presented at the CDF b-tagging meeting, May 17th (2010).

[40] S. Leone, *SecVtx (and JetProb) b-tagging scale factor for up to p28 data using the muon $p_{T,rel}$ method and prel GRL*, Presented at the CDF b-tagging meeting, May 17th (2010).

[41] S. Rappoccio, *Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, Ph.D. thesis, Harvard University, USA (2005).

[42] T. Aaltonen *et al.* (The CDF Collaboration), Phys. Rev. D **82**, 052002 (2010).

[43] T. Aaltonen *et al.* (The CDF Collaboration), *Combination of CDF Top Quark Pair Production Cross Section Measurements with up to 4.6 fb^{-1}*, Public CDF Note 9913 (2009).

[44] T. Aaltonen *et al.* (The CDF Collaboration), *Measurement of the Top Pair Cross Section in Lepton Plus Jets Decay Channel with 2.7 fb^{-1}*, Public CDF Note 9462 (2010).

[45] M. Franklin *et al.*, *Simultaneous Measurement of the $t\bar{t}$ Production Cross Section and b-tagging Efficiency in 1.12/fb*, Unpublished Internal Note CDF/ANAL/TOP/CDFR/9447 (2008).

[46] W. Press *et al.*, *Numerical Recipes in C, 2nd Ed* (Cambridge University Press, Cambridge, 1988).

[47] J. Guimarães da Costa and D. Sherman, *Combined Measurement of the Top Pair Production Cross Section and SecVtx Scale Factor in 695 pb^{-1}*, Unpublished Internal Note CDF/ANAL/TOP/CDFR/8106 (2006).

[48] T. Schwarz, *Method 2 for You Current Selection Information*, `http://www-cdf.fnal.gov/internal/physics/top/m2fu.shtml`.