A SEARCH FOR THE HIGGS BOSON AND A SEARCH FOR DARK-MATTER PARTICLE WITH JETS AND MISSING TRANSVERSE ENERGY AT COLLIDER DETECTOR AT FERMILAB

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Qiuguang Liu

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2013

Purdue University

West Lafayette, Indiana

To my wife Wu Min and our son Max.

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to Daniela Bortoletto for her guidance and support through my graduate studies. I have been fortunate to have her as my advisor. Her enthusiasm for particle physics and the way she energetically pursue research inspire me deeply.

My deepest gratitude also goes to Fabrizio Margaroli, my other daily advisor. The Purdue group at CDF experiment has been very fortunate to have him as the driver. I would like to thank him for being a teacher, an older brother, and a close friend, and for supporting me all the way.

In addition, I would like to thank Karolos Potamianos. Starting research work with Karolos has been a great experience. I learned a lot from him in computing technologies and in how to deliver high-quality analysis fast.

My work at CDF can be very different if I don't have help from my fellow students from Purdue, Roma and Pisa: Artur Apresyan, Marco Bentivegna, Fabio Anza, Maria Teresa Grippo, Jessica Green, Jenna Walrath, Matteo Cremonesi and Ludovico Bianchi.

Also, I would like to thank Kyle Knoepfel, Tom Junk, Oscar Gonzalez Lopez, Daniel Whiteson, Eric James, Tom Schwarz, Ben Kilminster, Hyun Su Lee, Shalhout Shalhout, Mousumi Datta, Yen-Chu Chen, Craig Group, for all the help I received in my thesis work. I would like to express my gratitude to the CDF spokespersons, Costas Vellidis, Luciano Ristori, Giovanni Punzi, and Rob Roser, for their help in publicizing our research results. There are many people who have helped me, in the work of this thesis, or in the life at Fermilab. To name a few: Zhenbin Wu, Yu Zeng, Jian Tang, Tingjun Yang, Jian Wang, Marco Trovato, Francesco Pupillo, Silvia Amerio, Cheryl Bentham, Benedetto di Ruzza, Martin Frank ...

Finally, I would like to thank my wife, Min. We met at college. The day we started to walk and chat along the road in the campus changed everything and every meaning of my life. I love you.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Liu, Qiuguang Ph.D., Purdue University, May 2013. A search for the Higgs boson and a search for dark-matter particle with jets and missing transverse energy at Collider Detector at Fermilab. Major Professor: Daniela Bortoletto.

Finding the standard model Higgs boson and discovering beyond-standard model physics phenomena have been the most important goals for the high-energy physics in the last decades. In this thesis, we present two such searches. First is the search for the low mass standard model Higgs boson produced in association with a vector boson; second is the first search for a dark-matter candidate (D) produced in association with a top quark (t) in particle colliders. We search in events with energetic jets and large missing transverse energy – a signature characterized by complicated backgrounds – in data collected by the CDF detector with proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV. We discuss the techniques that have been developed for background modeling, for discriminating signal from background, and for reducing background resulting from detector effects.

In the Higgs search, we report the 95% confidence level upper limits on the production cross section across masses of 90 to 150 GeV/c^2. The expected limits are improved by an average of 14% relative to the previous analysis. The Large Hadron Collider experiments reported a Higgs-like particle with mass of 125 GeV/c^2 by studying the data collected in year 2011/12. At a Higgs boson mass of 125 GeV/c^2, our observed (expected) limit is 3.06 (3.33) times the standard model prediction, corresponding to one of the most sensitive searches to date in this final state.

In the dark matter search, we find the data are consistent with the standard model prediction, thus set 95% confidence level upper limits on the cross section of the process $p\bar{p} \to t + D$ as a function of the mass of the dark-matter candidate. The

upper limits are approximately 0.5 pb for a dark-matter particle with masses in the range of $0 - 150$ GeV/c^2.

1. THEORETICAL BACKGROUND

1.1 The Standard Model

The standard model (SM) is a gauge theory for particle physics based on the gauge group $SU(3) \times SU(2)_L \times U(1)_Y$. SU(3) represents the strong force, which is described by the quantum chromodynamics (QCD), with 3 denoting the number of colors in which the quarks appear. Gluons, massless gauge bosons, are the exchange particles for the strong force between quarks. $SU(2)_L \times U(1)_Y$ describes the electroweak force – a unified form of the electromagnetic interaction and the weak interaction which was the work by Weinberg, Salam and Glashow in 1970s [1]. The subscript L indicates that the weak interaction only act on the left-handed fermions. Y is the generator of $U(1)_Y$, and is called the weak hypercharge. The $W^{\pm}$ and Z, and photon are the exchange particles for the weak and electromagnetic interactions, respectively.

Fundamental Particles

According to the SM [2], all fundamental particles are either bosons or fermions depending on their spin.

Particles associated with the matter fields – leptons and quarks – are fermions. They have half-integer spin and are divided into three families and twelve flavors. The lightest particles, among these fermions, are neutrinos with mass less than 2 eV; the heaviest is the top quark with mass of 172 GeV. The quarks have fractional electric charges, color charges, baryon numbers and flavor quantum numbers, and they are the only fundamental particles in the SM to experience all four fundamental forces. There are two kinds of leptons: charged leptons like the electron and neutral leptons like the neutrinos.

Gauge bosons are the force carriers for the interactions of the matter fields. Fundamental particles, whose interactions are described by a gauge theory, interact with each other by the exchange of gauge bosons – usually as virtual particles. Gauge bosons have spin of one. There are three kinds of gauge bosons in the SM: photons, the massless particles that give us light, which carry the electromagnetic interaction; $W^{\pm}$ and Z bosons, with mass in the order of 100 GeV/c^2, which carry the weak force; and gluons, believed to be massless, which carry the strong force.

A critical element of the standard model is the Higgs boson – a consequence of the Higgs mechanism [3]. The scalar Higgs boson is the only SM particle that is not confirmed. The Large Hadron Collider (LHC) experiments, ATLAS and CMS, recently reported the observation of a new particle which could be the standard model Higgs boson [4].

Overall, the SM presents 6 types of quarks and 6 of leptons and their anti-partners, photon, $W^{\pm}$ and Z bosons, 8 types of gluons, and the Higgs boson, which gives a total of 37 types of fundamental particles. A summary of the properties of these particles is shown in Tab. 1.1.

Table 1.1: The standard model of fundamental particles, with the gauge bosons in the rightmost column. The Higgs boson, not included here, is a yet to be observed SM particle.

	Observed fundamental particles			Charge	Spin
Quarks	u (2.4 MeV)	c (1.27 GeV)	t (171.2 GeV)	2/3	1/2
	d (4.8 MeV)	s (104 MeV)	b (4.2 GeV)	-1/3	1/2
Leptons	ν_e (< 2.2 eV)	ν_μ (< 0.17 MeV)	ν_τ (< 15.5 MeV)	0	1/2
	e (0.511 MeV)	μ (105.7 MeV)	τ (1.777 GeV)	-1	1/2
Gauge bosons	γ, gluons, $W^{\pm}$ (80.4 GeV) and Z (91.2 GeV)				1

Fundamental Forces

There are four known fundamental forces of nature: gravity, electromagnetic force, weak force, and strong force. The first two kinds, gravity and electromagnetic forces, are responsible for practically all the phenomena encountered in daily life. The weak and strong forces are only accessible in subatomic scope. All these forces, except the gravity, are described by the SM. The properties of these fundamental forces, or interactions, are summarized in Tab. 1.2.

Table 1.2: Properties of the four fundamental forces of nature: the strengths of interactions are shown relative to the strength of the electromagnetic force for two u quarks separated by the specified distances.

Property		**Gravitational interaction**	**Weak interaction**	**Electromagnetic interaction**	**Strong interaction**
Acts on:		Mass – Energy	Flavor	Electric charge	Color charge
Particles experiencing		All	Quarks Leptons	Electrically charged	Quarks Gluons
Particles mediating		Graviton (not yet observed)	Z^0, $W^{\pm}$	γ	Gluons
Strength at	10^{-18}m	10^{-41}	0.8	1	25
	3×10^{-17}m	10^{-41}	10^{-4}	1	60

The electromagnetic force acts between electrically charged particles.

The strong force holds quarks together to form hadrons by acting on the "color charges". The strong force also binds protons and neutrons to form the nuclei. It is the strongest force at the atomic scale.

The weak force was discovered in the beta decay ($n \to p + e^- + \bar{\nu}_e$), in which a continuous energy spectrum of electron is observed. The strength of the weak force is comparable to the electromagnetic force (EM) at 10^{-18} m scale but only $10^{-4}\times$ EM at 3×10^{-17} m scale. Weak propagators ($W^{\pm}$ and Z gauge bosons) are introduced

to solve the problem of infinite cross section at high energies. The large mass of the gauge bosons also make the force "weak".

1.1.1 The Electroweak Model and the Higgs Mechanism

The data on weak and electromagnetic processes suggest that the interactions are invariant under weak isospin $SU(2)_L$ and weak hypercharge $U(1)_Y$ transformation. In analog to quantum electrodynamics, where the electromagnetic amplitudes are calculated using an interaction,

$$-ie j_\mu^{em} A^\mu = -ie(\bar{\Psi}\gamma_\mu Q\Psi)A^\mu \qquad U(1)_{em}, \tag{1.1}$$

we write the weak processes in two basic interactions; first, an isotriplet of weak currents J_μ is coupled to three vector bosons W^μ,

$$-igJ_\mu W^\mu = -ig\bar{\chi}_L\gamma_\mu T W^\mu \chi_L \qquad SU(2)_L, \tag{1.2}$$

and second, a weak hypercharge current j_μ^Y is coupled to a fourth vector B^μ,

$$-i\frac{g'}{2}j_\mu^Y B^\mu = -ig'(\bar{\Psi}\gamma_\mu\frac{Y}{2}\Psi)B^\mu \qquad U(1)_Y. \tag{1.3}$$

The operators T and Y are the generators of the $SU(2)_L$ and $U(1)_Y$ groups of gauge transformations, respectively. The left-hand fermion fields,

$$\psi_i = \begin{pmatrix}\nu_i \\ \ell_i\end{pmatrix}_L \text{ and } \begin{pmatrix}u_i \\ d_i'\end{pmatrix}_L, \tag{1.4}$$

with $T = \frac{1}{2}$, $Y = -1$, of the ith fermion family transform as doublets under $SU(2)_L$, where $d_i' = \sum_j V_{ij}d_j$, and V is the *Cabbibo-Kobayashi-Maskawa mixing matrix* [5].

The right-hand fields,

$$\psi_i = \ell_{i_R} \text{ and } u_{i_R} \textit{ or } d_{i_R} \tag{1.5}$$

with $T = 0$, $Y = -2$.

The coupling constants for $SU(2)_L$ and $U(1)_Y$ are g and g', respectively, with relations of

$$\cos\theta_{\rm w} = \frac{g}{\sqrt{g^2+g'^2}} \text{ and } \sin\theta_{\rm w} = \frac{g'}{\sqrt{g^2+g'^2}}, \tag{1.6}$$

or,

$$\theta_{\rm w} = \tan^{-1}\frac{g}{g'}, \tag{1.7}$$

where $\theta_{\rm w}$ is the *Weinberg angle* or *weak mixing angle*.

The underlying fields, W_μ^i and B_μ are massless for retaining gauge invariant. To generate the particle masses in a gauge invariant way, we must use spontaneous symmetry breaking through the Higgs mechanism.

The most economical choice of arranging four fields in an isospin doublet – the Higgs fields – with weak hypercharge $Y = 1$ is:

$$\phi \equiv \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix} \text{ with } \begin{array}{l}\phi^+ = (\phi_1 + i\phi_2)/\sqrt{2}\\ \phi^0 = (\phi_3 + i\phi_4)/\sqrt{2}\end{array} \tag{1.8}$$

This is infact the choice originally made in 1967 by Weinberg.

As such, four gauge fields, the photon and massive bosons, can be written as linear combinations of W_μ^i, B_μ and $\theta_{\rm w}$.

$$\begin{array}{ll}\text{photon} & A = B\cos\theta_{\rm w} + W^3\sin\theta_{\rm w}\\ \text{neutral weak boson} & Z = -B\sin\theta_{\rm w} + W^3\cos\theta_{\rm w}\\ \text{charged weak boson} & W^\pm = (W^1 \mp iW^2)/\sqrt{2}\end{array} \tag{1.9}$$

The same Higgs doublet which generates $W^\pm$ and Z masses is also sufficient to give mass to the leptons and quarks.

As a result, an invariant Lagrangian energy density is written as follow,

$$\mathcal{L} = \frac{g}{\sqrt{2}}(J_\mu^- W_\mu^+ + J_\mu^+ W_\mu^-) + \frac{g}{\cos\theta_{\rm w}}(J_\mu^{(3)} - \sin^2\theta_{\rm w} J_\mu^{em} Z_\mu) + g\sin\theta_{\rm w} J_\mu^{em} A_\mu. \tag{1.10}$$

The first and second terms represent the weak charge current and the neutral current, respectively. The third term represents the electromagnetic neutral current.

1.1.2 The Higgs Boson

The electroweak model consists of a weak isospin triplet W^μ and an isospin singlet B^μ plus an isospin doublet of scalar Higgs fields ϕ. The Higgs doublet are denoted by

complex fields ϕ^+ and ϕ^0 with four real components such that $\phi^+ = (\phi_1 + i\phi_2)/\sqrt{2}$ and $\phi^0 = (\phi_3 + i\phi_4)/\sqrt{2}$. The antiparticles are $\phi^- = (\phi^+)^*$ and $\bar{\phi}^0 = (\phi^0)^*$. The ϕ^+, ϕ^- and $(\phi^0 - i\bar{\phi}^0)/\sqrt{2}$ act as the longitudinal third polarization component of the massive $W^\pm$ and Z bosons. The remaining neutral component, $H = (\phi^0 + i\bar{\phi}^0)/\sqrt{2}$, is predicted to exist as a free particle – the Higgs boson.

The Higgs boson is its own antiparticle and is CP-even. It has zero spin, no electric and color charge. The mass of the Higgs boson is unknown. In the SM, the Higgs boson has a number of indirect effects, such as Higgs loops result in tiny corrections to masses of the W boson and the top quark. Thus, the mass of the Higgs boson can be constrained experimentally via studies of these indirect effects.

Precision electroweak measurements, performed at LEP and SLD, CDF, and D0 experiments, have put constraint of the mass of the Higgs boson [6]. Figure 1.1 shows the $\Delta\chi^2$ curve derived from these measurements as a function of the Higgs-boson mass. The preferred value for its mass, corresponding to the minimum of the curve, is at 94 GeV/c^2, with an experiment uncertainty ($\Delta\chi^2 < 1$) of +29 and −24 GeV/c^2.

At LEP (Large Electron-Positron collider), the main process of the SM Higgs boson (H) production is the Higgsstrahlung process, $e^+e^- \to HZ$, which has a kinematic threshold at $m_H = \sqrt{s} - m_Z$. The main search topologies are therefore dictated by the dominant Higgs decay modes (mostly $b\bar{b}$, some $\tau\bar{\tau}$) and the Z decay modes. All four LEP experiments carried out searches for $(H \to b\bar{b})(Z \to l^+l^-, \nu\bar{\nu}, q\bar{q})$ (respectively: the leptonic, missing energy, and four-jet topologies), and for the main topologies with τs: $(H \to \tau^+\tau^-)(Z \to q\bar{q})$ and $(H \to b\bar{b})(Z \to \tau^+\tau^-)$. The combined LEP results [7] are consistent with the background-only hypothesis with m_H up to 114.4 GeV/c^2 being excluded at 95% confidence level (CL). Figure 1.2 shows the upper limits for the SM Higgs search from the LEP results.

At the Tevatron proton-antiproton collider, the main processes of Higgs boson production are gluon fusion, and vector boson associated production. Figure 1.3 (a) shows the relevant cross sections of these processes as a function of the Higgs boson mass. At low masses, below the WW and ZZ decay threshold, the Higgs boson

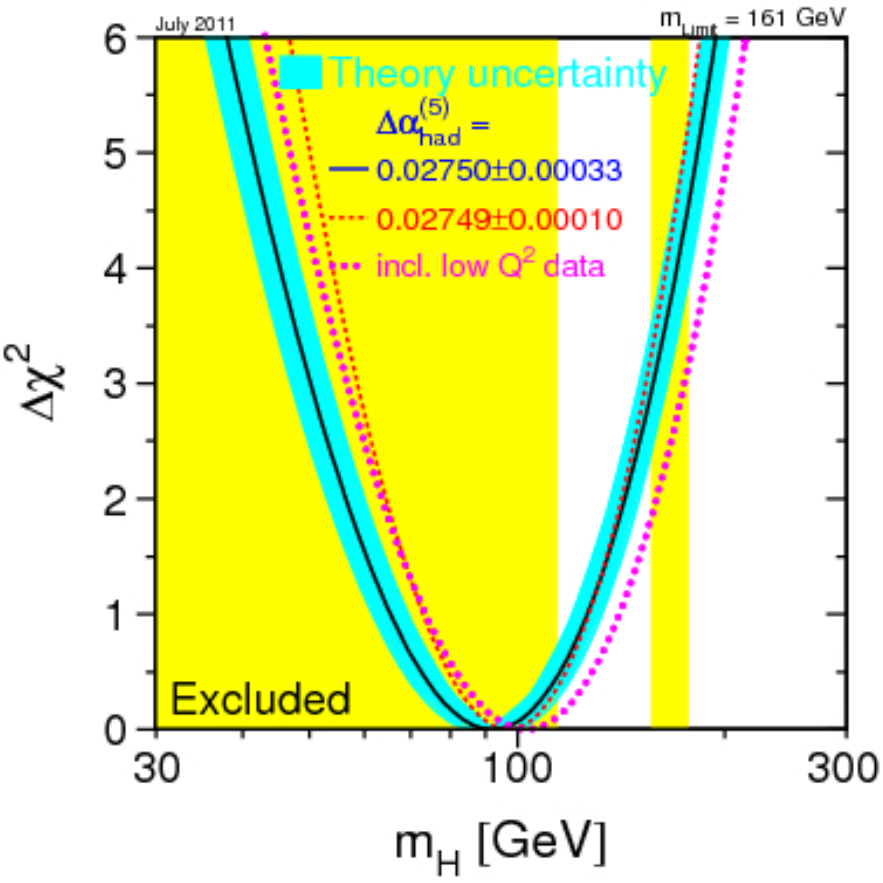


Figure 1.1.: The $\Delta\chi^2$ curve derived from high-Q^2 precision electroweak measurements, performed at LEP and by SLD, CDF, and D0, as a function of the Higgs-boson mass. The yellow band shows the Higgs masses excluded by LEP and Tevatron experiments.

mainly decays to $b\bar{b}$ quark pairs, as shown in Fig. 1.3 (b). The branching ratio to vector boson pairs increase at Higgs masses above 135 GeV/c^2.

The searches for the Higgs boson at Tevatron are carried at two different mass regions. Searches for the low mass Higgs boson are mainly conducted in the vector boson associated production; searches for the high mass Higgs boson are conducted in the gluon fusion production by also considering the associated production. The

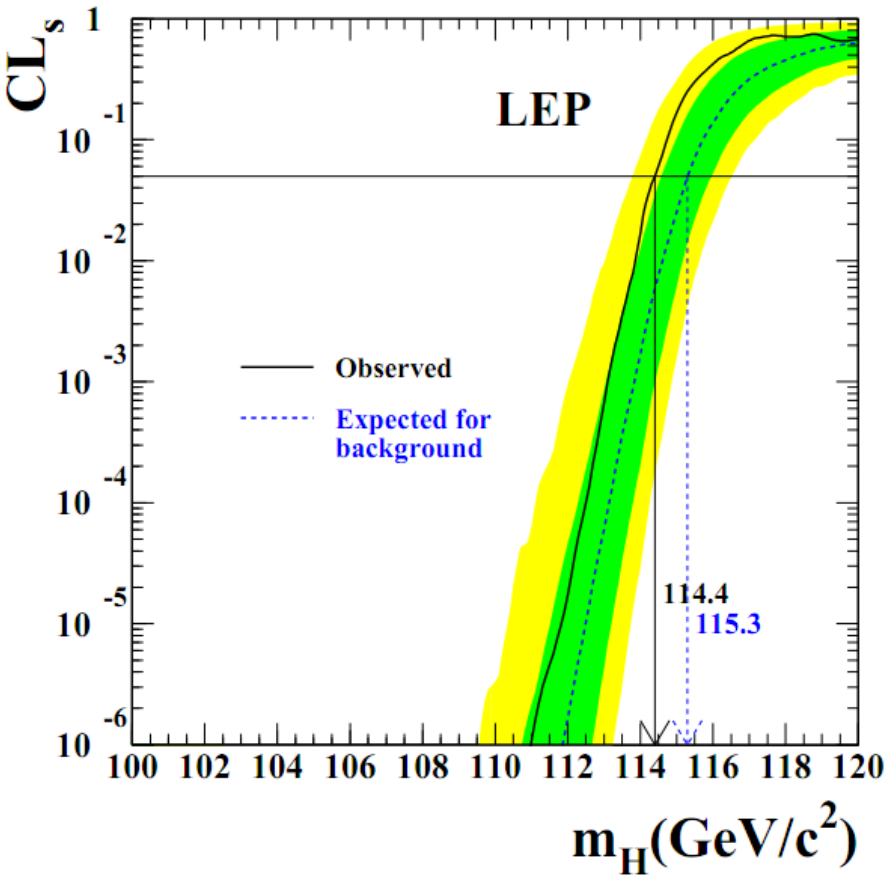
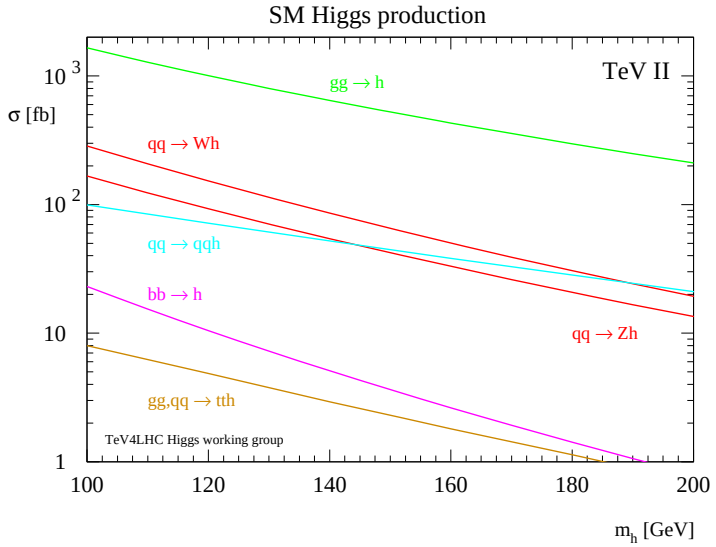
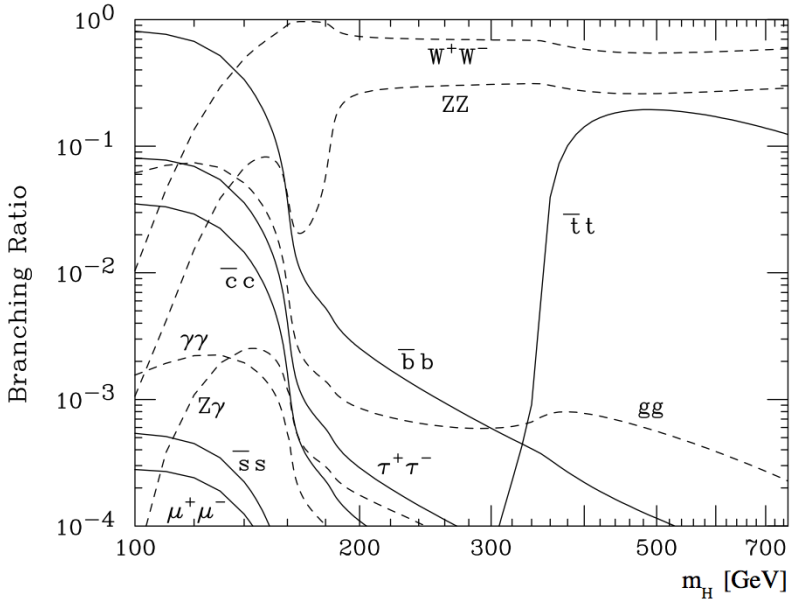


Figure 1.2.: The ratio $CL_s = CL_{s+b}/CL_b$ for the signal plus background hypothesis. Solid line: observation; dashed line: median background expectation. The dark and light shaded bands around the median expected line correspond to the 68% and 95% probability bands. The intersection of the horizontal line for $CL_s = 0.05$ with the observed curve is used to define the 95% confidence level lower bound on the mass of the standard model Higgs boson.

combined results [8] from the CDF and D0 experiments, along with the LEP, ATLAS and CMS results, indicate that the expected exclusion regions are, given the current sensitivity, $100 < m_H < 120$ GeV/c^2 and $139 < m_H < 184$ GeV/c^2, as shown in Fig. 1.4. In particular, both LHC experiments report local ~ 3 standard deviation excesses at approximately 125 GeV/c^2 [9].



(a) The SM Higgs boson production cross sections for $p\bar{p}$ collisions at 1.96 TeV



(b) Branching ratios for the main decays of the SM Higgs boson

Figure 1.3.: Higgs productions and decay modes at Tevatron

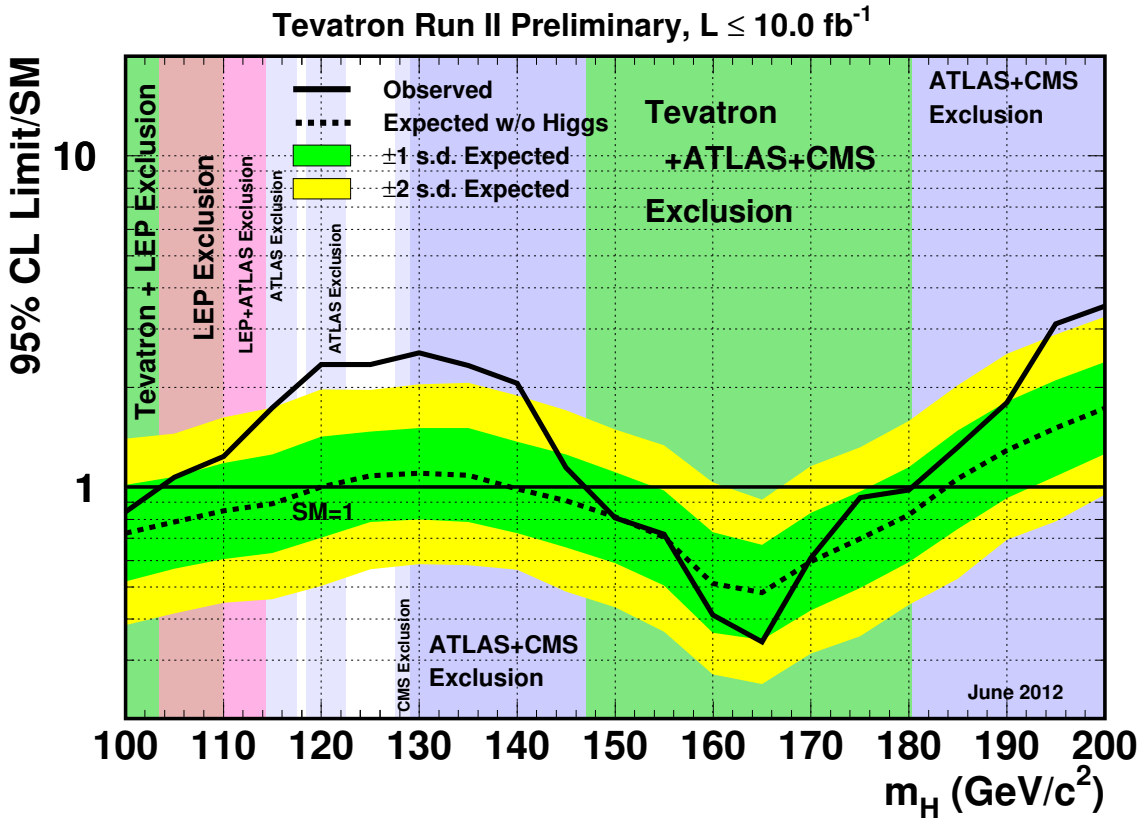


Figure 1.4.: Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on the ratios to the SM cross section, as functions of the Higgs boson mass for the combined CDF and D0 analyses. The limits are expressed as a multiple of the SM prediction for test masses (every 5 GeV/c^2) for which both experiments have performed dedicated searches in different channels. The points are joined by straight lines for better readability. The bands indicate the 68% and 95% probability regions where the limits can fluctuate, in the absence of signal. The limits displayed in this figure are obtained with the Bayesian calculation.

A main focus of this thesis is the search for the low mass Higgs boson ($90 < m_H < 150$ GeV/c^2) produced in association with a vector boson (W or Z), where the Higgs boson decays to $b\bar{b}$ pair and the vector boson decays leptonically with at least one neutrino in the final state.

1.2 Physics Beyond the Standard Model: Dark Matter Searches

The standard model of particle physics is not a complete or ultimate theory of the universe. Questions lying beyond-the-standard model (BSM) include, for instance, the origin of mass, the strong CP violation, neutrino oscillations, matter–antimatter asymmetry, and the nature of dark matter (DM) and dark energy.

In the universe we live in, more is unknown than known. Fits of theoretical models of the composition of the universe to the combined set of cosmological observations yield that only $\sim$ 5% of it is made of normal matter (stars, galaxies, neutrinos, electrons, protons, neutrons, etc.), with the remaining 72% of dark energy and 23% of dark matter. In particular, dark energy is proposed to explain the phenomenon that the universe is expanding at an accelerating rate [10]; dark matter is proposed to account for the missing mass observed in the gravitational curves of the galaxies.

In this section we review the astrophysical motivation for dark matter. We then proceed to discuss the various candidates and the potential to search them at accelerator experiments.

1.2.1 Dynamic Evidence

Why are we so sure that there are large amounts of dark matter lurking everywhere in the universe? The flat rotation curves of spiral galaxies provide perhaps the most direct and surely the most impressive evidence for the existence of dark matter. Spiral galaxies consist of a central bulge and a very thin disk which is stabilized against collapse by angular momentum conservation. The orbital velocity of the disk can be obtained by using the Doppler shift of spectral lines.

For a galaxy filled with luminous matter, the surface luminosity of the disk falls off exponentially with radius [11]

$$I(r) = I_0\, e^{-r/r_{\rm D}}, \tag{1.11}$$

where $r_{\rm D}$ is the "disk scale-length." Therefore, one would expect that most of the galactic mass is concentrated within a few scale-lengths and that the orbital velocity $v_{\rm rot}$ of the disk material is determined by this mass just as the orbital velocity of the planets in the solar system is dominated by the mass of the Sun. Because in such a system we have $v_{\rm rot} = \sqrt{G_{\rm N} M/r}$ (central mass M, Newton's constant $G_{\rm N}$) one expects the Keplerian $v_{\rm rot} \propto r^{-1/2}$ behavior in analogy to the solar system (Fig. 1.5).

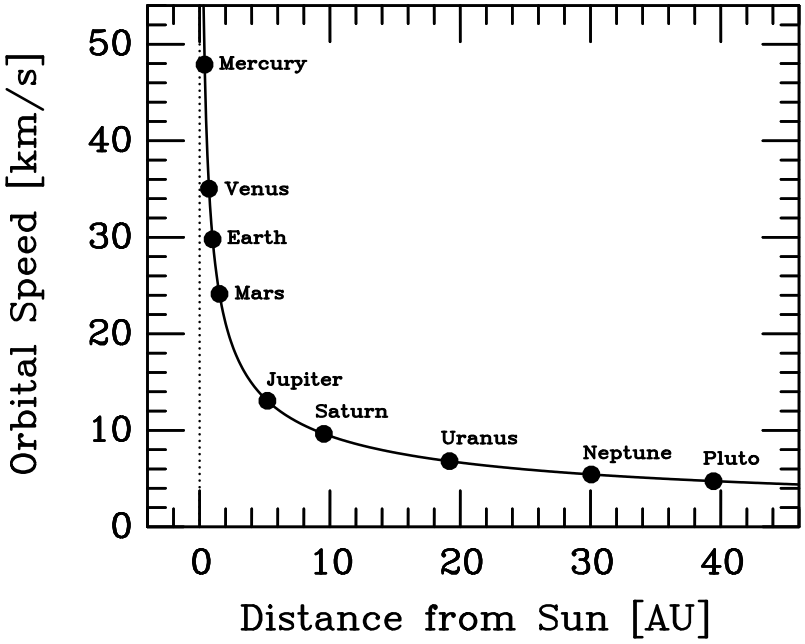


Figure 1.5.: Rotation curve of the solar system which falls off as $1/\sqrt{r}$ in accordance with Kepler's law. The astronomical unit (AU) is the Earth-Sun distance of $1.50 \times 10^{13}\,$cm.

Observations on many spiral galaxies found that the orbital velocity rises roughly linearly from the center outward in small radii. The rotation curve then stayed flat at the maximum velocity out to the largest measured radii. A case in point is the galaxy NGC 6503 where $r_{\rm D} = 1.73\,$kpc while the last measured hydrogen point is at $r = 22.22\,{\rm kpc} = 12.8\,r_{\rm D}$. The measured rotation curve is shown in Fig. 1.7 together with the relative components ascribed to the gravity of the disk alone and gas alone.

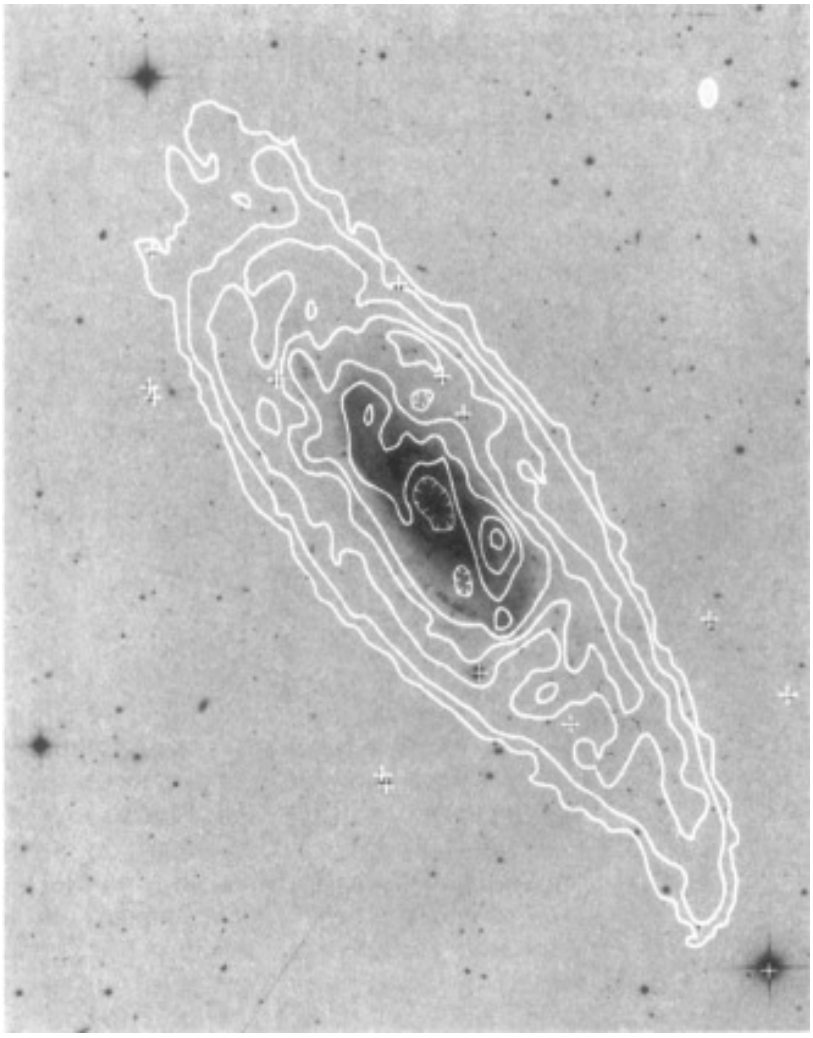


Figure 1.6.: Image of the spiral galaxy NGC 3198 with a superimposed contour map of the column density of hydrogen gas [12].

It is not easy to determine how much dark matter there is from these systems since the hydrogen gas runs out and we do not know how far the flat rotation curve extends. Nonetheless, it shows the galaxy to be made up of a nearly spherical "halo" of dark matter in which the disk is embedded. A review for details can be found in Ref. [14].

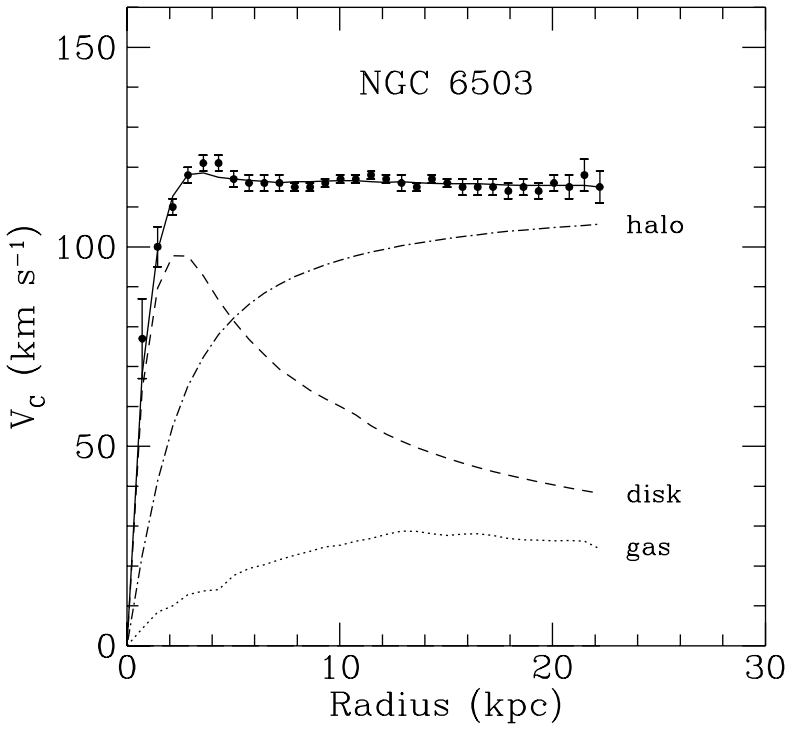


Figure 1.7.: Rotation curve of the spiral galaxy NGC 6503 as established from radio observations of hydrogen gas in the disk [13]. The last measured point is at 12.8 disk scale-lengths. The dashed line shows the rotation curve expected from the disk material alone, the dot-dashed line from the dark matter halo alone.

1.2.2 Candidates

Dark matter is matter that neither emits nor scatters light or other electromagnetic radiation; it is not observable through normal means, thus is dark. Several possibilities for dark matter candidates are explored in astronomy and physics. The two main categories of objects being considered as possibilities for dark matter are

- MACHOs (MAssive Compact Halo Objects),
- WIMPs (Weakly Interacting Massive Particles).

In the MACHOs theories, dark matter is due to extremely compact massive objects are subject to the gravitational effects. Due to the very large mass, it absorbs everything that comes to it, including light. Such models are explored by the astronomy experiments.

In the WIMPs models, dark matter (DM) can interact with the SM particles through the weak force, other than the gravity. Similar to neutrinos, these particles are relatively stable and leave very little trace and thus are difficult to be detected. In addition to the weak interaction, some BSM theories predict DM-SM interaction through more exotic mechanisms.

1.2.3 Searches

The non-gravitational interactions between dark matter and the SM particles open the way to three types of searches, as shown in Fig 1.8, besides of the astronomy searches.

Indirect Detection

Indirect detection experiments search for the products – in form of normal matter – of WIMPs annihilation or decay. Signatures, depending on theoretical models, can yield a variety of final states: high-energy gamma rays or SM particle-antiparticle

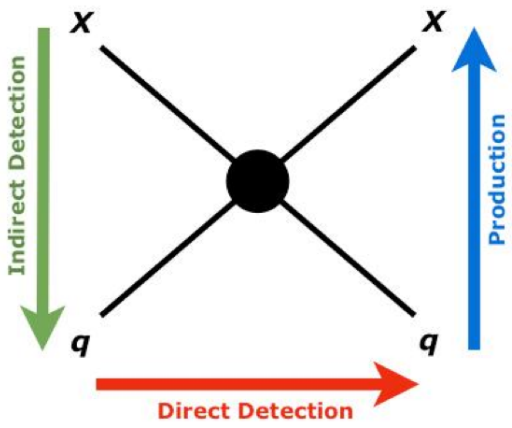


Figure 1.8.: Schematic view of three types of searches for DM-SM non-gravitational interactions; X represents dark matter particle and q is SM particle

pairs. These processes could be detected indirectly through an excess of gamma rays, antiprotons or positrons emanating from regions of high dark matter density. An example for such experiments is the Fermi Gamma-ray Space Telescope, which searches for gamma rays from dark matter annihilation and decay.

Direct Detection

Direct detection experiments search for DM-SM interactions through, for example, elastic scattering, typically by measuring the recoil energy of a scattered nucleus in a target detector. These experiments require a very low background rates, thus are usually performed in underground laboratories.

It was shown [15] that the motion of the earth around the sun introduces an annual modulation in the flux of dark matter particles reaching the earth (see Fig. 1.9). This is the type of signal searched for by the DAMA experiment [16] with a set of scintillating high purity NaI(Tl) crystals located at the Gran Sasso National Laboratory in Italy. The DAMA experiment found a peculiar annual modulation of the single-hit events in the (2-6) keV energy region. Based on 13 years of data, the con-

fidence level for the observed effect is cumulatively about 9σ CL. Meanwhile, the CoGeNT experiment [17], using a high-purity germanium crystal cooled to nitrogen temperature located at Soudan Underground Laboratory, seems to be able to detect an excess of events and an annual modulation with statistical significance of 2.8σ. There are still controversies on the compatibility of DAMA and CoGeNT results in comparison with the measurements of the other experiments (such as XENON100 [18] and EDELWEISS [19]). Uncertainties from both astrophysics and detector response open a non-conflicting window in the (mass - cross section) phase space, favoring a DM candidates of mass of about 5-10 GeV/c^2 and an interaction cross section with ordinary matter of about 100-500 fb, as shown in Fig. 1.10.

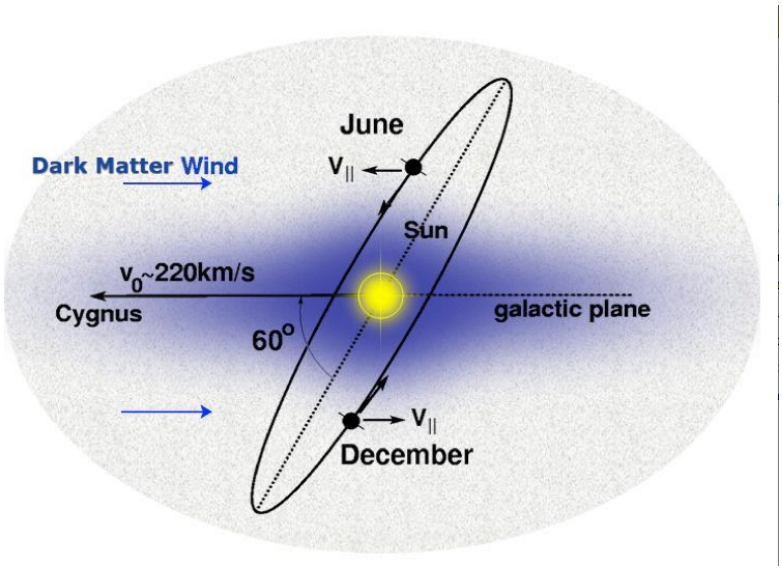


Figure 1.9.: Schematic explanation of the annual modulation signature reported by DAMA and CoGeNT

Search Dark Matter in Collider Experiments

An alternative approach to the direct dark matter searches is to search them in collider experiments. The interactions of DM and SM particles give the chance of producing dark matter at the particle colliders. Search for the production of dark

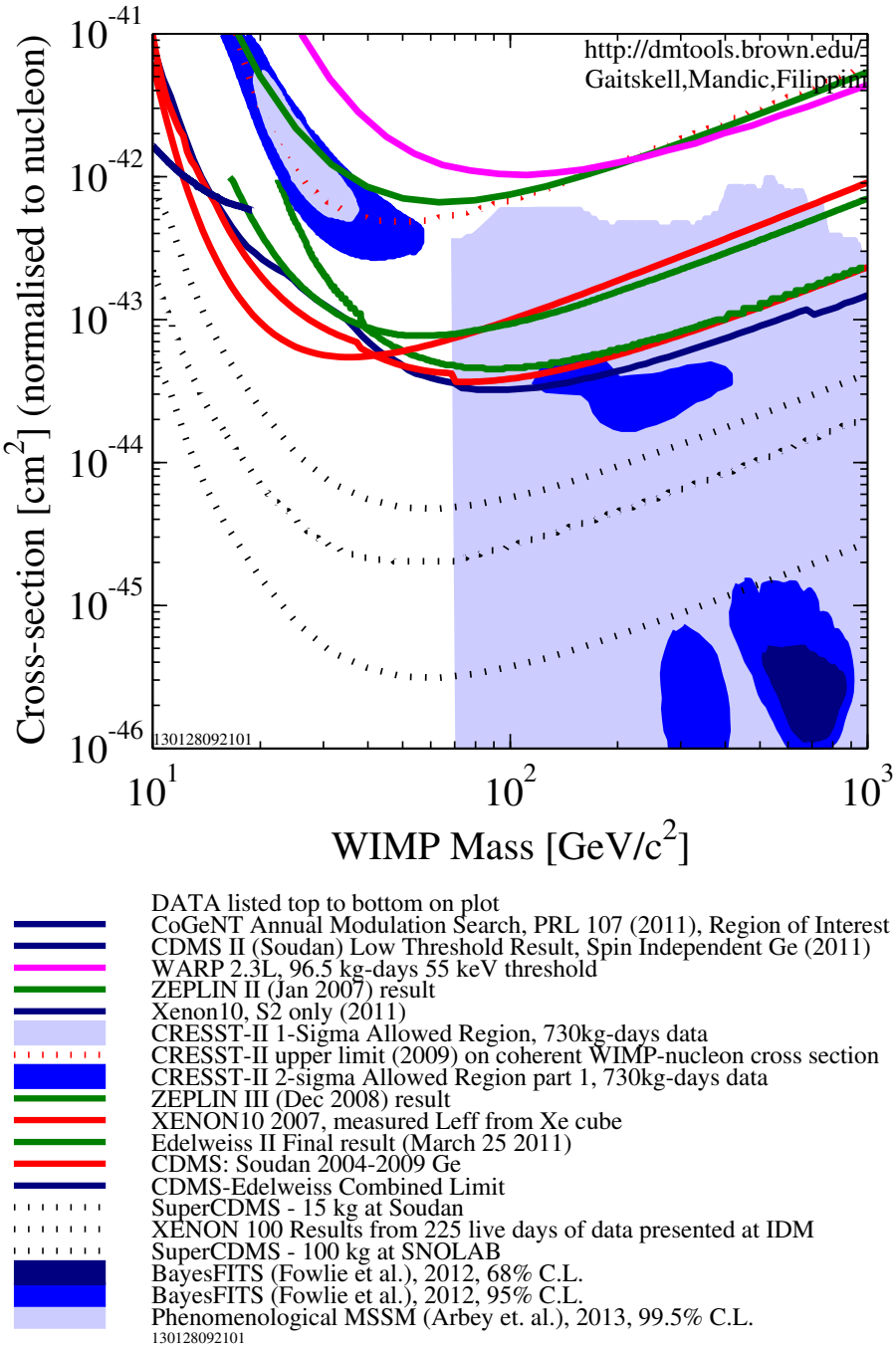


Figure 1.10.: Direct detection of WIMP dark matter, sensitivity plots, generated with the dark matter tools from Brown University [20]

matter in collider experiments consists primarily of looking for events with missing energy. For instance, in the Supersymmetry theory, if the R-parity preserves, the Lightest Supersymmetric Particle is stable and is a Weakly Interacting Massive Particle. Such a dark matter candidate is neutral and weakly interacting, once created, the particle will escape detection without leaving a track.

If the direct detection experiments are to prove the existence of dark matter, the collider experiments can further explain how the dark matter is created. This could be very important in understanding how the 25% of the Universe formed after Big Bang.

A major effort of this thesis is the search for a dark-matter production with the Tevatron proton-antiproton collision data. A model-independent theory, which predicts a final state of a dark matter particle and a top quark, has been investigated the first time at hadron colliders.

2. THE CDF EXPERIMENT

In this thesis studies of high p_T physics in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV are presented, using data collected by the CDF detector at the Tevatron collider at Fermilab.

2.1 Tevatron

The Tevatron is a proton-antiproton collider with the center-of-mass energy of 1.96 TeV. The protons and antiprotons are produced and accelerated through the Fermilab accelerator complex, as illustrated in Fig. 2.1.

At the first stage, the Cockcroft-Walton preaccelerator is used to ionize hydrogen gas and accelerate the negative ions to 750 keV. The ions are then passed into the 150-meter long linear accelerator (Linac) which uses oscillating electric fields to accelerate the ions to 400 MeV. The ions then pass through a carbon foil to remove the electrons and protons, and are moved into the Booster. The protons are accelerated to 8 GeV by the Booster and passed to the Main Injector.

The Main Injector has several operating modes: It can accelerate protons up to 150 GeV; it can produce 120 GeV protons for antiproton creation ($p+p \rightarrow p+p+p+\bar{p}$); it can increase antiproton energy to 120 GeV; it can inject protons and antiprotons into the Tevatron.

The Tevatron is a synchrotron that accelerates protons and antiprotons in a 6.28 km ring to energies of 980 GeV. The protons and antiprotons are accelerated in opposite directions crossing paths in the CDF and D0 detectors to collide at 1.96 TeV.

The Tevatron luminosity depends on several parameters,

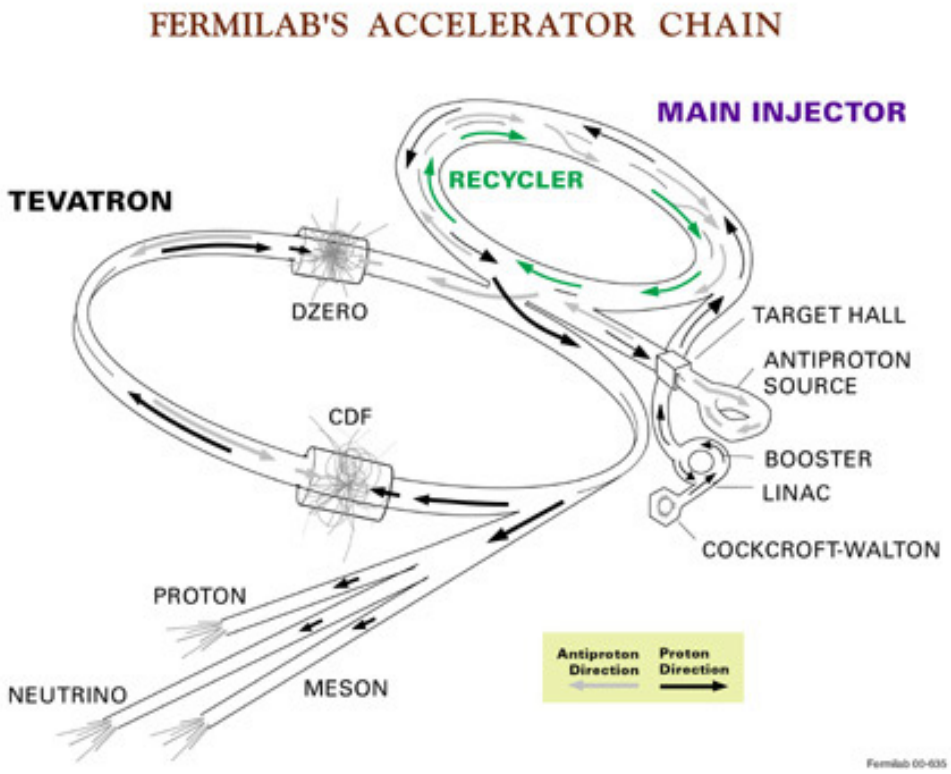


Figure 2.1.: Fermilab's accelerator chain

$$L = \frac{10^{-6} f B N_p N_{\bar{p}(6\beta_r\gamma_r)}}{2\pi\beta^*(\epsilon_p + \epsilon_{\bar{p}})} H(\sigma_l/\beta^*)\,(10^{31}\mathrm{cm}^{-2}\mathrm{s}^{-1}), \tag{2.1}$$

where $f = 47.7$ kHz is the revolution frequency; $B = 36$ is the number of bunches in a store; N_p and $N_{\bar{p}}$ are the bunch intensities for proton and antiproton bunches; $\beta_r\gamma_r = 1045$ is the *relativistic* $\beta \times \gamma$; $\beta^* = 35$ cm is the β function at the interaction region (IR); $H = 0.65 \sim 0.75$ is the hourglass factor; ϵ_p and $\epsilon_{\bar{p}}$ are the transverse emittances; σ_l is the bunch length (cm).

The Tevatron design luminosity was $2.86 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The actual luminosities evolve with time. Figure 2.2 shows the instantaneous peak luminosities of the Tevatron Run II as a function of time. The best instantaneous luminosity of $4.31 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ was reached on 3 May, 2011.

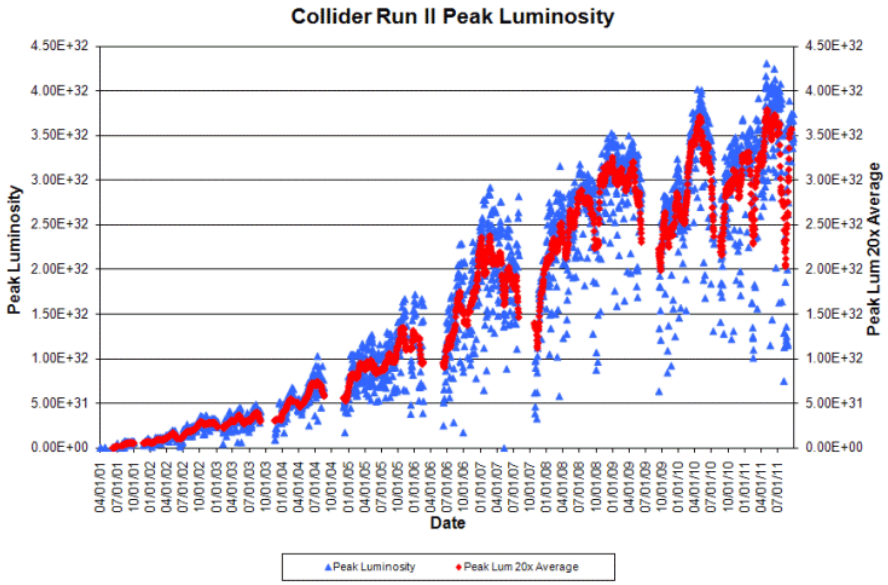


Figure 2.2.: The instantaneous peak luminosity as a function of time of the Tevatron.

The integrated luminosity (the size of cumulative data) is shown in Fig. 2.3. In Run II from March 2001 to September 2011, the Tevatron has delivered a total amount of data of about 12 fb^{-1} to both CDF and D0 experiments.

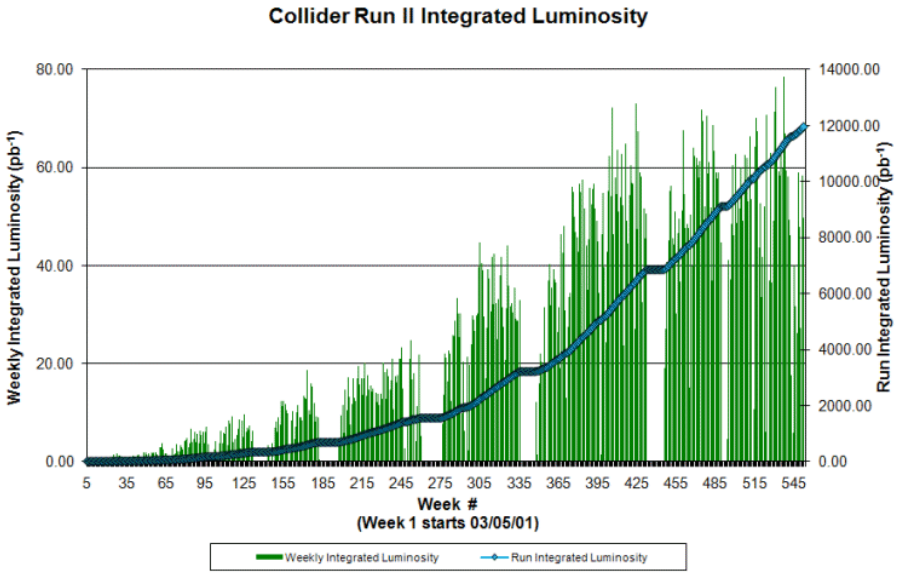


Figure 2.3.: The integrated luminosity as a function of time of the Tevatron.

2.2 The CDF Detector

CDF is a general purpose detector system. It combines precision charged particle tracking with fast projective calorimetry and fine grained muon detection. Such a sophisticated detector is required to detect the decay products of the $p\bar{p}$ collisions such as photons, electrons, muons, pions, kaons, protons, and neutrons. Neutrinos, or other weakly interacting particles, are undetectable in such system, but are indirectly measurable as an imbalance of energy.

The detector system is shown in an elevation view in Fig. 2.4, and a cutaway view in Fig. 2.5. The innermost part is the tracking system. The superconducting solenoid magnet is containing the tracking system, and provides a 1.4 T magnetic field parallel to the beam axis. Calorimeter and muon detector are all outside the solenoid.

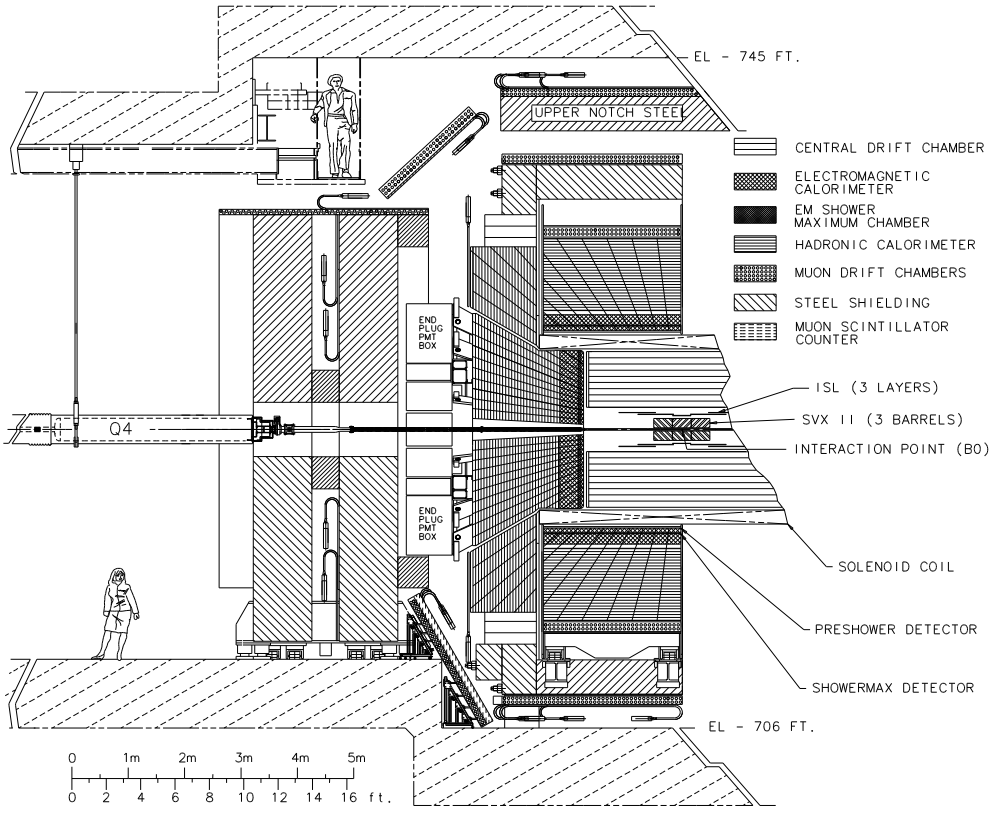


Figure 2.4.: Elevation view of one half of the CDF II detector

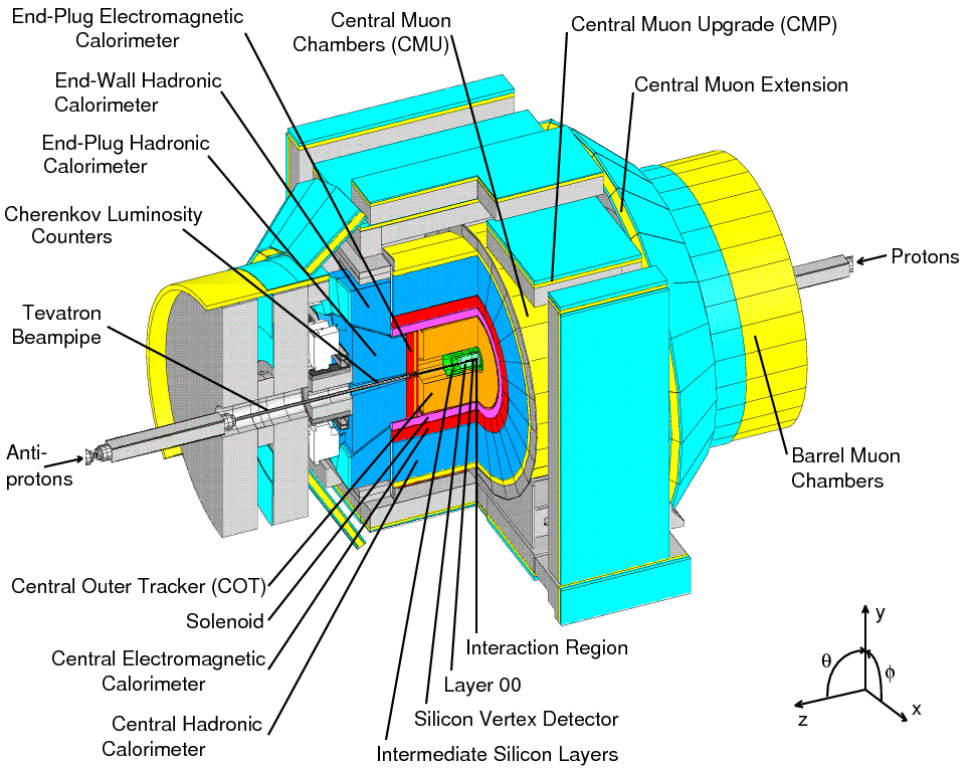


Figure 2.5.: A cutaway view of the CDF II detector with quadrant cut to expose the different subdetectors.

The detector elements are listed below and characterized by their functionalities:

- Tracking systems measure the momenta of charged particles, such as electrons, muons, charged pions and kaons, and protons, and performs precise measurement of particle impact parameters.
- Calorimeters measure the energies of photons and electrons through electromagnetic interactions and hadrons through hadronic interactions;
- Muon detectors identify muons, and also provides limited improvement to the muon momentum measurement.

We use a coordinate system where the polar angle θ is measured from the proton direction (also the z direction), the azimuthal angle ϕ is measured from the Tevatron plane. In a hadron collision, the momentum carried by each quark or gluon

is uncertain. To make the translation along the z axis easier, the pseudo-rapidity $\eta = -\ln(\tan(\theta/2))$ is defined to replace the use of θ, as the differences in pseudo-rapidity is invariant under boosts along the z axis with $\eta_1 - \eta_2 = \eta_1' - \eta_2'$.

2.2.1 Tracking Systems

A charged particle moves in a bending trajectory in the magnetic field, and the curvature indicates the particle momentum. In units used by nuclear and particle physicists, the relation is,

$$p_T = 0.3Br = 0.42r, \tag{2.2}$$

where p_T is the transverse momentum measured in GeV/c, B is the magnetic field in tesla (1.4 T for CDF) and r is the radius in meters. The track of a charged particle is reconstructed through the measurement of the interaction between the particle and the layers of detecting materials, also known as "hits".

The CDF II tracking systems consist of inner silicon detectors and a central outer tracker (COT). The silicon detectors provide high resolution for the vertex reconstruction. They include a single layer radiation-harden detector mounted on the beam pipe (L00), a double-sided five layer inner silicon detector (SVX II), and a double-sided two layer intermediate silicon tracker (ISL). The COT is a gaseous drift chamber filled with tens of thousands of gold wires arranged in layers and argon gas.

The CDF II tracking systems are shown schematically in Fig. 2.6. The main parameters of the CDF II tracking systems are summarized in Tab. 2.1 and Tab. 2.2.

The Inner Silicon Detectors

Silicon detector works, in principle, by doping narrow strips of silicon to make them into diodes, which are then reverse biased. As charged particles pass through these strips, they cause small ionization currents which can be detected and measured. Silicon detectors provide high resolution for the vertex detection while cost very high.

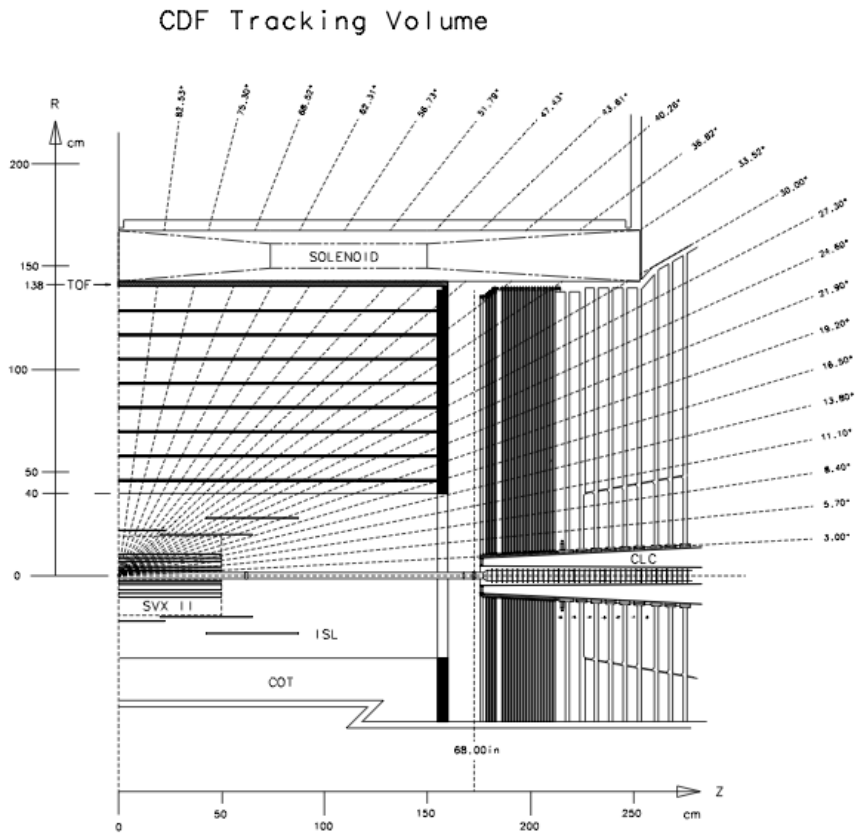


Figure 2.6.: Longitudinal View of the CDF II Tracking Systems

Figure 2.7 and 2.8 illustrate the structure of the inner silicon detectors at CDF in side view and end view, respectively.

Layer 00 (L00) is the innermost detector with a layer of single-sided axial strips located between a radius of 1.35 cm and 1.62 cm from the beam line. Each sensor is 7.84 cm in length. The implants have a pitch of 25 μm. An alternate-strips readout gives a readout pitch of 50 μm and a hit resolution of approximately 6 μm.

L00 serves two primary purposes. First, it sits inside the silicon tracker SVX II, and enhances the impact parameter resolution and pattern recognition and improves

Table 2.1: Design parameters of the CDF II inner silicon detectors.

Layer 00	
Radial coverage	1.35 to 1.65 cm
Resolution per measurement	6 μm (axial)
Number of channels	13,824
SVX II	
Radial coverage	2.4 to 10.7 cm, staggered quadrants
Number of layers	
Readout coordinates	r-ϕ on one side of all layers
Stereo side	r-z, r-z, r-sas, r-z, r-sas (sas $\equiv \pm 1.2^\circ$ stereo)
Readout pitch	60-65 μm r-ϕ, 60-150 μm stereo
Resolution per measurement	12 μm (axial)
Total length	96.0 cm
Rapidity coverage	$\lvert\eta\rvert \leq 2.0$
Number of channels	423,900
ISL	
Radial coverage	20 to 28 cm
Number of layers	one for $\lvert\eta\rvert < 1$; two for $1 < \lvert\eta\rvert < 2$
Readout coordinates	r-ϕ and r-sas (sas$\equiv \pm 1.2^\circ$ stereo) (all layers)
Readout pitch	110 μm (axial); 146 μm (stereo)
Resolution per measurement	16 μm (axial)
Total length	174 cm
Rapidity coverage	$\lvert\eta\rvert \leq 1.9$
Number of channels	268,800

significantly the ability to identify b-quark jets. The improvements in impact parameter resolution are (p_T is in units of GeV/c)

$$\sigma = 9 \oplus \frac{34}{p_T} \rightarrow 6 \oplus \frac{22}{p_T} \mu\text{m} \tag{2.3}$$

Table 2.2: Design parameters of the CDF II central outer tracker

COT	
Radial coverage	44 to 132 cm
Number of superlayers	8
Measurements per superlayer	12
Readout coordinates of SLs	+3° 0° -3° 0° +3° 0° -3° 0°
Maximum drift distance	0.88 cm
Resolution per measurement	180 μm
Rapidity coverage	$\|\eta\| \leq 1.0$
Number of channels	30,240

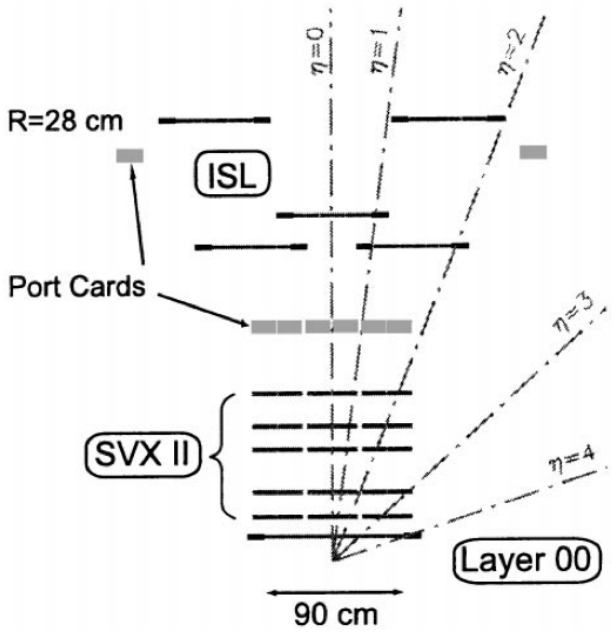


Figure 2.7.: A side view of half of the CDF II silicon detectors on a scale in which the z coordinate is highly compressed.

for tracks that do not pass through SVX II hybrids and

$$\sigma = 9 \oplus \frac{66}{p_T} \rightarrow 6 \oplus \frac{27}{p_T} \mu\text{m} \tag{2.4}$$

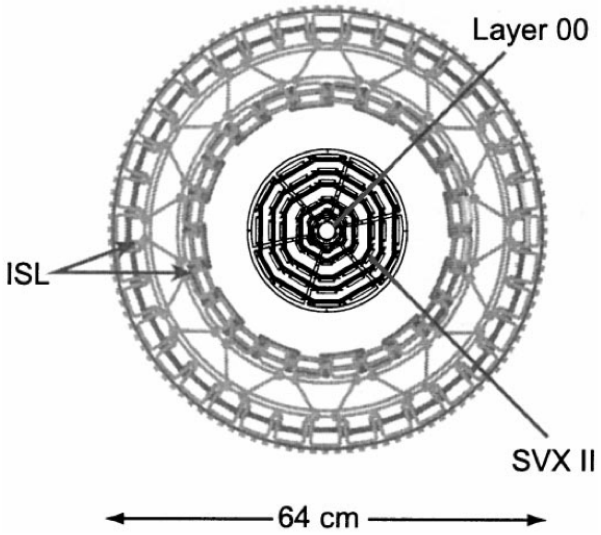


Figure 2.8.: An end view of the CDF II silicon detectors including the SVX II cooling bulkheads and ISL support structure.

for tracks that pass through hybrids in the innermost layer of SVX II.

Second, the sensor design was based on the R&D conducted for the development of the LHC detectors that operate at higher bias voltages, enabling good signal-to-noise even after extreme radiation doses. The detector can continue to operate after significant radiation damage.

Silicon Vertex Detector (SVX II) is a redesign and upgrade of the SVX to deal with the higher luminosity and shorter bunch crossing of the Tevatron Run II upgrade. SVX II is a silicon microstrip detector, made by thin silicon wafers. On the wafer surface several narrow lines of impurity atoms are implanted to form closely spaced strips. Diode junctions are thus created between the wafer and the strips. Applying a voltage with the proper polarity to the diode increases the depletion depth. The depletion junction, a free charged zone with strong electric field, is the actual sensitive region of the microstrip detector. When an ionizing particle passes through this region electrons are promoted to the conduction band which generates an electrical signal on a few strips. The strips are then read out by fast electronics

revealing, to within a fraction of the strip spacing, where the particle is intercepted the wafer along one dimension.

The whole SVX II coverage along z is driven by the spread of the primary interaction along this axis. The interactions are distributed approximately as a gaussian shape along the beam direction, with an average standard deviation of $\sim$30 cm. To provide a geometrical acceptance greater than 70%, the length of SVX II is $\sim$45 cm on both sides of the nominal interaction point. The pseudorapidity coverage is $|\eta| < 2$. SVX II is characterized by 12-fold symmetry in ϕ and consists of three 32 cm long cylindrical barrels, with 5 layers each. These layers consist of double side silicon wafers (ladders) mounted with staggered radii to provide some overlap between adjacent ladders. The innermost layer is placed at a radius of 2.4 cm, while the outermost is at a radius of $\sim$10.7 cm. Three of the 5 layers have on one side the microstrips aligned to the beam while on the other side orthogonal, the remaining two have instead the strips on the two sides with a small-angle stereo. These strips are spaced in $r - \phi$ by approximately 60 to 65 microns, depending on layer, and have implant widths of 14 to 15 microns. The stereo strips of the SVX II are spaced by (141, 125.5, 60, 141, 65) microns, and have implant widths of 20 microns for the 90° strips and 15 microns for the small-angle stereo layers. This design allows both a good resolution on the z-position of secondary vertices and an enhanced 3D pattern recognition. The length of each ladder is 29 cm, each divided into two half-ladders that are read out independently. The readout electronics consist of hybrid chips that are mounted directly to the silicon surface at each end of the half-ladder. The choice of electronics located on the silicon sensors was made to fulfill the requirement of a fast response and low occupancy detector. The electronics inside the detector implied the addition of material within the active sensitive volume like cables and cooling tubes. This increases the multiple scattering of a particle worsening the pattern recognition capability. To partially mitigate this effect another layer of silicon at small radius (Layer 00) was added to SVX II.

Intermediate Silicon Layers (ISL) is composed of five barrels in total, each barrel being composed of single layer of double-sided silicon microstrip sensors. An isometric view of the ISL barrels is shown in Fig. 2.9. The barrels are positioned at radius of 22.6/23.1 cm (central barrel), 19.7/20.2 cm (forward/backward inner barrels) and 28.6/29.0 cm (forward/backward outer barrels) such that there are coordinate measurements at two (one) positions for forward (central) tracks. Each barrel is made up from ladders, each ladder consisting of 6 silicon mircrostrip sensors which are ganged into two 3-sensor groups so that signals are read out from either end of the ladder. The silicon sensor is AC coupled and double sided having two planes with 112 micron pitch readout strips. The strips are running at a stereo angle of $\sim 1.2°$. The main function of the ISL is to measure the particle momentum in the forward regions where the outer tracker, COT, can not fully cover and to provide another hit points from which track segments in SVX II/ISL detector are searched for.

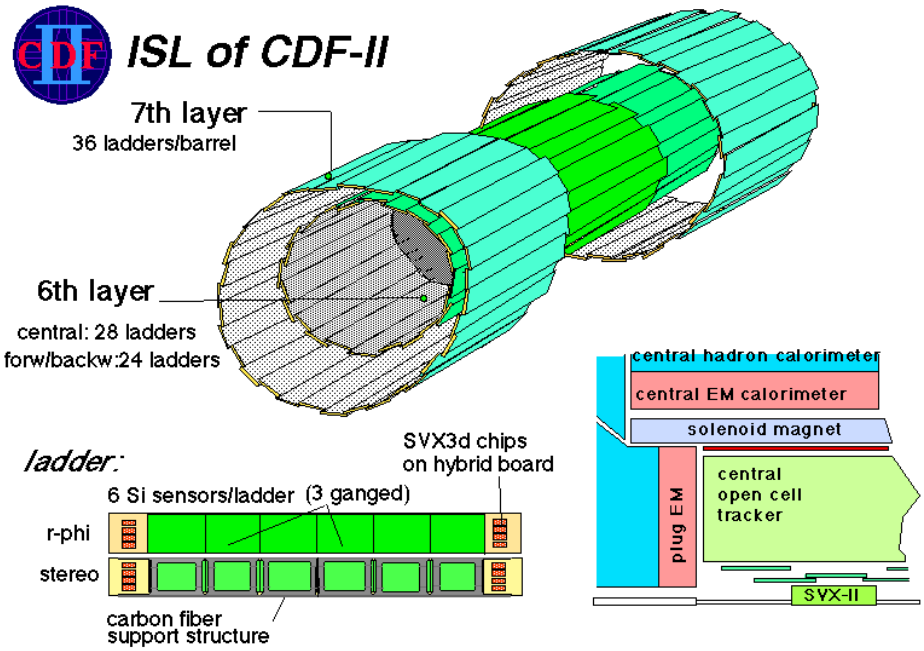


Figure 2.9.: An isometric view of the ISL barrels

Central Outer Tracker (COT)

The Central Outer Tracker, or COT, is a large cylindrical drift chamber constructed to replace the original central drift chamber for the higher luminosity expected for Run II. It is located in the radial region outside the silicon detectors and inside the time-of-flight (TOF) scintillators. The active volume of COT spans 310 cm in the beam (axial) direction, z; between 43.4 cm and 132.3 cm in radius, r; and the entire azimuth, ϕ.

The COT contains 30240 sense wires that run the length (in z) of the chamber between two end plates. Approximately half of the wires are axial and half are small angle (2°) stereo. The 96 sense wire layers in radius are grouped into eight "superlayers", as inferred from the end plate section shown in Fig. 2.10. Each superlayer is divided in ϕ into "supercells", and a maximum drift distance is approximately the same for all superlayers.

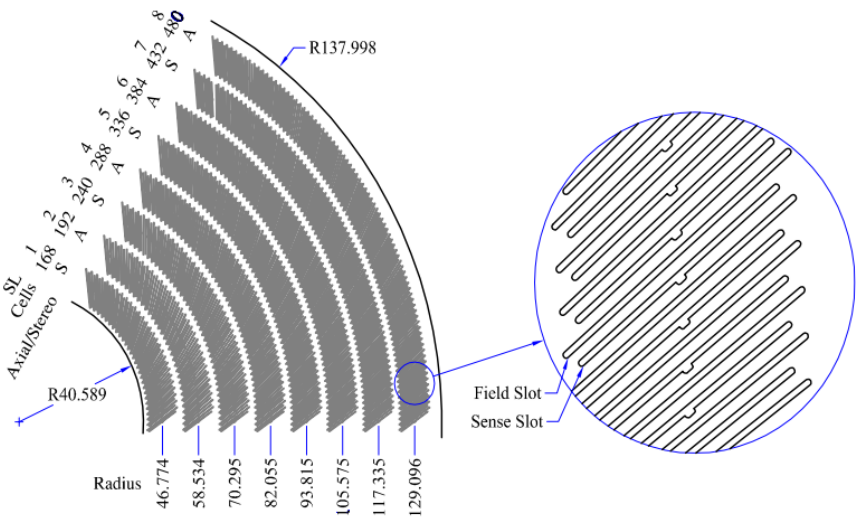


Figure 2.10.: 1/6 section of the COT end plate. For each superlayer is given the total number of supercells, the wire orientation (axial or stereo), and the average radius. The enlargement shows the sense and field slot geometry in detail. Dimensions are in cm.

COT is designed to operate with a maximum drift of 100 nsec by reducing the maximum drift distance and by using a gas mixture with a faster drift velocity. A gas mixture containing 50:35:15 Ar-Et-CF_4 has drift velocity of ~ 100 μm/ns which implies a maximum drift distance of ~ 1 cm.

The achieved performance of the integrated CDF tracking systems is a transverse momentum resolution $\sigma(p_T)/p_T^2 = 0.15\%$ (GeV/c)$^{-1}$ and an impact parameter resolution $\sigma(d) = 35$ μm at 2 GeV/c.

2.2.2 Time-of-Flight Detector

The Time-of-Flight detector, or TOF, is based on plastic scintillators and fine-mesh photomultipliers, as shown in Fig. 2.11. The primary physics motivation is to provide charged kaon identification to improve neutral B meson flavor determination. The distance from the beam pipe, which coincides with an axis of the barrel, is 140 cm, corresponding to about 5 ns flight time for the fastest particles. The TOF resolution of 100 ps provides at least two standard deviation separation between $K^{\pm}$ and $\pi^{\pm}$ for momenta $p < 1.6$ GeV/c, complementing the specific ionization energy loss, dE/dx, measured in the drift chamber. Besides that, the TOF detector is also a powerful tool for triggering: it has an advantage of a good ϕ segmentation associated with a high speed. The TOF hit information is already available to the first-level trigger, which generates a decision within 5.5 μs: It is used to trigger on magnetic monopoles, based on the fact that monopoles heavily ionize material they travel through due to the large coupling to the photons. In addition, the TOF detector has a limited amount of material between the beam line and scintillators, which is important in searches for the monopoles.

2.2.3 Calorimeters

Calorimeters measure the energy of particles. They complement the tracking systems by measuring the neutral particles, which leave no traces in the trackers.

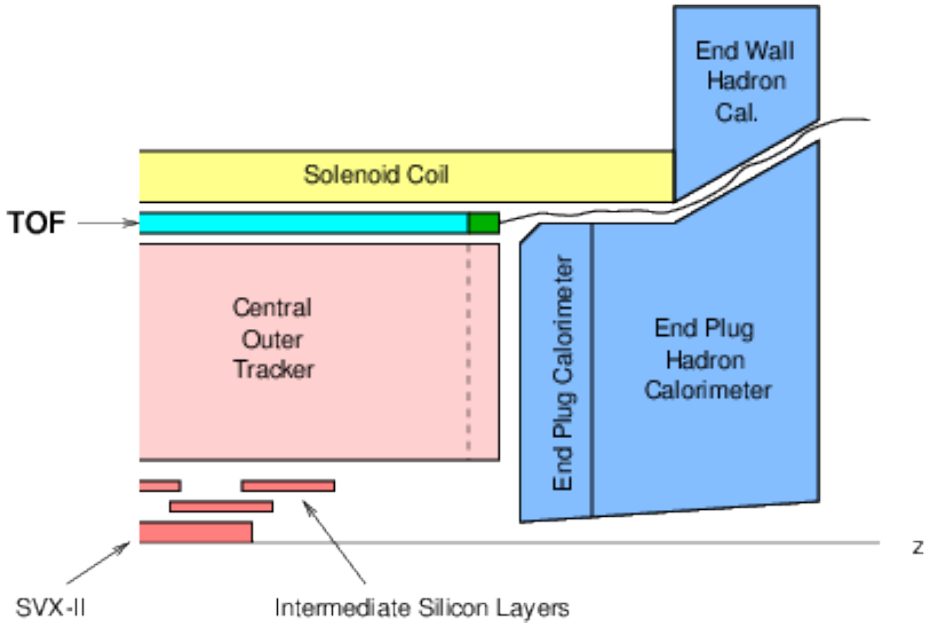


Figure 2.11.: The Time-of-Flight detector is a barrel of scintillators almost 3 m long that was installed in a few centimeters clearance just between the Central Outer Tracker and the solenoid.

They also provide measurement to the high p_T charged particles, for which as the p_T increases, the tracking resolution worsens. There are electromagnetic calorimeters and hadronic calorimeters. An electromagnetic calorimeter is specifically designed to measure the energy of particles that interact primarily via the electromagnetic interaction, such as the photons and electrons. A hadronic calorimeter is designed to measure particles that interact via the strong nuclear force.

The CDF calorimeter system consists of lead/scintillator sampling electromagnetic (EM) and iron/scintillator sampling hadronic (HAD) calorimeters covering the range $|\eta| < 3.64$. Both the central ($|\eta| < 1.1$) and plug ($1.1 < |\eta| < 3.64$) electromagnetic calorimeters have fine grained shower profile detectors at electron shower maximum, and pre-shower pulse height detectors at approximately $1 \cdot X_0$ depth, where X_0 is the radiation length. Electron identification is accomplished using E/p from the EM calorimeter; using $HAD/EM \sim 0$ – little energy leak in hadronic calorimeter; and

using shower shape and position matching in the shower max detectors. Together with the dE/dx information provided by the COT, CDF achieves $\sim 10^{-3}$ pion/electron rejection in the central region. The parameters for the central and plug calorimeters are given in Table 2.3. A side view of the east end plug, the wall hadronic calorimeter (WHA), and portions of the solenoid, central electromagnetic calorimeter (CEM) and central hadronic calorimeter (CHA) is shown in Fig. 2.12.

The CDF calorimeters have played a key role in the physics program by measuring electron and photon energies, jet energies, and net transverse energy flow.

Table 2.3: Central and Plug upgraded calorimeter comparison

	Centeral	Plug
EM Calorimeter		
Thickness	$19X_0$, 1λ	$21X_0$, 1λ
Sample (Pb)	$0.6X_0$	$0.8X_0$
Sample (scint.)	5 mm	4.5 mm
WLS	sheet	fiber
Light yield	160 pe/GeV	300 pe/GeV
Sampling res.	$11.6\%/\sqrt{E_T}$	$14\%/\sqrt{E}$
Stoch. res.	$14\%/\sqrt{E_T}$	$16\%/\sqrt{E}$
Shower Max. seg. (cm)	$1.4\phi \times (1.6 - 2.0)$ Z	0.5×0.5 UV
Pre-shower seg. (cm)	$1.4\phi \times 65$ Z	by tower
Hadron Calorimeter		
Thickness	4.5λ	$7\ \lambda$
Sample (Fe)	1 to 2 in.	2 in.
Sample (scint.)	10 mm	6 mm
WLS	finger	fiber
Light yield	$\sim$40 pe/GeV	39 pe/GeV

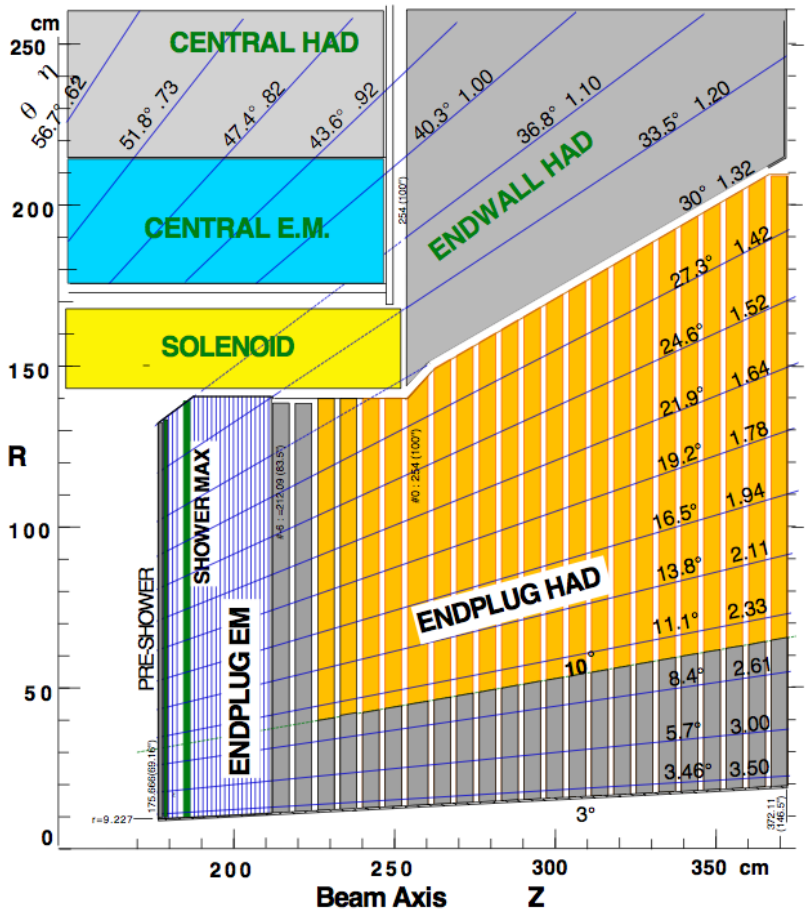


Figure 2.12.: A side view of the east end plug, the wall hadronic calorimeter (WHA), and portions of the solenoid, central electromagnetic calorimeter (CEM) and central hadronic calorimeter (CHA)

Central Calorimeter

The CDF central calorimeter is a scintillator sampling system with tower segmentation; each tower is 15 degree in azimuth by about 0.11 in pseudorapidity. Each wedge consists of a lead/scintillator EM section backed by an iron/scintillator central hadron calorimeter. The wedges are assembled into central arches. In the EM calorimeter, for each 15 degree wedge, it has alternating lead and scintillator with an imbedded two dimensional readout strip chamber at shower maximum. Wavelength shifting fiber (WLS) at the ϕ surfaces carry the light out to photomultiplier tubes

(PMTs) located on the back plane of each endplug. The EM calorimeter, along with the tracking and hadron calorimeter, has provided effective identification of electron and photons. The central hadron calorimeters are composed of alternating layers of iron and scintillators.

Endwall Hadron Calorimeter

The endwall hadron calorimeter consists of modules mounted to the solenoid flux return to provide hadronic coverage in polar angle θ from 30° to 45° on both side. Its composition is the same to the central hadron calorimeter.

Plug Calorimeter

The plug calorimeter covers $1.1 < |\eta| < 3.64$, and consists of an electromagnetic section followed by a hadronic section. The basic structure of the calorimeter is a lamination of scintillator plates segmented into tiles, interleaved with absorber plates. The EM calorimeter has lead plates, while the HAD calorimeter has steel plates. The EM and HAD is segmented into the same projective towers.

Shower Maximum Detector

To measure the position of electron and photon showers, and separate electron and photon decaying from neutral pion, a position sensitive shower-maximum detector (SMD) has been placed inside the EM section of the plug calorimeter at a depth of approximately 6 radiation lengths. The SMD is divided into eight 45° sectors, each covering the region from 11 cm to an outer radius of 130 cm from the beam axis. Each sector contains two layers (U and V) of 5 mm wide scintillator strips. The U and V layers are held together by two 1/8 inch thick lexan covers, and inserted into the fifth sampling slot within the EM calorimeter. The two scintillator laters U and V

are aligned at +22.5° and -22.5° with respect to the symmetry axis of the 45° sector, to provide a two-dimensional position measurement.

2.2.4 Muon System

The calorimeter steel serves as a filter for muon detection in the central (CMU) and extension (CMX) muon proportional chambers, over the range $|\eta| < 1$, $p_T > 1.4$ GeV/c. Additional iron shielding, including the magnet yoke, provides a muon filter for the upgrade muon chambers (CMP) in the range $|\eta| < 0.6$, $p_T > 2.2$ GeV/c. The (non-energized) forward toroids from Run I provide muon filters for intermediate $1.0 < |\eta| < 1.5$ muon chamber (IMU) for $p_T > 2$ GeV/c. Scintillators for triggering are included in CMP, CMX, and IMU. Muon identification is accomplished by matching track segments in the muon chambers with COT/SVX tracks; matching is available in $r\phi$ for all detectors and in the z views in CMU and CMX. The muon systems' parameters are summarized in Tab. 2.4.

2.2.5 Electronics and Triggering

The CDF II electronics systems are redesigned to handle the reduced separation between particle bunches and the increased instantaneous luminosity of the Tevatron Run II. The bunch crossing rate at the Tevatron is about 2.5 MHz, since the bunch spacing is 396 ns. The actual cross rate, diminished due to the train structure of the beam, is about 1.7 MHz. Given a event size of $\sim 1/4$ Megabyte, and the tape writing speed of 20 MB/s, it is required to significantly reduce the size of events for recording, from 17 million events per second to less than 100 events per second.

A triggering system is designed to preselect events online and decide if a collision event is unwanted or of interest. Figure 2.13 shows the functional block diagram of such system. A three level of pipeline decision system is selected for this purpose. Level 1 (L1) and Level 2 (L2) are hardware based systems while the Level 3 filters run on a dedicated computer farm. L1 and L2 hold only a part of event data, from

Table 2.4: Design parameters of the CDF II Muon detectors. Pion interaction lengths and multiple scattering are computed at a reference angle of $\theta = 90^\circ$ in CMU and CMP, at an angle of $\theta = 55^\circ$ in CMX, and for a range of angles for the IMU.

	CMU	CMP	CMX	IMU
Pseudo-rapidity coverage	$\|\eta\| \leq 0.6$	$\|\eta\| \leq 0.6$	$0.6 \leq \|\eta\| \leq 1.0$	$1.0 \leq \|\eta\| \leq 1.5$
Drift tube cross-section	2.68×6.35 cm	2.5×15 cm	2.5×15 cm	2.5×8.4 cm
Drift tube length	226 cm	640 cm	180 cm	363 cm
Max drift time	800 ns	1.4 μs	1.4 μs	800 ns
Total drift tubes (Run II)	2304	1076	2208	1728
Scintillation counter thickness		2.5 cm	1.5 cm	2.5 cm
Scintillation counter width		30 cm	30-40 cm	17 cm
Scintillation counter length		320 cm	180 cm	180 cm
Total counters (Run II)		269	324	864
Pion interaction lengths	5.5	7.8	6.2	6.2-20
Minimum muon p_T	1.4 GeV/c	2.2 GeV/c	1.4 GeV/c	1.4-2.0 GeV/c
Multiple scattering resolution	12 cm/p	15 cm/p	13 cm/p	13-25 cm/p

certain sub-detectors, on which they make decision either to pass it for subsequent processing or to reject it. Level 3 (L3) trigger consists of a Linux PC farm, which read out entire event data, where the events are reconstructed and a final recording decision is made. Events that satisfy the L3 selection are then transferred to the mass storage. A trigger system block diagram is shown is Fig 2.14.

To accommodate a 4 μs decision time window, the L1 trigger is designed to find physics objects based on a subset of the detector. Data from the calorimeters, the central tracking chamber, and the muon detectors are send to the Level-1 trigger system, which determines whether a $p\bar{p}$ collision is sufficiently interesting to hold the data for the Level-2 trigger hardware. The Level-1 trigger is a synchronous system with a decision reaching each front-end card at the end of the 42-crossing pipeline. Upon a Level-1 trigger accept, the data on each front-end card are transferred to one of four local Level-2 buffers. The Trigger Supervisor System (TSI) is responsible for

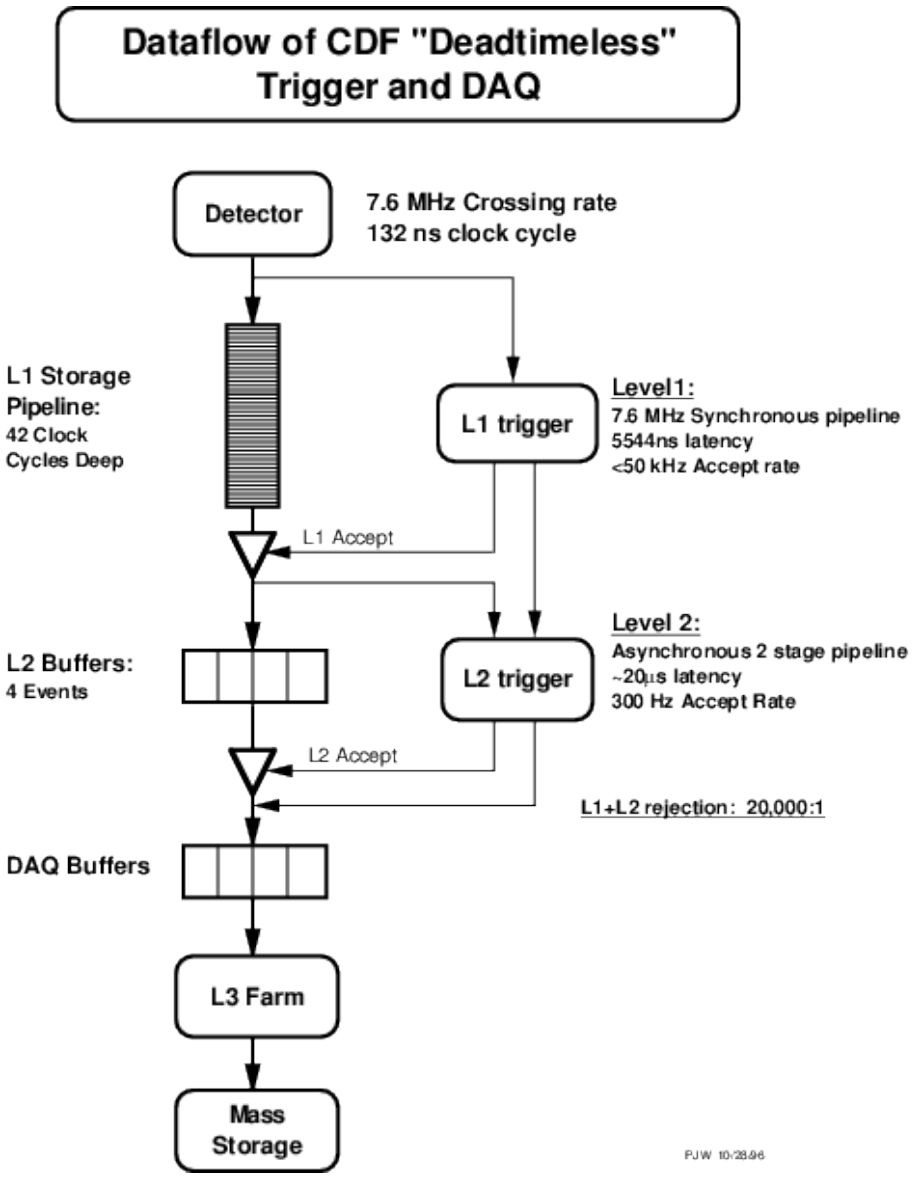


Figure 2.13.: Functional block diagram of the CDF II data flow

maintaining synchronization and allocating buffer space for each event accepted at Level-1. The typical L1 accept rate is 25 kHz.

The second trigger level is an asynchronous system with an average decision time of 20 μs. At L2 the information from calorimeter, tracker, and muon systems, CES and SVX is used to reconstruct physics objects. The L2 cluster finding algorithm combines contiguous regions of calorimeter to form jet candidates. Additionally,

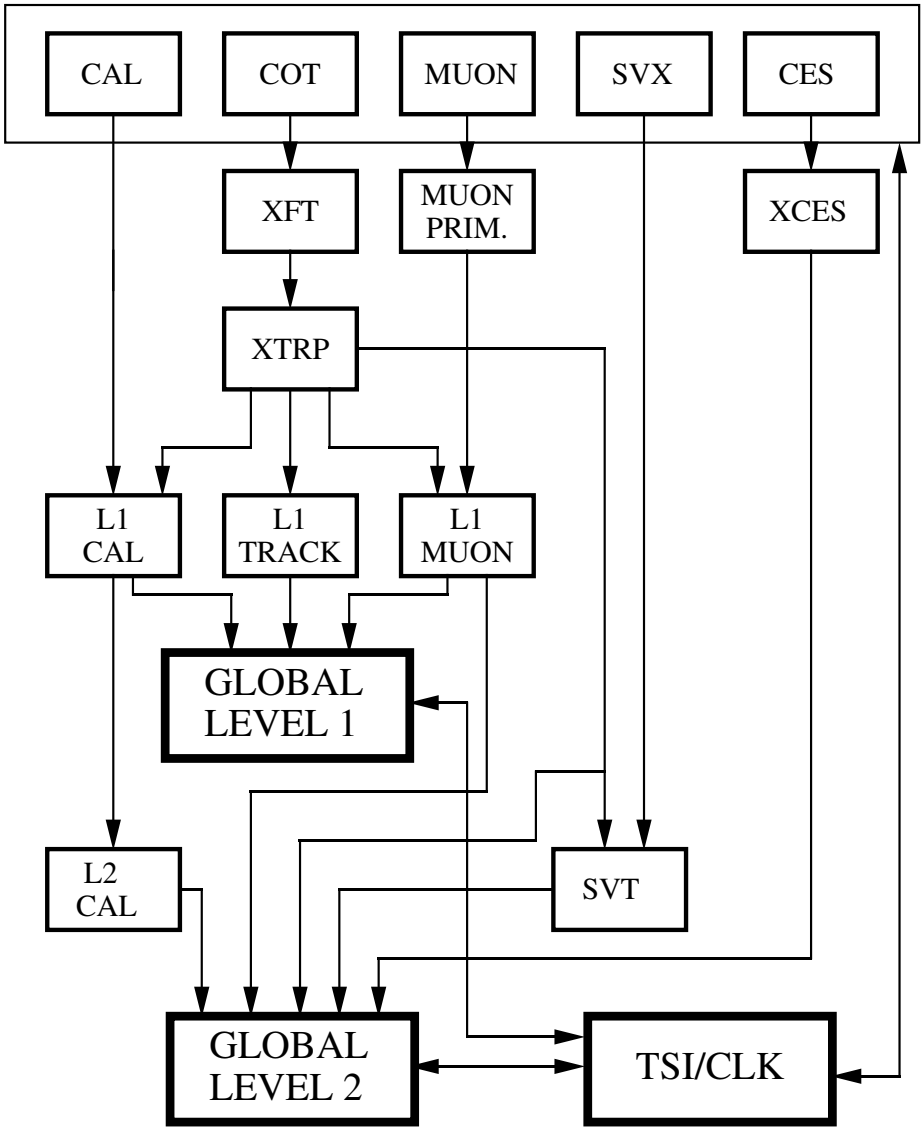


Figure 2.14.: CDF II Trigger system block diagram

CES and showermax detectors allow to obtain further positional information on the calorimeter clusters with a better spatial resolution than from calorimeters alone. The Silicon Vertex Trigger (SVT) [22] combines SVX information with the Level 1 tracking, allowing selection of tracks with large impact parameter. Information from the muon system is combined with with tracking system for the muon triggers. The typical L2 accept rate is 350 Hz.

A L2 trigger accept flags an event for readout. Data are collected in DAQ buffers and then transferred via a network switch to a L3 CPU node, where the complete event is assembled, analyzed, and, if accepted, with $\sim$100 Hz written out to permanent storage. These events can also be viewed by online monitoring programs running on other workstations.

3. PHYSICS OBJECTS

In this chapter we describe the physics objects most relevant to this analysis such as jets (especially jets originated from b-quarks), high p_T leptons, missing transverse energy as well as missing transverse momentum.

3.1 Jet

A jet is a narrow spray of hadrons and other particles produced by the hadronization of a quark or gluon, and is the experimental signature of quark and gluon. Quarks and gluons, confined by their color charges, cannot exist in free form. Therefore they fragment into hadrons before they can be directly detected. After hadronization and decay of short lifetime hadrons, a jet is mostly composed of hadrons like pions, kaons, protons and neutrons, as well as electrons and muons. The hadrons in a jet have small transverse momenta relative to their parent patron's direction and the sum of their longitudinal momenta roughly gives the parent parton's momentum. An illustration of the formation and the structure of a jet is shown in Fig. 3.1.

3.1.1 Jet Clustering

The energy of a jet can be calculated from the energy deposited in the calorimeter towers using different clustering algorithms. In this thesis, jets are clustered using a cone algorithm [23] with a fixed cone size in which the center of the jet is defined as (η^{jet}, ϕ^{jet}) and the size of the jet cone as

$$\Delta R \equiv \sqrt{(\eta^{tower} - \eta^{jet})^2 + (\phi^{tower} - \phi^{jet})^2} < 0.4. \tag{3.1}$$

The jet clustering algorithm groups calorimeter towers with $E_{T_i} > 1$ GeV into jets. $E_{T_i} = E_i \sin\theta_i$ is the transverse energy of a tower with respect to the z-direction

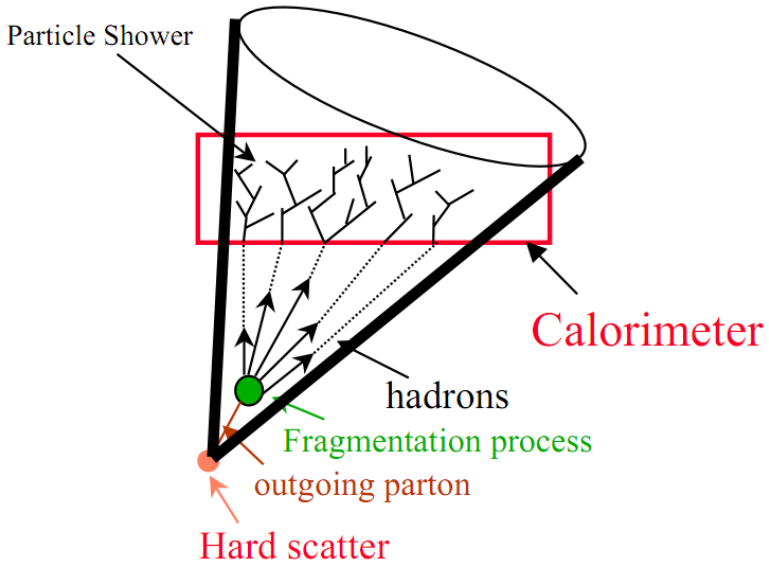


Figure 3.1.: Sketch of jet shower with fragmentation and hadronization.

of a $p\bar{p}$ interaction, and the energy E_i is the sum of the energies measured in the electromagnetic and hadronic compartments of that tower.

Firstly *seed towers* are defined in order of decreasing E_{T_i}. For each seed tower the towers within a radius of size $\Delta R = 0.4$ with respect to its position are used to build "clusters". Once we have an initial list of clusters, the cluster transverse energy and the location (center) of the cluster are calculated as follows:

$$\begin{aligned} E_T^{jet} &= \sum_{i=0}^{N_{tower}} E_{T_i} \\ \phi^{jet} &= \sum_{i=0}^{N_{tower}} \frac{E_{T_i}\phi_i}{E_T^{jet}} \\ \eta^{jet} &= \sum_{i=0}^{N_{tower}} \frac{E_{T_i}\eta_i}{E_T^{jet}} \end{aligned} \tag{3.2}$$

where N_{tower} is the number of towers inside the radius R with $E_T > 1$ GeV.

This procedure is repeated iteratively and a new list of towers around the new center is determined. The jet E_T and direction are recalculated until the list of towers assigned to the clusters is stable, that is, when the geometrical center of the tower corresponds to the cluster centroid. Overlapping jets are merged if they overlap by

more than 50%. If the overlap is smaller than 50%, each tower in the overlap region is assigned to the nearest jet.

The final jet energy and momentum coordinates are computed from the final list of towers:

$$
\begin{aligned}
E_{jet} &= \sum_{i=0}^{N_{tower}} E_i, \\
p_x^{jet} &= \sum_{i=0}^{N_{tower}} E_i \sin\theta_i \cos\phi_i, \\
p_y^{jet} &= \sum_{i=0}^{N_{tower}} E_i \sin\theta_i \sin\phi_i, \\
p_z^{jet} &= \sum_{i=0}^{N_{tower}} E_i \cos\theta_i, \\
p_T^{jet} &= \sqrt{(p_x^{jet})^2 + (p_y^{jet})^2}, \\
\phi_{jet} &= \tan\frac{p_y^{jet}}{p_x^{jet}}, \\
\sin\theta_{jet} &= \frac{p_T^{jet}}{\sqrt{(p_x^{jet})^2 + (p_y^{jet})^2 + (p_z^{jet})^2}}, \\
E_T^{jet} &= E_{jet}\sin\theta_{jet}
\end{aligned}
\tag{3.3}
$$

These are called the raw jet informations.

3.1.2 Jet Energy Correction

For a sampling calorimeter system, the energies deposited in the calorimeter towers are only part of the parton energies. To reconstruct the true parton energy, we need to correct the jet energy for a number of detector and collision effects.

The CDF jet energy corrections are divided into different levels to accommodate different effects that can distort the measured jet energy, such as, response of the calorimeter to different particles, non-linearity response of the calorimeter to the particle energies, un-instrumented regions of the detector, spectator interactions, and

energy radiated outside the jet clustering algorithm. Depending on the physics analyses, a subset of these corrections can be applied.

The original parton transverse energy can be estimated by correcting the jet for instrumental effects and for radiation and fragmentation effects:

$$p_T^{parton} = (p_T^{jet} \times C_\eta - C_{MI}) \times C_{Abs} - C_{UE} + C_{OOC} = p_T^{particle} - C_{UE} + C_{OOC} \quad (3.4)$$

where

- p_T^{parton} is the transverse momentum of the parent parton the procedure is aimed at;
- p_T^{jet} is the transverse momentum measured in the calorimeter jet, $p_T^{particle}$ is the transverse momentum of the particle jet, that is, a jet corrected by all instrumental effects which corresponds to the sum of the momenta of the hadrons, leptons, and photons within the jet cone;
- C_η, "η-dependence" correction, ensures homogeneous response over the entire angular range;
- C_{MI}, "Multiple Interaction" correction, is the energy to subtract from the jet due to pile-up of multiple $p\bar{p}$ interactions in the same bunch crossing;
- C_{Abs}, "Absolute" correction, is the correction of the calorimeter response to the momentum of the particle jet. Particle jets can be compared directly to data from other experiments or theoretical predictions which include parton radiation and hadronization.
- C_{UE} and C_{OOC} , the Underlying Event and Out-Of-Cone corrections, correct for parton radiation and hadronization effects due to the finite size of the jet cone algorithm that is used. Note that these corrections are independent of the experimental setup, i.e., the CDF detector environment.

The precision of jet energy determines the precision of many physics measurements. For example, a 1% uncertainty on the energy scale of jets results in an

uncertainty of 10% on the cross section for jet production at transverse momenta of 500 GeV/c [24] and in a 1 GeV uncertainty on the top-quark mass measurement [25].

Detail of the energy correction for each of the above effects is described in [23]. In this thesis, we use the correction up to level 5, which includes online/offline calibrations, η-dependence correction, multiple interaction correction, and the absolute correction. Corrections on underlying event and out-of-cone effect are not considered.

3.1.3 The H1 Algorithm

The vast majority of analyses at CDF employ the generic jet energy corrections described in the previous section, which relies only on calorimeter information. For analyses in this thesis, we also apply jet energy corrections that are further improved with the tracking information.

The main constituents of jet are hadrons such as $\pi^{\pm}$, π^0, $K^{\pm}$, K_S, K_L, protons and neutrons. Decays of π^0 to a pair of photons are reconstructed in the EM calorimeter. K_S usually decay to $\pi^+\pi^-$ or $\pi^0\pi^0$, hence can be reconstructed either in the tracking or calorimeter system. Only K_L and the neutrons must usually be reconstructed by the hadronic calorimeter. The other charged particles, $\pi^{\pm}$ and $K^{\pm}$, leave information in both the tracking and calorimeter systems.

Roughly 2/3 of the energy of a jet is carried by charged hadrons as a result of isospin symmetry. Due to the fact that the majority of these charged hadrons in a jet have low momenta, the tracking detectors are able to measure their momenta to a much higher precision with respect to the calorimeter measurement. This gives the possibility of improving the estimate of the jets' true energy using the tracking information. We apply an algorithm based on a method developed and successfully used by the H1 collaboration. The algorithm is denoted as H1 algorithm. The corrections to the jet energy obtained from this method are applied in addition to the generic jet energy corrections.

Tracks used in H1 algorithm must satisfy the following requirements, in order to have a reliable momentum measurement,

- $0.5 < p_T^{track} < 15$ GeV/c
- $N_{COT} > 25$ for $|\eta_{track}| < 0.8$
- $Z_0 < 60$ cm

where p_T^{track} is the transverse momentum of the track, N_{COT} is the number of hits in the COT detector associated to the track, η_{track} is the pseudo-rapidity of the track and Z_0 is the distance of the closest approach to the beam line.

The list of tracks satisfying the above conditions is then sorted on the basis of ascending p_T^{track} and extrapolated to the surface of the calorimeter. Towers with $|\eta_{tower} - \eta_{track}| < 0.1$ and $|\phi_{tower} - \phi_{track}| < 0.2$ are considered as possibly containing energy from the track, based on a study of simulation of single pion events.

The selected towers are sorted in order of distance to the track. If the track energy is greater than the total energy of the selected towers, the energy in the towers is replaced with the track energy and the towers are removed from further consideration ("locked"). If this is not true, the energy in the first n towers:

$$\sum_{i=1}^{n} E_i^{tower} \leq E^{track} \tag{3.5}$$

$$\sum_{i=1}^{n+1} E_i^{tower} > E^{track} \tag{3.6}$$

are locked, where E_i^{tower} is the energy measured in the i^{th} tower. The energy of the $(n+1)^{th}$ tower is scaled such that the total locked energy is equal to the track energy. The process is repeated for each track. Energy already locked by a previous track is not considered for subsequent tracks. The total energy of jet is the scalar sum of the p_T of all selected tracks and all non-locked tower energies.

There is possibility that many particles in a jet fall into a single tower, and there could be neutral particles along with charged particles. This would result in the

calorimeter tower energy exceeding the track energy. The H1 algorithm will then return the calorimeter energy for these high density regions, resulting in no improvement in the measurement. Usually, a jet contains regions of both high and low densities. Therefore, the H1 algorithm provides modest improvements in resolution compared to calorimeter only jets. The results would be even better for a detector with a finer granulated calorimeter, and in fact, the calorimeters for the next linear collider are designed to be able to separate individual particles. However, even small improvements in energy resolution are crucial for Higgs boson searches in low mass where $H \to b\bar{b}$. As it has been reported in a low mass Higgs search analysis [26], the use of H1 algorithm substantially improves the sensitivity of the search. Additionally, the H1 algorithm helps to increase the acceptance to signal events by $\sim 10\%$ in that analysis. Energetic jets may sometime appear as low p_T jets, since part of the energy is lost in, e.g., uninstrumented regions of the calorimeter. The H1 algorithm helps to recover this kind of events by properly reconstructing their energy, allowing to retain events that would have otherwise been discarded by the analysis event selection.

3.2 *b*-quark Jet

Jets originating from the bottom quarks are highly related to interesting physics topics. For example, the top quark nearly always decay to a b quark in association with a W boson; the standard model Higgs boson is expected to decay to b-quark pair more than any other particles if its mass is below 135 GeV/c^2. Identifying b-quark jets, or b-tagging, helps to identify the decays of these particles and to reduce the light-flavor background.

The methods for b-tagging are based on the unique features of bottom-flavored hadrons. These hadrons usually have sufficient lifetime that they can travel some distance before decaying. For example, a B^+ meson has $c\tau = 492$ μm; with the initial momentum from a Tevatron collision, it can travel about 3 mm in the detector. But,

their lifetimes are not so long as some of light-quark hadrons (K_L has $c\tau = 15.34$ m), so that b-hadrons decay inside the detector rather than escape.

The bottom quark is much more massive than anything it decays into, thus its decay products tend to have higher transverse momentum (the momentum perpendicular to the original direction of the b-quark jet). This causes b-jets to be wider, and to have higher multiplicities (numbers of constituent particles) and invariant masses, and also to contain low-energy leptons with momentum perpendicular to the jet.

These unique features can be measured, and used to identify b-jets. Figure 3.2 is a cartoon to illustrate the behavior of jets originated from light-flavor quarks and b quarks.

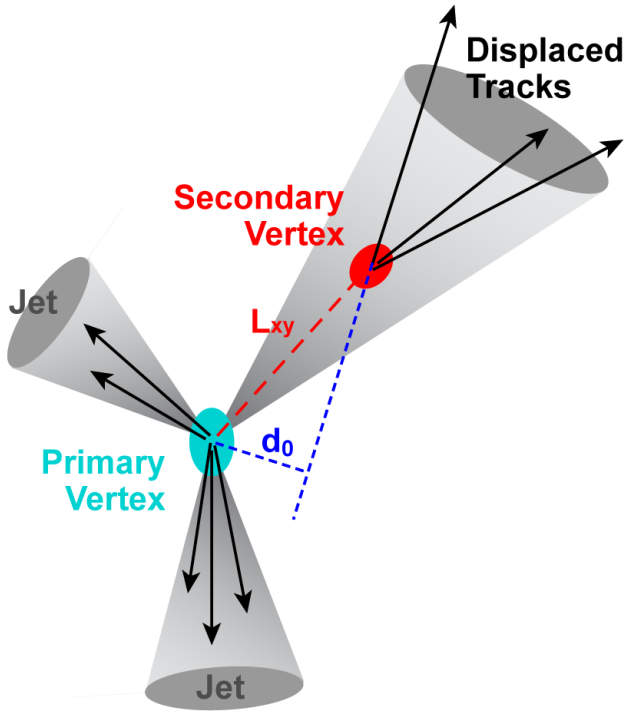


Figure 3.2.: A cartoon to show jets from light-flavor quarks and b quarks

In this thesis, three kinds of b-tagging algorithm are used: SecVTX, JetProb, and HOBIT.

3.2.1 The SecVTX Algorithm

The SecVTX algorithm [27] relies on the displacement of secondary vertices relative to the primary event vertex to identify b hadron decays.

In order to select displaced tracks coming from decays of b-hadrons, precise knowledge of the collision point is necessary. If there is an identified high-momentum electron or muon in the event, the primary vertex is the nearest vertex to these objects. For events without high-momentum leptons, the algorithm uses the vertex which has the highest total scalar sum of transverse momentum of associated tracks. The position of the primary vertex is then determined by vertex fitting a set of tracks within a ± 1 cm window in z around this vertex and with impact parameter significance (relative to the beam line) $d_0/\sigma_{d_0} < 3$, where σ_{d_0} includes the uncertainties on both the track and the beam line positions. The transverse profile of the beam line at the z of the original vertex estimate is also used as a constraint in the fit.

Secondary vertex tagging operates on a per-jet basis, where only tracks within the jet cone are considered for each jet in the event. A set of cuts involving the transverse momentum, the number of silicon hits attached to the tracks, the quality of those hits, and the χ^2/ndf of the final track fit are applied to reject poorly reconstructed tracks. Only jets with at least two of these good tracks are considered being able to produce a displaced vertex; jets are defined as "taggable" if they have two good quality tracks.

Once a secondary vertex is found in a jet, the two-dimensional decay length of the secondary vertex L_{2D} is calculated as the projection onto the jet axis, in the $r - \phi$ view only, of the vector pointing from the primary vertex to the secondary vertex. The sign of L_{2D} is positive (negative) if the absolute difference $|\Delta\phi|$ between the jet axis and the secondary vertex vector is less (greater) than 90°.

The secondary vertices corresponding to the decay of b- and c-hadrons are expected to have large positive L_{2D}. For light-flavor hadrons, L_{2D}s peak at zero and the width of the distribution is determined by the tracking resolution. To reduce the background

from the false secondary vertices (*mistags*), a good secondary vertex is required to have $L_{2D}/\sigma_{L_{2D}} > 7.5$, where $\sigma_{L_{2D}}$, the total estimated uncertainty on L_{2D} including the error on the primary vertex, is estimated vertex-by-vertex but is typically around 190 μm. A tagged jet is defined to be a jet containing a good secondary vertex; the SecVTX algorithm will find at most one good vertex per jet.

The negative tags, with vertex of $L_{2D}/\sigma_{L_{2D}} < -7.5$, are useful in calculating the mistag rate. The mistags will be further discussed in Section 5.2.2.

The requirement on the track selection and the secondary vertex quality varies for the *loose*, *tight*, and *ultratight* SecVTX to accommodate different analyses at CDF. For example, for a top-pair measurement, with two b-quark and several other high E_T object in the final state, a *loose* selection of b-jets allows large acceptance of signal with limited background contamination. For the Higgs search analyses, we use *tight* SecVTX selections, for which the mistag rate vs. tagging efficiency is optimal.

3.2.2 The JetProb Algorithm

The JetProb (jet probability b-tagging) algorithm [28] uses tracks associated with a jet to determine the probability for the jet that originates from a b quark. The calculation of the probability is based on the impact parameters (d_0) of the tracks in the jet and their uncertainties. The impact parameter is assigned a positive or negative sign depending on the position of the tracks point of closest approach to the primary vertex with respect to the jet direction, as shown in Fig. 3.3.

The particles in a jet coming from a light parton originate at the primary vertex, but some of these tracks are reconstructed with a nonzero impact parameter due to the finite tracking resolution. They have an equal probability of being positively or negatively signed. Jets which originate from a heavy parton contain long-lived hadrons giving rise to tracks displaced in the jet direction, which preferentially populate the positive side of the signed impact parameter distribution. Therefore the

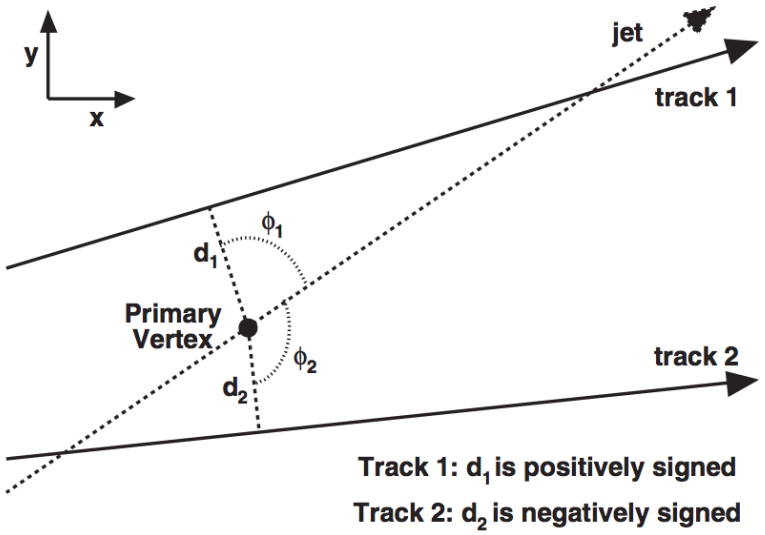


Figure 3.3.: The sign of the impact parameter of a track. The impact parameter is positive (negative) if the angle ϕ between the jet axis and the line connecting the primary vertex and the tracks point of closest approach to the primary vertex itself is smaller (greater) than $\pi/2$.

width of the negative impact parameter distribution is solely due to the tracking detector resolution, beam spot size, and multiple scattering.

In the JetProb algorithm, the tracks in a jet are divided into 72 different categories according to the number and quality of SVX hits, detector η, and p_T. The signed impact parameter significance, $S_{d_0} \equiv d_0/\sigma_{d_0}$, is parametrized for each track category. An example of the distribution of the impact parameter significance, S_{d_0}, is shown in Fig. 3.4. As we discussed, the negative impact parameter is a result of the tracking detector resolution. Therefore we can define a probability, $P_{tr}(S_{d_0})$, for a given track,

$$P_{tr}(S_{d_0}) = \frac{\int_{-\infty}^{-|S_{d_0}|} R(S)dS}{\int_{-\infty}^{0} R(S)dS}, \tag{3.7}$$

where the $R(S)$ is a resolution function fitted with the negative side of the S_{d_0} distribution.

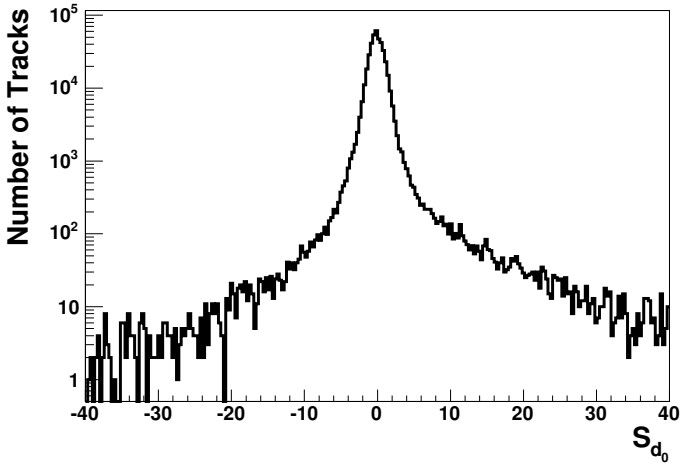


Figure 3.4.: Distribution of the impact parameter significance for tracks in an inclusive jet sample with at least 5 good SVX hits, $p_T > 5$ GeV/c, and $|\eta| < 0.6$.

The jet probability P_J that a jet is consistent with a zero lifetime hypothesis is defined as,

$$P_J = \prod \times \sum_{k=0}^{N_{trk}-1} \frac{(-\ln \prod)^k}{k!}, \tag{3.8}$$

where

$$\prod = \prod_{l=1}^{N_{trk}} P_{tr} \tag{3.9}$$

and N_{trk} is the number of jet probability tracks with positive impact parameter. By construction, the probability for tracks originating from the primary vertex is uniformly distributed from 0 to 1. For a jet coming from heavy flavor hadronization, the distribution peaks at 0, due to tracks from long-lived particles that have a large impact parameter with respect to the primary vertex.

Figure 3.5 shows the P_J distributions for Monte Carlo events and data. For heavy-flavor jets, the P_J peaks around 0, and for light-flavor jets, the probability is flat over interval from 0 to 1. In this thesis, we choose $P_J < 0.05$ as to define JetProb-tagged b jets.

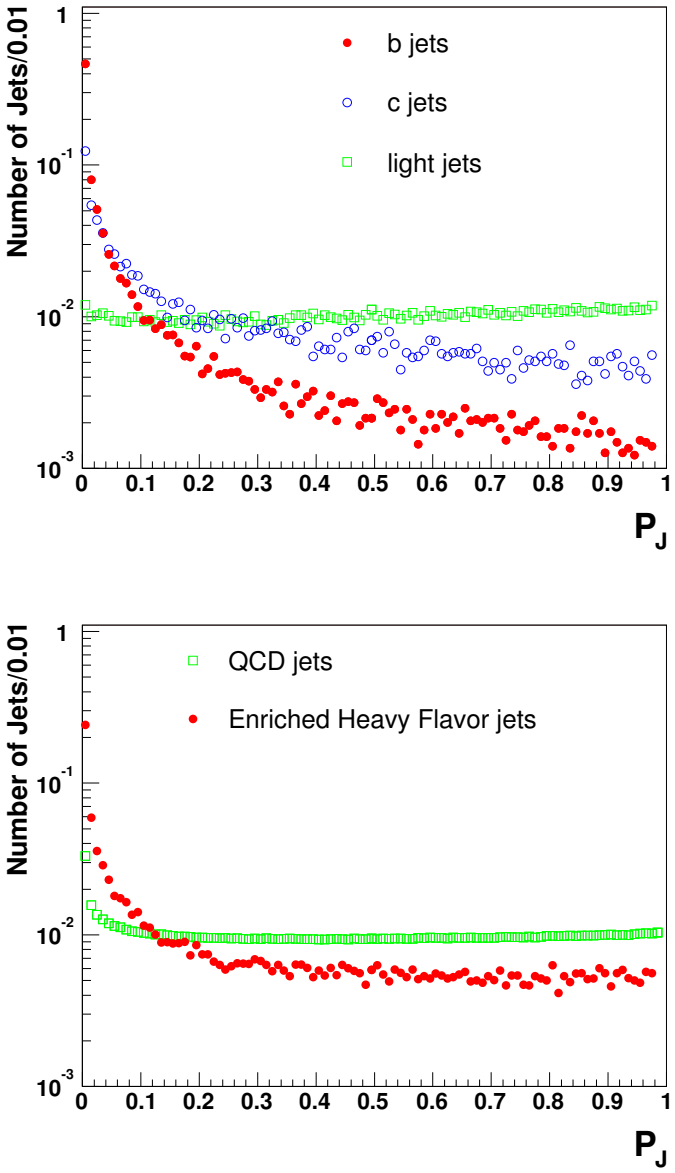


Figure 3.5.: Top: jet probability distributions for jets matched to b (full circles), c (empty circles) and light (empty squares) quarks in Monte Carlo simulated events. Bottom: jet probability distributions for electron jets in inclusive electron data (full circles) and for generic QCD jets in Jet50 data (empty squares).

3.2.3 The HOBIT Algorithm

The Higgs-Optimized b-Identification Tagger (HOBIT) [29] is an algorithm that has been optimized to identify the b-jets from the decay of the Higgs bosons.

We have discussed that the SecVTX and JetProb algorithms determine b-flavored jet based on the displacement of secondary vertices or tracks to the primary vertex. The HOBIT algorithm, however, uses a machine learning technique to find the difference between the light- and heavy-flavor jets. A neural-network technique is employed for the HOBIT algorithm; Section 6.1 will have more discussion on neural-network techniques.

HOBIT uses 25 input variables for the neural network. In particular, HOBIT explicitly uses as inputs the output of the SecVTX algorithm set to its *loose* operating point, the output of CDF's soft muon tagger, and inputs to the earlier RomaNN [30] and Bness [31] multivariate taggers. A full list of these variable is as follow,

- Jet E_T – transverse energy corrected to L5
- SecVTX *loose* tag flag
- SecVTX mass of jet
- RomaNN inputs
  - The invariant mass, pseudo-$c\tau$, 3-d displacement and 3-d displacement significance of the most heavy-flavor-like vertex.
  - The number of tracks both in heavy-flavor-like vertices and standalone heavy-flavor-like tracks associated to a displaced vertex, as well as their combined invariant mass, and the ratio of the scalar sum of the p_T's of these tracks to the scalar sum of the p_T's of all tracks in the jet.
  - The *loose* SecVTX tag status, as well as the mass of the tracks used in the *loose* SecVTXvertex fit.

- BNess inputs: the ten highest evaluated track-by-track neural-network outputs for tracks in a jet serve as inputs to HOBIT ($BNess\ Track^{9}_{i=0}$).

Distributions of the inputs to HOBIT are shown in 3.6.

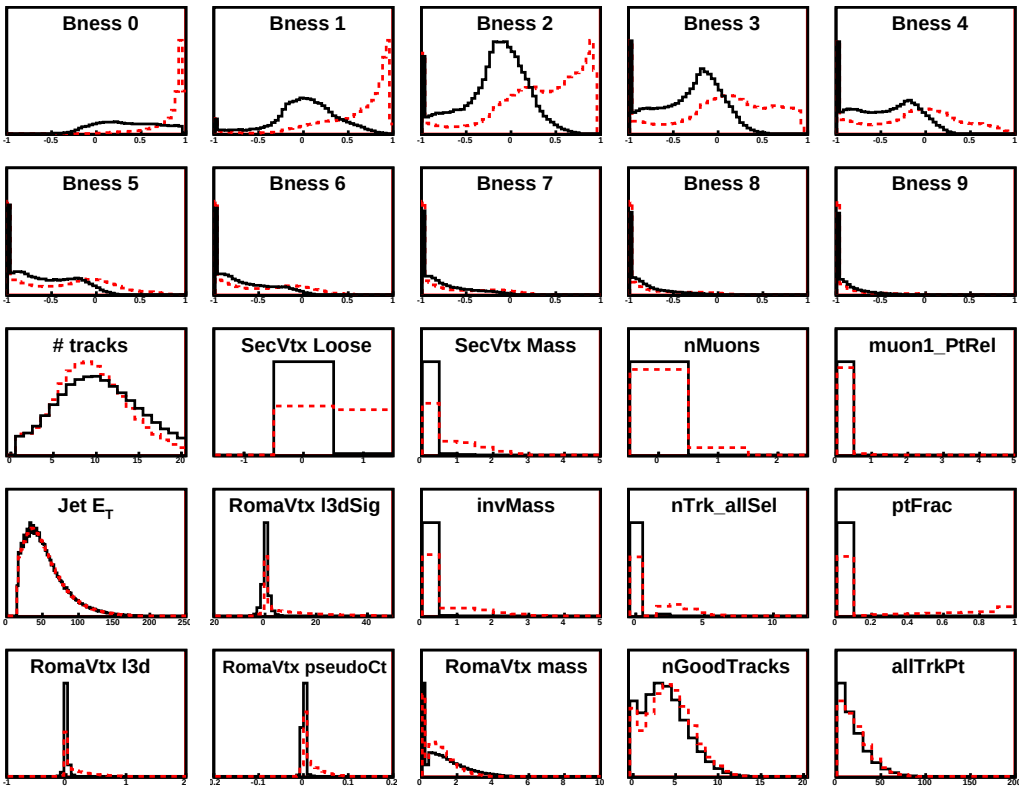


Figure 3.6.: Inputs to HOBIT. The solid histogram is for light-flavor jets and the dashed (colored) histogram is for b jets. Taken from MC, the distributions are normalized to one another. Left to right, top to bottom: the Bness value for the 10 highest Bness tracks; the number of Bness-selected tracks; the *loose* SecVtx tag status and the mass of its fitted vertex; the number of SLT-tagged muons and the momentum transverse to the jet axis of the most SLT-favored muon; jet E_T ; the 3-d displacement significance of the most heavy-flavor-like vertex in RomaNN; the invariant mass, number, and fraction of total track p_T of heavy-flavor-like tracks; the 3-d displacement, pseudo-$c\tau$ and invariant mass of the most heavy-flavor-like vertex; the number of RomaNN-selected tracks and their total p_T.

The neural network was trained using b-jets in WH ($m_H = 120$ GeV/c^2) Monte Carlo and light jets from W+jets Monte Carlo samples. The light-flavor jets in the W+jets samples were reweighted to have the same E_T spectrum as the b-jets from Higgs decays.

The neural-network output ranges from -1 to 1, as shown in Fig. 3.7. In Fig. 3.8, the b-jet efficiencies and the light-flavor jet efficiencies (mistag rates) as a function of jet E_T and η are shown for two HOBIT operating points – a requirement of a HOBIT output > 0.72(*loose*) and a requirement of a HOBIT output > 0.98 (*tight*).

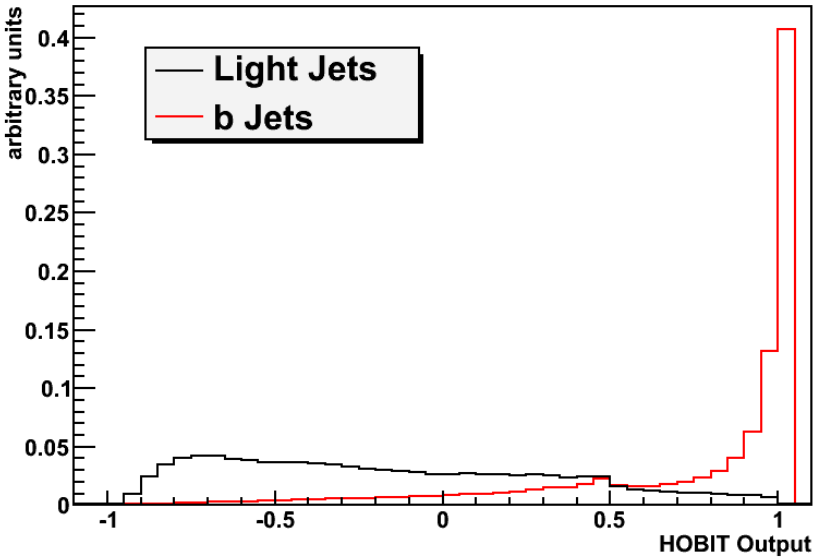


Figure 3.7.: The HOBIT output distribution. The output is trained so that 1 is b jet-like and -1 is targeted to be light-flavor jet-like. The black histogram is for light-flavor jets and the colored histogram is for b jets. Taken from MC, the distributions are normalized to one.

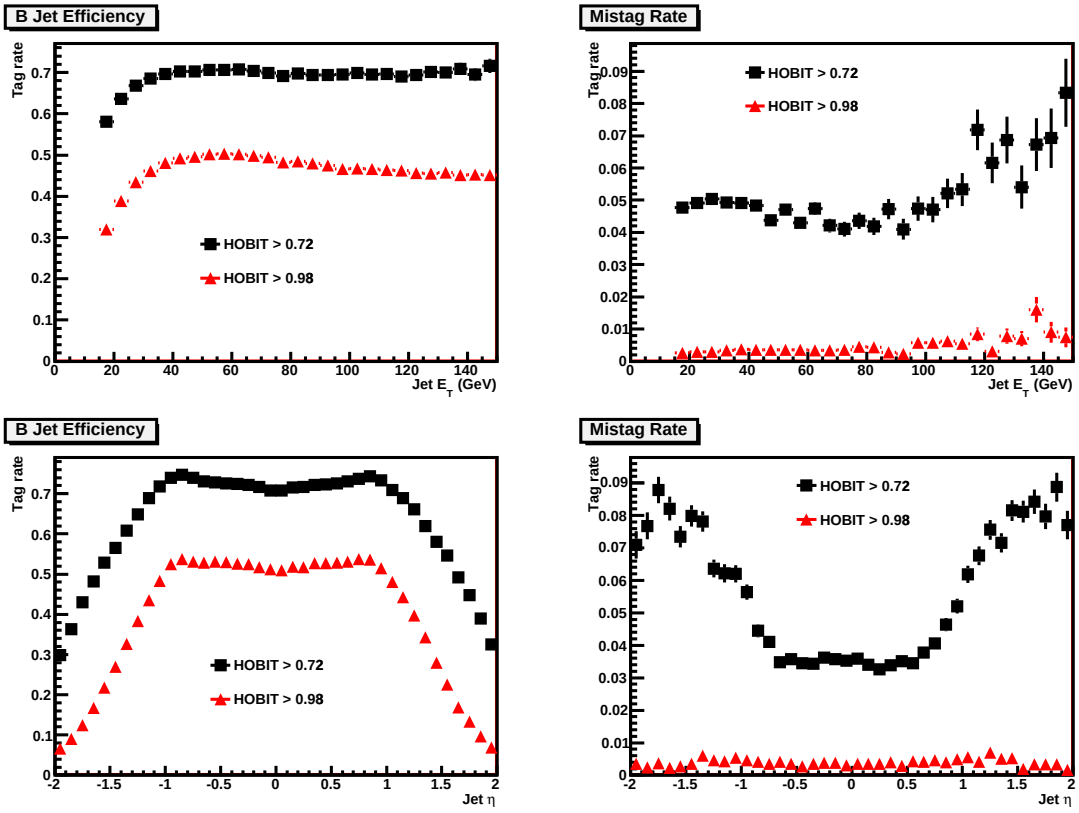


Figure 3.8.: The b-jet and light-jet efficiencies in Monte Carlo before scale-factor corrections as a function of η and E_T. The black rectangles are for the looser operating point and the colored triangles are for the tighter operating point.

3.3 Missing Transverse Energy

Particles like neutrinos barely leave trace in the collider detectors, which has limited amount of detecting material, thus are referred as non-detectable objects. The energy carried by these particles is denoted as missing energy.

In hadron colliders, the energy of each proton (antiproton) is split, and constantly exchanged, between its constituents – u, d quarks and gluons. The initial momenta of the colliding partons along the beam axis are uncertain. Therefore the amount of total missing energy cannot be determined in the z direction. The initial momentum in the $r\phi$ plane, however, is *zero*; any net momentum in the transverse direction indicates missing transverse energy ($\vec{\not\!\!E}_T$).

Any presence of significant amount of missing transverse energy is an indication of production of non-detectable neutrino-like particles or mis-measurements. $\vec{\not\!\!E}_T$-related physics channels become very important as the $\vec{\not\!\!E}_T$ is highly associated to many interesting physics and is a very useful tool to discriminate these physics from background.

The raw $\vec{\not\!\!E}_T$ is defined as the negative vector sum of the transverse energy measured in the calorimeter. At the Level 3 trigger system, $\vec{\not\!\!E}_T$ is reconstructed as the negative vector sum of all calorimeter towers, both electromagnetic and hadronic, and is defined as

$$\vec{\not\!\!E}_T = -\sum^{towers} (E_i \sin\theta_i)\hat{n}_i \tag{3.10}$$

where E_i is the energy of the ith tower, θ_i is the polar angle of the line pointing from the interaction point ($z = 0$) to the ith tower, and $\hat{n}_i$ is a transverse unit vector pointing to the center of each tower. The sum extends to $|\eta|$ of the detector smaller than 3.6; the region of the forward calorimeters is avoided as there is uninstrumented regions that are designed to accommodate the quadrupole focusing magnets.

After correcting the jet energy as described in Section 3.1.2, we recompute the $\vec{\not\!\!E}_T$ with the corrected jet energies,

$$\begin{aligned} \not\!\!E_{T_x}^{\ corr} &= \not\!\!E_{T_x}^{\ raw} - \sum_{i=1}^{N_{jets}} (E_x^{corr,i} - E_x^{raw,i}) \\ \not\!\!E_{T_y}^{\ corr} &= \not\!\!E_{T_y}^{\ raw} - \sum_{i=1}^{N_{jets}} (E_y^{corr,i} - E_y^{raw,i}) \\ \phi_{\not\!\!E_T^{\ corr}} &= \tan^{-1}(\frac{\not\!\!E_{T_y}^{\ corr}}{\not\!\!E_{T_x}^{\ corr}}) \end{aligned} \tag{3.11}$$

In particular, we have L5-jet corrected $\vec{\not\!\!E}_T$ and H1-jet corrected $\vec{\not\!\!E}_T$.

The resolution of $\not\!\!E_T$ depends on the response of the calorimeter to the total energy deposited in the event. It is parametrized in terms of the total scalar sum of E_T denoted as $\sum E_T$. The resolution is measured with minimum bias events, which are collected requiring a coincidence of hits in both the forward and backward beam-beam counter. Minimum bias events are dominated by inelastic $p\bar{p}$ collisions; no significant $\not\!\!E_T$ is expected in these events. Fits to these data yield [32]

$$\sigma(\not\!\!E_{T_x}) = -0.582 + 0.7418\sqrt{\sum E_T}. \tag{3.12}$$

The non-zero first term in the expression is due to run-by-run offsets and out-of-time accidental energy. For large $\not\!\!E_T$, the resolution is a function of $\not\!\!E_T$ and the above expression is not valid.

While large $\not\!\!E_T$ is expected in events with neutrino in the final states, significant amount of $\not\!\!E_T$ can also rise from detector effects:

- when protons and antiprotons of the Tevatron beam collide with nuclei of gas atoms or beam collimators, they produce a "halo" of muons, traveling roughly parallel to the beam. Some of these muons cross a row of calorimeter towers along the z-axis depositing energy in the calorimeters asymmetrically in ϕ.

- problems with some of the calorimeter tower electronics/calibrations may cause the tower to report a wrong value of the energy of the incident particles.

- cosmic muons traveling through the detector, promoting low $\not\!\!E_T$ to a higher value.

- muons carrying a large momentum can be created in the hard collision. Being minimum ionizing particles, these muons can pass through the calorimeter, without substantially losing energy.

- proton or antiproton beam remnants and beam losses in very forward regions.

- when the direction of an energetic jet is near an uninstrumented calorimeter region, the energy of the jet will be underestimated. A configuration of two jets that are produced back-to-back with the same momentum will appear to have a momentum imbalance.

A set of selections is recommended by the $\not\!\!E_T$ working group at CDF to use in all $\not\!\!E_T$-based analyses, to remove the instrumental sources of $\not\!\!E_T$, including beam halo muon events. These criteria are described in note [33]. Here is a summary of the selections:

- Pass 1

  - At least one central jet with $|\eta_{det}| < 0.9$ and $\not\!\!E_T > 10$ GeV.
  - Event Electromagnetic Fraction (EEMF):

$$EEMF \equiv \frac{\sum_{j=1}^{N_{jet}} E_T^j * EMF_j}{\sum_{j=1}^{N_{jet}} E_T^j} > 1, \tag{3.13}$$

    where EMF_j is the fraction of the jet energy deposited in the electromagnetic calorimeter. Only jets with $E_T > 10$ GeV are considered.

  - At least one COT track with $p_T > 0.5$ GeV/c and one axial superlayer with six or more hits on it.

- Pass 2

- Event Charge Fraction (ECHF):

$$ECHF \equiv \frac{\sum_{j=1}^{N_{jet}} CHF_j}{N_{jet}} > 0.1, \tag{3.14}$$

  where CHF_j is the jet charge fraction which is defined as the ratio of the sum of the p_T of the tracks matching to the jet energy E_T.

- At least one good primary vertex in the event.

- Pass 3

  - Exclude a geometrical region of the calorimeter, called the "chimney", where jets would mismeasured. This region hosts cryogenic and instrumental connections to the inner detector. The jets that fall into the chimney region ($60^\circ < \phi < 100^\circ$ and $0.5 < \eta < 1.0$) are almost certainly mismeasured. Hence, we exclude any event that has a jet with $E_T > 10$ GeV that falls into the chimney region.

  - Require that the event primary vertex falls within $|z| < 60$ cm of the nominal interaction point at the detector center.

All events that fail any of the Pass 1-3 requirements are discarded from the analysis in the initial stages of the analysis.

3.4 Missing Transverse Momentum

Beside the missing transverse energy, we define another quantity as the missing transverse momentum ($\vec{\not\!\!p}_T$), for which we only use the information from the tracking detector:

$$\vec{\not\!\!p}_T = -\sum_{tracks} \vec{p}_{T_i} \tag{3.15}$$

Despite the similarity in the definition to $\vec{\not\!\!E}_T$, $\vec{\not\!\!p}_T$ provides a very important tool to discriminate the instrumental background from the other physics backgrounds.

In order to reconstruct the $\vec{\not\!\!p}_T$ observable, we use as inputs to its computation the charged particle momentum reconstructed in the spectrometer. The basic track quality criteria for the calculation of $\vec{\not\!\!p}_T$ are the same used for the reconstruction of the primary vertex at CDF. Only tracks with 0.5 GeV/$c < p_T <$ 200 GeV/c, $|\eta| < 1.5$ and $|Z_{vtx}| < 2$ cm are used, where Z_{vtx} is the closest approach distance of the track from the primary vertex along the z axis. We then classify the tracks in four different categories, on the basis of the number of axial (N^{ax}_{COT} , N^{ax}_{SVX}) and stereo (N^{st}_{COT} , N^{st}_{SVX}) COT and SVX layers that have at least 5 hits, together the χ^2 of the track fit. The first category fulfills tight requirements on the track reconstruction in the COT; the second, third and fourth have decreasing requirements on the number of COT hits compensated by requirements on the χ^2 and on the SVX hit layers. If the first category requirements fail we check the second category, and so on. Table 3.1 summaries these selections.

In events where only charged particles and undetected particles are produced, the $\vec{\not\!\!p}_T$ is highly correlated in module and direction to the undetected particle(s) momentum, and thus provide a way to measure their energy with potentially better resolution than $\vec{\not\!\!E}_T$. The presence of quarks in the final state complicates the picture. In the radiation and hadronization process forming a jet, most particles produced are pions (with a 10-20% of kaons). Due to isospin symmetry, roughly 2/3 of the energy of a jet will be carried by charged pions, which will be measured with both calorimeter and tracking chamber. The $\vec{\not\!\!p}_T$ underestimates the undetected particles energy because it

Table 3.1: Quality requirements for tracks used in $\not\!\!\!p_T$ calculation. The rows are listed in the order the selections are checked. If the first row requirements fail, we check the second row, etc.

Category	Quality requirements
Tight COT part	$N_{COT} > 0$ $(\chi^2 - \chi^2_{SVX})/(N_{COT} - 5) < 4.0$ $N^{St}_{COT}(\geq 5\ hits) \geq 3, N^{Ax}_{COT}(\geq 5\ hits) \geq 3$
COT part is loose, good χ^2	$N_{COT} > 0$ $(\chi^2 - \chi^2_{SVX})/(N_{COT} - 5) < 4.0$ $N^{St}_{COT}(\geq 5\ hits) \geq 2, N^{Ax}_{COT}(\geq 5\ hits) \geq 2$ $N^{Ax}_{SVX} \geq 4,\ N^{St}_{SVX} \geq 3$ $\chi^2_{SVX}/(N_{SVX} - 5) < 8.0$
COT part really loose, good χ^2	$N_{COT} > 0$ $(\chi^2 - \chi^2_{SVX})/(N_{COT} - 5) < 4.0$ $N^{Ax}_{SVX} \geq 5, N^{St}_{SVX} \geq 3$ $\chi^2_{SVX}/(N_{SVX} - 5) < 8.0$
Inside-Out or silicone stand-alone	$N^{Ax}_{SVX} \geq 5, N^{St}_{SVX} \geq 3$ $\chi^2_{SVX}/(N_{SVX} - 5) < 8.0$

does not take into account the energy carried by the neutral components of the jets, and for the same reason has a worse angular resolution. In addition, while calorimeter coverage is $|\eta| < 3.6$, the spectrometer coverage in collider detector is generally far more limited with $|\eta| < 1.0$. For these reasons, $\not\!\!\!p_T$ cannot substitute the role of the $\vec{\not\!\!E}_T$ tool as a measurement of momentum and direction of undetected particles; still, the $\not\!\!\!p_T$ can provide informations complementary to those given by $\vec{\not\!\!E}_T$.

In events resulting from instrumental effects, usually the QCD multijet events, the $\not\!\!E_T$ arises from the mismeasurement of jet energies in the calorimeter. For this type of events, $\not\!\!\!p_T$ provides a relatively independent measurement to the calorimeter

information. For instance, a QCD dijet event with one jet being under-measured would have a significant amount of $\not\!\!E_T$ aligning to the jet with less measured energy. $\not\!\!p_T$, however, only depends on the momenta carried by the charged particles in each jet. The charge fraction for these jets are arbitrary, hence $\not\!\!p_T$ could aligning to the jet with less charged fraction. An illustration of this explanation is shown in Fig. 3.9.

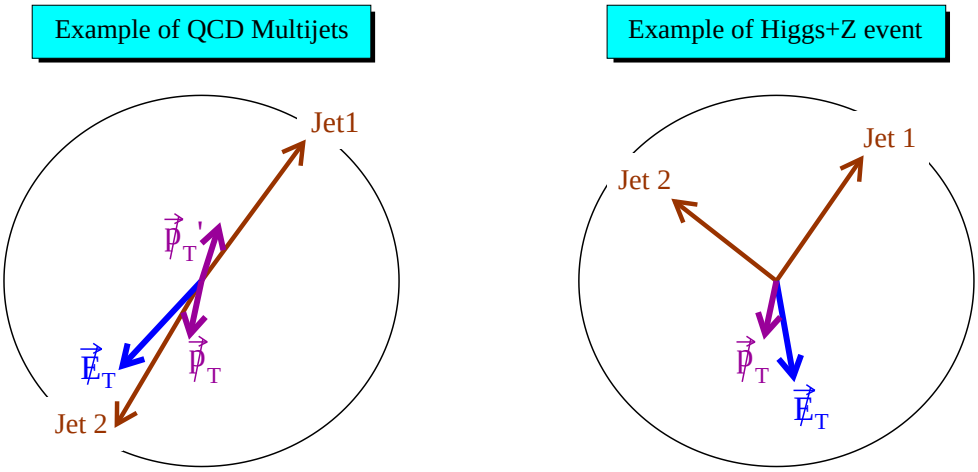


Figure 3.9.: Left: an example of QCD dijet event. $\not\!\!E_T$ aligns to the jet with less measured energy. $\not\!\!p_T$, however, could align to both jet depending their charge fraction. Right: an example of ZH events, in which there is real neutrino in the final state. $\not\!\!p_T$ is highly correlated to the $\not\!\!E_T$ in both module and direction.

This particular topology is reflected in the distribution of the azimuthal distance between the two vectors, $\Delta\phi(\vec{\not\!\!E}_T, \vec{\not\!\!p}_T)$. As shown in Fig. 3.10, events containing undetected particles like neutrinos are concentrated near 0, while QCD dijet events have almost equal chance to populate the region around 0 or around π. This special distribution allows to effectively suppress the QCD contribution in samples with $\not\!\!E_T$ and jets. Dijet (and 3 jets) events are important because have the most simple kinematic, so that they can be also studied through QCD MC. Fig. 3.11 shows the $\Delta\phi(\vec{\not\!\!E}_T, \vec{\not\!\!p}_T)$ distributions for data and simulated samples of QCD and events with neutrinos. The

presence of b quarks in the final state is expected not to alter significantly the situation, because their semileptonic decays would give rise to low-energetic neutrinos.

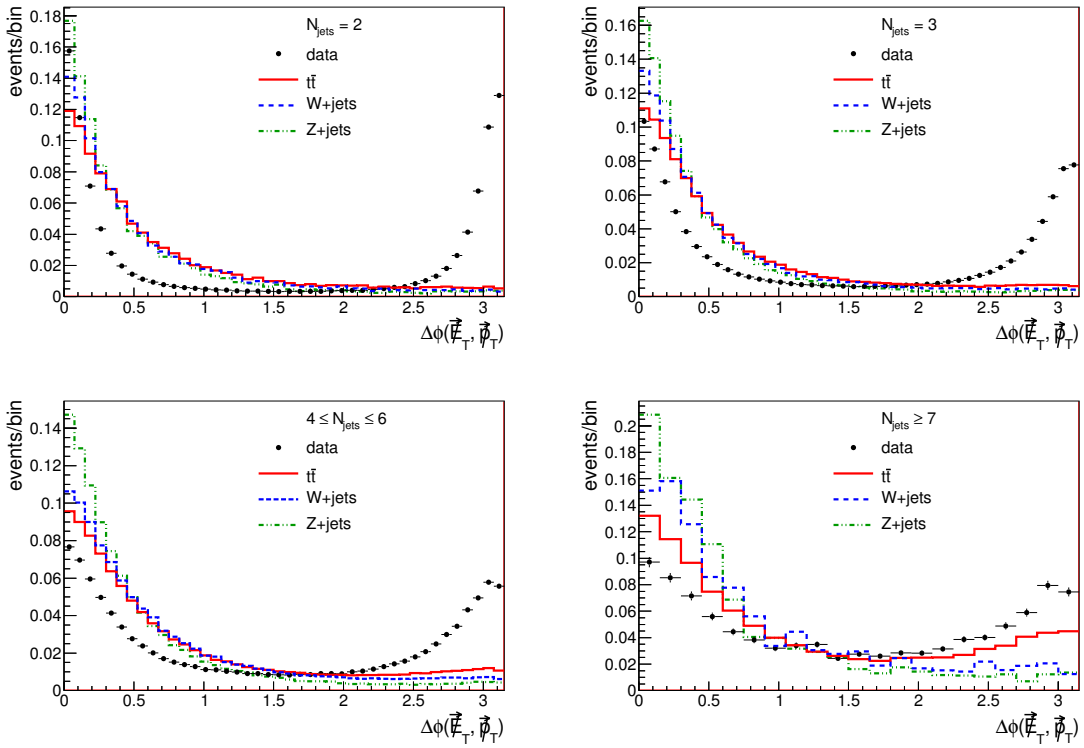


Figure 3.10.: The distribution of $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T)$ data and major Standard Model processes giving rise to neutrinos. The events are selected with $\not{E}_T > 50$ GeV and number of jet N_{jet} varying from 2 (top-left plot) to 7 or more (bottom-right plot). All distributions are normalized to unit area.

The QCD background can be reduced by a factor of two by requiring $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T) < \pi/2$, with minimal signal loss. Its implementation in more complex methods, like multivariate analysis, would maximize its impact.

An important consequence of the angular correlation between $\not{E}_T$ and $\not{p}_T$ is that QCD is supposed to populate symmetrically the $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T)$ distribution, with the symmetry centered at $\pi/2$. By rejecting the events with $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T) > \pi/2$ and defining as signal region the one with $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T) < \pi/2$, one can infer the contribution

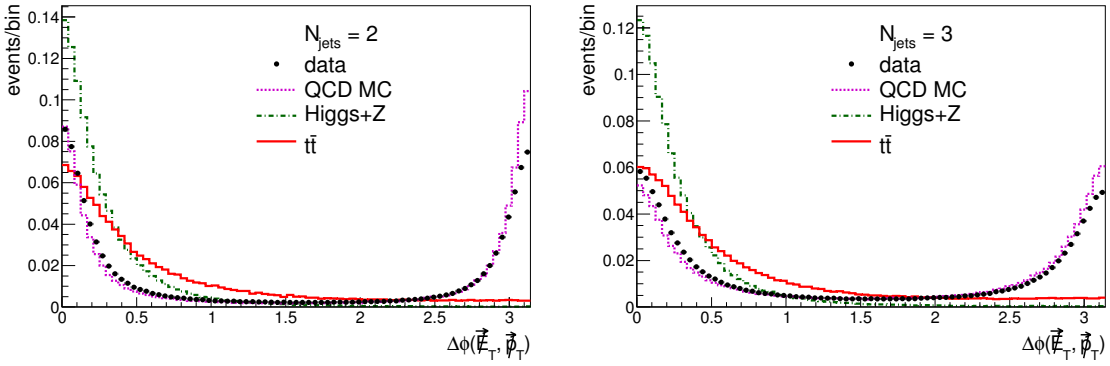


Figure 3.11.: The $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T)$ distribution in events with 2 (left) and 3 (right) jets. Lines represents QCD, Higgs+W/Z and $t\bar{t}$ MC, while dots represents data. All plots are normalized to unit area.

of QCD in the signal region in a data-driven fashion by counting the events populating the rejected region, i.e. $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T) > \pi/2$. The robustness of this method has been tested by varying the jet multplicity, by placing cuts on the angular correlation among $\not{E}_T$ and the jets direction, varying $\not{E}_T$ cuts, etc. In all instances, by using MC simulation the above assumption has been found true to a level of $\sim 20\%$.

Following the same reasoning, it is legitimate to assume that QCD events in the $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T) > \pi/2$ could be used to model the kinematic features of QCD events in the complementary region. The systematic uncertainties associated with the simulation of high multiplicity QCD jet production are very large. For these reasons a reliable data-driven modeling method for this category of events is of fundamental importance in collider physics analysis.

More details about the $\not{p}_T$ study is documented in [34, 35].

3.5 Lepton Identification

The high p_T physics analyses are signature oriented, which are usually divided by the number of identified high p_T leptons, the type of leptons, the number of jets, etc., in the final state. In this thesis, we selecte with energetic jets and large missing transverse energy; we reject events with identified leptons, either electrons or muons. Rejecting events with identified leptons allows selecting a sample that is statistically independent from the other CDF analyses probing the same physics. This has facilitated the combination of the results presented in this thesis to maximize the CDF reach.

In this thesis we reject events with leptons using the standard CDF selections described in the Joint Physics webpage [36].

3.5.1 Electron Identification

The central electrons (electrons in the coverage of central calorimeters) are identified with *loose* selections that consider both the EM calorimeter and COT informations.

- In the central EM calorimeter region with $|\eta| < 1.1$.
- $E_T \geq 20$ GeV.
- Not from the photon conversion.
- Isolation $(\Delta R = 0.4)/E_T \leq 0.1$, where $(\Delta R = 0.4)$ represents the scalar sum of the p_T of all tracks in the cone radius $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.4$.
- Very small hadronic calorimeter deposition with $E_{Had}/E_{EM} \leq 0.055 + 0.00045 \times E$.
- The corresponding COT track has:

  - $Z_0 \leq 60$ cm, where Z_0 is the z-coordinate of the track at the point of closest approach to the detector origin.
  - $p_T \geq 10$ GeV/c.
  - With 3 axial superlayers (SLs) with 5 hits/SL, and 2 stereo SLs with 5 hits/SL.

The plug electrons are identified with selections that consider the plug EM calorimeter (PEM) information.

- $1.2 \leq |\eta| \leq 2.8$
- $E_T \geq 20$ GeV.
- Isolation $(\Delta R = 0.4)/E_T \leq 0.1$.
- $E_{Had}/E_{EM} \leq 0.05$.
- PEM 3×3 $\chi^2 \leq 10$, where PEM 3×3 is the result of the comparison of the PEM shower profile of the electron candidate with the measurements with the test beam electrons. The test beam results for PEM were obtained using 3×3 cluster size.
- Requirements on the Plug Electromagnetic Shower Maximum Detector (PES).
  - PES 5×9 U ≥ 0.65
  - PES 5×9 V ≥ 0.65
  - ΔR(PES centroid and PEM centroid) ≤ 3.0 cm.

3.5.2 Muon Identification

The central muons are identified with a combination of selections from the COT, muon counter and calorimeter:

- $p_T > 20$ GeV/c.

- EM energy < 2 GeV $+ max(0, 0.0115 \times (p - 100))$.
- Had energy < 6 GeV $+ max(0, 0.028 \times (p - 100))$.
- Isolation $(\Delta R = 0.4)/p_T \leq 0.1$.
- With 3 axial SLs with 5 hits/SL, and 2 stereo SLs with 5 hits/SL.
- COT track $Z_0 \leq 60$ cm.
- Track with no silicon hits attached: $|d_0| < 0.2$ cm.
- Track with silicon hits attached: $|d_0| < 0.02$ cm.
- for Central Muon Detector (CMU) and Central Muon Upgrade Detector (CMP) muons, $|\delta_X(\text{CMU})| < 7$ cm and $|\delta_X(\text{CMP})| < 5$ cm.
- for Central Muon Extension Detector (CMX) muons, run number > 150144, $|\delta_X(\text{CMX})| < 6$ cm and COT track radius less than 140 cm.

4. PHYSICS WITH JETS AND LARGE MISSING TRANSVERSE ENERGY

The focus of this thesis is studying physics phenomena in events with a signature of energetic jets and large missing transverse energy. Two analyses are presented. The first is the search for the low mass SM Higgs boson produced in association with a vector boson (Z or W). The second is the search for a dark matter candidate produced in association with a single-top quark, or the *monotop* search. This chapter is dedicated to describe these two physics production mechanisms.

Although we only focus on these two physics topics, the techniques developed for this thesis actually have also been applied in several other studies, which include the measurement of the top-quark pair production [37], the measurement of the single-top quark production [38], the measurement of the diboson production using the dijet mass spectrum and the search for a dijet-resonance [39], and the search for a dark matter candidate produced in association with a top-quark pair [40].

4.1 The Search for the Low Mass SM Higgs

Section 1.1.2 has introduced the status of the search for the Higgs boson at colliders. At the Tevatron $p\bar{p}$ collider, the searches for the SM Higgs boson are divided into the low mass categories with the Higgs boson dominantly decaying into fermion pairs, of which the decay $H \to b\bar{b}$ has the largest branching ratio, and the high mass categories with the Higgs boson decaying to WW or ZZ pair. In the low mass categories, the most copious production mode is gluon-gluon fusion via a heavy quark loop ($gg \to H$), producing a single Higgs boson. However, the search for the Higgs boson in this mode is prohibited by the overwhelming QCD multijet background.

Therefore, the major searches are carried through Higgs boson associated production with a W or Z boson, as shown in Fig. 4.1.

Therefore, the major searching channels are,

- The low mass Higgs boson search ($90 < m_H < 150$ GeV/c^2)
  - one lepton: $q\bar{q} \to WH \to \ell\nu b\bar{b}$
  - two leptons: $q\bar{q} \to ZH \to \ell\ell b\bar{b}$
  - no lepton: $q\bar{q} \to ZH \to \nu\nu b\bar{b}$ and $q\bar{q} \to WH \to \not{\ell}\nu b\bar{b}$
- The high mass Higgs boson search ($100 < m_H < 200$ GeV/c^2)
  - two leptons: $gg \to H \to WW \to \ell\nu\ell\nu$

A combination of these channels, along with the other smaller channels, is performed in achieving the best sensitivity for the Tevatron data [41].

Multiplying the production rate and the Higgs boson decay branching ratio, the sensitivity to each channel can be ranked by the acceptance, as it is shown in Fig. 4.2.

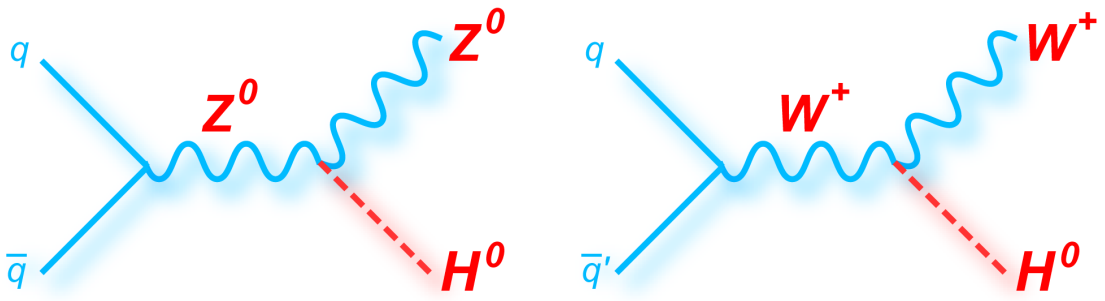


Figure 4.1.: The diagram for the Higgs boson produced in association with a Z and W boson.

In this thesis, we are interested in the *no lepton* channel. We are sensitive to the Higgs boson produced in association with a Z boson decaying to neutrinos ($ZH \to \nu\nu b\bar{b}$). We are also sensitive to the Higgs boson produced in association with a W boson decaying to a lepton plus neutrino, where the lepton is a hadronically decaying

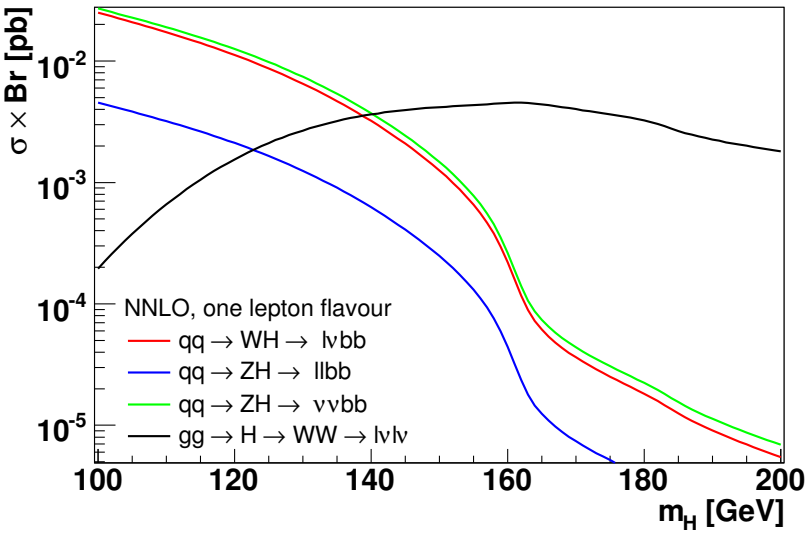


Figure 4.2.: The Higgs production cross sections times branching ratios.

τ or is a miss-identified electron or muon, ($WH \to \tau\nu b\bar{b}$ or $WH \to \ell\nu b\bar{b}$). With the benefit of the large production ratio of signal events, the search of the Higgs boson in this signature provides one of the most sensitive results at the Tevatron experiments.

The Higgs boson productions are modeled with the PYTHIA generator [42]. We include the productions

$$\begin{aligned} q\bar{q} &\to ZH \to \nu\bar{\nu} + b\bar{b} \\ q\bar{q} &\to ZH \to \ell\ell + b\bar{b} \\ q\bar{q} &\to WH \to l\bar{\nu} + b\bar{b} \end{aligned} \tag{4.1}$$

in which the m_H ranges from 90 to 150 GeV at a step of every 5 GeV. Table 4.1 presents the cross section and branching ratio used for the Higgs boson search for $90 < m_H < 150$ GeV/c^2.

The samples with additional/less ISR and FSR are generated for m_H of 120 GeV to count the radiation systematic effects; ISR and FSR are initial and final state radiation, respectively.

Table 4.1: The cross section (σ) and branching ratio (BR) of the Higgs boson production for $90 < m_H < 150$ GeV

m_H (GeV)	σ_{ZH} (fb)	σ_{WH} (fb)	$\mathrm{BR}_{H\to b\bar{b}}$ (%)
90	224.0	394.7	81.2
95	190.3	332.1	80.4
100	162.7	281.1	79.1
105	139.5	238.7	77.3
110	120.2	203.7	74.5
115	103.9	174.5	70.5
120	90.2	150.1	64.9
125	78.5	129.5	57.8
130	68.5	112.0	49.4
135	60.0	97.2	40.4
140	52.7	84.6	31.4
145	46.3	73.7	23.1
150	40.8	64.4	15.7

4.2 The Search for Dark Matter Particle

The search for a dark-matter candidate in $\not\!\!E_T$+jets events is based on an idea proposed by J. Andrea *et al.* [43]. The authors suggest a signature-based model where an invisible particle is produced in association with a top quark. The invisible particle, or the dark-matter candidate, escapes detection leaving significant amount of $\not\!\!E_T$. Hence the final state contains only the top quark decaying products. In analogous to a *monojet* production model [44], this signature-based model is dubbed as *monotop* production.

No process in the SM can lead to monotop production at tree level, as the dominant production mode being suppressed both by a loop factor and by two powers of

nondiagonal Cabibbo-Kobayashi-Maskawa matrix elements. Many beyond-SM theories, however, allow such production with a large cross section. For examples, two classes of processes can lead to such a final state signature, through baryon number-violating and flavor-changing neutral interactions, as shown in Fig 4.3.

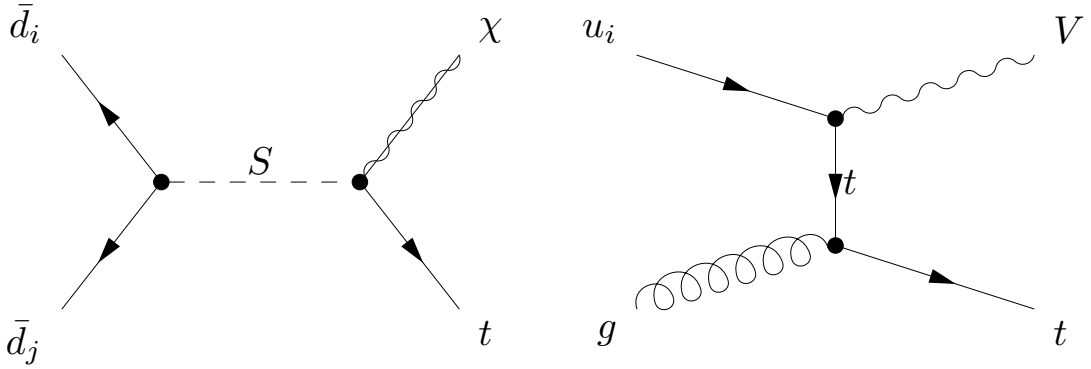


Figure 4.3.: Representative Feynman diagrams leading to monotop signatures, through the resonant exchange of a colored scalar field S (left) and via a flavor-changing interaction with a vector field V (right). In these two examples, the missing energy is carried by the V and χ particles. More diagrams with, for example, t-channel and s-channel exchanges for the two type of processes respectively, are possible.

This signature-based model is generalized to accommodate several different beyond-SM theories. In the framework of effective field theories, the monotop DM production is described by a set of Lagrangians incorporating all possible types of DM particles (scalar, vector, fermion, *etc.*). This effective theory can inclusively describe many beyond-the-standard models [45–49]. For example, new physics processes with a monotop final state can arise from the decay of a supersymmetric squark into a neutralino and a top quark, from the decay of a vector leptoquark into a massless neutrino and a top quark, or through flavor-changing neutral interactions (FCNC) with a new vector state escaping detection. FCNC are very interesting features of many new-physics models that predict, as suggested by the Yukawa interactions, the existence of an interaction that break flavor symmetry, falling that, i.e., in the SM, the monotop production is strongly suppressed, both by loop factor and by two powers

of a non-diagonal element of the CKM matrix. Following this insight Kamenik and Zupan [50] propose a model that couple the DM with quarks, through a new flavor violating interaction. They compare the flavor violating production of DM with the flavor conserving one, finding that in the first case the cross section is substantially enhanced. In the same way, J. Andrea *et al.* propose a new model where they takes into account for different new physics framework, such as SU(5) theories, R-parity violating SUSY, UED, via an EFT Lagrangian (same spirit of the [51]) with a new FCNC that couples particles of the new physics theories to the quarks, giving rise to the monotop final state.

J. Andrew *et al.* present five different scenarios for monotop productions, corresponding to the two different mechanisms: a scalar (vector) resonant monotop production with $m_\chi = 50$ (300) GeV, flavor changing neutral production with a 300 GeV scalar (50 GeV vector) invisible state and, finally, a scenario including four-fermion interactions with a massless invisible state χ. The values for the masses of the invisible states are inspired by present collider data and lie right above the lower bound on the mass of the lightest neutralino in typical SUSY scenarios and of the lightest Kaluza-Klein excitation in extra-dimensional scenarios, respectively.

We decide to test the monotop scenario for different mass points of the dark matter candidate. The predicted cross-sections for different mass points are shown in Table 4.2 and Fig. 4.4. In the production of a massive DM particle (X), we can either produce produce it in a transverse mode or in a longitudinal mode. The cross section associated to the latter are proportional to $1/m_X$. Hence, the cross section becomes smaller as the X mass increasing ($m_X > 5$ GeV/c^2region). Nevertheless, for massless DM particle, we can only produce them transversely polarized. Consequently, the cross section drops ($0 < m_X < 5$ GeV/c^2), since the dominant $1/m_X$ channels are absent.

We model monotop DM production in the flavor-violating process ($ug \to tD$) with MadGraph [52]. Additional showering and hadronization are described by Pythia. We have generated 11 signal samples assuming various DM mass in steps of 5 GeV/c^2

from 0 to 25 GeV/c^2, and then in steps of 25 GeV/c^2 from 25 to 150 GeV/c^2. In our simulation we assume a top-quark mass of 172.5 GeV/c^2, consistent with the world's best determination at the time this analysis was published [53, 54].

Table 4.2: The predicted cross-sections of monotop for different dark matter masses. The first column is the mass of the DM candidate. The second column (cross section (σ) A) shows the cross-section corresponding to different DM masses, before the cut at the generation level and before the top decay; instead the third column σ B are the cross-sections afterwards.

Mass (GeV)	σ A (pb)	σ B (pb)
0	6.89	1.57
5	392.1	96.7
10	100.9	24.8
15	46.66	11.5
20	27.50	6.74
25	18.48	4.508
50	5.59	1.370
75	2.67	0.658
100	1.46	0.350
125	0.86	0.217
150	0.52	0.126

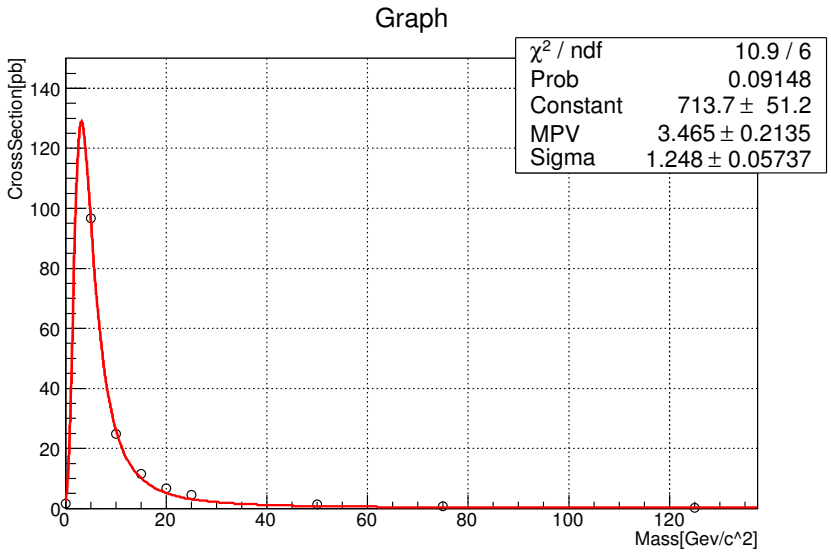


Figure 4.4.: Dependence of the cross-section (after the generation level cuts, etc.) on the mass of the DM candidate, fitted with a Landau distribution

5. DATA SAMPLE AND BACKGROUND MODELING

5.1 Data Sample

We use the data collected by the CDF II detector with the Tevatron Run II proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV. The data is collected starting from March 2002 to the shut down of the Tevatron collider at September 30, 2011.

Data collected by CDF experiment is organized into sets, each set being part of a "run" [56]. A run correspond to a store in the Tevatron, which normally last for about 16 hours. When a store of protons and antiprotons are injected in the Tevatron for collision, waited until the detector and beam conditions are stable, an incremental 6-digit run number is assigned as a unique identifier in the database.

If the detector was functional during an entire run or a section of the run, then these run/sections are marked as good, and the data collected during that period can be used for an analysis. The *good run list* used for this analysis requires good EM calorimeter and silicon detectors since these detectors are used for detecting jets, b-tagging and lepton identification.

The full dataset recorded in Run II from February 2002 to September 2011 is used. The luminosities measured by the Cherenkov Luminosity Counter (CLC) are shown period by period in Table 5.1. The corrected luminosity is 1.019× CLC luminosity. The full dataset corresponds to an integrated luminosity of 9.45 fb^{-1}.

5.1.1 Trigger Selections

The signature we are searching for requires the presence of large imbalance in the transverse energy. Therefore, it is natural to choose a trigger, which is most efficient

Table 5.1: The luminosities for the dataset used in this thesis, presented period by period.

period	Lumi(pb)	period	Lumi(pb)	period	Lumi(pb)	period	Lumi(pb)
0	319.631	10	246.306	20	224.507	30	414.527
1	102.238	11	236.276	21	436.154	31	132.422
2	118.148	12	164.053	22	270.428	32	427.703
3	100.141	13	234.449	23	198.545	33	357.514
4	82.792	14	38.871	24	264.144	34	278.256
5	133.978	15	160.901	25	209.667	35	325.713
6	97.953	16	114.996	26	167.136	36	399.263
7	28.309	17	182.073	27	403.969	37	172.042
8	100.417	18	307.777	28	257.632	38	244.260
9	158.673	19	212.37	29	329.225		

for collecting events with large $\not\!\!E_T$. Three classes of $\not\!\!E_T$-based triggers are selected for this thesis work, as summarized in Tab. 5.2.

Table 5.2: The data passed the $\not\!\!E_T$-trigger paths are used for the analyses in this thesis.

Trigger paths	Level 1	Level 2	Level 3
MET DIJET	JET10, MET28	MET28, TWO JET3	$\not\!\!E_T > 30$
MET35 CJET	MET28	MET30, CJET20, JET15, DPS	$\not\!\!E_T > 35$
MET40/45	MET28	MET40	$\not\!\!E_T > 40/45$

Section 2.2.5 has introduced the CDF trigger systems. CDF has a three-level trigger system. The $\not\!\!E_T$-based triggers select events with the detector information at these three levels.

The Level 1 trigger system is based on custom electronics designed for fast decision making, and allows the reconstruction basic physics object primitive that can be triggered on. Physics calorimeter towers are organized into a 24×24 array of trigger towers in $\eta - \phi$ plane. To reduce the complexity and the processing time, the Level 1 trigger uses only a 8-bit trigger tower energy information, by dropping the least significant bit and the most significant bit. The transverse projection of the tower energies are calculated with the assumption that the event primary vertex is located at $z = 0$, and the missing energy at Level 1 is calculated as a vector sum of trigger tower pairs. The missing energy at Level 1 has poor resolution, due to a limited available information and the need to make a fast decision. The $\not\!\!E_T$ calculated by the Level 1 trigger system is usually underestimated. Therefore, the $\not\!\!E_T$ threshold at Level is chosen to be as low as possible in order to maximize efficiency. Our L1 selection is as follow,

- JET10: with at least one jet by requiring central jet ($|\eta| < 1.1$) or plug jet ($1.1 < |\eta| < 3.6$) with $E_T > 10$ GeV
- MET28: $\not\!\!E_T > 28$ GeV

Jets at Level 2 are reconstructed as "clusters" of energy deposits, using a simple algorithm from Run I. The algorithm finds a "seed" tower (threshold 3 GeV), then attaches adjacent towers (threshold 1 GeV) until no more energetic adjacent towers are found. The cluster position is defined as the seed tower position. While this algorithm worked well at the low instantaneous luminosity of the initial phase of Run II, it become problematic as the Tevatron luminosity increased. The effect of the multiple interactions in high luminosity events significantly increases the detector occupancy and leads to the presence of many energetic towers around the seed towers. As a consequence, large number of energetic towers could be erroneously clustered around a seed into a single jet. Since the MET+Jets trigger requires the presence of at least two clusters in an event, the erroneous merging of separate clusters into a single cluster results in a significant efficiency loss. The CDF calorimeter trigger

was upgraded to solve the problems associated with data-taking in high luminosity regime [55]. A new trigger was implemented after the upgrade was completed. We have L2 selections:

- MET28: $\not\!\!E_T > 28$ GeV
- TWO JET3: at least two jets ($|\eta| < 3.6$) with $E_T > 3$ GeV
- CJET20: at least one central jet ($|\eta| < 1.1$) with $E_T > 20$ GeV
- DPS: dynamic pre-scale between 1 and 20
- MET30: $\not\!\!E_T > 30$
- TWO JET15: at least two jets ($|\eta| < 3.6$) with $E_T > 15$ GeV
- MET40: $\not\!\!E_T > 40$

In the Level 3 trigger system the full detector information is available, leading to a much improved reconstruction of physics objects. The entire 10-bit information from the trigger towers is available, leading to an enhanced $\not\!\!E_T$ resolution. Table 5.2 lists the $\not\!\!E_T$ cuts for the three different trigger paths.

MET DIJET Trigger

The MET DIJET trigger path is a trigger introduced in Run IIb (with higher luminosity), after the completion of the Level 2 trigger upgrade, to collect events with some $\not\!\!E_T$ and at least two jets. This is the most efficient trigger in selecting the $\not\!\!E_T$ candidates.

MET35 CJET Trigger

The MET35 CJET trigger path is the main trigger path for analyses looking at final states with $\not\!\!E_T$ and at least 2 jets. It was implemented before the CDF calorime-

ter trigger upgrade. The requirement on the $\not\!\!E_T$ and jet E_T is tighter compared to the MET DIJET trigger.

MET 40/45 Triggers

The MET 40/45 trigger path is an inclusive MET trigger which is constructed by applying cuts on $\not\!\!E_T$ at the three levels of the trigger. No additional selections are required.

5.1.2 Trigger Efficiency

Detailed comparisons with experimental data show that offline quantities provide the most reliable description of the simulated physics processes and the detector response. While the event selection at the trigger level is based on online measurements, the data analysis is performed using offline quantities, with many backgrounds estimated from simulation. Therefore, in order to estimate the trigger acceptance for simulated physics processes, a trigger efficiency needs to be estimated as a function of offline quantities.

A recently developed parametrization method [59], by employing neural-network multivariate technique, for the trigger efficiency has significantly improved the modeling of the trigger turn-on outside of the fully efficient region – with $\not\!\!E_T > 50$ GeV and jet $E_T s > 30$ GeV. This allows us to set the minimum $\not\!\!E_T$ selection as low as 35 GeV, which significantly increases the acceptance of the signal. The uncertainty of the neural network parametrization is addressed by applying the same technique to both the high-p_T muon sample and the JET-50 sample: the high-p_T muon sample to give a good estimation for real-$\not\!\!E_T$ events and a medium-E_T sample, which gives the case in which the $\not\!\!E_T$ is basically produced (and smeared) by mismeasurement of jets and dependencies of the jet energy scale. The trigger efficiency ranges from 40% for events having $\not\!\!E_T = 35$ GeV to 100% for events with $\not\!\!E_T > 80$ GeV.

5.2 Background Modeling

In both the search for the low mass SM Higgs boson and the one for the dark-matter monotop production, we select events with energetic jets and large missing transverse energy. Many SM physics processes also produce such events, and are referred as backgrounds in this thesis. These SM processes include the QCD multijet productions, the W/Z boson produced in association with additional jets, top-quark (single and pair) productions, boson-pair (diboson) production. Figure 5.1 shows the relative size of the contributions of these backgrounds.

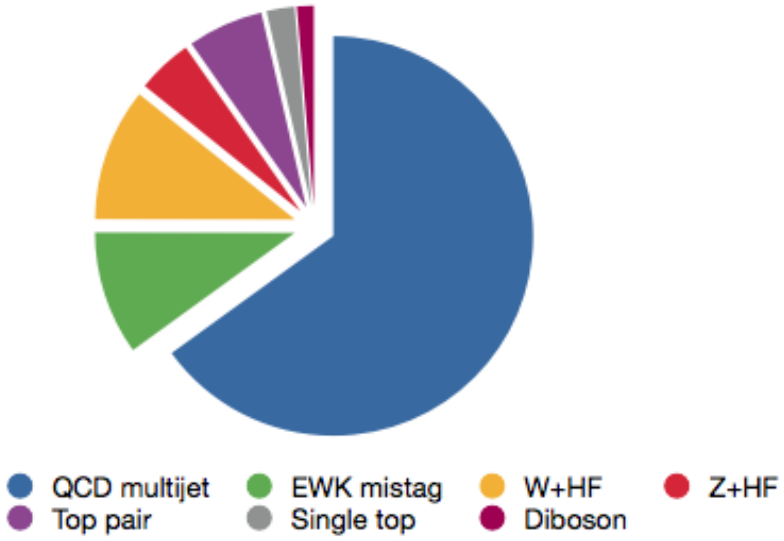


Figure 5.1.: A schematic view of SM physics contributions to the $\not\!\!E_T$ + jets signature; the area of each process roughly corresponds to their relative contribution.

The QCD multijet production is the dominant background, then followed by W+jets and Z+jets, top, and diboson. We denote all the processes that are not QCD multijet as EWK (you may argue that some of them, e.g., the top-quark pair production, can be produced via QCD processes). We choose this notation since QCD multijet is mainly instrumental background with fake $\not\!\!E_T$, while the EWK productions are processes have real neutrino thus real $\not\!\!E_T$ in the final states.

We select events with at least one b-tagged jet. In each background process, there are events with real heavy-flavor (HF) jets which can be b-tagged, and light-flavor (LF) jets which can be mistagged. As shown in Fig.5.2, we have developed different strategies to model these processes. Overall, we model the EWK heavy-flavor events with Monte Carlo simulation, the mistags from EWK light-flavor events with Monte Carlos weighted by a *mistag matrix* (data-driven), the QCD multijet (both HF and mistags from LF) with a data-driven Tag Rate Matrix method.

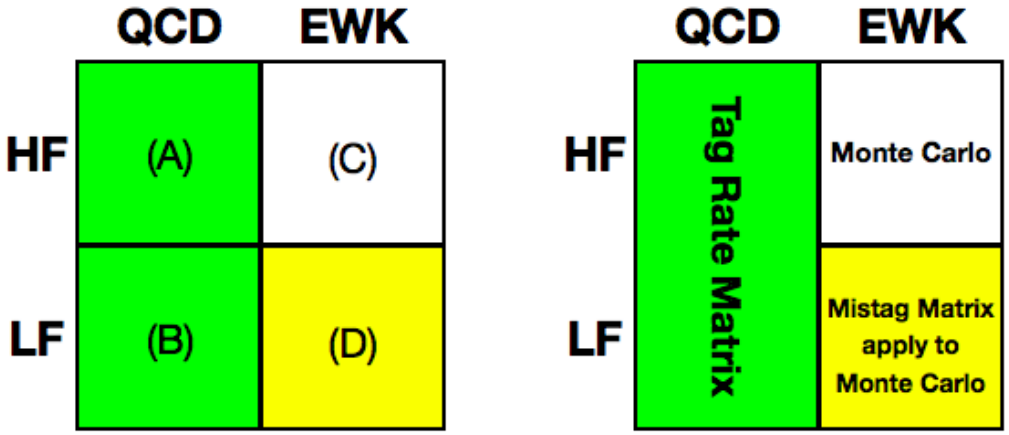
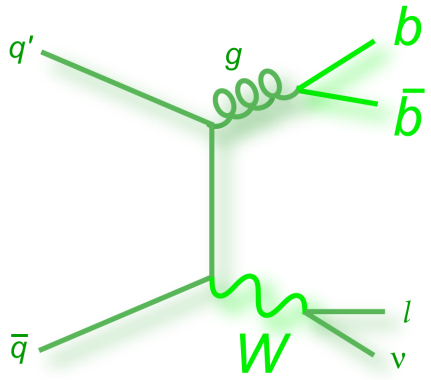


Figure 5.2.: Background modeling strategies in $\not\!\!E_T + b$-jets analyses. HF: events with heavy-flavor jet; LF: events with light-flavor jet.

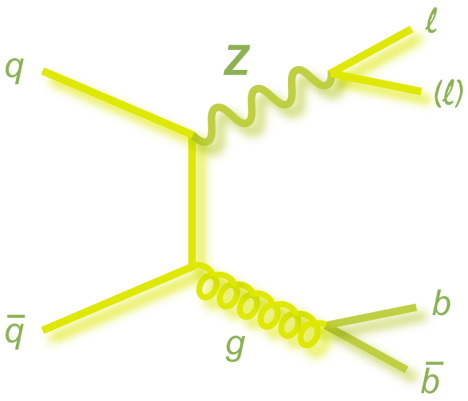
5.2.1 Modeling of EWK Heavy-Flavor Background

The productions of W/Z plus jets, top-quarks and diboson can have neutrinos and jets in the final state. Figure 5.3 shows examples Feynman diagram for these modes. These EWK processes are relatively well understood, we model them with samples generated by Monte Carlo.

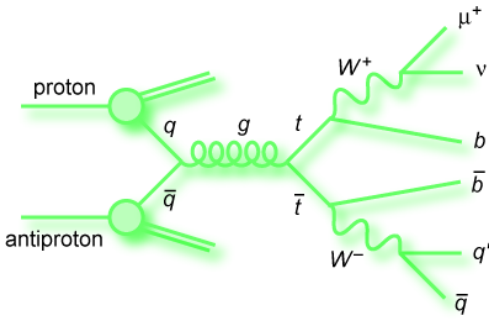
The production of W/Z plus light-flavor and heavy-flavor jets are simulated by Alpgen [60] with showering and hadronization performed by Pythia. The processes are split by the boson decaying channels and the number of associated partons (parent



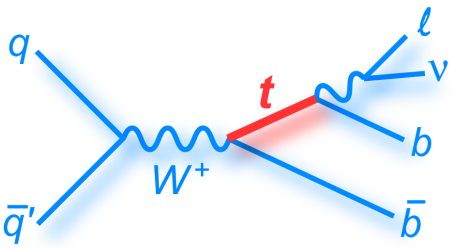
(a) $W + b\bar{b}$



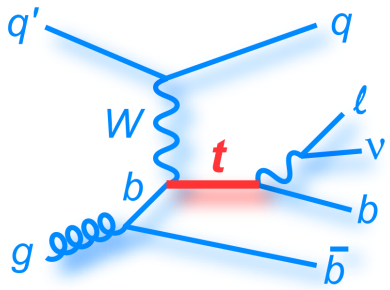
(b) $Z + b\bar{b}$



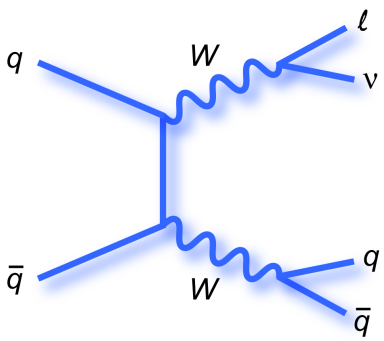
(c) Top-quark pair production



(d) Single top-quark production (s-channel)



(e) Single top (t-channel)



(f) WW pair

Figure 5.3.: Selected Feynman diagrams for W/Z plus jets, top-quarks, and diboson productions.

of jets). A total of 88 samples are generated to describe $W(\to \ell\nu) + b\bar{b} + X$, $W(\to \ell\nu) + c\bar{c} + X$, $W(\to \ell\nu) + c + X$, $W(\to \ell\nu) + X$, where X is the additional parton(s). A total of 69 samples are generated for $Z(\to \nu\nu) + b\bar{b} + X$, $Z(\to \nu\nu) + c\bar{c} + X$, and $Z(\to \nu\nu) + X$. We also generated a sample of $Z(\to b\bar{b}) + X$ with PYTHIA. All these processes are normalized to the next-to-leading order (NLO) cross sections with a k factor of 1.3. An uncertainty of 30% is assigned to the cross section of these processes.

Diboson (WW, WZ, and ZZ) and $t\bar{t}$ production are generated by PYTHIA and normalized to NLO cross section predicted using the MCFM program [61, 62] and the approximate next-to-next-to-leading order cross section [63], respectively. The boson decays are set to be inclusive. A 6% uncertainty on the cross section is assigned to the diboson normalization [61, 62]. A 6.5% on the cross section is assigned to the top pair normalization [64].

Single top, both s- and t-channel production, are modeled using MADGRAPH with PYTHIA and normalized to NLO cross sections [65, 66]. The uncertainties on cross section is assigned to be 10%.

Table 5.3 summarizes the modeling methods and the cross section used in normalization for these processes. When using Monte Carlo to estimate background containing b-quarks we require the b-tagged jet is originating from a heavy-flavor (b or c) quark at the generator level. The mistagged light flavor events are estimated using a data-driven mistag model.

5.2.2 The Mistag Model for EWK Light-Flavor

A light-flavor jet can be mistagged due to the tracking resolution, or long-lived particle decay (like K_S, Λ). This type of contribution is referred as *mistags*, and is classified in two categories: First, the mistagged light-flavor QCD multijet events; Second the mistagged light-flavor events from W/Z plugs jets, top (negligible), and diboson production.

Table 5.3: Modeling for the production of W/Z in association with jets, the top-quark productions, and the Diboson production.

Process	Modeling	σ (pb)
W/Z+jets	ALPGEN + PYTHIA	1.3× LO
$t\bar{t}$	PYTHIA	7.04
single top (s-channel)	MADGRAPGH	1.05± 0.17
single top (t-channel)	MADGRAPGH	2.12± 0.32
WW	PYTHIA	12.4 ± 1.4
WZ	PYTHIA	3.7 ± 0.4
ZZ	PYTHIA	3.6 ± 0.4

This section described the modeling for the second type of mistags. The mistags from QCD multijet will be discussed in section 5.2.3.

A data-driven mistag model is used to estimate the mistag rate for light-flavor jets. The mistag models for SECVTX, JETPROB, and HOBIT are all based on a set of multi-dimension matrices, which divide the data sample by their kinematic quantities, to calculate the mistag rate for each bin of data.

Mistag Matrices for SECVTX and JETPROB

We have introduced how the SECVTX and JETPROB algorithms identify a jet originated from a b-quark in section 3.2.1 and 3.2.2. In these algorithms, the secondary vertices/tracks that are "behind" the event primary vertex are assigned as negative, and are used to defined negatively tagged jets. Examples of the positively/negatively tagged jets are shown in Fig. 5.4.

Mistagging a light-flavor jet is a result of tracking resolution. The chance of positively tagging or negatively tagging a light-flavor jet is about equal, while almost only light-flavor jets can be negatively tagged. Both SECVTX and JETPROB use

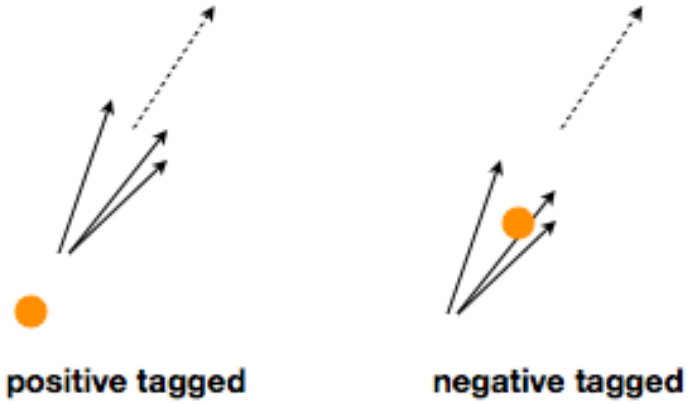


Figure 5.4.: Sketch of positive tagged and negative tagged jets; the yellow spot is the event primary vertex.

the negatively tagged jets (e.g., $L_{2D}/\sigma_{L_{2D}} < -7.5$ for *tight* SecVTX) to estimate the mistag rate. A set of *mistag rate matrices* is constructed. These matrices are defined in bins of jet E_T and η, the number of good tracks in a jet, the number of good vertices, the z position of the primary vertex, etc. The mistag rate is the number of negatively tagged jets divided by the number of taggable jets in that bin of the matrix. To correct the bias due to the long-live light-flavor jets (e.g., K_L), additional asymmetry factors are applied citeref:mistag-secvtx.

The high E_T jet data are used for the mistag rate estimation. The negative tag rates have been varying from period to period, and the standard parameterization does not predict them perfectly. We use the matrices that match the data used for the thesis.

The HOBIT Mistag Matrix

The HOBIT algorithm tags a jet based on a machine learning method. An important difference between HOBIT and SecVTX is the absence of negative tags in HOBIT.

Two new techniques are developed in HOBIT to provide the mistag rates: the "$t\bar{t}$ cross section method" and the "electron conversion method" [29].

The $t\bar{t}$ cross section method seeks to calibrate the predicted b-tagging efficiency and mistag rate in MC to match those measured in data using $t\bar{t}$ candidate events in a W + 3-or-more-jets sample under the assumption the the $t\bar{t}$ cross section is known. A two-dimensional fit was performed to maximize the likelihood of observing the data counts as functions of the b-tag scale factor and the mistag rate. The mistag rate is parameterized in a similar way as the SecVTX with a 5-dimensional matrix, where each element is the measured rate within a bin of the following five variables: jet E_T , jet η, the number of tracks in the jet, the number of primary vertices in the event, and the z location of the primary vertex from which the jet is calculated to have originated. For different HOBIT operating points, separate matrices are constructed for b-, c-, and light-flavor jets. The $\sigma_{t\bar{t}}$ is assumed to be equal to 7.04 ± 0.70 pb.

The electron conversion method takes the advantage of the heavy-flavor enhancement among jets containing electrons, discriminating between heavy- and light-flavor jets based upon whether the electron is identified as coming from a photon conversion. A dijet-event sample with one jet containing electron candidate is used to estimate the b-tag scale factor and mistag rate.

These two methods are then compared and combined. The results for the mistag rates are 1.331 ± 0.130 and 1.492 ± 0.277 for the *loose* and *tight* HOBIT tagging, respectively. The combination also results in a greater than 25% reduction in the size of the uncertainty on the b-tag efficiency in comparison to the SecVTX.

The Mistag Matrix Applications

The mistag matrices yield mistag probability per jet. We use the following equations to find the mistag probabilities for different event tagging categories,

$$\begin{aligned} mistag\ rate_{1S} &= M_{S_{j_1}} \times (1 - M_{S_{j_2}}) + M_{S_{j_2}} \times (1 - M_{S_{j_1}}) \\ mistag\ rate_{SS} &= M_{S_{j_1}} \times M_{S_{j_2}} \\ mistag\ rate_{SJ} &= M_{S_{j_1}} \times (1 - M_{S_{j_2}}) \times M_{J_{j_2}} + M_{S_{j_2}} \times (1 - M_{S_{j_1}}) \times M_{J_{j_1}} \end{aligned} \tag{5.1}$$

where the 1S is the category with one SecVTX-tagged jet and one non-tagged jet, the SS is with two SecVTX-tagged jets, and the SJ is with one SecVTX-tagged jet and the other jet is JetProb-tagged and non-SecVTX-tagged. The $M_{S_{j_i}}$ ($M_{J_{j_i}}$) is the SecVTX (JetProb) mis-tag rate for j_i.

For the HOBIT tagged categories,

$$\begin{aligned} mistag\ rate_{1T} &= M_{T_{j_1}} \times (1 - M_{L_{j_2}}) + M_{T_{j_2}} \times (1 - M_{L_{j_1}}) \\ mistag\ rate_{TT} &= M_{T_{j_1}} \times M_{T_{j_2}} \\ mistag\ rate_{TL} &= M_{T_{j_1}} \times (M_{L_{j_2}} - M_{T_{j_2}}) + M_{T_{j_2}} \times (M_{L_{j_1}} - M_{T_{j_1}}) \end{aligned} \tag{5.2}$$

The T and L in subscripts 1T, TT, and TL stand for the tight HOBIT cut (greater than 0.98) and loose HOBIT cut (between 0.72 and 0.98). The notations, 1T, TT, and TL, define events with 1 tight, 2 tight, and 1 tight and 1 loose HOBIT tags, respectively. The $M_{T_{j_i}}$ and $M_{L_{j_i}}$ are the mistag rates for the tight and loose tagged HOBIT jets.

The above formulas allow us to estimate the mistags coming from EWK samples with the two leading jets originated from light-flavor quarks. We also consider the mistags for double tagged events that are promoted from the single tagging categories. For example, the mistag-SS also includes the single tagged (1S) events with the non-tagged light-flavor jets weighed by its mistag rate.

We apply these mistag rates to (and only to) light-flavor MC samples, which include W/Z+light-flavor jets, the diboson events, plus very small contribution from top events. Since the mistag rate is well determined by other independent studies, we don't apply additional scale factors.

5.2.3 The QCD Multijet Model

The event selection for $\not\!\!E_T$+jets gives a data sample dominated by QCD multijet events. A QCD multijet event, as shown in Fig. 5.5, doesn't have intrinsic missing transverse energy that results from undetectable particles. The false $\not\!\!E_T$ arises mainly from the mis-measurement of jet energies.

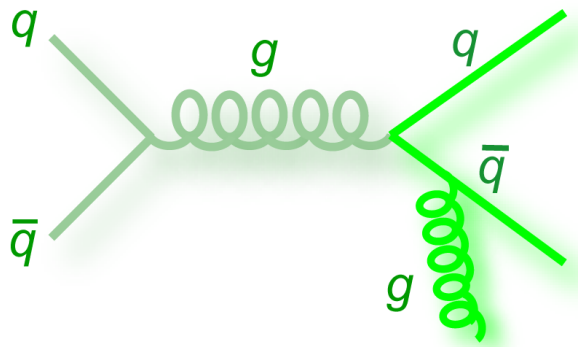


Figure 5.5.: An QCD multijet production example. Usually no neutrino is presented in the final states, except when it appears because of the decay of hadrons.

The QCD multijet events have very large production rate of 10 orders of magnitude larger than the sum of other SM processes. In a QCD dijet event, as shown in Fig. 5.6, the jets should carry same amount of momenta and give a net momentum balance of zero. Mismeasurements of jet energies could be due to the limited calorimeter energy resolution ($> 10\%$ for jet with E_T of 50 GeV) or if part of the jet energy is lost in uninstrumented regions, thus resulting in imbalance of measured energies. Such mismeasurements become problematic when combined with the large QCD multijet production rate.

Simulations of the QCD multijet production is prohibitive due to the high production rate and large theoretical uncertainties. Instead, we have developed a method which relies on data.

Since both the Higgs and the dark matter analyses focus on the b-tagged events, we only need to model the tagged QCD multijet events. The method we developed

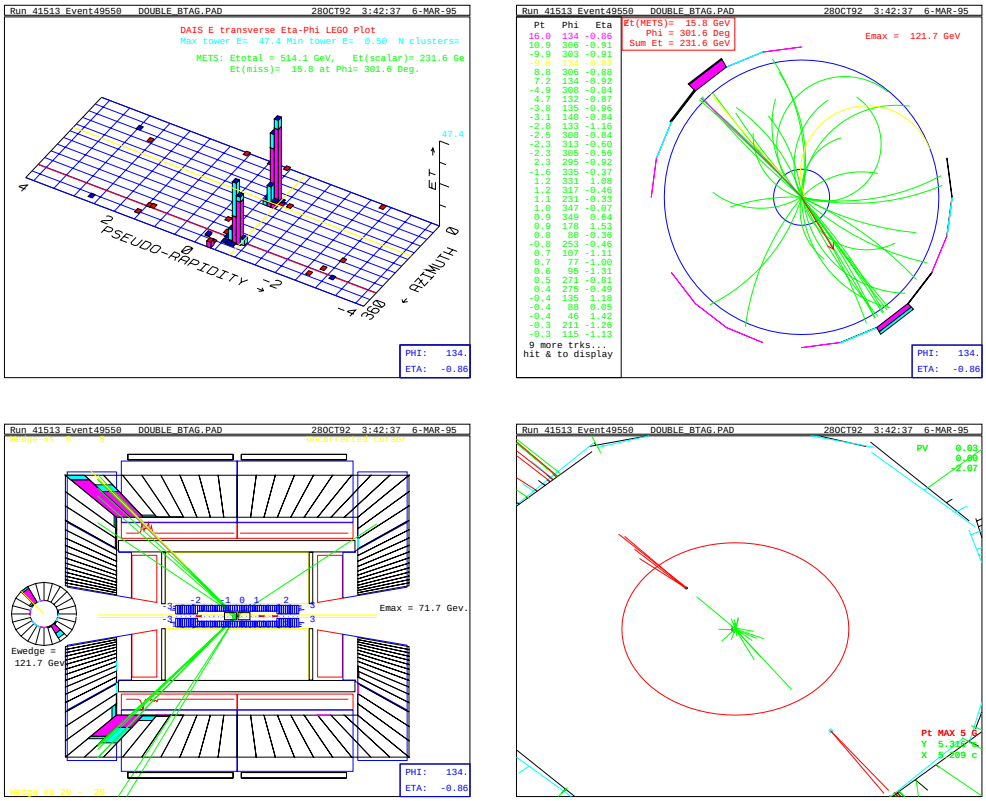


Figure 5.6.: A display of a dijet event production at CDF

is a Tag Rate Matrix (TRM) method. The TRM method utilizes an estimate of the probability for QCD multijet events to have tagged jets. The probability is derived in a control region dominated by QCD multijet events with negligible other contributions. We then apply these probability weights to the data in the signal region to estimate the tagged multijet background.

The validation of this model, along with the EWK models, is performed in several control regions, as described in Section 5.3.

To estimate the multijet tag rate, we select the events that are purely from QCD production. A very big part of the data passing the $\not\!\!E_T$ triggers are QCD multijet events. These events generally have back-to-back jets, with the missing transverse energy aligning to the jet with less measured energy, i.e., the second jet, j_2. We

define a region with $\Delta\phi(\not\!\!E_T, j_2) < 0.4$ and $35 < \not\!\!E_T < 70$ GeV (events with $\not\!\!E_T > 70$ GeV are reserved for model validation) to estimate the tagging probability of QCD multijet events.

The TAG RATE MATRIX is constructed as a 4-D matrix, where each element is the measured rate within a bin of the following variables:

- H_T: the scalar sum of E_T of all the reconstructed jets;
- $\not\!\!p_T$: the imbalance of transverse momenta of tracks;
- $j_{1,2}$ Z: fraction of sum of track p_T passed SECVTX *pass1* selections and the jet transverse energy.

The first two variables, H_T and $\not\!\!p_T$, are events variables; the j_i Z is a jet variable. The binning of the 4-D matrix is shown in Tab. 5.4.

Table 5.4: The binning setup for the TAG RATE MATRIX

	bin sets
H_T (GeV)	45, 80, 100, 120, 150, 200, 1000
$\not\!\!p_T$(GeV/c)	0, 8, 15, 25, 35, 45, 60, 1000
j_1 Z	0, 0.025, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.6, 0.8, 10.0
j_2 Z	0, 0.025, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.6, 0.8, 10.0

If it is for the HOBIT tagged categories (1T, TL, TT), we have 2 extra dimensions: N_{vtx} and $p_\perp^\mu$.

- N_{vtx}: number of reconstructed vertices.
- $p_\perp^\mu \equiv p_{\mu_1} \sin(\hat{\mu}_1, \hat{j}_1) + p_{\mu_2} \sin(\hat{\mu}_2, \hat{j}_2)$: relative transverse momentum of muons within first and second jets.

The bin boundaries of these two dimensions are as follows:

$$\begin{aligned} p_\perp^\mu &\in [-1 \times 10^5, 1 \times 10^{-4}, 0.3, 0.6, 0.9, 1.2, 1.4, 1.7, 2.0, 100] \text{GeV}/c \\ N_{\text{vtx}} &\in [-0.5, 1.5, 2.5, 3.5, 100] \end{aligned}$$

Two matrices are filled for each tagging category (1S, SJ, SS, 1T, TL, TT): one to count taggable[1] events, one to count tagged events. The tagging rate element is calculated from these two matrices,

$$Tag\,Rate = \frac{number\ of\ tagged\ events}{number\ of\ taggable\ events} \tag{5.3}$$

In the application of the Tag Rate Matrix, we weight the pretag data with the calculated *tag rate*, and subtract the weighted electroweak process contributions.

$$QCD\ multijet = (tag\ rate \times pretag\ data - tag\ rate \times pretag\ electroweak) \times SF. \tag{5.4}$$

For the control regions, a scale factor (SF), is used to normalize the QCD multijet to the data subtracted with the other SM contributions. The SF is calculated individually for each tagging category, as the Tag Rate Matrix is derived category by category. The SF is defined as,

$$SF = \frac{data - electroweak}{tag\ rate \times pretag\ data - tag\ rate \times pretag\ electroweak}. \tag{5.5}$$

The uncertainty of the SF is,

$$\left(\frac{\delta SF}{SF}\right)^2 \simeq \left(\frac{\delta(data - electroweak)}{data - electroweak}\right)^2 + \left(\frac{\delta(tag\ rate \times pretag\ data - tag\ rate \times pretag\ electroweak)}{tag\ rate \times pretag\ data - tag\ rate \times pretag\ electroweak}\right)^2 \tag{5.6}$$

In the calculation of SF uncertainty, the systematic errors from electroweak heavy flavor and mistag of electroweak light flavor also have been included, thus all the errors are being properly propagated to the determination of the QCD multijet background.

In order to maintain a blind analysis, we compute the SF for the signal region in an independent control region.

In a summary, the probability estimated by the Tag Rate Matrix is applied as a per-event weight to all events meeting our analysis selections excluding the b-tagging requirement. From this sample of weighted events, we subtract the expected

[1] SecVTX and JetProb define a jet with more than 2 or 1 good displaced tracks as a taggable jet, respectively. HOBIT defines a jet with at least one good SecVTX track as taggable.

electroweak components (as modeled by applying the same TRM probability to simulated samples). The resulting events form our model of the QCD multijet background component of our analyses.

5.3 Validating the Background Models

The background modeling is a combination of Monte Carlo methods and data-driven methods. It is very important to validate the background modeling in different kinematic regions and for a variety of variables, for the following reasons (but not limited to):

- The signal we are searching for usually has small cross section, and is negligible in many kinematic regions. For those regions, we expect the data to be perfectly reproduced by the background models. Otherwise, any discovery in the signal region could be a result of mismodeling.

- For a signal such as the Higgs boson production, a resonance appearing in the dijet invariant mass distribution is expected. The invariant mass is a variable calculated from the jet energies and directions which is expected to provide discrimination between signal and background. A good modeling of the jet energy and direction variables ensures confidence of finding evidence for the Higgs or new physics in the invariant mass distribution.

- When the multivariate methods are used, it is more important to validate all the used variables, as any mismodeling in the input variables could lead to wrong conclusion of discovery of new physics.

There are common selections between the Higgs search analysis and the dark matter search analysis, as well as the similarity in control region definitions. In this section, we discuss the validation of background models regardless of the analysis.

We select events that contain 2 or 3 energetic jets and large missing transverse energy. Specifically, we have the following cuts:

- $E_T^{j_1} > 25$ GeV, $E_T^{j_2} > 20$ GeV, and $E_T^{j_3} > 15$ GeV (if exist); jet energy cuts apply to both L5-corrected jet and H1-corrected jet
- $|\eta^{j_i}| < 2.4$, $|\eta^{j_1}|$ and $|\eta^{j_2}| < 2$, and $|\eta^{j_1}|$ or $|\eta^{j_2}| < 0.9$
- $\not\!\!E_T > 35$ GeV (both L5 and H1 corrected $\not\!\!E_T$)
- For data, we also apply the $\not\!\!E_T$ clean-up cuts, which include removing *halo* events and events with jets in the chimney area.

We further separate the events passed the above selections into signal-sensitive region (*Preselection*) and control regions (*TRM, QCD1, and EWK*).

- **Preselection**: signal-sensitive region that vetoes the obvious QCD multijet events as well as the lepton events.
  - $\Delta\phi(\vec{\not\!\!E}_T, j_{2,3}) > 0.4$, $\Delta\phi(\vec{\not\!\!E}_T, j_1) > 1.5$
  - no identified electron or muon with, e.g., $p_T > 20$ GeV/c (see Section 3.5)
- **TRM**: a QCD multijet dominated region, used to derive the Tag Rate Matrix and to perform the self validation
  - $\Delta\phi(\not\!\!E_T, j_2) < 0.4$
  - $35 < \not\!\!E_T < 70$ GeV
- **QCD1**: an independent QCD multijet dominated region, used to validate the Tag Rate Matrix model
  - $\Delta\phi(\not\!\!E_T, j_2) < 0.4$
  - $\not\!\!E_T > 70$ GeV
- **EWK**: a region requires the presence of high p_T lepton, thus is dominated by MC-modeled EWK processes. It is used to validate the MC model as well as the QCD model.

- $\Delta\phi(\not{E}_T, j_{2,3}) > 0.4$, $\Delta\phi(\not{E}_T, j_1) > 1.5$
- with identified electron or muon with, e.g., $p_T > 20$ GeV/c

The validation on a variety of used variables in these control regions are presented in the Appendix A.

6. REVIEW OF TECHNIQUES

Data analysis involves inspecting, cleaning and modeling data. Other than the techniques developed for data reconstruction and background modeling that have been discussed in the previous chapters, many other techniques are also employed for the analyses in this thesis. In this chapter, we will discuss the principle and application of the multivariate techniques and the binned maximum likelihood fit method. We will also briefly discuss the ideas that have been applied in developing a versatile programming framework for the $\not\!\!E_T$+jets analyses.

6.1 Artificial Neural Network Data Analysis

Conventionally, data cleaning is done by applying certain selections on particular physics observables, e.g., selecting events with a high p_T lepton when searching for signal with a W boson in the final state; or selecting events with a b-tagged jet when searching for signal with a b-quark; or make a cut on a reconstructed quantity like the invariant mass of two jets when looking for events with a resonance decay to two jets. This approach is sometimes referred as the cut-base analysis. The cut-base method is more straightforward, when these quantities can provide a clear discrimination between signal and background. There are, however, many other quantities that show only limited differences between signal and background. To improve their discrimination power, it is necessary to treat these variables in a fully *multivariate way*.

In this thesis, we use the Artificial Neural Network technique to pursue the multivariate approach. Many other multivariate methods are also popular for data analysis, like the Boost Decision Tree, Matrix Element method, etc.

In particular, we use the Feed-forward Neural Network, which is also known as multilayer perceptron (MLP), as shown in Fig. 6.1. A MLP [68] consists of an interconnected group of neurons or nodes arranged in layers; each node processes the information it receives with an activation (or transformation) function, then passes the result to the next layer of nodes. The first layer, known as the input layer, receives the variables. This is followed by one or more hidden layers of nodes. The last layer outputs the final response of the network. Each interconnection is characterized by a weight, and each processing node may have a bias or a threshold. The weights and thresholds are the network parameters, often referred to collectively as weights, whose values are learned during the training phase. The activation function is generally a nonlinear function that allows for flexible modeling. Neural Networks with one hidden layer are sufficient to model the posterior probability to arbitrary accuracy. Although Neural Networks are typically described, as above, in terms of neurons and activation, it is useful to think of them as simply a specific class of nonlinear functions.

For a MLP with one hidden layer of nodes and a data set with d variables $\mathbf{x} \equiv \mathbf{x_1}, \mathbf{x_2}, ...\mathbf{x_d}$, the output of the network is

$$O(\mathbf{x}) = f(\mathbf{x}, \mathbf{w}) = g(\theta + \sum_j \mathbf{w}_j \mathbf{b}_j) = p(s|\mathbf{x}), \tag{6.1}$$

where $\mathbf{b}_j$ is the output from the hidden nodes:

$$\mathbf{b}_j = g(\theta_j + \sum_i \mathbf{w}_{ij}(\mathbf{x_i})). \tag{6.2}$$

The nonlinear activation function g is commonly taken as:

$$g(a) = \begin{cases} a & \text{Linear,} \\ \frac{1}{1+e^{-a}} & \text{Sigmoid,} \\ \frac{e^a - e^{-a}}{e^a + e^{-a}} & \text{Tanh,} \\ e^{-a^2/2} & \text{Radial.} \end{cases} \tag{6.3}$$

The Neural Network software used in this thesis is provided by TMVA [69] (The Toolkit of Multivariate Analysis), which has a ROOT-integrated [70] environment for

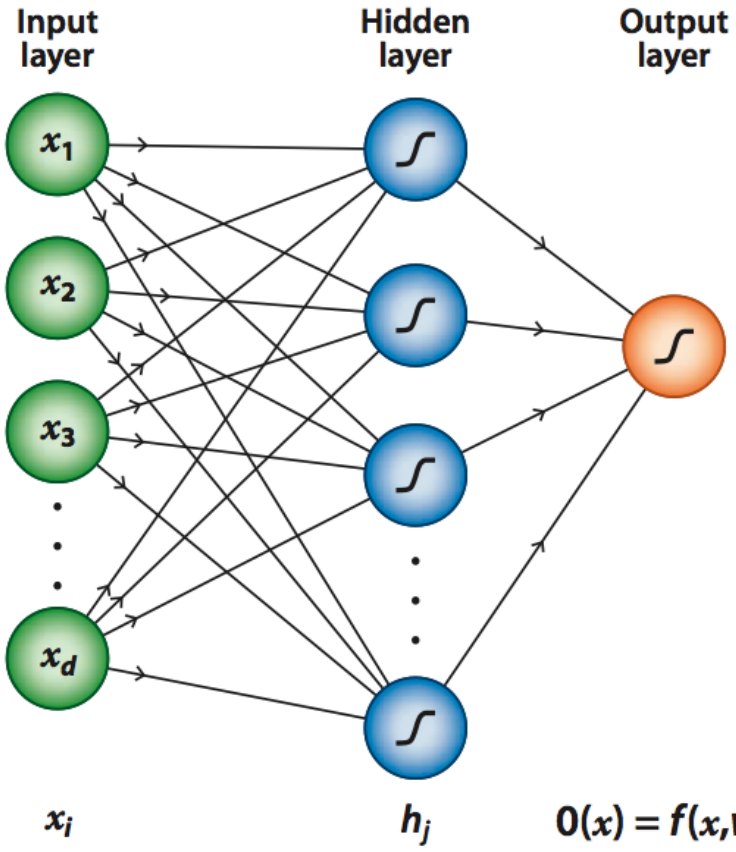


Figure 6.1.: A schematic representation of a three-layer Feed-forward Neural Network

the processing, parallel evaluation and application of multivariate classification and – since TMVA version 4 – multivariate regression techniques.

Neural network techniques can be used not only to reduce background, but also to provide better discriminant than some conventional quantities like the *dijet invariant mass* or the *missing transverse energy* to extract the signal component from the data. Neural network regression techniques can also be used to provide combined trigger parameterization to increase data acceptance.

6.2 Limit Calculation: Binned Maximum Likelihood Fit

Conventionally, we measure certain process by counting the observed events in a signal region, and subtracting the corresponding standard model processes to estimate the potential signal events. In this manner, the data is treated as one bin of a histogram. The cross section for the process (s) to be measured is,

$$\sigma_s = \frac{\epsilon n - b}{AL}, \tag{6.4}$$

where ϵ is the experimental efficiency for data taking, such as the trigger and reconstruction efficiencies, pile-up correction, etc., n is the raw number of observed data passing event selection, b is the number of expected background events, A is the acceptance of signal events, and L is the integrated luminosity.

When the observed data n is compatible with the average number of background events b expected from known processes, there is no hint for new physics. To quantify this, the experimental result needs to be translated into an *upper limit* on the cross section σ for the hypothetical new process.

Let us assume an experiment is conducted, and n events (Poisson distributed) are observed, with a mean expectation $s\epsilon + b$, where s is the number of expected signal events, ϵ is the acceptance (product of branching fraction, detector efficiency, luminosity factor, etc.) and b is the number of expected background events. Let us assume that the background is known precisely, and the signal acceptance is known with a precision σ_ϵ from a subsidiary measurements. The Bayesian theorem then states that the posterior for s is:

$$p(s, \epsilon|n) = \frac{P(n|s, \epsilon)\pi(s)\pi(\epsilon)}{\int\int P(n|s, \epsilon)\pi(s)\pi(\epsilon)dsd\epsilon} \tag{6.5}$$

where $P(n|s, \epsilon)$ is the probability of observing n events given a signal s with acceptance ϵ. The functions $\pi(s)$ and $\pi(\epsilon)$ are the prior probability densities for s and ϵ respectively. If there is no experimental information about the signal, in the Bayesian

technique a flat prior is assigned to s (i.e. $\pi(s) = 1$). The number of observed events in data follows Poisson distribution, therefore:

$$P(n|s,\epsilon) = \frac{(s\epsilon + b)^n}{n!} e^{-(s\epsilon+b)} \tag{6.6}$$

We can obtain the posterior for s by "marginalizing" over ϵ:

$$p(s|n) = \int_0^\infty p(s,\epsilon) d\epsilon. \tag{6.7}$$

The effect of systematic uncertainties are incorporated by 'smearing' with a Gaussian function, e.g., the efficiency ϵ:

$$p(s|n) = \int_0^\infty p(s,\epsilon) \cdot g(\epsilon, \Delta\epsilon) d\epsilon. \tag{6.8}$$

Finally, to obtain the limit on s, the posterior density function $p(s|n)$ is integrated until the desired confidence level (C.L.), i.e., by solving the following equation for s_u:

$$\int_0^{s_u} p(s|n) ds = 0.95 \tag{6.9}$$

an upper limit for s_u at 95% C.L. can be found.

To increase the sensitivity, we can expand the above method to the binned distribution, *i.e.* a histogram. In such approach, each bin of a histogram represents a separate search channel, with its own background acceptances and uncertainties.

In the case of a binned distribution of the discriminant variable, with a total number of bins N, the k^{th} bin contains n_k observed events and $s\epsilon_k + b_k$ expected events. Here, s is the total number of expected signal events, ϵ_k and b_k are the signal acceptance and the amount of background events in k^{th} bin. All of the ϵ_k and b_k have uncertainties and are considered nuisance parameters. Similar to the procedure above, they are assigned priors, that may be correlated, and the joint prior is written as:

$$\pi(\epsilon_1, b_1 ... \epsilon_N, b_N) \tag{6.10}$$

and the marginalized posterior for s is proportional to:

$$\pi(s) \int_{2N} ... \int \pi(\epsilon_1, b1 ... \epsilon_N, b_N) [\prod_{k=1}^{N} \frac{e^{-(s\epsilon_k + b_j)} (s\epsilon_k + b_k)^{n_k}}{n_k!}] d\epsilon_1 db_1 ... d\epsilon_N db_N \tag{6.11}$$

The software we used to compute the posterior, integrate the posterior, and calculate limits is described in [84]. The software is written with C++ language. The ensemble of histograms (templates) for the signals, backgrounds, and their corresponding systematic uncertainties is provided by the analyzer.

6.3 The Analysis Framework

Analyzing the high-energy physics data is a collaborative effort as well as an individual effort. The data from the particle factory are first analyzed online, then are reconstructed and calibrated offline. These online and offline processes are a collaborative effort that involves groups of experts to design the data processing systems. After the reconstructed data is stored, special analysis tools are developed for broad users. Examples of these tools are the b-tagging algorithms and the jet energy corrections. When it comes to a particular physics measurement, it is more like an individual effort, although usually with involvement of a small group of physicists.

Take an example at the CDF experiment. The reconstructed data are stored as complex ROOT files in the form of StNtuple (or TopNtuple, BNtuple, Duke Ntuple, etc.). These complex files have relatively complete event information, such as the reconstructed jets, tracks, as well as hits in tracking system, fired towers in calorimeters. Accessing these files usually requires load of particular libraries. These files are very large (multi petabytes) and contain too much information for final users. An individual analyzer needs to *skim* these files into regular ROOT files that only contain event information that are useful for a particular analysis. The individual analyzer, who often is not a professional computer scientist, has the responsibilities of developing the skimming code and the physics analysis code. In such process, code developed by final users often are only understandable and usable by the authors. As a result, an analysis could take a PhD student several years to develop the analysis framework and finish the physics measurement. Transferring the framework to new analyzers is also hard.

An early effort of this thesis was to develop a generalized analysis framework that can be easily adapted for different analyses and easily used by newcomers. Although the technique details on coding are out of the scope of this thesis, it is worth to note the ideas that have been conceived in the framework development.

Regularize the ROOT Files

There are more than 150 Monte Carlo samples used for this thesis. In addition, the pre-skim large-$\not\!\!E_T$ data sample is close to the size of exabyte. After skimming, the analyzed ROOT files are still more than 100 gigabyte, which is as much a tenfold increase to the size of previously analyzed samples. An inefficient structure of the ROOT file could take larger storage space as well as more time to analyze the data. It is even more time consuming if some additional physics information is found to be needed in the middle of an analysis; re-skimming all the samples could take weeks or months. Things such as bad naming of physics variables, duplicated informations, could also cause inefficiency in data analysis.

The following ideas are applied in regularizing the ROOT files:

- Have a well organized structure of the physics variables and group the related information.
- Use easy-to-read name for variables and follow certain conventions. Always comment the variables.
- Remove duplicated information and easy-to-calculate variables (e.g., $\not\!\!E_T/H_T$ is not needed when we have $\not\!\!E_T$ and H_T variables).
- Use array as often as possible, thus accessing variables is easier in the analysis code development. The size of arrays should be dynamic to save storage space.
- Any future modification should not affect the current structure, thus the analysis code could be always backward compatible.

Some rules may sound intuitive, however a regularized ROOT file is the base of the analysis code development.

A User-friendly Analysis Code

The object-oriented programming not only improves the performance of code but also increase the readability of code. The latter is even more important for physicists, as it can largely increase the possibility of reuse of code and broaden the user base.

The following ideas are applied in developing a user-friendly analysis code:

- Apply modular programming.
- The functionality sometimes may be hard to understand, the function name has to be clear.
- Develop accessing tools. The user only needs to change some external variables to run for different configurations, even different analysis. This is not only for convenience but also requires less modification of code and reduces complication and human mistakes.
- Reduce complexity so that only a few button needs to be pushed to finish an analysis.
- Regularize input and output files.

The framework that I have developed was used successfully in over ten CDF analyses by variety of users from undergraduates to postdocs.

7. SEARCH FOR THE LOW MASS STANDARD MODEL HIGGS BOSON

One of the most sensitive SM Higgs boson search channels at the Tevatron is the $VH \to \not\!\!E_T + b\bar{b}$ final state, where the V is the W or Z boson, $\not\!\!E_T$ represents the missing transverse energy resulting from neutrinos or unidentified charged leptons in the event. This chapter reports an update to the previous CDF analysis in the $\not\!\!E_T + b\bar{b}$ search channel [85]. The full CDF data, corresponding to an integrated luminosity of 9.5 fb^{-1}, are analyzed. A recently developed b-tagging strategy is used; that significantly improves the signal acceptance.

7.1 Event Preselection

In addition to the event *preselection* that has been introduced in Section 5.3, we require jets to have transverse energies that satisfy $25 < E_T^{j_1} < 200$ GeV, $20 < E_T^{j_2} < 120$ GeV, and $15 < E_T^{j_3} < 100$ GeV (if the 3^{rd} jet exists). The requirements on the maximum energy of the jets are applied to avoid the over efficiency of the HOBIT algorithm for the very-high-energy jets. The ratio of HOBIT to SecVTX tagged jets in Fig. 7.1 shows that the HOBIT tagging efficiency is very large for jets with large E_T. The reason for the jump is that in the HOBIT neural-network training the Monte Carlo samples don't contain many jets with E_T above 100 GeV and therefore the HOBIT output is not reliable for those jets.

The requirements that we have described select candidate events consistent with the $ZH \to \nu\bar{\nu}b\bar{b}$ process. Because τ leptons are not explicitly reconstructed and some electrons and muons escape detection or reconstruction, events from the $WH \to \ell\nu b\bar{b}$ process are also expected to contribute significantly. We are also sensitive to the $ZH \to \ell\bar{\ell}b\bar{b}$ as a signal sample, despite its very small contribution.

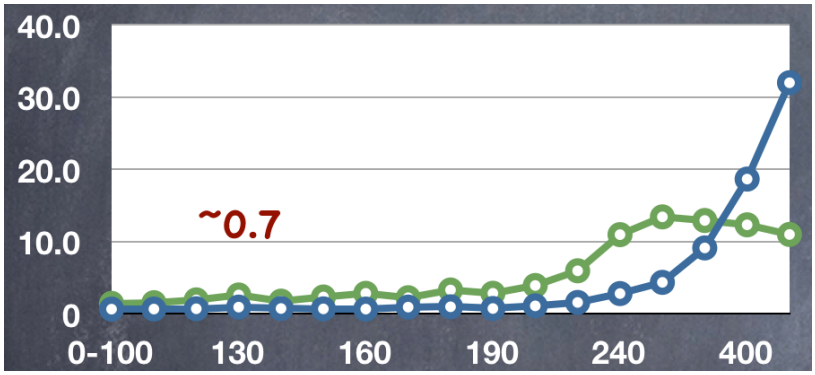


Figure 7.1.: The blue (green) points are the ratio of number of tight (loose) HOBIT and SecVTX tagged jets versus the jet E_T for the leading jet in the event.

7.2 Tagging Categories

This analysis employs a multivariate b-tagging algorithm (HOBIT) specifically optimized for $H \to b\bar{b}$ searches. Jets initiated by b quarks tend to cluster at values (v) close to 1, whereas those initiated by light-flavor quarks are more likely to populate the region near -1. Two operating regions are used: jets with $v \geq 0.98$ are considered to be tightly tagged (T), whereas jets with $0.72 < v < 0.98$ are loosely tagged (L). We accept events assigned to one of three categories based on the tag quality of the two leading-E_T jets: both jets are tightly tagged (TT); one jet is tightly tagged, and the other loosely tagged (TL); and only one jet is tightly tagged (1T). The tag categories used in both analyses and the associated tagging efficiencies of Higgs boson signal events are given in Tab. 7.1. As can be seen, the HOBIT algorithm achieves a 32% (11%) relative improvement in the tagging efficiency of signal events into the double-tight (tight-loose) category.

Table 7.1: Comparison of b-tagging efficiencies per signal event in the tag categories of this analysis and the previous one [85]. Jets tagged by the SECVTX b-tagging algorithm are labeled "S", and those that are tagged by the JETPROB algorithm but not SECVTX are labeled "J". There is no overlap between the tag categories of a given analysis by design.

Tag category	b-tagging efficiency per event	
	Ref. [85]	This analysis
Two tight b tags	13.7% (SS)	18.1% (TT)
One tight and one loose b tag	13.1% (SJ)	14.6% (TL)
Only one tight b tag	31.4% (1S)	31.6% (1T)

7.3 Multivariate Discriminants

7.3.1 QCD Discriminant for the Signal Region

To optimally separate Higgs boson signal from background, a multivariate approach is used. First a neural network $\mathrm{NN_{QCD}}$ is trained to discriminate between QCD multijet and signal processes. Events that satisfy a minimum $\mathrm{NN_{QCD}}$ threshold requirement are subjected to a second neural network $\mathrm{NN_{SIG}}$, designed to separate the signal from the remaining SM backgrounds.

The $\mathrm{NN_{QCD}}$ discriminant is trained using the data passed *preselection* and weighted with the TAG RATE MATRIX as the background sample and the combination of tagged events of 50% of $ZH(125\ \mathrm{GeV})$ and 50% of WH MC as the signal sample. The collection of input variables to the $\mathrm{NN_{QCD}}$ includes kinematic, angular, and event-shape quantities, each of which is validated with tagged data in the control regions. The input variables are listed in Tab. 7.2.

Figure 7.2 shows the $\mathrm{NN_{QCD}}$ distribution for tagged events satisfying the preselection criteria. By imposing a minimum $\mathrm{NN_{QCD}}$ requirement of 0.6 (which defines the

Table 7.2: The input variables for the $\mathrm{NN_{QCD}}$ development

Variable	Description
$\not\!\!E_T$	Absolute amount of the missing transverse energy
$\not\!\!p_T$	Absolute amount of the missing transverse momentum
$\not\!\!E_T/H_T$	Ratio of $\not\!\!E_T$ to H_T
$\not\!\!H_T$ / $\not\!\!E_T$	Ratio of $\not\!\!H_T$ to $\not\!\!E_T$
$M(\vec{\not\!\!E}_T, \vec{j}_1, \vec{j}_2)$	Invariant mass of $\not\!\!E_T$, $\vec{j}_i$ and $\vec{j}_2$
$\Delta\phi(\vec{\not\!\!E}_T, \vec{\not\!\!p}_T)$	Azymuthal difference between $\not\!\!E_T$ and $\not\!\!p_T$
$Max(\Delta\phi(\vec{j}_i, \vec{j}_k))$	Maximum of $\Delta\phi$ between any two jets $\vec{j}_i, \vec{j}_k$
$Max(\Delta R(\vec{j}_i, \vec{j}_k))$	Maximum of ΔR between any two jets $\vec{j}_i, \vec{j}_k$
$Min(\Delta\phi(\vec{\not\!\!E}_T, \vec{j}_i))$	Minimum of $\Delta\phi$ between $\vec{\not\!\!E}_T$ and any jet $\vec{j}_i$
$Min(\Delta\phi(\vec{\not\!\!p}_T, \vec{j}_i))$	Minimum of $\Delta\phi$ between $\vec{\not\!\!p}_T$ and any jet $\vec{j}_i$
ϕ^*	$\Delta\phi$ of $(\vec{j}^1, \vec{j}^2)$ axis in their rest frame and their vector sum in the lab frame
Sphericity	$S = \frac{3}{2}(\lambda_2 + \lambda_3)$ [86]

signal region), 87% of the signal is retained while 90% of the QCD multijet background is rejected.

Table 7.3 shows the expected number of signal and background events and the observed data events in the signal region. For Higgs boson mass of 125 GeV, we expect 19 signal events in the 1T category and roughly 11 signal events in both the TL and TT categories.

7.3.2 Final Discriminant for Limit Calculation

A second neural-network variable, $\mathrm{NN_{SIG}}$, is developed as a final discriminant for the limit calculation. Since the background composition is different in events with 2 or 3 jets, we train separate networks in each category. The outputs of these networks are combined in the end, when searching for the signal. For the neural-network training of

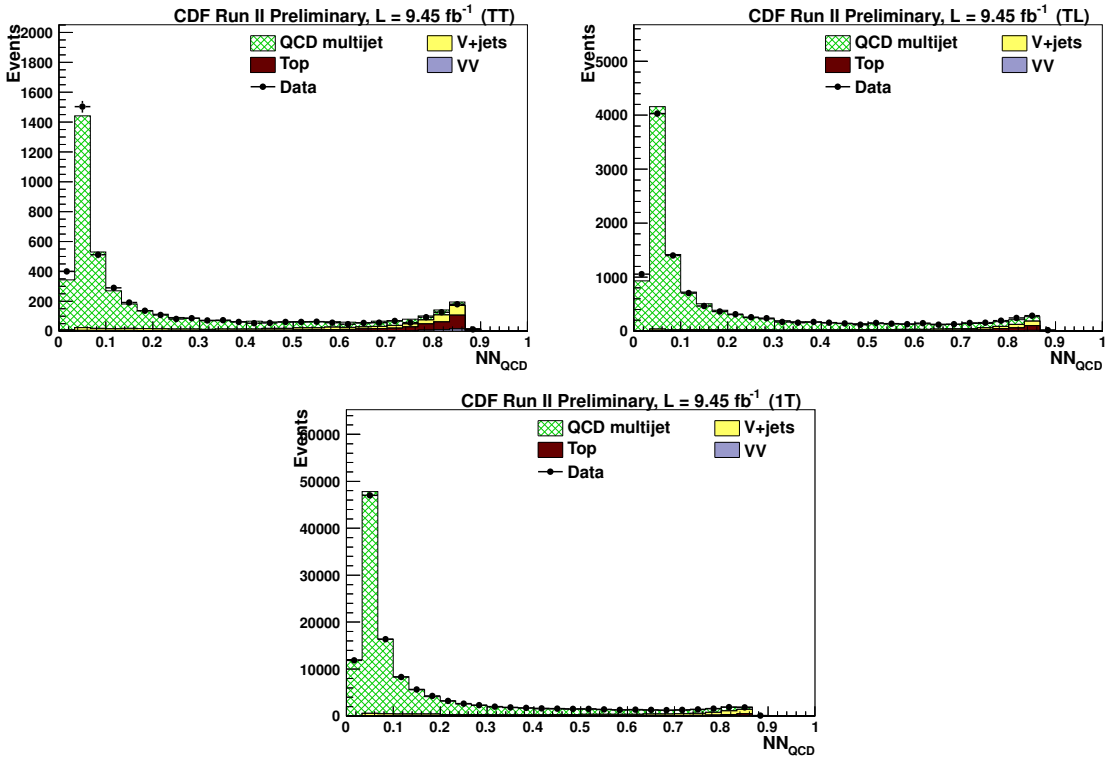


Figure 7.2.: The distribution of tagged data events and the corresponding expected backgrounds for the $\mathrm{NN_{QCD}}$ discriminant function for all tagged events in the preselection sample.

2-jet (3-jet) events we use a background sample made of 75% (50%) of untagged data (none of the jets in the events are b-tagged) and 25% (50%) of $t\bar{t}$ events. We choose the untagged data because they are mainly QCD multijet and W/Z+jets events, they are not contaminated by signal events. The Higgs signal used for the training is a mixture of 50% WH events and 50% ZH events. We train one network for each of the 12 mass points we probe (from 90 to 150 GeV/c^2 in steps of 5 GeV/c^2).

The neural network chosen here is once again the Multi Layer Perceptron. The 7 input variables are presented in Tab. 7.4. In order to increase the separating power of the $\mathrm{NN_{SIG}}$, we implement a track-based discriminant, TrackMET, which was trained to optimize the separation of both ZH and WH events from QCD and $t\bar{t}$ backgrounds.

Table 7.3: Comparison of the number of expected and observed events in the signal region for different b-tagging categories. The uncertainties shown include systematic contributions and (when appropriate) statistical uncertainties on the simulation samples, added in quadrature for a given process. The quoted uncertainties for the total expected background prediction take into account the appropriate correlations among the systematic uncertainties for each background process. Signal contributions are given for an assumed Higgs boson mass of 125 GeV/c^2.

Process	1T	TL	TT
QCD multijet	5941 ± 178	637 ± 25	222 ± 16
Top	1174 ± 158	302 ± 40	271 ± 34
V + heavy-flavor jets	3124 ± 718	286 ± 83	211 ± 65
Electroweak mistags	1070 ± 386	55 ± 21	13 ± 6
Diboson	305 ± 46	48 ± 6	41 ± 5
Total expected background	11612 ± 949	1329 ± 112	759 ± 86
Observed data	11955	1443	692
$ZH \to \nu\bar{\nu}b\bar{b}, \ell\bar{\ell}b\bar{b}$	9.7 ± 1.0	5.4 ± 0.5	5.4 ± 0.5
$WH \to \ell\nu b\bar{b}$	9.8 ± 1.0	5.3 ± 0.5	5.3 ± 0.5

Figure 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, 7.9 show the final discriminant distributions for different Higgs boson masses. They are also the templates used to extract the limits at each Higgs boson mass.

7.4 Results

We perform a binned likelihood fit to search for the presence of a Higgs boson signal. A combined likelihood is formed from the product of Poisson probabilities of the event yield in each bin of the $\mathrm{NN_{SIG}}$ distribution for each tag category. Systematic uncertainties are treated as nuisance parameters and incorporated into the limit by

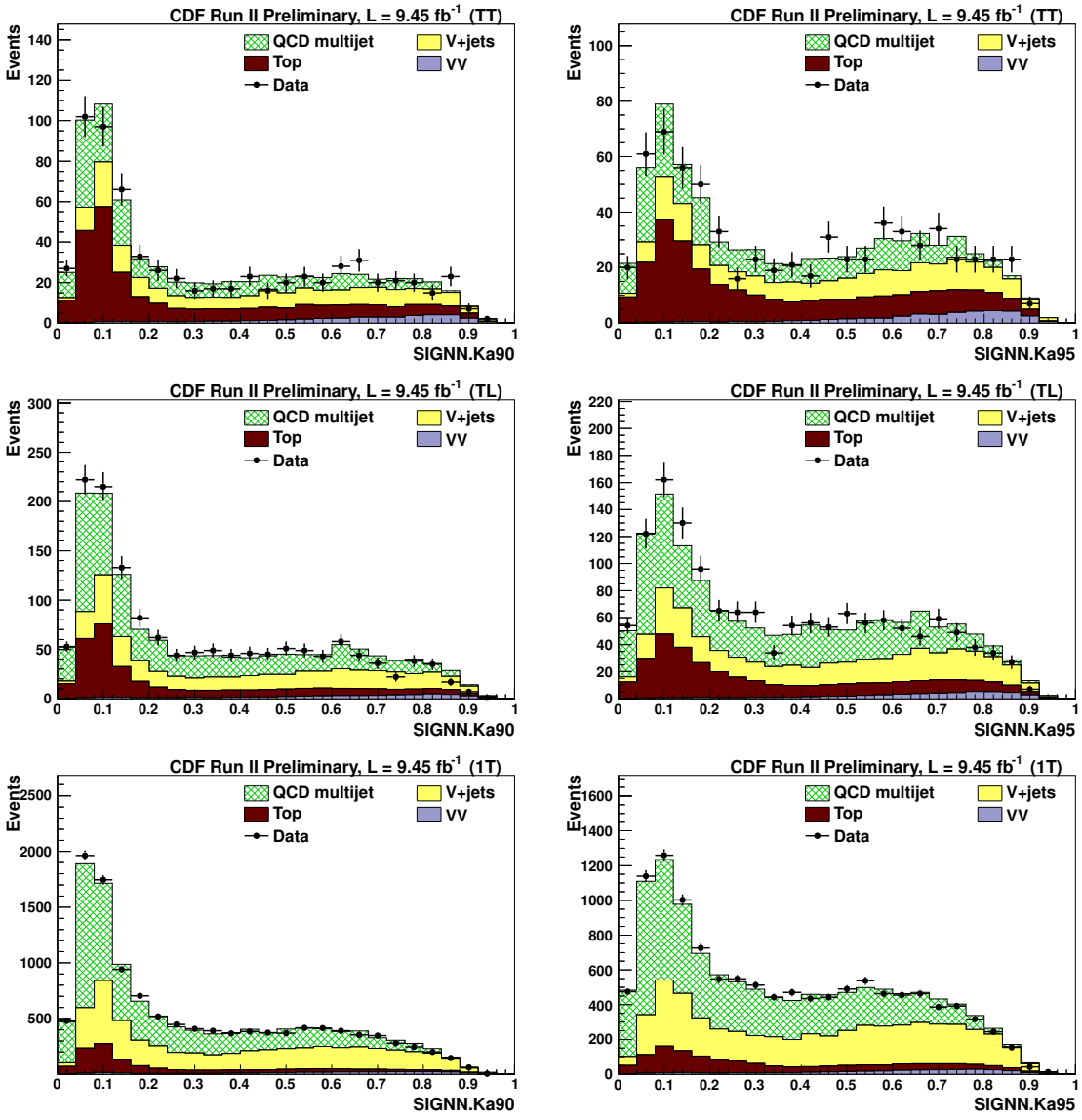


Figure 7.3.: Final discriminants for $m_H = 90$ GeV/c^2 (left) and 95 GeV/c^2 (right)

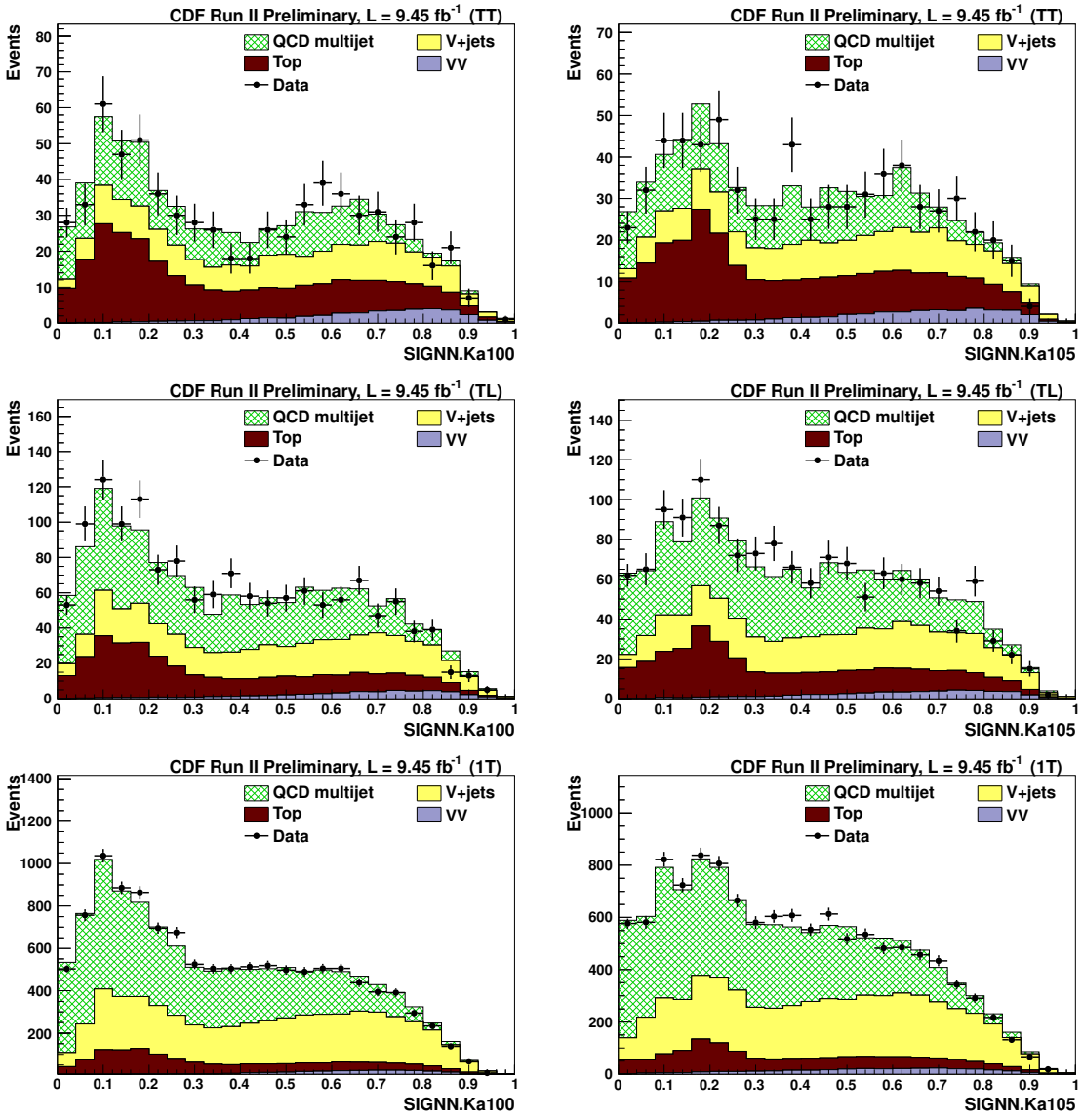


Figure 7.4.: Final discriminants for $m_H = 100$ GeV/c^2 (left) and 105 GeV/c^2 (right)

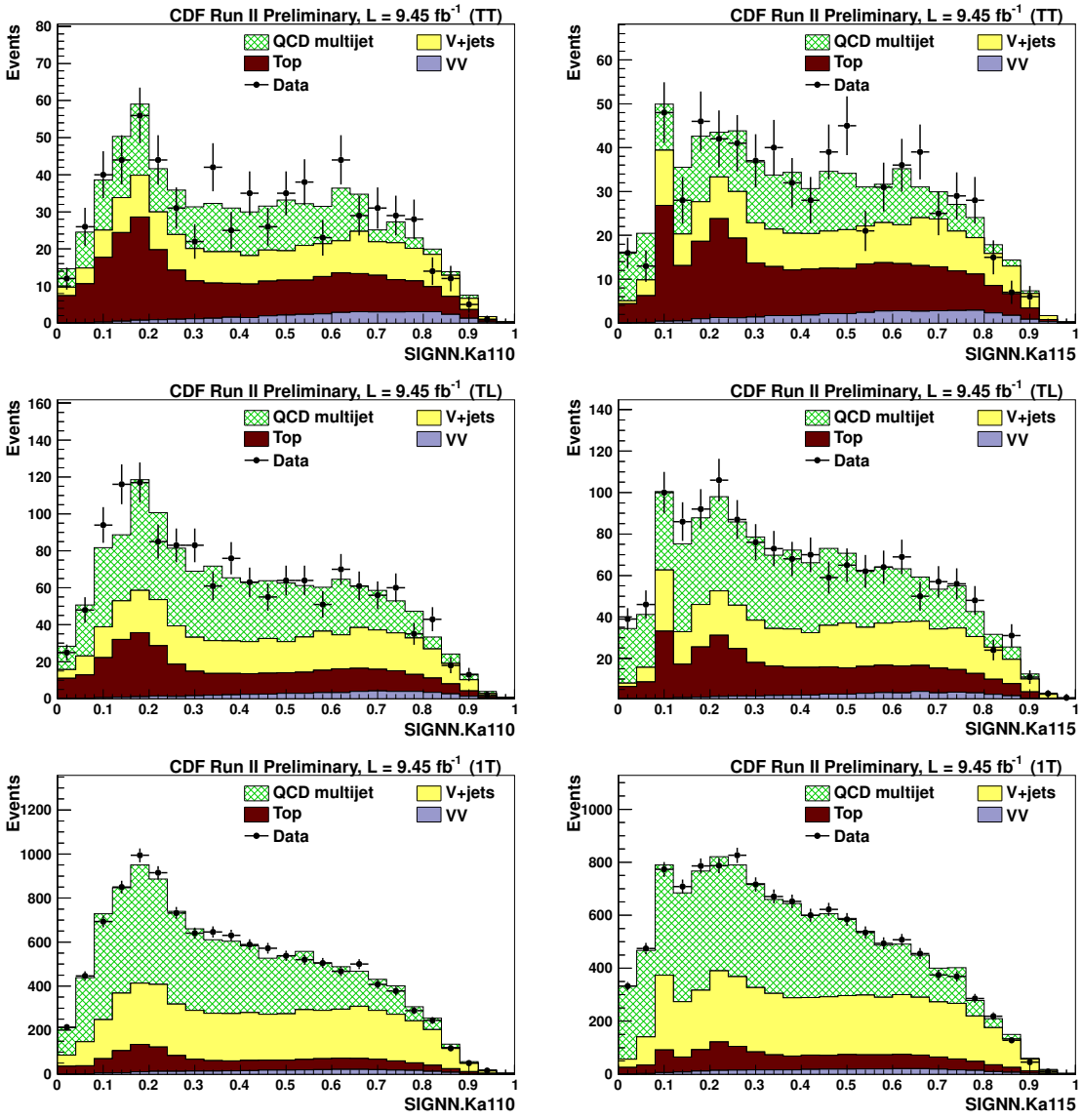


Figure 7.5.: Final discriminants for $m_H = 110$ GeV/c^2 (left) and 115 GeV/c^2 (right)

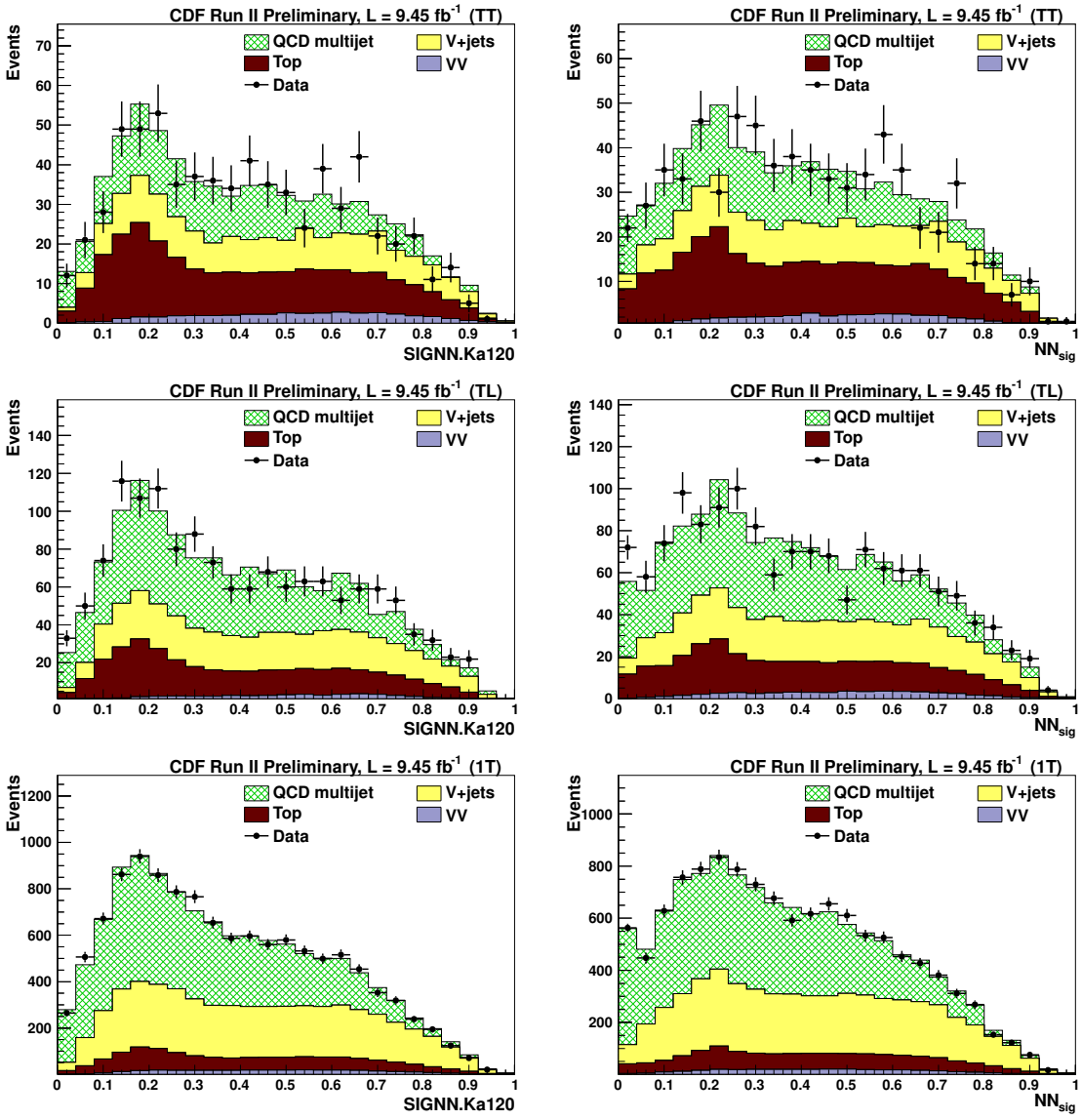


Figure 7.6.: Final discriminants for $m_H = 120$ GeV/c^2 (left) and 125 GeV/c^2 (right)

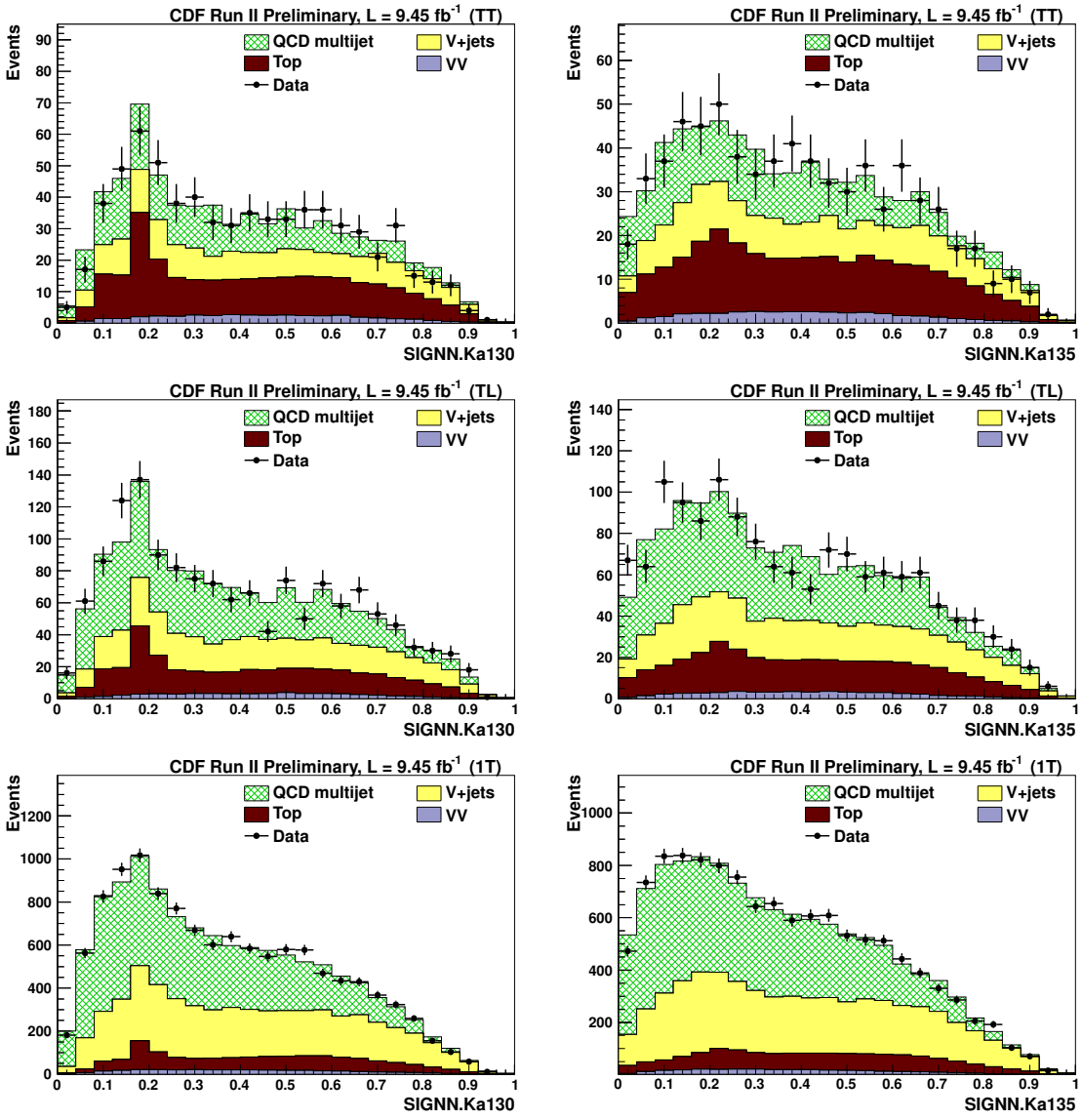


Figure 7.7.: Final discriminants for $m_H = 130$ GeV/c^2 (left) and 135 GeV/c^2 (right)

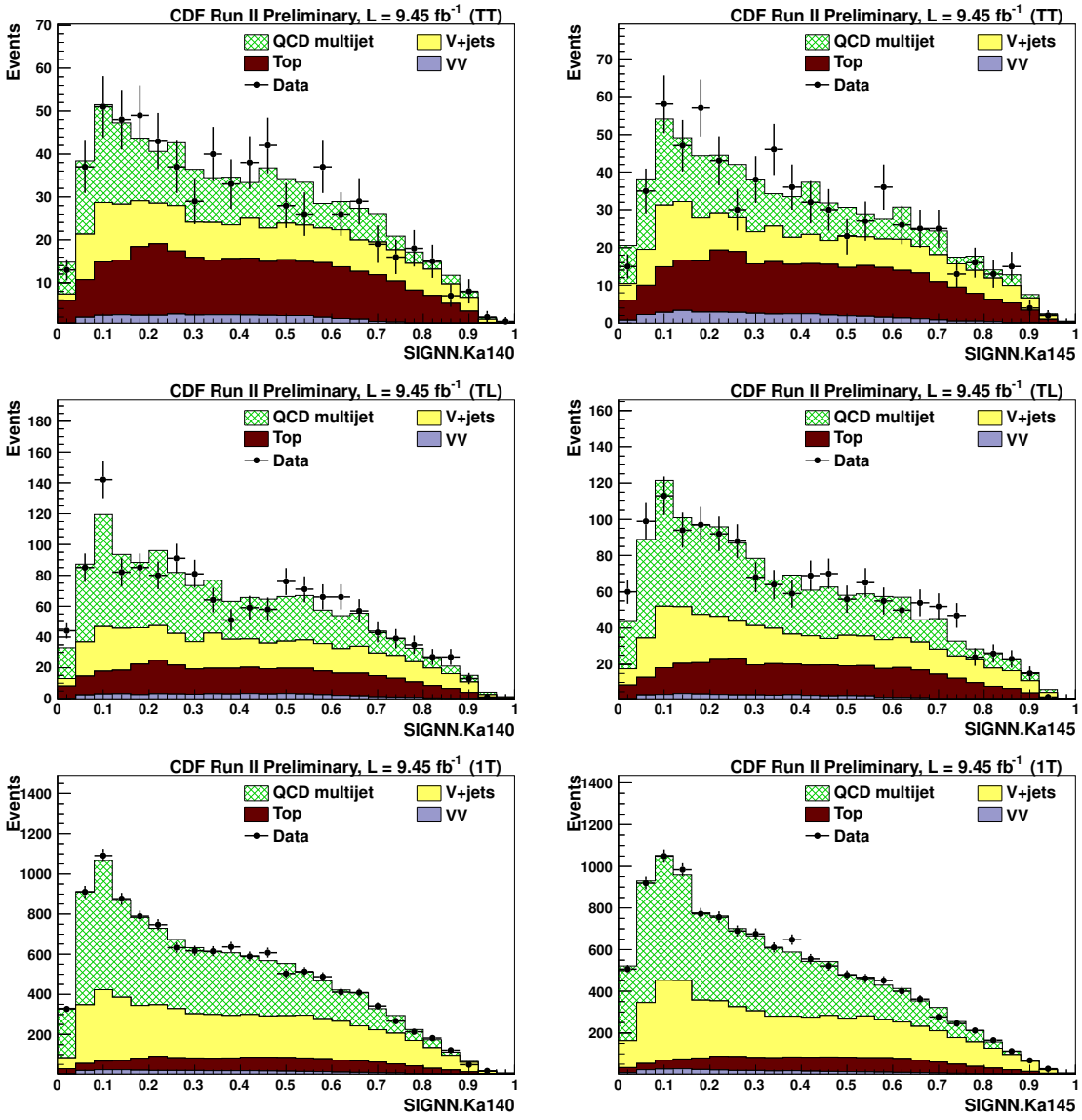


Figure 7.8.: Final discriminants for $m_H = 140$ GeV/c^2 (left) and 145 GeV/c^2 (right)

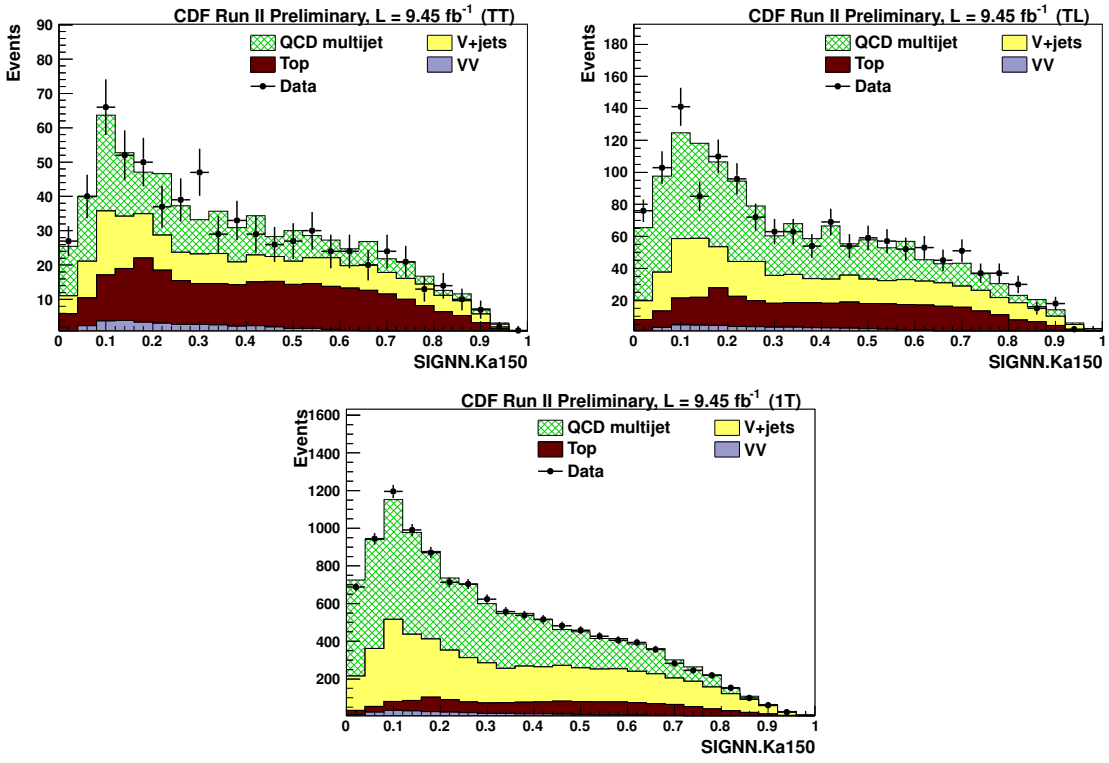


Figure 7.9.: Final discriminants for $m_H = 150$ GeV/c^2

Table 7.4: Input variables to the final discriminant neural network.

Variable	Description
$H_T - \not\!\!E_T$	Scalar sum of the jet energies minus $\not\!\!E_T$
$\not\!\!H_T$ - $\not\!\!E_T$	$\not\!\!H_T$ subtract $\not\!\!E_T$
TrackMET	TrackMET
$M(\vec{\not\!\!E}_T, \vec{j_1}, \vec{j_2})$	Invariant mass of $\not\!\!E_T$, $\vec{j_i}$ and $\vec{j_2}$
$M(j_1, j_2)$	Invariant mass of two leading jets
$Max(\Delta\phi(\vec{j_i}, \vec{j_k}))$	Maximum of $\Delta\phi$ between any two jets $\vec{j_i}, \vec{j_k}$
$\mathrm{NN_{QCD}}$	output of the $\mathrm{NN_{QCD}}$

assuming Gaussian prior probabilities, centered at the nominal value of the nuisance parameter, with an rms width equal to the absolute value of the uncertainty. The dominant systematic uncertainties arise from the normalization of the V plus heavy flavor background contributions (30%), differences in b-tagging efficiencies between data and simulation (8–16%) [29], uncertainty on the top (6.5–10%) and diboson (6%) cross sections, normalizations of the QCD multijet background (3–7%), luminosity determination (6%) [87], jet-energy scale (6%) [23], trigger efficiency (1–3%), parton distribution functions (2%), and lepton vetoes (2%). Additional uncertainties applied only to signal include those on the Higgs boson production cross section (5%) [88] and on initial- and final-state radiation effects (2%). Also included are uncertainties in the $\mathrm{NN_{SIG}}$ shape, which arise primarily from variations in the jet-energy scale and the QCD multijet background model.

A Bayesian likelihood method is used to set 95% credibility level (C.L.) upper limits on the SM Higgs boson production cross section times branching fraction $\sigma(\mathrm{VH}) \times \mathcal{B}(H \to b\bar{b})$. For the signal hypothesis, a flat, non-negative prior probability is assumed for the number of selected Higgs boson events. The Gaussian priors of the nuisance parameters are truncated at zero to ensure non-negative event yield predictions in each $\mathrm{NN_{SIG}}$bin. The 95% C.L. limits for the observed data and the

median-expected outcomes assuming only SM backgrounds are shown in Fig. 7.10 and Table 7.5. An average improvement of 14% is obtained in expected upper limits relative to the previous analysis [85]. The observed limits lie below the expected values at the level of roughly one standard deviation for $m_H \geq 120$ GeV/c^2, and at the level of approximately two standard deviations for lower Higgs boson masses. In constrast, the observed limits of the previous analysis exceed the median-expected limits by roughly one standard deviation for $m_H > 120$ GeV/$c^2$and are in approximate agreement with expected limits for lower masses. These differences correspond to a decrease of roughly 55% in the observed limits relative to those of the previous analysis [85] independent of m_H.

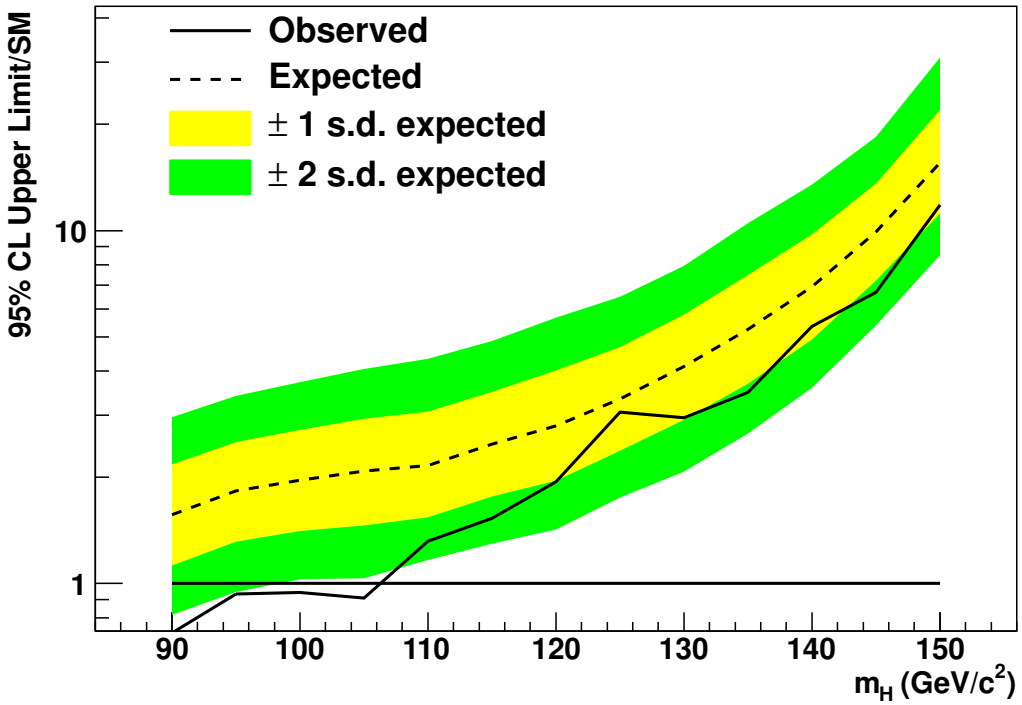


Figure 7.10.: Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on VH cross section times $\mathcal{B}(H \to b\bar{b})$ divided by the SM prediction, as a function of the Higgs boson mass. The bands indicate the 68% and 95% credibility regions where the limits can fluctuate, in the absence of signal.

Table 7.5: Expected and observed 95% C.L. upper limits on the VH cross section times $\mathcal{B}(H \to b\bar{b})$ divided by the SM prediction [8].

m_H (GeV/c^2)	90	95	100	105	110	115	120	125	130	135	140	145	150
Expected	1.57	1.83	1.96	2.08	2.16	2.48	2.80	3.33	4.13	5.26	6.93	9.91	15.55
Observed	0.72	0.94	0.94	0.91	1.32	1.53	1.94	3.06	2.95	3.49	5.35	6.69	11.82

8. SEARCH FOR DARK MATTER PARTICLE PRODUCED IN ASSOCIATION WITH A SINGLE TOP QUARK

In the SM, top quarks are primarily produced in pairs at particle colliders. They can also be produced singly via weak interactions, resulting in a final state consisting of a single top quark with additional lighter-flavor quarks. SM single top-quark events in the missing energy plus jets channel have been studied within the standard-model hypothesis [89]. As the beyond-the-SM monotop theory predicts production of a single top quark in association with a DM particle, the published SM single top-quark results do not provide any conclusive information on the existence of monotop. In addition, searches for the associated production of top quarks with DM particles have only been performed in the context of events containing a pair of top quarks [90–92]. Therefore, a dedicated search for monotops produced in colliders is needed, as the observation of monotops would be a clear sign of new physics. In this chapter, we report the first direct search for monotop signatures at particle colliders, assuming the top quark to be produced through flavor-changing interactions of up and top quarks, in association with a DM candidate D. We assume that the D particle has a mass in the range of $0 - 150$ GeV/c^2; we do not consider decays of the D particle to up and top quarks in a higher mass range.

8.1 Event Selection

We use a data sample corresponding to an integrated luminosity of 7.7 ± 0.5 fb^{-1}. The sample is collected from Feb. 2002 to Mar. 2011 with the Tevatron Run II collisions.

In order to retain only those events for which the trigger system is fully efficient, we select events with $\not\!\!E_T > 50$ GeV and three jets. Exactly one jet is identified as a b-jet by using the SecVTX algorithm. We require the jet transverse energy $E_T^{j_i}$, to be $E_T^{j_1} > 35$ GeV, $E_T^{j_2} > 25$ GeV, $E_T^{j_3} > 15$ GeV, where the jets j_i $(i = 1, 2, 3)$ are ordered by decreasing energy. We require that either j_i or j_2 have $|\eta| < 0.9$, and that all three jets have $|\eta| < 2.4$. We veto events with identified high-p_T electrons or muons, removing monotop events inconsistent with a hadronically-decaying top quark.

After the selection described above, we are left with 6471 data events. We expect that approximately 70% of these events come from QCD multijet production. In order to further suppress the QCD contamination and the other SM backgrounds, we require the azimuthal distances between the $\not\!\!E_T$ and j_2, $\Delta\phi(\not\!\!E_T, j_2) > 0.7$, as the $\not\!\!E_T$ in QCD multijet background tends to align to the jet with less measured energy. We also require the invariant mass of the three jets to be consistent with the reconstructed top-quark mass, $110 < m_{jjj} < 200$ GeV/c^2, large $\not\!\!E_T$ significance ($\not\!\!E_T/\sqrt{\sum E_T} > 3.5\sqrt{\text{GeV}}$, where $\sum E_T$ is the scalar sum of transverse energy deposited in the calorimeter) and $E_T^{j_3} > 25$ GeV. Figure 8.1 shows these variables in comparison between data and signal.

All selections have been chosen to optimize the significance $S/\sqrt{S+B}$, where S and B are the expected number of signal and backgrounds events, respectively. Table 8.1 shows the number of events in the signal region for the data, the number of events for SM backgrounds, and the expected signal assuming different values of the DM particle's mass. The events that fail these signal-region selections are used to form a control region that is used to validate the background models, as well as to determine the normalization of the QCD multijet background.

Figure 8.2 and 8.3 shows the validations plots for the reconstructed top mass distributions and the $\not\!\!E_T$ distributions, respectively.

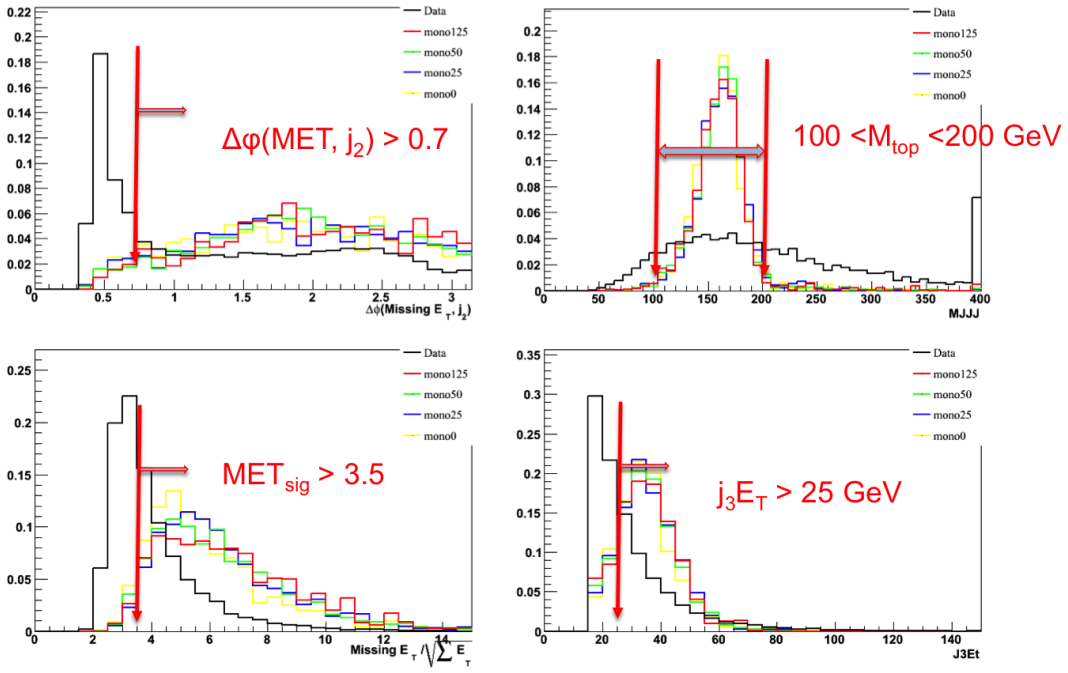
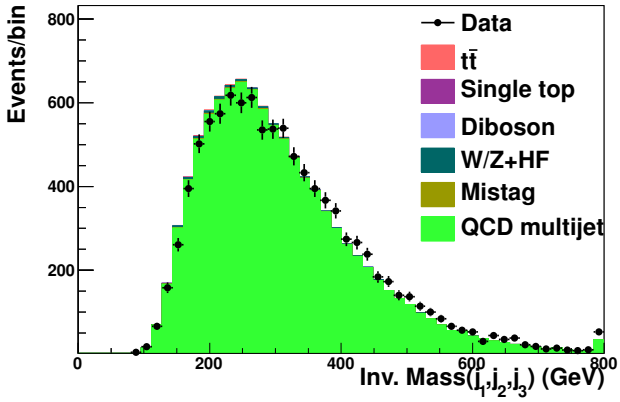


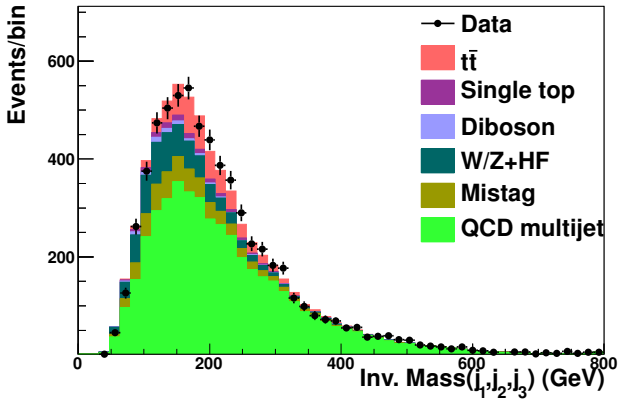
Figure 8.1.: Event selections for the signal region: the black histogram is the distribution of data, which is mostly QCD multijet production, the colored histograms are four signal samples.

8.2 Systematics

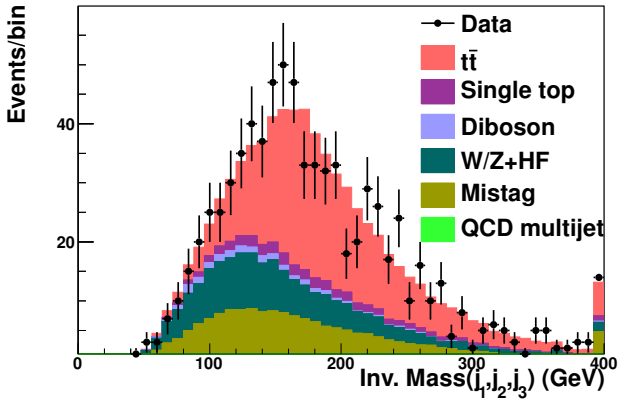
We consider several systematic uncertainties affecting the sensitivity of this search. The dominant systematic sources are the uncertainties on multijet normalization (25.5%), the mistag rate (16.6%) and the background cross sections (6.5% – 30%). We also consider uncertainties from the jet energy scale [23] (2.8% – 10.7%), the luminosity measurement [87] (6%), parton density functions (2%), lepton veto (2%), b-tagging efficiency (5.2%), trigger efficiency (0.4% – 0.9%), and from the initial-state and final-state radiation (4%). We also assign systematic uncertainties, based on the variation in the shape of the distribution of kinematic quantities, under a $\pm 1\sigma$ variation of the jet energy scale and the uncertainty on the efficiency of the data acquisition system.



(a) QCD region

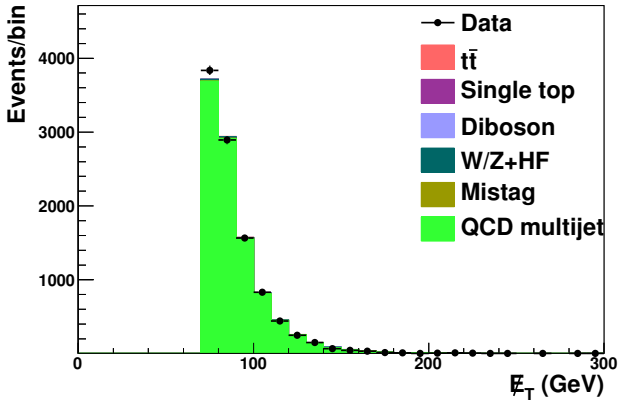


(b) Preselection region

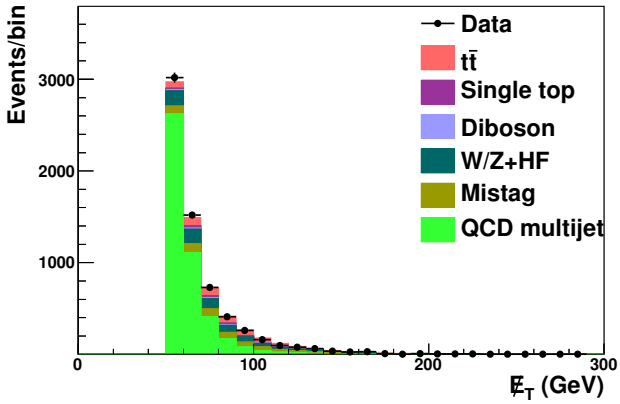


(c) EWK region

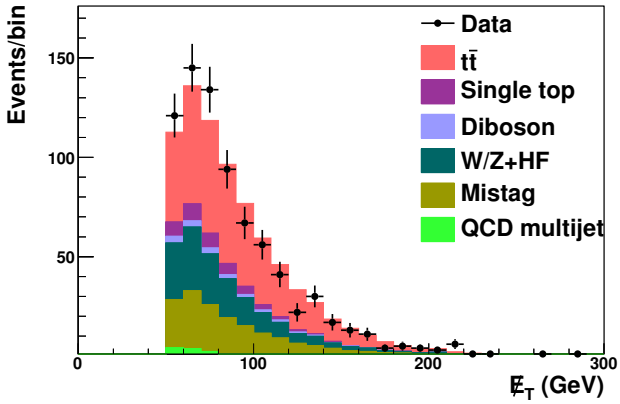
Figure 8.2.: The reconstructed top mass distributions in control regions



(a) QCD region



(b) Preselection region



(c) EWK region

Figure 8.3.: The missing transverse energy distributions in control regions

Table 8.1: Number of expected signal and background events compared to data in the signal region. The expected signals, assuming different values for the mass of the DM particle, are also presented. The errors include statistical and systematic uncertainties.

Processes	Events
$p\bar{p} \to t + D$	
$m_D = 20$ GeV/c^2	2116.9 ± 121.4
$m_D = 75$ GeV/c^2	232.3 ± 22.9
$m_D = 100$ GeV/c^2	129.8 ± 12.5
$m_D = 125$ GeV/c^2	94.5 ± 9.3
$t\bar{t}$	182.8 ± 20.2
Single top	24.3 ± 4.5
Diboson	15.7 ± 2.7
W/Z+HF	130.5 ± 33.8
Mistag	96.9 ± 39.4
QCD multijet	210.2 ± 54.5
Total background	660.2 ± 78.1
Data	592

8.3 Results

The $\not\!\!E_T$ is chosen to discriminate the signals from the backgrounds. The $\not\!\!E_T$ distribution due to a DM particle of mass of 125 GeV/c^2 and the SM backgrounds are shown in Fig. 8.4. The signal is expected to contribute significantly at high values of $\not\!\!E_T$. We find no significant excess of signal-like events in the data analyzed, and thus proceed to set 95% confidence level (C.L.) upper limits on the monotop DM production cross section. The limits are calculated with the $\not\!\!E_T$ distribution as the shape discriminant using a Bayesian maximum likelihood method assuming a flat

prior for the signal cross section [84]. We treat systematic uncertainties using a Bayesian marginal likelihood method. Figure 8.5 shows the calculated upper limits on the monotop cross section as a function of the mass of the DM candidate compared to the theoretical predictions.

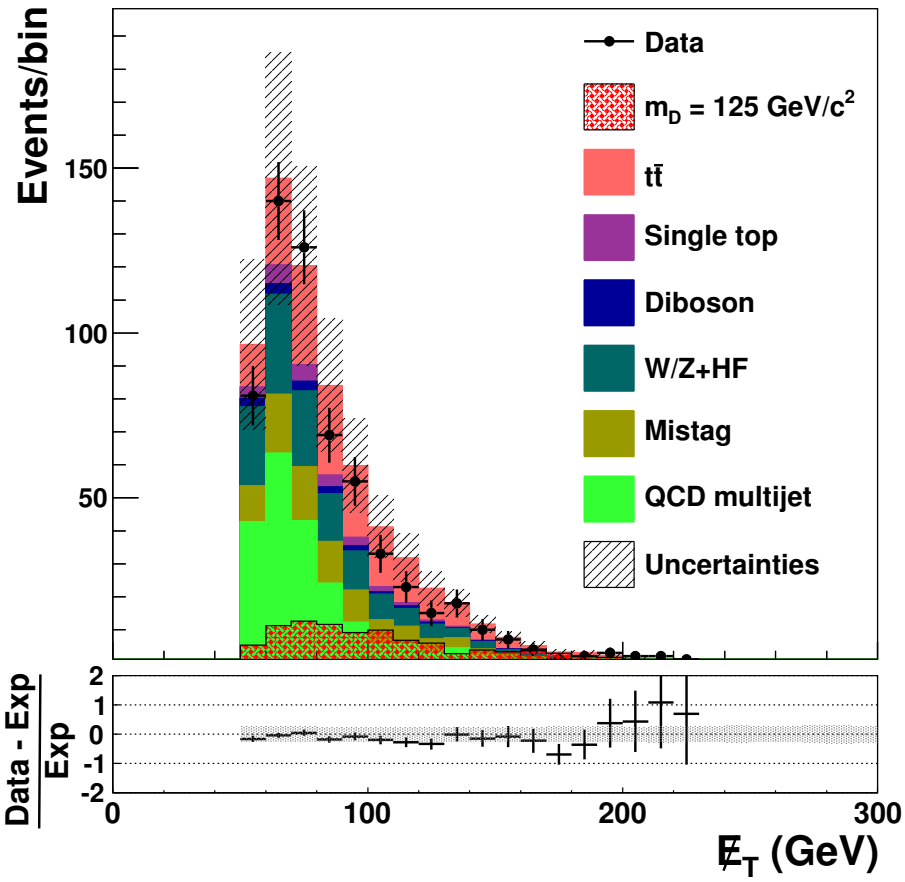


Figure 8.4.: The $\not\!\!E_T$ distribution in the signal region. The data is compared to the sum of the SM contributions. The distribution of signal events with a DM mass of 125 GeV/c^2 is also shown.

In conclusion, we have performed the first search for the production of DM in association with a single top quark at hadron colliders. In an analysis of 7.7 fb^{-1} of CDF II data we have found that the observed data is consistent with the expectation from SM backgrounds. We set 95% C.L. upper limits on the cross section of $p\bar{p} \to D{+}t$

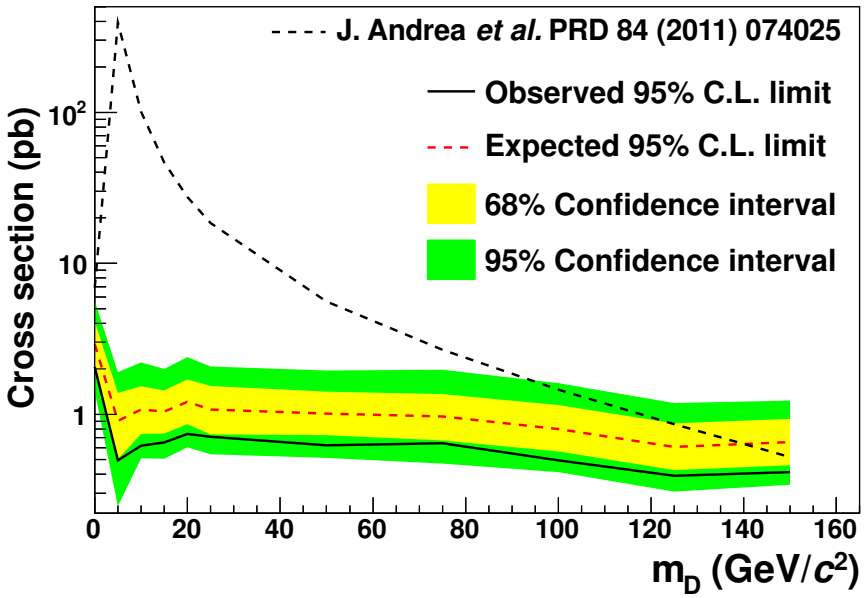


Figure 8.5.: Exclusion curve of the monotop cross section as a function of the mass of DM particle

as a function of the DM mass in the range of $0 - 150$ GeV/c^2. Future searches for new physics in monotop final states can probe resonant production of top quarks and DM candidates with exotic mediators. While these processes are predicted to have low production rates (making them difficult to probe with Tevatron data), they are expected to be within the reach of LHC experiments with sufficient data.

9. CONCLUSION

We have presented a result of searching for the Higgs boson and a result of searching for a dark-matter candidate. We use data collected by the CDF detector with proton-antiproton collisions at a center-of-mass energy of 1.96 TeV. We search in events with energetic jets and large missing transverse energy – a signature characterized by complicated backgrounds. We have discussed the techniques that have been developed for background modeling, for improving discrimination be tee signal and background, and for reducing background resulting from detector effects.

The Higgs results are based on searching for the Higgs boson produced in association with vector boson that the Higgs boson decays to b-quark pair and the vector boson decays leptonically with neutrino(s) in the final state. The previous CDF analysis in this channel uses the conventional SecVTX and JetProb algorithms to identify b-jets. The results presented in this thesis applies a multivariate-technique based method – HOBIT – to identify b-jets. With higher efficiencies in b-jet identification, the expected limits on the Higgs searches across mass of 90 to 150 GeV/c^2 have been improved by 14%.

The dark matter results are based on searching for a dark-matter candidate produced in association with a top quark that the dark-matter escapes detection and the top quark decays to three jets. We search for the dark-matter production in the distribution of missing transverse energy – a quantity that represent the amount of energy carried by the dark-matter particle. We have performed the first search of this signature in the hadron collider experiments. We find the data collected at the the Tevatron are consistent with the standard model predictions, and have estimated the upper limits for the dark-matter production theory.

LIST OF REFERENCES

LIST OF REFERENCES

[1] S. Glashow, A. Salam, S. Weinberg, The Nobel Prize in Physics (1979).

[2] S.F. Novaes, Standard Model: An Introduction. arXiv:hep-ph/0001283v1.

[3] P.W. Higgs, Phys. Rev. Lett. **12**, 132 (1964);
idem., Phys. Rev. **145**, 1156 (1966);
F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964);
G.S. Guralnik, C.R. Hagen, and T.W. Kibble, Phys. Rev. Lett. **13**, 585 (1964).

[4] ATLAS Collaboration, Combined search for the Standard Model Higgs Boson using up to 4.9 fb^{-1} of pp collision data at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC, Phys. Lett. B **710**, 49 (2012).
CMS Collaboration, Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV, Phys. Lett. B **710** 26 (2012).

[5] M. Kobayashi, T. Maskawa, CP-Violation in the Renormalization Theory of Weak Interaction, Progress of Theoretical Physics, Vol. 49, No. 2, February 1973.

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

[7] G. Abbiendi *et al.*, Search for the Standard Model Higgs Boson at LEP, Phys.Lett.B **565**:61-75 (2003).

[8] The CDF Collaboration, the D0 Collaboration, Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data, arXiv:1207.0449 (2012).

[9] ATLAS Collaboration, Combined measurements of the mass and signal strength of the Higgs-like boson with the ATLAS detector using up to 25 fb^{1} of proton-proton collision data, ATLAS-CONF-2013-014.
CMS Collaboration, Combination of Standard Model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV, CMS-HIG-12-045.

[10] S. Perlmutter, B.P. Schmidt, A.G. Riess, The Nobel Prize in Physics (2011).

[11] K.C. Freeman, Astrophys. J. **160** (1970) 811.

[12] T.S. van Albada, J.N. Bahcall, K. Begeman and R. Sancisi, Astrophys. J. **295** (1985) 305.

[13] K.G. Begeman, A.H. Broeils and R.H. Sanders, Mon. Not. R. Astr. Soc. **249** (1991) 523.

[14] G.G. Raffelt, Dark Matter: Motivation, Candidates and Searches, arXiv:hep-ph/9712538 (1997).

[15] A.K. Drukier, K. Freese, D.N. Spergel, Detecting cold dark-matter candidates, Phys. Rev. D **33**, 3495 (1986).
K. Freese, J. Frieman, A. Gould, Signal modulation in cold-dark-matter detection, Phys. Rev. D **37**, 3388 (1988).

[16] R. Bernabei *et al.*, arXiv:1007.0595 (2010).

[17] C.E. Aalseth *et al.*, Phys. Rev. Lett. **107** 141301 (2011).

[18] E. Aprile *et al.*, XENON100 Collaboration, Phys. Rev. Lett. **107** 131302 (2011).

[19] E. Armengaud *et al.*, EDELWEISS Collaboration, Phys. Lett. B**702** 329 (2011).

[20] `http://dmtools.brown.edu`

[21] The CDF Collaboration, The CDF Detector Technical Design Report, FERMILAB-Pub-96/390-E;
The CDF Collaboration, The CDF IIb Detector Technical Design Report;
M. Paulini, A.B. Wicklund, The CDF Detector.

[22] B. Ashmanskas *et al.* The CDF Silicon Vertex Trigger. Nucl. Instrum. Meth. A, 518(1-2):532–536 (2004).

[23] A. Bhatti *et al.*, Nucl. Instrum. Meth. A 566:375-412 (2006).

[24] T. Affolder *et al.*, Phys. Rev. D **64**, 032001 (2001).

[25] T. Affolder *et al.*, Phys. Rev. D **63**, 032003 (2001).

[26] T. Aaltonen *et al.*, Search for the Higgs boson Using Neural Networks in Events with Missing Energy and b-Quark Jets in ppbar collisions at $\sqrt(s) = 1.96$ TeV, Phys. Rev. Lett. **104**, 141801 (2010).

[27] D. Acosta *et al.*, 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, Phys. Rev. D **71**, 052003 (2005).

[28] A. Abulencia *et al.*, 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 jet probability b-tagging, Phys. Rev. D **74**, 072006 (2006).

[29] J. Freeman *et al.*, Introduction to HOBIT, a b-Jet Identification Tagger at the CDF Experiment Optimized for Light Higgs Boson Searches, arXiv:1205.1812 (2012).

[30] C. Ferrazza, Identificazione di quark pesanti in getti adronici in interazioni $p\bar{p}$ con il rivelatore CDF al Tevatron, Master's thesis, Universita "La Sapienza" Roma (2006).
P. Mastrandrea, Study of the heavy flavour fractions in Z+jets events from $p\bar{p}$ collisions at energy = 1.96 TeV with the CDF II detector at the Tevatron collider, FERMILAB-THESIS-2008-63.

[31] J. Freeman *et al.*, An Artificial Neural Network Based B-Jet Identification Algorithm at the CDF Experiment, Nucl. Instrum. Meth. A, Vol. 663 (2012), pp. 27-37.

[32] A. Affolder *et al.*, The CDF Collaboration, Measurement of the W Boson Mass with the Collider Detector at Fermilab, Phys.Rev.D **64**, 052001 (2001).

[33] D. Acosta, *et al.*, Cleaning Up of the Missing Transverse Energy Datasets, CDF Note 7410

[34] D. Bortoletto, Q. Liu, F. Margaroli, Everything you wanted to know on the missing transverse momentum, CDF Note 9843

[35] M. Bentivegna, Q. Liu, F. Margaroli, K. Potamianos, Detection of invisible particles at hadron collider experiments through the magnetic spectrometer, arXiv:1205.4470 (2012).

[36] Lepton ID information: `http://www-cdf.fnal.gov/internal/physics/joint_physics/instructions/`

[37] Q. Liu *et al.*, CDF Collaboration, Measuring the Top Pair Background to the Higgs Search in the MET+b-jets Channel, CDF public note 10237 (2010)

[38] M. Cremonesi, Q. Liu *et al.*, Measurement of Single Top Production Cross Section in $\not\!\!E_T$ plus Jets Sample with the Full CDF Data, CDF note 10969 (2013)

[39] Q. Liu *et al.*, CDF Collaboration, Search for a Dijet Resonance in Events with Jets and Missing Transverse Energy, CDF public note 10968 (2013)

[40] T. Aaltonen *et al.*, CDF Collaboration, Search for New T$'$ Particles in Final States with Large Jet Multiplicities and Missing Transverse Energy in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$TeV, Phys. Rev. Lett. 107, 191803 (2011)

[41] The TEVNPH Working Group, Updated combination of CDF and D0 searches for standard model Higgs boson production with up to 10.0 fb^{-1} of data, arXiv:1207.0449 (2012).

[42] T. Sjostrand, S. Mrenna, and P. Skands, J. High Energy Phys. 05, 026 (2006).

[43] J. Andrea, B. Fuks, and F. Maltoni, Phys. Rev. D **84**, 074025 (2011).

[44] Y. Bai, P. J. Fox, and R. Harnik, J. High Energy Phys. 12, 048 (2010).

[45] J. L. Feng, J. Kumar, D. Marfatia, and D. Sanford, Phys. Lett. B, **703**, 2 (2011).

[46] B. Batell, J. Pradler, and M. Spannowsky, arXiv:1105.1781.

[47] J. Kile and A. Soni, Phys. Rev. D **84** 035016, (2011).

[48] P. Agrawal, S. Blanchet, Z. Chacko, and C. Kilic, arXiv:1109.3516.

[49] S. Chen and Y. Zhang, arXiv:1106.4044.

[50] J. Kamenik and J. Zupan, Phys. Rev. D **84**, 111502 (2011).

[51] D. Alves *et al.*, arXiv: 1105.2838.

[52] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, J. High Energy Phys. 06, 128 (2011).

[53] Tevatron Electroweak Working Group for the CDF and D0 Collaborations, arXiv:1107.5255.

[54] A. B. Galtieri, F. Margaroli, and I. Volobuev, arXiv:1109.2163.

[55] A. Bhatti *et al.*, IEEE Transactions on Nuclear Science 56, 3 (2009).

[56] J. Antos *et al.* Data processing model for the CDF experiment. IEEE Trans. Nucl. Sci., 53:2897-2906, 2006.

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

[58] T. Aaltonen *et al.*, CDF Collaboration, First Observation of Vector Boson Pairs in a Hadronic Final State at the Tevatron Collider, Phys. Rev. Lett. **103**: 091803 (2009).

[59] O. Gonzalez, K. Potamianos, M. Vidal, H. Wolfe, Neural-Network Parameterizations of the Trigger Efficiency for the MET+JET Sample, CDF note 10538.

[60] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, JHEP 0307:001,2003, hep-ph/0206293.

[61] J. M. Campbell and R. K. Ellis, Phys. Rev. D **60**, 113006 (1999).

[62] J. M. Campbell and R. K. Ellis, Phys. Rev. D **62**, 114012 (2000).

[63] U. Langenfeld, S. Moch, and P. Uwer, Phys. Rev. D **80**, 054009 (2009).

[64] CDF and D0 Collaborations, Combination of the $t\bar{t}$ production cross section measurements from the Tevatron Collider,CDF public note 10926 (2012).

[65] B. W. Harris *et al.*, Phys. Rev. D **66**, 054024 (2002).

[66] Z. Sullivan, Phys. Rev. D **70**, 114012 (2004).

[67] SecVTX Mistag Matrices for 5.6 fb^1 of Data Up to P28, CDF note 10179, 2010.

[68] Pushpalatha C. Bhat, Multivariate Analysis Methods in Particle Physics, Annu. Rev. Nucl. Part. Sci. 2011. 61:281309

[69] A. Hoecker *et al.*, TMVA - Toolkit for Multivariate Data Analysis, arXiv:physics/0703039

[70] R. Brun and F. Rademakers, ROOT - An Object Oriented Data Analysis Framework, Nucl. Inst. Meth. in Phys. Res. A 389, 81 (1997).

[71] A. Abulencia *et al.* [CDF Collaboration], "Measurement of the t anti-t production cross section in $p\bar{p}$ collisions at $sqrts = 1.96$-TeV using missing E(t) + jets events with secondary vertex b-tagging," Phys. Rev. Lett. **96** (2006) 202002

[72] G. Compostella, S. Amerio, D. Lucchesi, "Measurement of the ttbar production cross section in the MET+jets channel in 2.2 fb-1" CDF conference note 9988;

[73] T. Aaltonen *et al.*, (CDF Collaboration), Phys. Rev. Lett. **106**, 171801 (2011).

[74] P. Fox, J. Liu, D. Tucher-Smith, N. Weiner, Phys. Rev. D 84, 115006 (2011)

[75] E. Eichten, K. Lane, A. Martin, arXiv:1104.0976 (2011)

[76] V.M. Abazov *et al.*, (D0 Collaboration), Phys. Rev. Lett. **107**, 011804 (2011).

[77] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **109**, 251801 (2012)

[78] T. Aaltonen *et al.*, (CDF Collaboration), Phys. Rev. Lett. **109**, 111804 (2012)

[79] F. Abe, *et al.* (CDF collaboration), Phys Rev. D **45**, 001448 (1992).

[80] CDF uses a cylindrical coordinate system with the z axis along the proton beam axis. Pseudorapidity is $\eta = -\ln(\tan(\theta/2))$, where θ is the polar angle relative to the proton beam direction, and ϕ is the azimuthal angle while $p_T = |p|\sin\theta$, $E_T = E\sin\theta$.

[81] T. Aaltonen *et al.* [CDF and D0 Collaborations], Phys. Rev. D **86** (2012) 092003

[82] S. Alioli *et al.*, J. High Energy Phys. 06, 001 (2010).

[83] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. D **82**, 112005 (2010).

[84] http://www-cdf.fnal.gov/ trj/mclimit/production/mclimit.html.

[85] Phys. Rev. Lett. **109**, 111805 (2012).

[86] A momentum tensor is defined as $M_{lm} = \frac{\sum_o j_l^o j_m^o}{\sum_o |\vec{j^o}|}$, where $\vec{j^o}$ is the momentum of a reconstructed jet, and l and m are Cartesian coordinates. The index o runs over the number of jets in the event. The sphericity in an event is defined as $S = \frac{3}{2}(\lambda_2 + \lambda_3)$, where λ_2 and λ_3 are the smallest two eigenvalues of the normalized momentum tensor.

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

[88] J. Baglio and A. Djouadi, J. High Energy Phys. 10 (2010) 064; O. Brien, R. V. Harlander, M. Weisemann, and T. Zirke, Eur. Phys. J. C **72**, 1868 (2012).

[89] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. D **81**, 072003 (2010).

[90] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **106**, 191801 (2011).

[91] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **107**, 191803 (2011).

[92] G. Aad *et al.* (ATLAS Collaboration), arXiv:1109.4725.

APPENDIX

A. VALIDATION PLOTS

A.1 TRM Region

1T	TL	TT

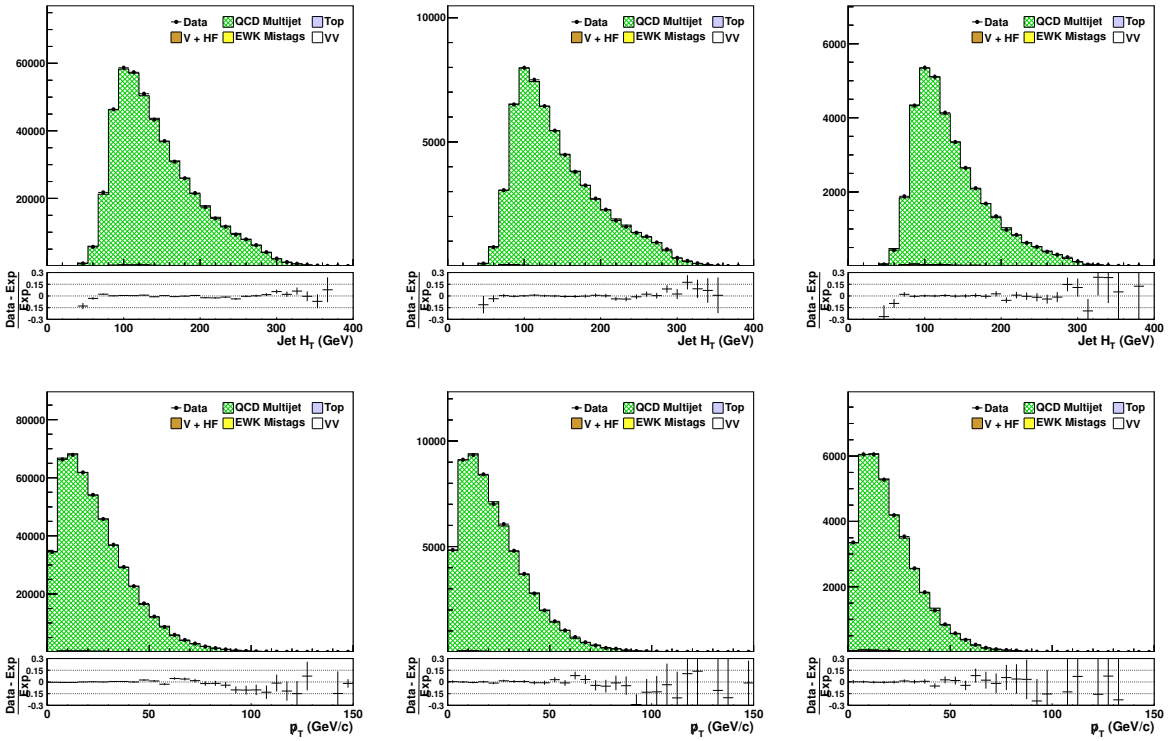


Figure A.1.: Comparisons between data and modeling of TRM parameters in TRM region.

1T	TL	TT

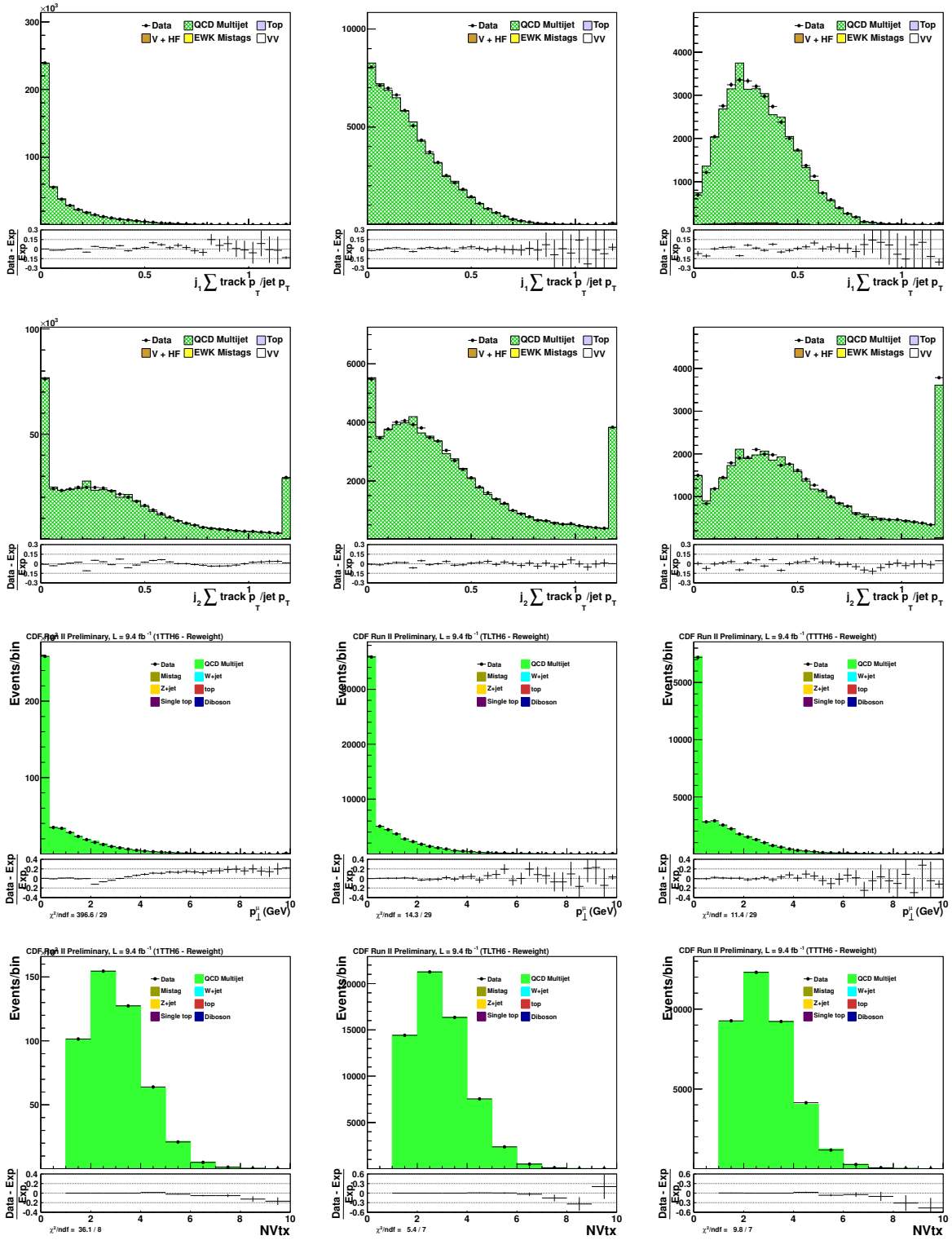


Figure A.2.: Comparisons between data and modeling of TRM parameters in TRM region.

A.2 Excursus on QCD Jet-E_T Reweighting

As mentioned in CDF Note 10786, the different angular requirements in the preselection and electroweak regions, relative to the TRM region, change the kinematic acceptance of the sample. The QCD multijet model, derived in a region where $\Delta\phi(\vec{\not\!\!E}_T, \vec{j}_2) < 0.4$, therefore needs to be adjusted to be applicable to a region where $\Delta\phi(\vec{\not\!\!E}_T, \vec{j}_2) > 0.4$. This is done by constructing a 2D matrix, binned as a function of the transverse energies of jet 1 and jet 2. To derive the QCD adjustment factors (or Jet-E_T reweights), the electroweak contributions are subtracted from tagged data in the preselection region. This 2D (binned) shape is then compared with that as predicted from the QCD multijet model, and a 2D matrix of correction factors are derived to be applied in addition to the TRM weights as mentioned in Sec. **??**. The reweights are shown in histogram form in Fig. A.3. All remaining plots in Sec. **??** include these reweights.

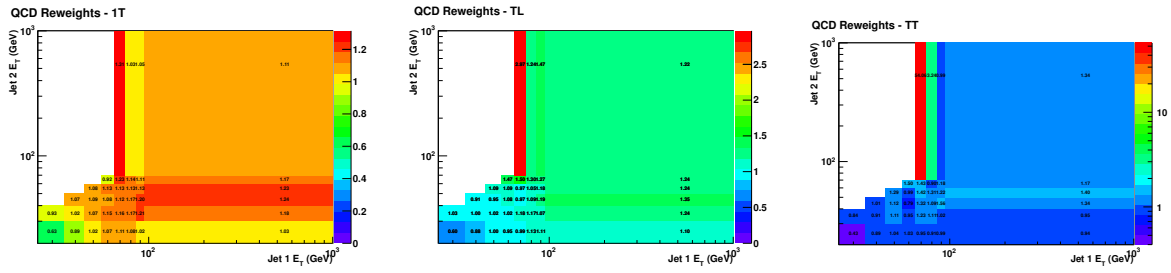


Figure A.3.: QCD jet-E_T reweighting correction factors for each tag category. The bin with a weight of 54 corresponds to a very low-statistics region.

A.3 Electroweak Region

1T	TL	TT

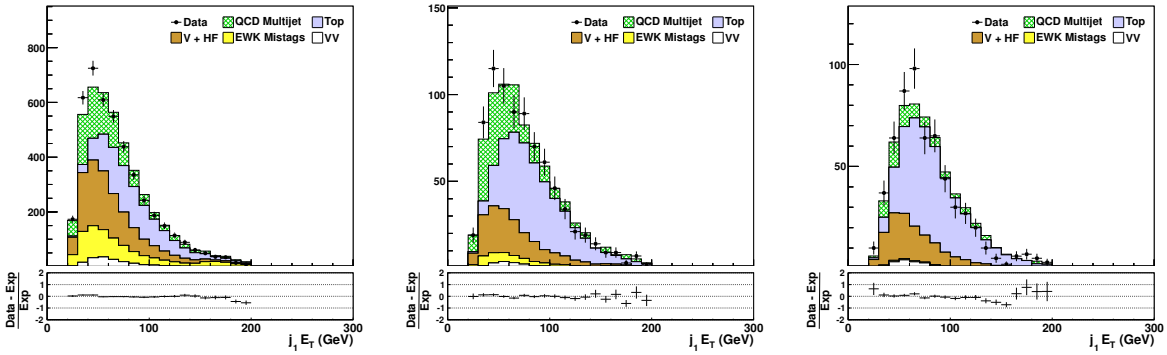


Figure A.4.: Comparisons between data and modeling of jet-1 E_T in the electroweak region.

1T	TL	TT

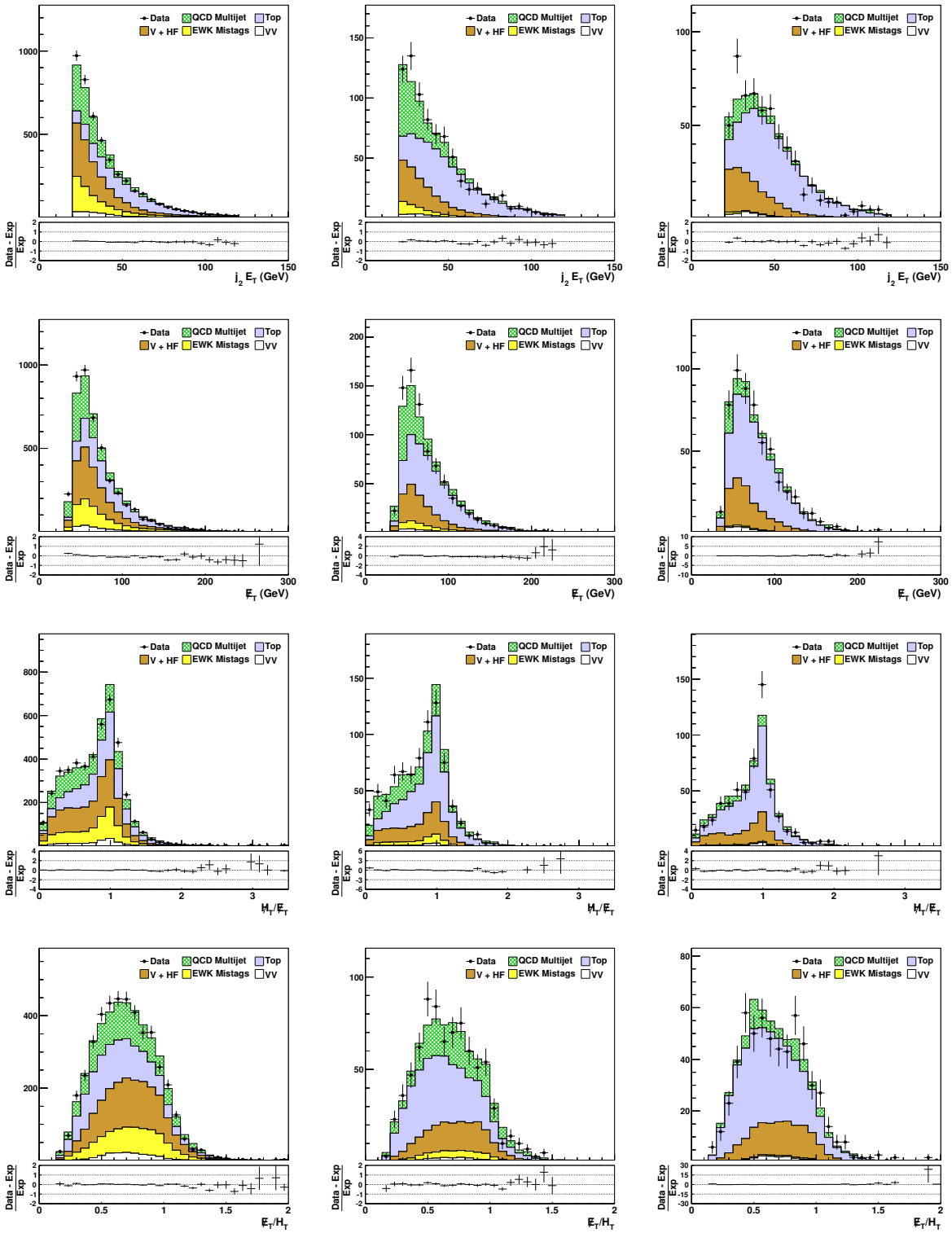


Figure A.5.: Comparisons between data and modeling of kinematic variables in the electroweak region.

1T	TL	TT

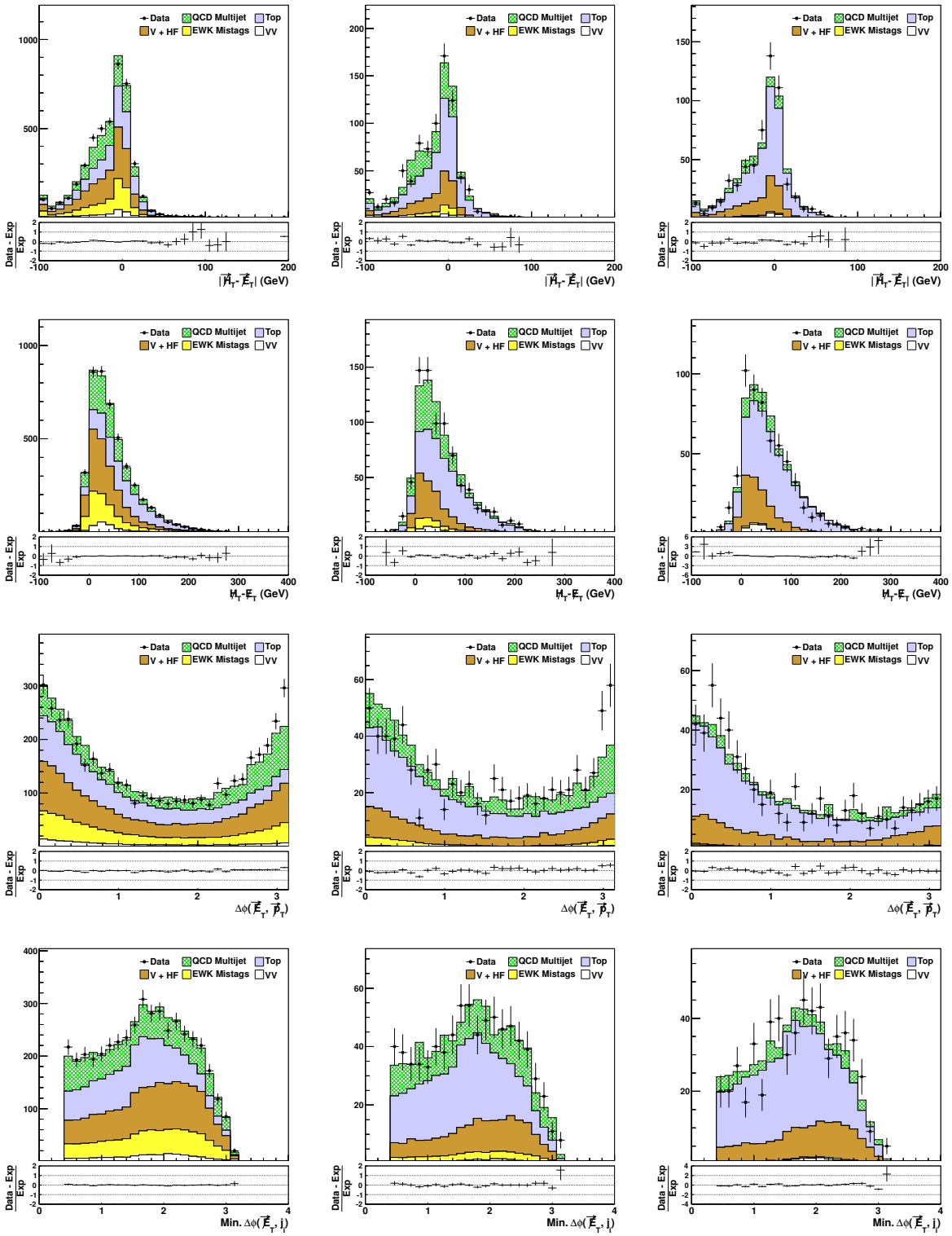


Figure A.6.: Comparisons between data and modeling of kinematic and angular variables in the electroweak region.

1T	TL	TT

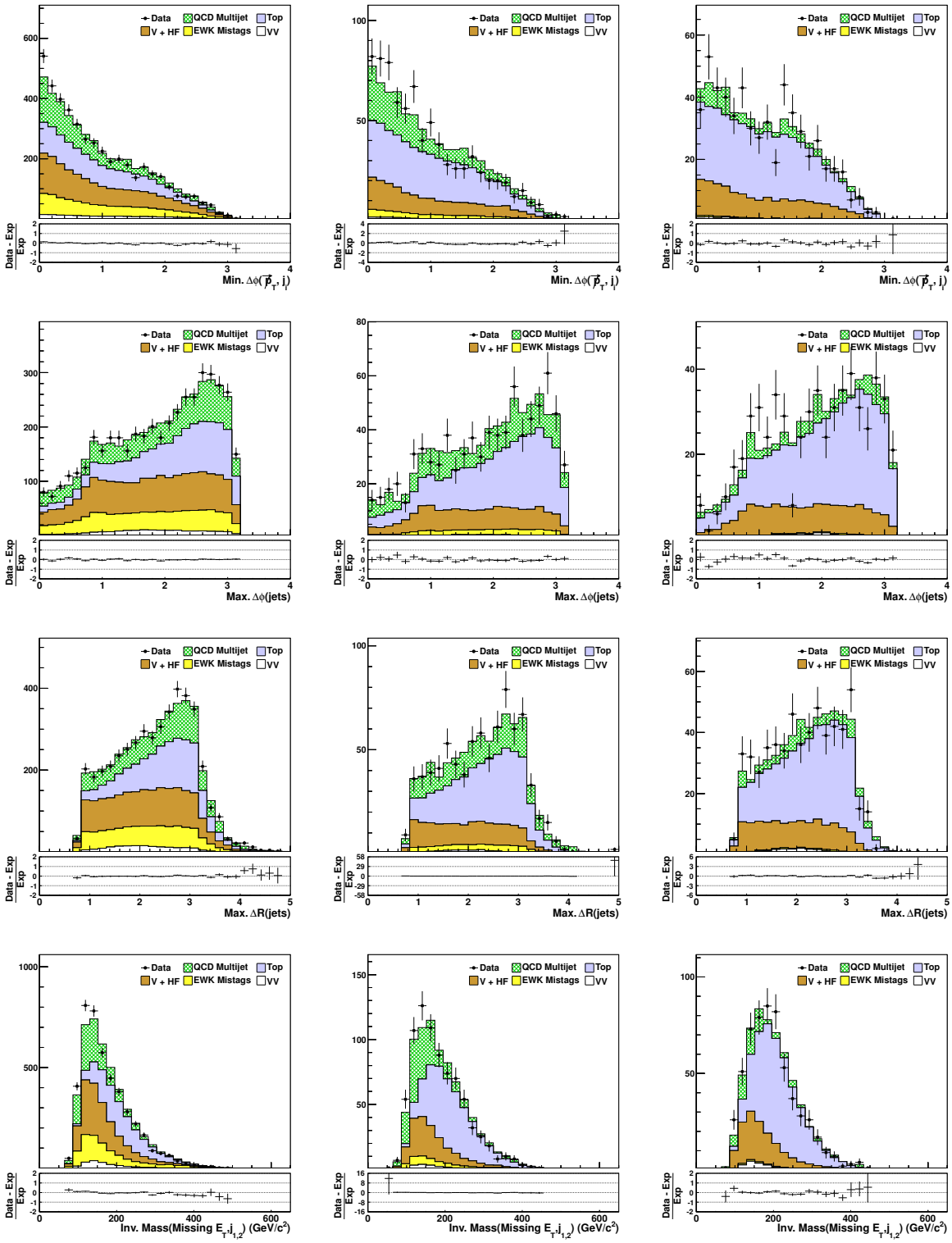


Figure A.7.: Comparisons between data and modeling of kinematic and angular variables in the electroweak region.

1T	TL	TT

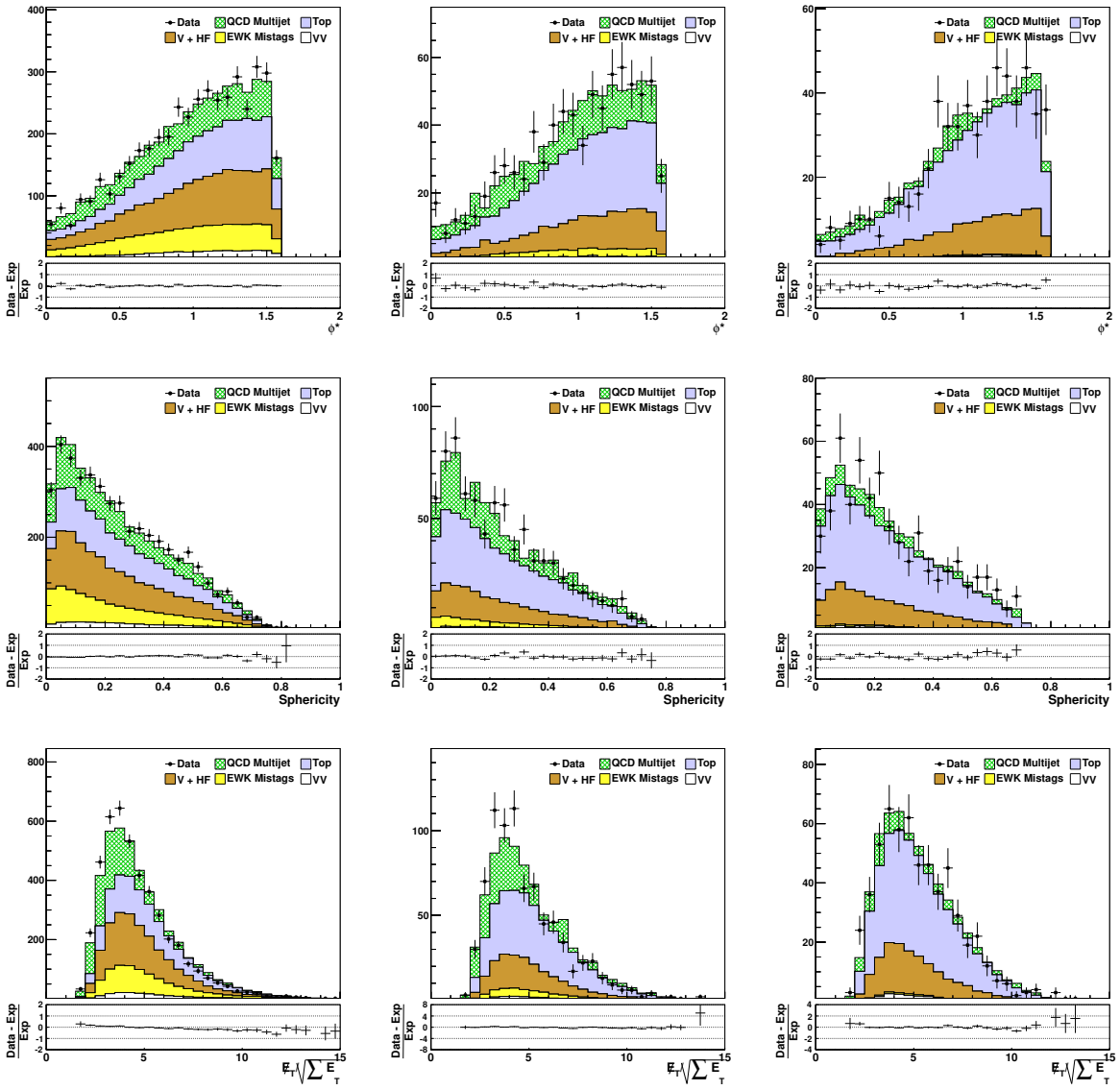


Figure A.8.: Comparisons between data and modeling of angular and topological variables in the electroweak region.

A.4 Preselection Region

1T	TL	TT

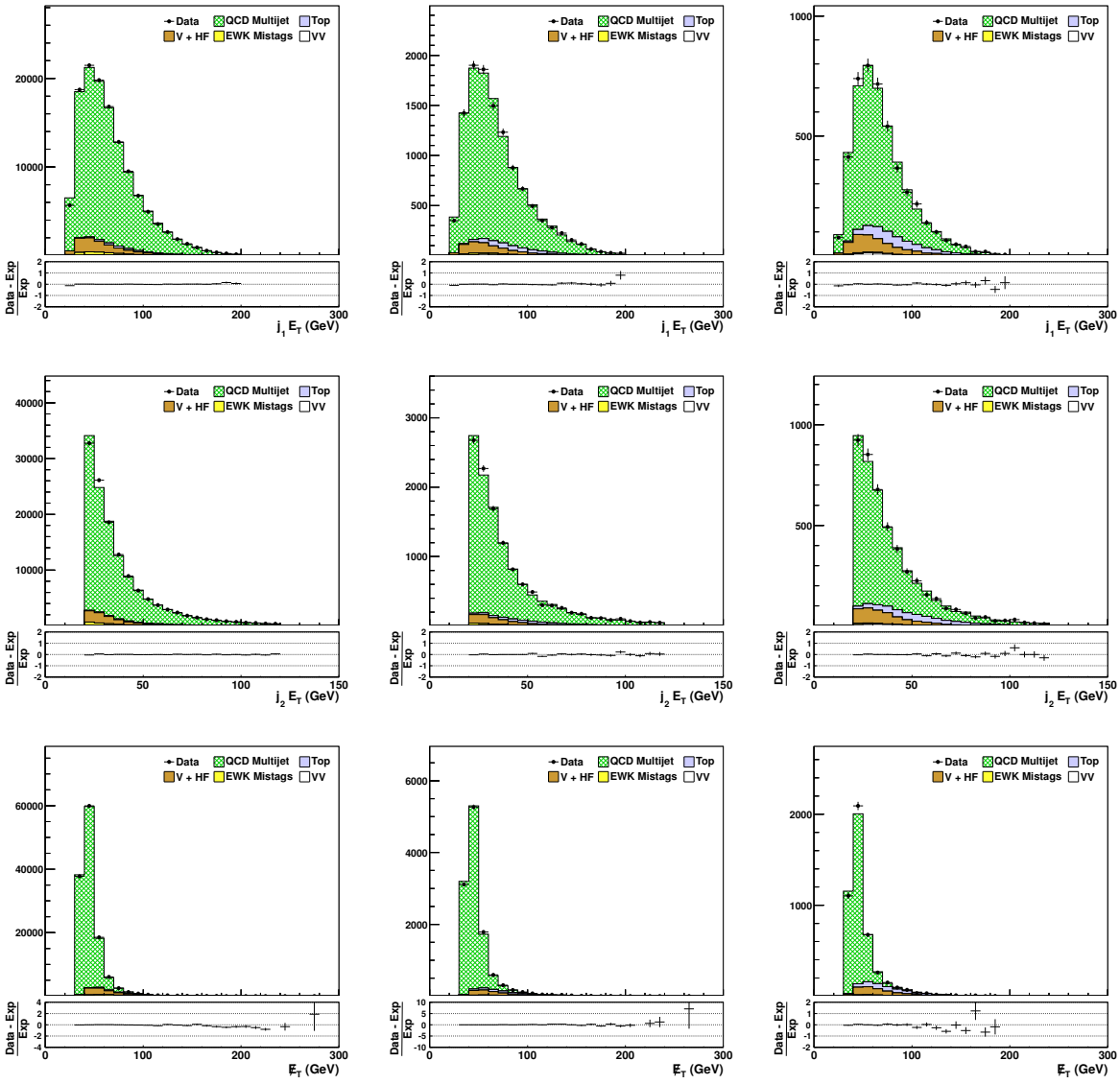


Figure A.9.: Comparisons between data and modeling of kinematic variables in the preselection region.

1T	TL	TT

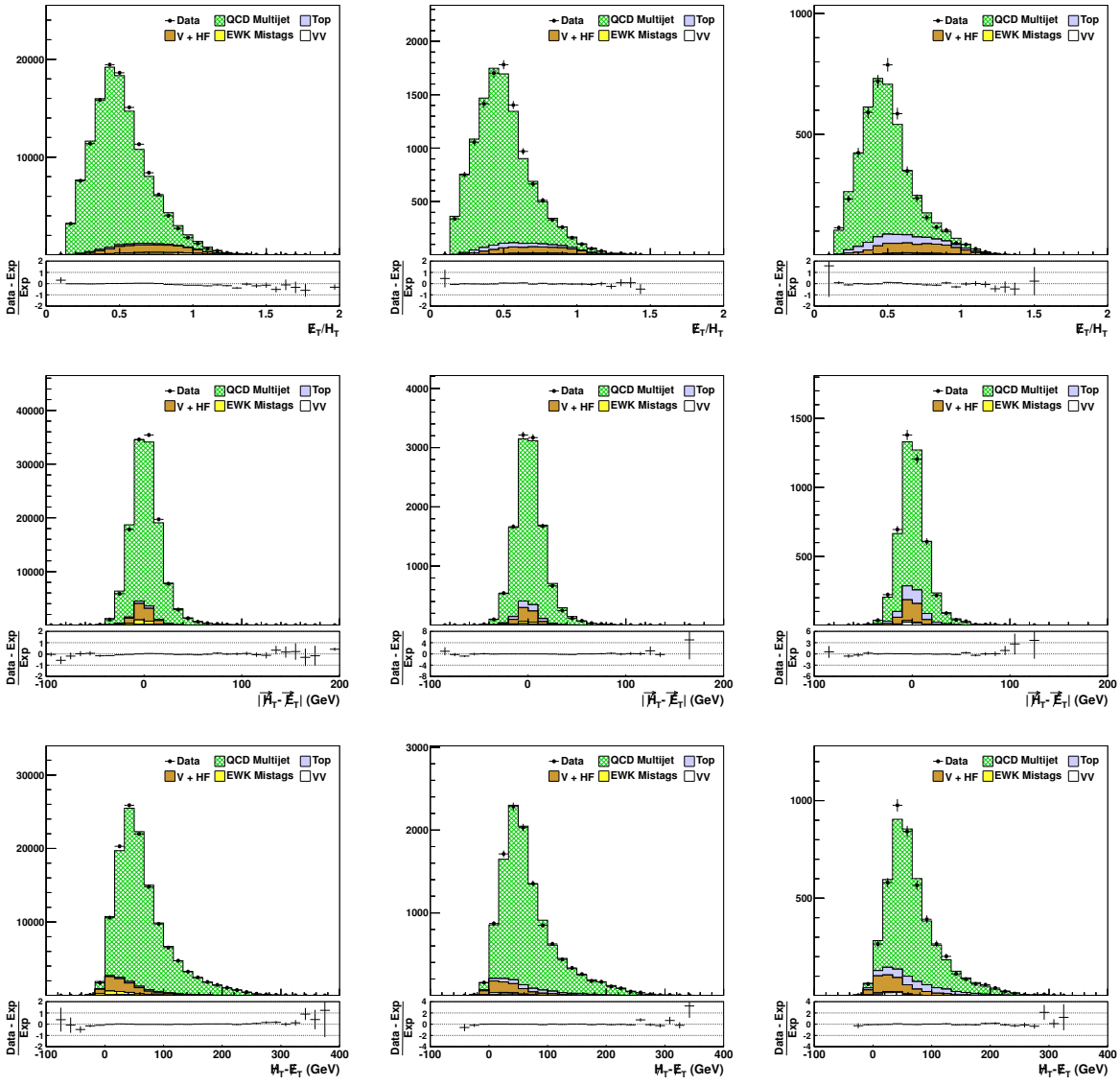


Figure A.10.: Comparisons between data and modeling of kinematic variables in the preselection region.

1T	TL	TT

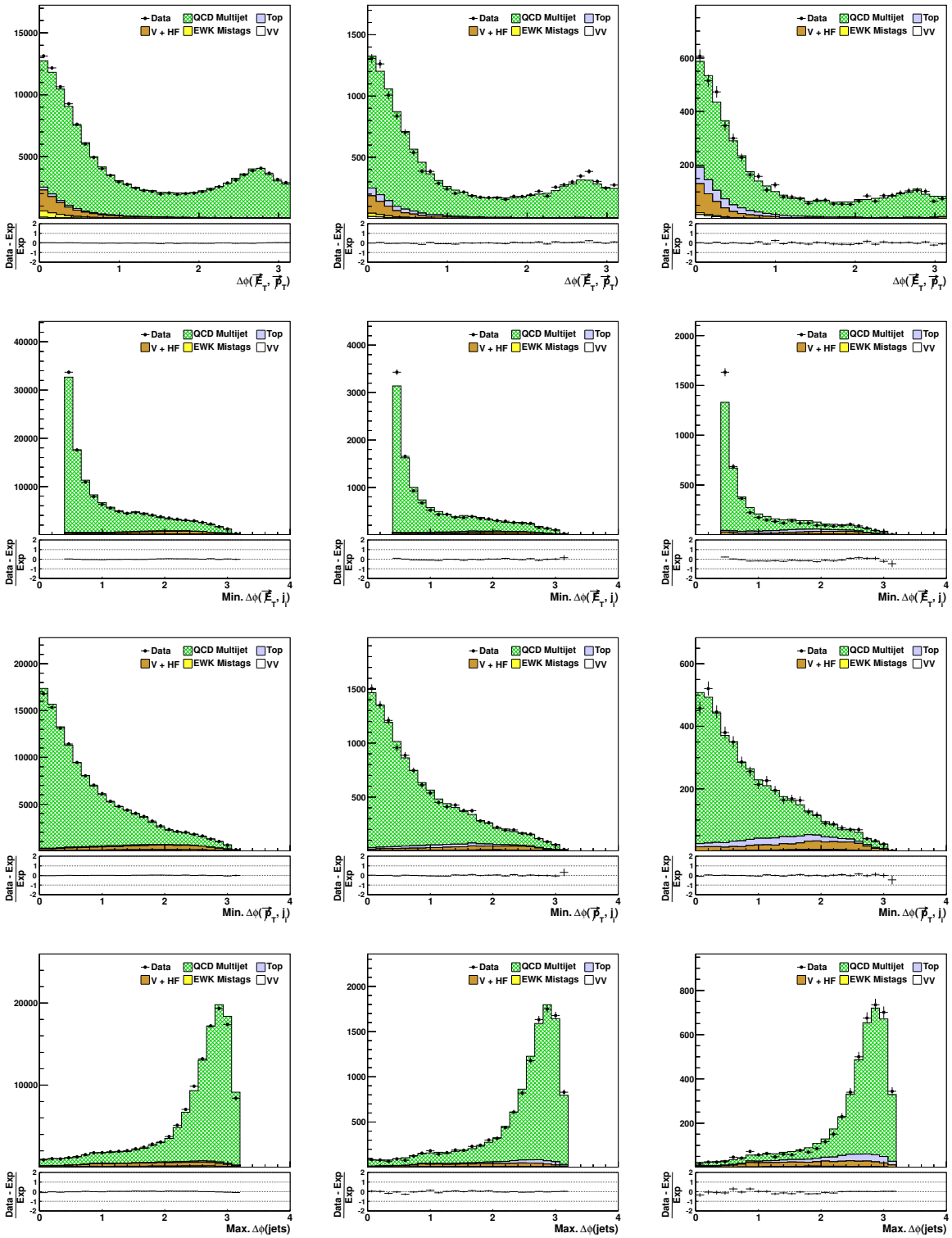


Figure A.11.: Comparisons between data and modeling of angular variables in the preselection region.

1T	TL	TT

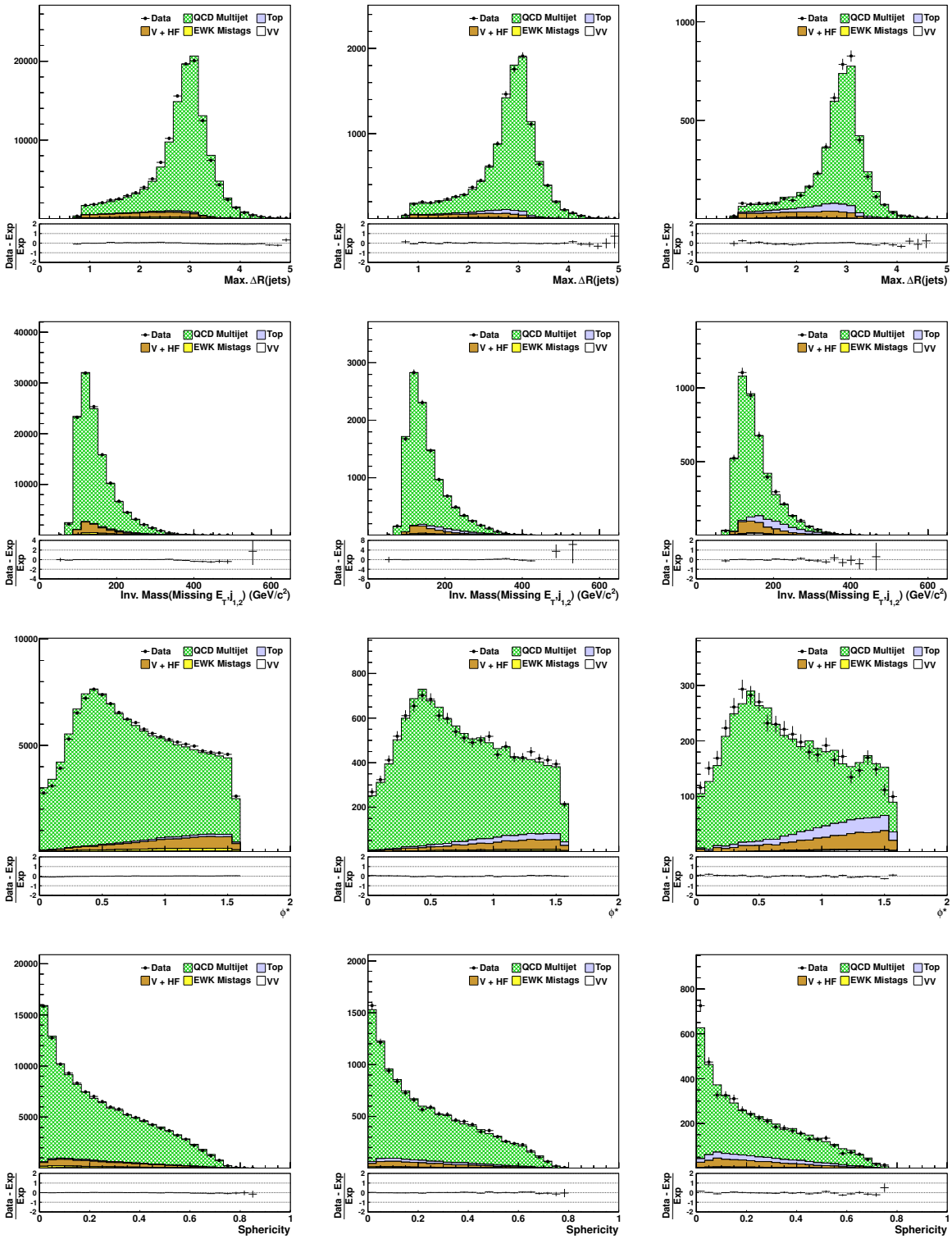


Figure A.12.: Comparisons between data and modeling of kinematic, angular, and topological variables in the preselection region.

1T	TL	TT

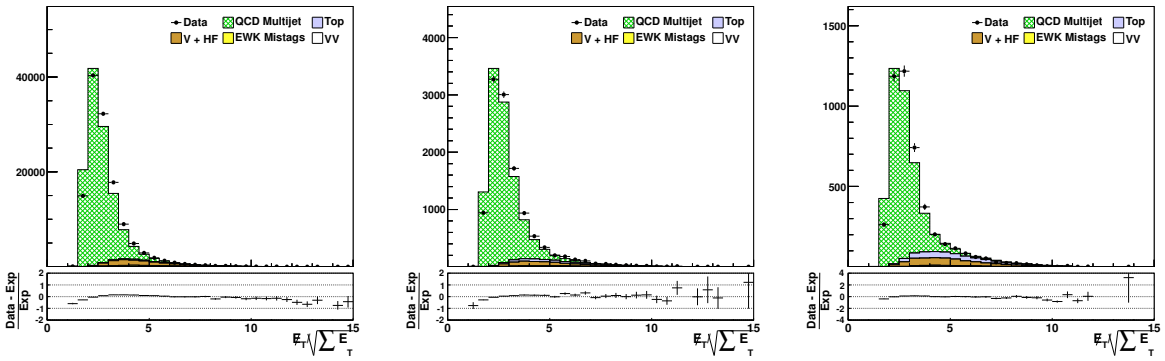


Figure A.13.: Comparisons between data and modeling of the $\not\!\!E_T$ significance in the preselection region.

B. SEARCH FOR A DIJET RESONANCE IN EVENTS WITH JETS AND MISSING TRANSVERSE ENERGY

B.1 Introduction

A study of the dijet invariant mass (m_{jj}) distribution in events with jet pairs produced in association with a W boson was recently performed by the CDF collaboration using data corresponding to an integrated luminosity of 4.3 fb^{-1} [73]. That analysis focused on leptonic W boson decays to $\ell\nu$ ($\ell = e$ or μ) , where an identified electron or muon is required in the event selection. Ref [73] reported evidence of an excess of events in the neighborhood of $m_{jj} = 145$ GeV/c^2 corresponding to 3.2 standard deviations (s.d.) with respect to the standard model (SM) expectation. In that study, the excess was modeled assuming a Gaussian distribution, centered at 145 GeV/c^2 with an rms width of 14.3 GeV/c^2, corresponding to the expected experimental m_{jj} resolution of the CDF detector. The acceptance and selection efficiencies for events associated with such a resonance were estimated by simulating Higgs boson (H) production in association with a W boson for a Higgs boson mass of 150 GeV/c^2. Assuming the excess originated from particle X with a branching fraction to quarks, $BR(X \to q\bar{q})$, of one, an estimated production cross section of $\sigma(p\bar{p} \to WX) = 3.1 \pm 0.8$ pb was obtained.

Many theoretical models have been proposed to explain this excess. Among them, a Z' model [74] and a technicolor model [75] are relevant to the study in this letter. In these models, a hypothetical particle can be produced in association with either a W boson or a Z boson. While studies on WX production are presented in Ref [73] and [76, 77], no studies focusing on ZX production have been reported.

In this article, we present a search for a dijet resonance in the m_{jj} spectrum associated with a potential new particle by studying the invariant mass distribution

from the two jets with highest transverse energy (E_T) in events with only two or three jets and large missing transverse energy ($\not\!\!E_T$). We use the entire CDF II data corresponding to an integrated luminosity of 9.1 fb^{-1}. We also veto events containing one or more identified high p_T leptons. The resulting final states are sensitive to $WX \to \not\!\ell\nu jj$ and $ZX \to \nu\bar{\nu}jj$ production and decay, where $\not\!\ell$ represents a hadronically decaying τ or an unidentified e or μ. The search for WX and ZX production is analogous to the search for WH and ZH production [85] which has comparable sensitivity to the WH search in the lepton plus jets final state [78] but is based on a completely orthogonal data sample.

B.2 Data Sample and Event Preselection

The data are collected by CDF II [24], a general-purpose detector used to study Tevatron $p\bar{p}$ collisions at a center-of-mass energy of 1.96 TeV. CDF II contains a charged particle tracking system consisting of a cylindrical open-cell drift chamber and silicon microstrip detectors immersed in a 1.4 T magnetic field parallel to the beam axis. Electromagnetic and hadronic calorimeters surrounding the tracking system measure charged and neutral particle energies. Drift chambers and muon scintillators located outside the calorimeter identify muons.

We consider events selected online based on the presence of two calorimeter clusters with $E_T > 3$ GeV and $\not\!\!E_T > 30$ GeV. We also consider events that satisfy the inclusive online requirement of $\not\!\!E_T > 45$ GeV. Jets are reconstructed using the JETCLU algorithm [79] with a clustering radius of 0.4 in azimuth-pseudorapidity space (ϕ, η) [80]. Jet energies are corrected [23] for nonuniformities of the calorimeter response as a function of η, energy contributions from multiple $p\bar{p}$ interactions within the event, and the nonlinear response of the calorimeter.

In order to retain only events for which the online selection is fully efficient, we select those with two or three energetic jets and $\not\!\!E_T > 50$ GeV. We require the two leading jets to have $E_T > 35$ GeV and $E_T > 25$ GeV, respectively, and $|\eta(j_i)| < 2$,

where one satisfies $|\eta(j_i)| < 0.9$. We require the jets to be separated by $\Delta R(j_1, j_2) > 1$. By considering events containing a third jet with $E_T > 15$ GeV and $|\eta| < 2.4$, we accept additional signal events containing a jet from initial- or final-state radiation or a hadronically decaying τ in the final state. We reject events with an identified electron or muon with $p_T > 20$ GeV/c.

B.3 Background Modeling

We model SM background processes using a variety of Monte Carlo (MC) simulation programs. The diboson processes (WW, WZ and ZZ) are generated with PYTHIA [42] incorporating $\gamma^\star$ contributions to the Z boson components for masses above 2 GeV/c^2. The normalization of simulated samples is obtained from next-to-leading order calculations [61, 62] restricting $\gamma^\star/Z$ components to the mass range between 75 and 105 GeV/c^2. The cross section values are 11.3 pb for WW, 3.2 pb for WZ and 1.2 pb for ZZ. Top-quark production is generated by assuming a top-quark mass of 172.5 GeV/c^2 [81]. Top-quark pair production is generated with PYTHIA, and its contribution is normalized to the approximate next-to-next-to-leading order cross section [63]. Single top-quark production, both s- and t-channel, is modeled using POWHEG [82] and normalized to NLO cross sections [65, 66]. W/Z production in association with parton jets is modeled by ALPGEN [60] incorporating PYTHIA to simulate parton showering and hadronization. The normalization of the W/Z plus jets background is obtained from data and discussed in more detail below.

We model events from QCD multijet production, a major source of background in final states with jets and $\not\!\!E_T$, using a data-driven method. We define missing transverse momentum $\not\!\!\vec{p}_T$, a variable similar to $\vec{\not\!\!E}_T$, as the negative vector sum of charged particle momenta from the reconstructed tracks in an event. As shown in Fig. B.1, $\vec{\not\!\!E}_T$ and $\not\!\!\vec{p}_T$ are aligned in processes with neutrinos in the final state, such as diboson production, but aligned or anti-aligned in the data, which is dominated by QCD multijet production. In QCD multijet events, $\vec{\not\!\!E}_T$ originates from jet energy

mismeasurements and tends to align with the sub-leading jet. However, as the amount of summed track momentum ($\sum p_T$) within the jet cone is somewhat random, either the leading or sub-leading jet can have greater amount of $\sum p_T$. The distribution of the angular separation between the $\vec{\not{E}}_T$ and $\vec{\not{p}}_T$ in QCD multijet events thus peaks in the regions near 0 and π. Based on this behavior, one can suppress QCD multijet background by rejecting data events where $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T) > \pi/2$, and moreover use the rejected events to model the QCD multijet background contained within the selected data sample defined by $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T) < \pi/2$. The applicability of this model is confirmed in data control regions and more details can be found in Ref. [35]. Examples of other analyses based on this technique can be found in Refs. [**?**, 89, 91].

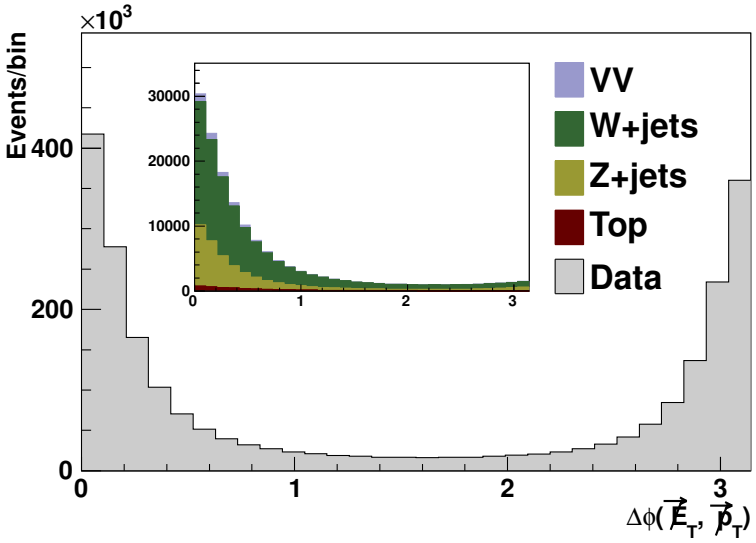


Figure B.1.: The $\Delta\phi(\vec{\not{E}}_T, \vec{\not{p}}_T)$ distribution for events that satisfy the preselection requirements. Data, for which 94% of events are estimated to originate from QCD multijet production, have $\vec{\not{E}}_T$ and $\vec{\not{p}}_T$ either aligned or anti-aligned, whereas productions with real neutrinos always have $\vec{\not{E}}_T$ and $\vec{\not{p}}_T$ aligned.

B.4 Analysis Method

To search for the presence of an anomalous resonance within the dijet invariant mass distribution, we fit the observed distribution in data to the modeled distributions for each contributing background processes. Such a resonance would appear as an additional feature within the distribution on top of the expected resonance from SM diboson production. As a first step, we extract a measurement of diboson production by fitting the dijet invariant mass distribution for the relative event contributions from known SM processes and compare the result with theoretical predictions. We then allow for an additional modeled contribution from WX and ZX production and set 95% credibility level (CL) upper limits on the cross section for such production using different theoretical constructs. The fits used to extract cross sections and upper limits are based on the Bayesian marginal likelihood method [83].

The event selection described in previous section yields over 2 million candidate events, of which 94% are estimated to originate from QCD multijet production. To reduce this background contribution, we require azimuthal separation between the direction of the $\not\!\!E_T$ and sub-leading jets, $\Delta\phi(\vec{\not\!\!E}_T, j_i) > 0.8$. We also require $\not\!\!p_T > 20$ GeV and large $\not\!\!E_T$ significance ($\not\!\!E_T/\sqrt{\sum E_T} > 3.5$ GeV$^{1/2}$, where $\sum E_T$ is the scalar sum of transverse energies deposited in the calorimeter), as well as $\not\!\!H_T/\not\!\!E_T < 1.2$, where $\not\!\!H_T$ is the magnitude of negative vector sum of jet transverse energies. These selections reduce the QCD multijet background by more than 99%. In terms of the sensitivity to diboson production, $S/\sqrt{B}$ increases to 11.7 from 3.3 after the application of the additional selection criteria, where S is the predicted number of SM diboson events and B is the predicted number of events from other SM processes.

Within the final fits, contributions from top quark production are constrained based on theoretical predictions. Initial normalizations for the W/Z plus jets and the QCD multijet background contributions are obtained by fitting the $\not\!\!E_T$ distribution, which provides good discrimination between signal- and background-like processes, using a χ^2 minimization technique. Figure B.2 shows the fitted $\not\!\!E_T$ distribution, where

the W/Z plus jets and QCD multijet contributions are initially treated as unconstrained and extracted from the fit. The resulting uncertainties on the QCD multijet and W/Z plus jets background contributions originating from this procedure are 19% and 30%, respectively.

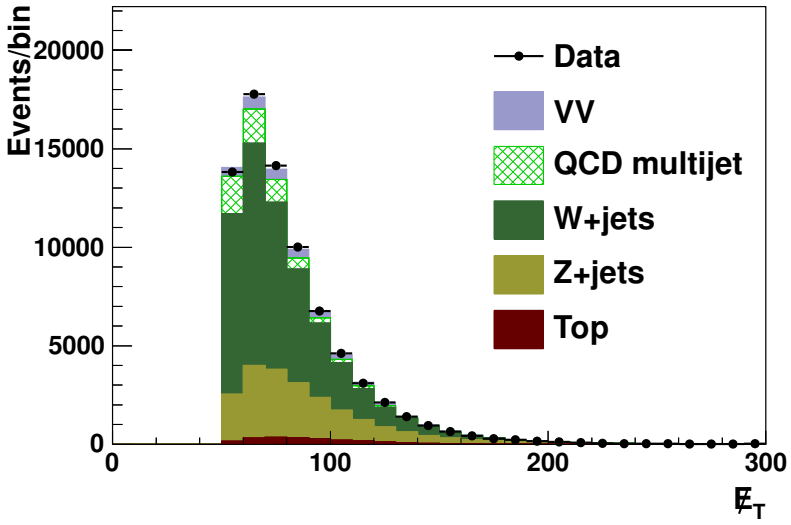


Figure B.2.: The $\not\!\!E_T$ distribution for events that satisfy the signal region definition. It is used to determine the normalization of the W/Z+jets and the QCD multijet backgrounds.

The estimated event contributions from diboson production and other SM backgrounds to the data sample, which are used as inputs to the final fits made to the dijet invariant mass spectrum, are shown in Table B.1. Figure B.3 compares the modeled distributions for leading jet E_T, sub-leading jet E_T, and $\Delta\phi(j_1, j_2)$, variables highly correlated to the dijet invariant mass, with the observed distributions in data.

When performing the Bayesian maximum likelihood fits, we consider several sources of systematic uncertainties. Dominant systematic uncertainties include those on the initial normalizations of the QCD multijet (19%) and other SM background (6.5 - 30%) contributions. Uncertainties associated with the jet energy scale (JES) [**?**] (1.4 - 12.9%), luminosity measurement [87] (5%), parton density functions (2%), lepton veto (2%), and trigger efficiency (0.4 - 1.5%) are also incorporated. These system-

Table B.1: Expected number of events from each contributing SM process in final data sample and total number of observed events. Errors include statistical and systematic uncertainties.

Process	Yield
WW	2058 ± 184
WZ	732 ± 66
ZZ	383 ± 34
Top	2197 ± 204
W+jets	45530 ± 13989
Z+jets	19765 ± 6073
QCD multijet	6155 ± 1170
Total expected	76820 ± 15302
Data	76861

atic sources affect the normalization of modeled background contributions and their effects are referred to as rate uncertainties.

We have also incorporated systematic uncertainties that result in variations in the shape of the modeled m_{jj} distributions of contributing background processes. For processes modeled via simulation, we include $\pm 1\sigma$ variations of the jet energy scale as shape uncertainties. In case of W/Z plus jets background contribution, we also include shape uncertainties resulting from changes to the Q^2 scale, a parameter in the perturbative expansion used to for calculating the matrix elements in the ALPGEN generator. Finally, for the modeled m_{jj} distribution from QCD multijet production, we obtain shape uncertainties by varying the normalization of the modeled contributions from other processes, which are subtracted from the data distribution obtained using events using $\Delta\phi(\not\!\!E_T, \not\!\!p_T) > \pi/2$.

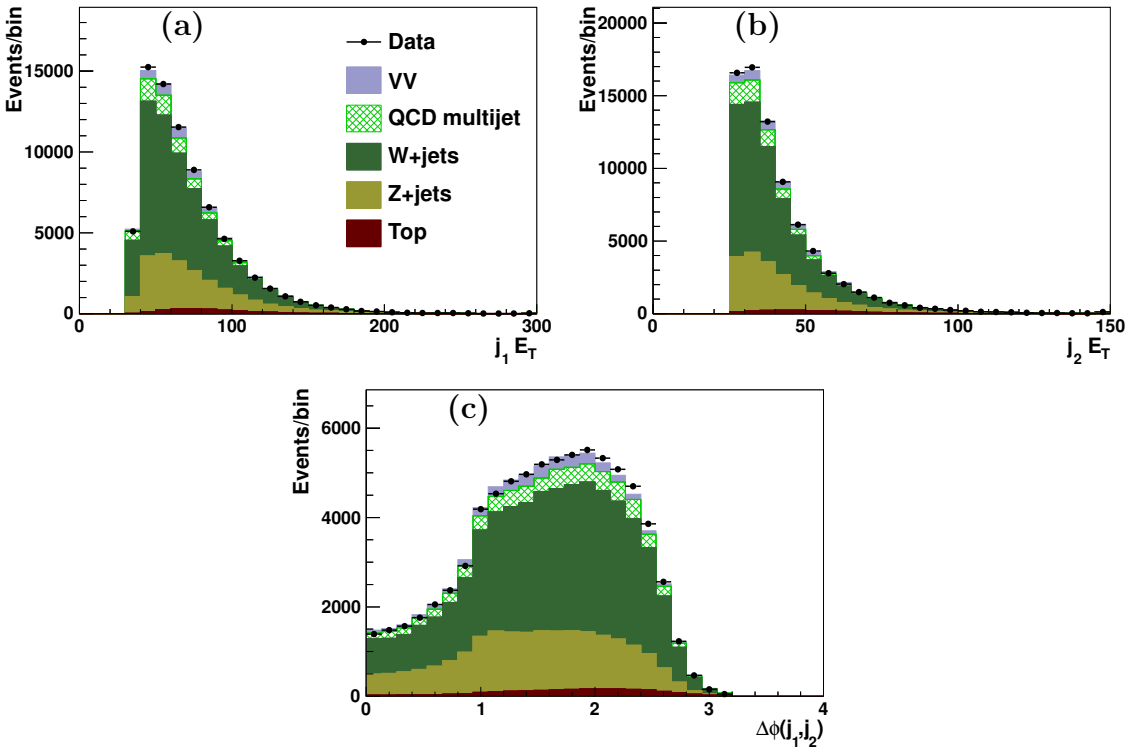


Figure B.3.: Distributions for (a) leading jet E_T, (b) sub-leading jet E_T and (c) $\Delta\phi(j_1, j_2)$ in the signal region.

B.5 Results

B.5.1 Diboson Measurement

We fit the invariant mass distribution obtained from the two leading jets contained within the events in our sample to extract a measurement of cross section for diboson production. When performing the Bayesian maximum likelihood fit, the diboson contribution are allowed float freely by assuming a flat prior for their cross sections. Figure B.4 shows the fitted m_{jj} distribution and a comparison of the fitted diboson contribution against the data with the other fitted background contributions having been subtracted. The cross section $\sigma(p\bar{p} \rightarrow WW,\ WZ,\ ZZ)$ is measured to be $11.7^{+2.8}_{-2.8}$ pb, which is in good agreement with the SM prediction of 15.8 ± 1.0 pb.

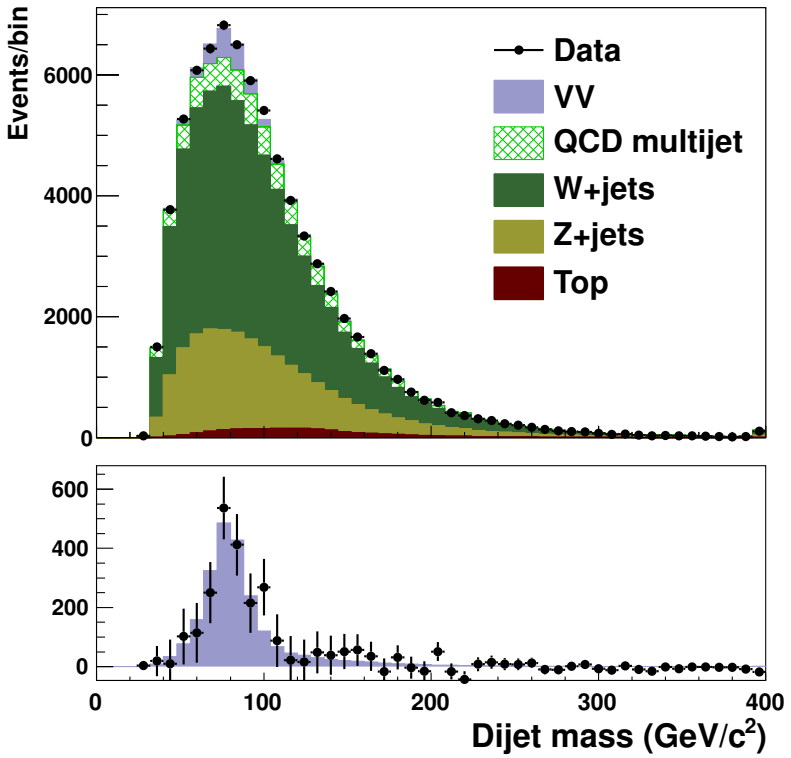


Figure B.4.: The m_{jj} distribution, as a result of post-maximum likelihood fitting, for events that satisfy the signal region definition. Top: Comparison between data and fitted signal and background. Bottom: Comparison of the fitted diboson signal (filled histogram) and the background-subtracted data (points).

B.5.2 Limits on Dijet Mass Resonance Cross Sections

To search for anomalous WX and ZX production, we perform a second fit, normalizing diboson contributions to their theoretical predictions and assuming a 6% uncertainty to the theoretical cross sections. We allow for an additional signal contribution, modeled assuming a Gaussian distribution, centered at 145 GeV/c^2 with an rms width of 14.3 GeV/c^2, in accordance with Ref. [73]. To extract cross section limits, we model signal acceptance from simulated Higgs boson production in association with a W or Z boson for a Higgs boson mass of 150 GeV/c^2. As the relative composition of WX and ZX production varies between theoretical models, we set

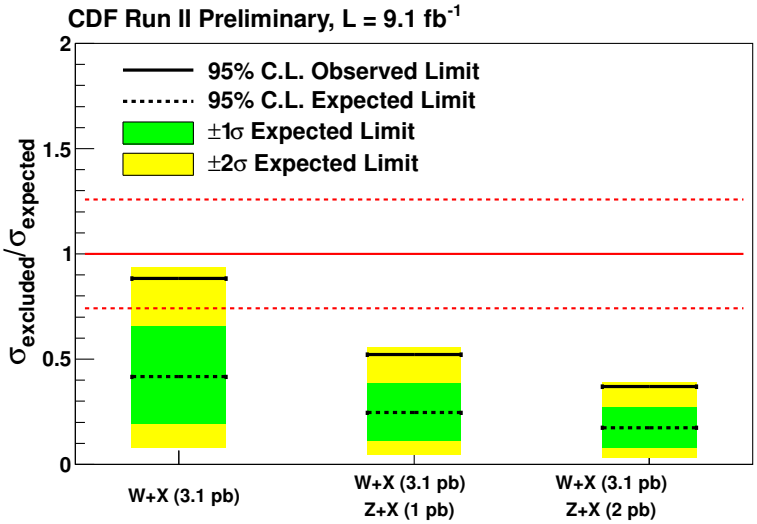


Figure B.5.: Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on cross sections of three signal scenarios divided by the expected cross sections. The red lines are for the expected cross sections (solid) and their undertainties (dash).

upper limits on the combined cross section for three scenarios: (1) $\sigma_{WX} = 3.1$ pb and $\sigma_{ZX} = 0$ pb, (2) $\sigma_{WX} = 3.1$ pb and $\sigma_{ZX} = 1$ pb, and (3) $\sigma_{WX} = 3.1$ pb and $\sigma_{ZX} = 2$ pb. The cross section limits obtained from the fit are presented in Tab. B.2 and Fig. B.5.

Table B.2: Expected and observed 95% C.L. upper limits on the combined WX+ZX production cross section under different hypotheses for the ratio of σ_{WX}/σ_{ZX}.

Signal scenarios	Expected upper limits	Observed upper limits
$\sigma_{WX} = 3.1$ pb and $\sigma_{ZX} = 0$ pb	1.29 pb	2.73 pb
$\sigma_{WX} = 3.1$ pb and $\sigma_{ZX} = 1$ pb	1.01 pb	2.14 pb
$\sigma_{WX} = 3.1$ pb and $\sigma_{ZX} = 2$ pb	0.89 pb	1.89 pb

B.6 Conclusion

We have studied the dijet invariant mass distribution in events with energetic jets and large missing transverse energy using the full CDF II data set corresponding to an integrated luminosity of 9.1 fb^{-1}. We find good agreement between data and the standard model expectation and measure a cross section of $\sigma(p\bar{p} \to WW,\ WZ,\ ZZ) = 11.7^{+2.8}_{-2.8}$ pb for diboson production. In the absence of a significant excess over background expectations in m_{jj} spectrum, we set 95% C.L. upper limits on the combined cross section for production of a new particle X in association with a W or Z boson, under several hypotheses for the ratio of σ_{WX}/σ_{ZX}.

VITA

VITA

Qiuguang Liu was born in Hunan, China on Sepetember 11, 1984. He completed his high school education at No. 1 Yueyang school in Yueyang county in 2001. He attended Sichuan University between Fall of 2001 and Spring 2005 where he earned his B.S. in Physics. Subsequently, he entered the graduate school at the Chinese Academy of Science and studied at the Institute of High Energy Physics to pursue the Master degree, which he received in Spring 2008. In the Fall of 2008 he entered the graduate program at Purdue University. He joined the CDF Collaboration at Fermilab in May 2009 under the supervision of Daniela Bortoletto. He was a URA Visiting Scholar at Fermilab during 2012-2013. He completed his Ph.D. in May 2013.