A Search for the Higgs Boson in Proton – Antiproton Collisions at $\sqrt{s} = 1.8\ TeV$

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Christopher Carl Neu, B.S., M.S.

* * * * *

The Ohio State University

2003

Dissertation Committee:

Prof. Brian L. Winer, Adviser

Prof. Richard E. Hughes

Prof. Eric Braaten

Prof. Linn Van Woerkem

Approved by

Adviser
Department of Physics

ABSTRACT

Although the Standard Model of fundamental particles and their interactions has enjoyed much success over the past quarter-century, a portion of the theory has eluded experimental verification. Mass is a manifest quality of the constituents of our universe; it remains to be understood however in the context of the Standard Model why the fundamental particles have the masses they possess. The imposition of mass is a consequence of the breaking of the symmetry that unifies the weak and electromagnetic forces; electroweak symmetry breaking is accomplished in the Standard Model via the Higgs mechanism. This elegant portion of the theory not only provides the dynamics for the symmetry breaking, but also predicts a physically observable scalar particle, the Higgs boson, which is yet to be discovered.

A new search for the Higgs boson has been performed in the proton – antiproton collisions at center-of-mass energy of 1.8 TeV provided by the Tevatron accelerator at Fermi National Accelerator Laboratory. In this analysis, a neural network was utilized to aid in the rejection of collision events that share the equivalent signature as Higgs events but are produced via other, less interesting production mechanisms. The neural network was implemented as part of an advanced event selection that in simulation studies was shown to provide a 34% increase in signal sensitivity over conventional methods. When the technique is applied to the data collected by the CDF collaboration during the Tevatron's Run 1 (1992—1995), an excess of events is identified above the background expectation. The limit on the Higgs production cross

section is calculated for six Higgs mass hypotheses in the range 100 $GeV/c^2 < M_H < 150$ GeV/c^2. The WH production cross section upper limit was determined to be 18–22 pb in the range $M_H < 130$ GeV/c^2 at 95% confidence; the limit in the range $M_H > 130$ GeV/c^2 is considerably larger. The measured limit is a factor of two larger than the limit from *a priori* studies. The Standard Model theory prediction is approximately two orders of magnitude lower than this upper limit in the M_H range.

ACKNOWLEDGMENTS

This thesis could not have been completed without the efforts, support and encouragement of countless people, a few of which I would like to mention.

First I would like to thank my thesis advisor Brian Winer for the invaluable guidance he has provided in my time with OSU. I have never met anyone with as many good ideas, several of which I have been taking credit for over the past few years. You have been a good friend, and there is no better person to have in one's corner. I would also like to thank Richard Hughes; your physics input has been valuable, but our daily debates — on politics, society, the NBA — have been even more enjoyable.

To my Columbus amigos, Catalin and Carlos, thank you for introducing me to Smith 2025 basketball, and for being people I could count on for a laugh at all times. I would also like to thank Evelyn, Phil and Jong-young for tolerating my incessant talk over the years about the White Sox, Bears, Bulls, or the Illini, depending on the season. And to Rob Roser, my first boss on this experiment, for always giving me great advice, usually bestowed upon me over beers that *he* bought. Thanks for all the words of wisdom.

Special thanks to Michelangelo Mangano for his patient responses to my naive questions; to Avi Yagil for countless helpful suggestions; and to John Conway for your support and enthusiasm.

To my parents, thank you for the unwavering support over the years and encouraging me in all my endeavors.

Last but certainly not least, I would like to thank Laura, my best friend and the love of my life, without whom none of what I have done would be possible.

VITA

15 August 1976	Born – Oak Lawn, IL
May 1998	B.S. in Physics, B.S. in Mathematics University of Illinois, Urbana, IL
June 1998 – May 1999	Fowler Fellowship The Ohio State University, Department of Physics, Columbus, OH
May 1999 – June 2002	Graduate Research Associate The Ohio State University, Columbus, OH
February 2001	M.S. in Physics The Ohio State University, Columbus, OH
July 2002 – June 2003	Presidential Fellow The Ohio State University, Columbus, OH

PUBLICATIONS

The CDF Collaboration, "Search for a W′ Boson Decaying to a Top and Bottom Quark Pair in 1.8 TeV $p\bar{p}$ Collisions", *Phys. Rev. Lett.* 90, 081802 (2003).

The CDF Collaboration, "Momentum Distribution of Charged Particles in Jets in Dijet Events in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV and Comparisons to Perturbative QCD Predictions", *Phys. Rev. D* 61, 012003 (2003). FERMILAB-PUB-02/096-E.

The CDF Collaboration, "Search for Long-lived Charged Massive Particles in $\bar{p}p$ Collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. Lett.* 90, 131801 (2003).

The CDF Collaboration, "Search for Associated Production of Upsilon and Vector Boson in $p\bar{p}$ at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. Lett.* 90, 221803 (2003).

The CDF Collaboration, "Search for Supersymmetric Partner of the Top Quark in Dilepton Events from $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. Lett.* 90, 251801 (2003).

The CDF Collaboration, "Central Pseudorapidity Gaps in Events with a Leading Antiproton at the Fermilab Tevatron $\bar{p}p$ Collider", submitted to *Phys. Rev. Lett.* February 15, 2003. FERMILAB-PUB-03/049-E.

The CDF Collaboration, "Measurement of the Mass Difference $m(D_s^+)$-$m(D^+)$ at CDF II" submitted to *Phys. Rev. D* March 19, 2003. FERMILAB-PUB-03/048-E.

The CDF Collaboration, "Search for Pair Production of Scalar Top Quarks in R-parity Violating Decay Modes in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV", submitted to *Phys. Rev. Lett.* May 7, 2003. FERMILAB-PUB-03/070-E.

The CDF Collaboration, "Search for Lepton Flavor Violating Decays of a Heavy Neutral Particle in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV", submitted to *Phys. Rev. Lett.* June 20, 2003. FERMILAB-PUB-03/203-E.

The CDF Collaboration, "Mesurement of Prompt Charm Meson Production Cross Sections in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV", submitted to *Phys. Rev. Lett.* July 28, 2003. FERMILAB-PUB-03/217-E.

The CDF Collaboration, "Search for new physics in photon-lepton events in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 89, 041802 (2002).

The CDF Collaboration, "Υ production and polarization in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 88, 161802 (2002).

The CDF Collaboration, "Diffractive dijet production at $\sqrt{s} = 630\ GeV$ and $1800\ GeV$ at the Fermilab Tevatron", *Phys. Rev. Lett.* 88, 151802 (2002).

The CDF Collaboration, "Search for new heavy particles in the WZ^0 final state in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 88, 071806 (2002).

The CDF Collaboration, "Study of $B^0 \to J/\psi K^{(*)0}\pi^+\pi^-$ decays with the Collider Detector at Fermilab", *Phys. Rev. Lett.* 88, 071801 (2002).

The CDF Collaboration, "Measurement of the strong coupling constant from inclusive jet production at the Tevatron $p\bar{p}$ collider", *Phys. Rev. Lett.* 88, 042001 (2002).

The CDF Collaboration, "Search for gluinos and scalar quarks in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$ using the missing energy plus multijets signature", *Phys. Rev. Lett.* 88, 041801 (2002).

The CDF Collaboration, "Measurement of the ratio of b quark production cross-sections in $\bar{p}p$ collisions at $\sqrt{s} = 630\ GeV$ and $\sqrt{s} = 1800\ GeV$", *Phys. Rev. D* 66, 032002 (2002).

The CDF Collaboration, "Search for new physics in photon-lepton events in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 66, 012004 (2002).

The CDF Collaboration, "Comparison of the isolated direct photon cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$ and $\sqrt{s} = 0.63\ TeV$", *Phys. Rev. D* 65, 112003 (2002).

The CDF Collaboration, "Search for the decay $B_s \to \mu^+\mu^-\phi$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 65, 111101 (2002).

The CDF Collaboration, "Measurement of B-meson lifetimes using fully reconstructed B decays produced in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 65, 092009 (2002).

The CDF Collaboration, "Charged jet evolution and the underlying event in proton-antiproton collisions at $1.8\ TeV$", *Phys. Rev. D* 65, 092002 (2002).

The CDF Collaboration, "Search for single top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 65, 091102 (2002).

The CDF Collaboration, "Soft and hard interactions in $p\bar{p}$ collisions at $\sqrt{s} = 1800\,GeV$ and 630 GeV", *Phys. Rev. D* 65, 072005 (2002).

The CDF Collaboration, "Searches for new physics in events with a photon and b-quark jet at CDF", *Phys. Rev. D* 65, 052006 (2002).

The CDF Collaboration, "Study of the Heavy Flavor Content of Jets Produced in Association with W Bosons in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. D* 65,052007 (2002).

The CDF Collaboration, "Measurement of the B^+ total cross section and B^+ differential cross section $d\sigma/dp_T$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 65, 052005 (2002).

The CDF Collaboration, "Limits on Extra Dimensions and New Particle Production in the Exclusive Photon and Missing Energy Signature in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. Lett.* 89 281801 (2002).

The CDF Collaboration, "Cross Section for Forward J/ψ Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. D* 66, 092001 (2002).

The CDF Collaboration, "Search for Radiative b-Hadron Decays in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. D* 66, 1122002 (2002).

The CDF Collaboration, "Search for gluinos and squarks using like-sign dileptons in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 87, 251803 (2001)

The CDF Collaboration, "Observation of diffractive J/ψ production at the Fermilab Tevatron", *Phys. Rev. Lett.* 87, 241802 (2001).

The CDF Collaboration, "Search for quark-lepton compositeness and a heavy W' boson using the electron neutrino channel from $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 87, 231803 (2001).

The CDF Collaboration, "Charged-particle multiplicity in jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 87, 211804 (2001).

The CDF Collaboration, "Double diffraction dissociation at the Fermilab Tevatron collider", *Phys. Rev. Lett.* 87, 141802 (2001).

The CDF Collaboration, "Measurement of $d\sigma/dM$ and forward-backward charge asymmetry for high-mass Drell-Yan e^+e^- pairs from $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 87, 131802 (2001).

The CDF Collaboration, "Measurement of the top quark P_T distribution", *Phys. Rev. Lett.* 87, 102001 (2001).

The CDF Collaboration, "Search for neutral supersymmetric Higgs bosons in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 86, 4472 (2001).

The CDF Collaboration, "Production of χ_{c1} and χ_{c2} in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 86, 3963 (2001).

The CDF Collaboration, "First measurement of the ratio $B(t \to Wb)/B(t \to Wq)$ and associated limit on the Cabibbo-Kobayashi-Maskawa element V_{tb}", *Phys. Rev. Lett.* 86, 3233 (2001).

The CDF Collaboration, "Cross-section and heavy quark composition of $\gamma + \mu$ events produced in $p\bar{p}$ collisions", *Phys. Rev. D* 65, 012003 (2001).

The CDF Collaboration, "Search for narrow diphoton resonances and for $\gamma\gamma + W/Z$ signatures in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 64, 092002 (2001).

The CDF Collaboration, "Observation of orbitally excited B mesons in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 64, 072002 (2001).

The CDF Collaboration, "Measurement of the W boson mass with the Collider Detector at Fermilab", *Phys. Rev. D* 64, 052001 (2001).

The CDF Collaboration, "Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$" , *Phys. Rev. D* 64, 032001 (2001).

The CDF Collaboration, "Measurement of the top anti-top production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 64, 032002 (2001).

The CDF Collaboration, "Measurement of the two-jet differential cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1800\ GeV$", *Phys. Rev. D* 64, 012001 (2001).

The CDF Collaboration, "Search for the supersymmetric partner of the top quark in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 63, 091101 (2001).

The CDF Collaboration, "Tests of enhanced leading order QCD in W boson plus jets events from 1.8 TeV $p\bar{p}$ collisions", *Phys. Rev. D* 63, 072003 (2001).

The CDF Collaboration, "Measurement of the top quark mass with the Collider Detector at Fermilab", *Phys. Rev. D* 63, 032003 (2001).

The CDF Collaboration, "Measurement of $\mathrm{d}\sigma/\mathrm{d}y$ for high mass Drell-Yan e^+e^- pairs from $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 63, 011101 (2001).

S. Holm, C. Neu, *et al.*, "System architecture and hardware design of the CDF XFT online track processor", *IEEE Transactions on Nuclear Science* 47, 895 (2000).

C. Neu *et al.*, "Online track processor for the CDF upgrade", *IEEE Transactions on Nuclear Science* 46, 933 (1999).

FIELDS OF STUDY

Major Field: Physics

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Curious humans have inquired for centuries about the nature of the world around us. Of the fundamental questions that have been the focus of much thought, one is particularly cogent: What is this world made of? Early thinkers considered all things in nature to be made from combinations of four fundamental elements: earth, wind, fire and water. Considering the everyday world evident to the early Greek philosophers that formulated this theory, one must acknowledge that this is a reasonable first attempt. Over the centuries, our understanding of the universe has advanced as our technological abilities have allowed us to further probe fundamental structure. However, as our view of the world has become more sophisticated, we continue to ask that same fundamental question — What is this world made of? — and to refine our response to that question to this very day.

The foundation of contemporary understanding of the fundamental world is a theory called the Standard Model. This theory can explain the phenomena encountered in our everyday experiences as well as explain interactions manufactured under the most harsh conditions achievable: at the collision point of two beams of charged particles in massive particle colliders. The theory's predictive abilities over such a large energy range are well-tested, and no significant discrepancies between theory

and experiment have been observed. The Standard Model is the culmination of the centuries-long attempts at understanding the building blocks of the universe.

The Standard Model has been the focus of intense scrutiny over the past two decades in which the theory has been tested to very high precision [3]. The discovery [1] of the top quark and observation [2] of the tau neutrino in experiments at the Fermi National Accelerator Laboratory are the most recently discovered Standard Model particles. According to the Standard Model, these particles were guaranteed to be present, but they eluded discovery for many years. The observations of the top quark and tau neutrino were several in a series of convincing arguments for the validity of the Model.

Despite the phenomenal success of the predictive powers of the Standard Model, there is a portion of the framework for which no evidence has been discovered. In the Model, the mechanism by which the fundamental particles acquire mass — the Higgs mechanism — has thus far eluded experimental verification. It is the charge of this and subsequent generations of experiments to elucidate this portion of the theoretical framework. A primary current experimental goal is a search for the Higgs boson, a physically observable particle that is an additional consequence of the Higgs mechanism upon which the tenets of the Standard Model rely. Unfortunately the Model does not predict the mass of the Higgs, M_H; lacking a clear prediction from theory, previous direct and indirect searches considered a wide range of possible M_H values. The results of these analyses indicate that a low-mass Higgs is favored, and its mass lies within the range $115\ GeV/c^2 < M_H < 211\ GeV/c^2$ at 95% confidence [18],[14].

This document chronicles a new search for the Higgs boson in the remnants of proton-antiproton collisions manufactured by the Tevatron accelerator complex at Fermi National Accelerator Laboratory, in Batavia, IL. This search uses data collected

by the Collider Detector at Fermilab (CDF) experiment during Run 1 of the Tevatron (1992 – 1995). We search for Higgs production in association with a leptonically decaying $W^{\pm}$ boson (referred to as WH production), one of the prominent production channels at the Tevatron. Two factors make this search difficult: small expected Higgs production rates limit the number of Higgs events in the Run 1 data set; and formidable levels of other, less-interesting interactions that mimic the signature of WH events hinder attempts at identification of a purely-Higgs event sample. It is crucial under such circumstances to develop identification techniques that balance the need to retain as many Higgs events as possible while rejecting a significant portion of the events from other sources. To this end, a neural network is employed in the advanced selection in this analysis to aid in the identification of the Higgs signal.

A neural network is a crude form of artificial intelligence, implemented in software, that learns to identify events consistent with the Higgs signature via patterns in simulated proton-antiproton collisions. After sufficient training the network is then asked to identify Higgs events in the actual Run 1 dataset. The advanced selection is completed by a requirement on an event level quantity sensitive to the Higgs mass. The employment of the advanced selection utilized in this analysis was shown to provide a 34% increase in sensitivity over conventional techniques in simulation studies. Such tools will be useful in future searches in which experimenters seek to reduce the overall amount of data necessary for Higgs discovery, thus providing physics insight in a shorter amount of collider running time and ultimately saving resources.

We use this method to establish a new upper limit on Higgs production at the Tevatron for six Higgs mass hypotheses in the regime $100\ GeV/c^2 < M_H < 150\ GeV/c^2$. In the Run 1 dataset, an excess of events survive our selection process over the expectation from non-Higgs sources. The cross section limit was determined to be

18-22 pb, approximately 100 times larger than the theoretical cross section, for 100 $GeV/c^2 < M_H < 130\ GeV/c^2$.

The contents of this document are organized as follows. Chapter 2 contains a description of the theoretical aspects behind the Higgs mechanism in the Standard Model, as well as present search limits. Chapter 3 describes the Tevatron accelerator complex and CDF detector, followed by a brief introduction to neural networks in Chapter 4. Chapters 5 and 6 describe how events consistent with Higgs production were selected in this search. Chapter 7 illustrates how the event yields after applying the selection criteria can be turned into limits on Higgs production rate and the systematic errors inherent in our estimates. The result of applying this method on the Run 1 data is described in Chapter 8, and the conclusions of the analysis are documented in Chapter 9.

CHAPTER 2

THEORETICAL MOTIVATION AND PRESENT UNDERSTANDING

The current knowledge of the fundamental world is summarized in a theoretical framework called the Standard Model. The 1950s and 1960s saw a flurry of activity in the experimental particle physics community; bubble chamber experiments regularly discovered new particles. A theory was needed that could explain the existence of all of these new states; and out of this necessity, the Standard Model was born, taking much of its current shape in the late 1970s.

2.1 The Standard Model

The Standard Model accommodates matter and forces. There are four known forces in the universe:

- the strong force, responsible for, among other things, holding nucleons together
- the weak force, responsible for such processes as nuclear beta decay
- the familiar electromagnetic force, whose effect can be seen in the attraction of two oppositely charged objects
- and the force of gravity, the attractive force between two massive objects.

The Standard Model successfully joins the strong, weak and electromagnetic forces into a single framework; its main tenets postulate the existence of matter particles and force-carrying particles.

The matter particles are divided into two classes, quarks and leptons. Both quarks and leptons are spin-1/2 fermions. There are six quarks: up(u), down(d), strange(s), charm(c), bottom(b) and top(t). The quarks are electrically charged, massive, and grouped into three generations:

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} \quad \begin{matrix} Q = 2/3 \\ Q = -1/3 \end{matrix}$$

Table 2.1: The three generations of quarks in the Standard Model, and their charges.

Masses of the quarks increase from the first generation to the third; the top quark is the heaviest by far, with $m_t \simeq 175\ GeV/c^2$, more than 40 times heavier than the bottom. Quarks participate in the strong, weak and electromagnetic interactions. All but the top quark form bound states known as hadrons. For example, the first generation up and down quarks constitute the hadrons known as protons (uud) and neutrons (udd) and thus most of the matter in the world around us.

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix} \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix} \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}$$

Table 2.2: The three generations of leptons in the Standard Model.

A similar arrangement exists for the leptons (see Table 2.2), which include the electron(e), muon(μ) and tau(τ) charged leptons, and their associated neutrinos (ν_e,ν_μ,ν_τ). The muon resembles the familiar electron but with ~ 200 times the mass. The charged leptons participate in the weak and electromagnetic interactions. Neutrinos only participate in the weak force. This feature allows cosmic neutrinos from deep within the reaches of space to pass through the Earth nearly unimpeded.

The Standard Model also demands force-carrying particles of integral spin. These integral-spin particles are called gauge bosons; they have this name because the underlying structure of the Standard Model is based on gauge theories, and it is through the construction of the Standard Model gauge theory that these bosons appear. The electromagnetic force is mediated by the exchange of massless photons (γ). The weak force is transmitted via the exchange of intermediate vector bosons, of which there are three types, $W^{\pm}$ and Z. Finally, in the Standard Model, the strong force is mediated by the eight massless gluons, which possess the charge of strong force, called *color*.

Force	Particle	Mass
Strong	gluon	0
Weak	$W^{\pm}$	80 GeV/c^2
	Z^0	91 GeV/c^2
Electromagnetic	photon	0

Table 2.3: The force-carrying gauge bosons of the Standard Model.

2.2 Broken Symmetry and the Higgs Mechanism

Despite the phenomenal success of the predictive powers of the Standard Model, there is a portion of the framework that has no experimental evidence. One particularly clever aspect of the Standard Model formalism is the unification of the electromagnetic and weak interactions through the exploitation of a local gauge symmetry. This symmetry is a quality of the theory in which the Standard Model Lagrangian is invariant to local gauge transformations. However, this elegant theory requires the $W^{\pm}$ and Z^0 to be massless, contradicting experiment.

It is through the spontaneous breaking of this electroweak gauge symmetry, initiated by what is called the Higgs mechanism [4], that the fundamental massive fermions and bosons acquire mass. The introduction of a scalar doublet self-interacting field implements the Higgs mechanism in the Standard Model. One member of the doublet acquires a nonzero vacuum expectation value (vev), v. From the Yukawa-type interactions of the doublet with the fundamental fermion fields, this nonzero vev causes nonzero mass terms in the final Lagrangian for the Standard Model for both the massive gauge bosons and fundamental fermions. For example, the masses of the electron and the $W^{\pm}$ gauge boson are given by

$$m_e = \frac{g_e v}{\sqrt{2}} \tag{2.1}$$

$$M_{W^{\pm}} = \frac{g_2 v}{2} \tag{2.2}$$

where g_2 is the coupling constant for the weak interaction that arises from the construction of the gauge theory, and g_e is the electron Yukawa coupling, a measure of the strength of the interaction between the Higgs and the electron. From the measured $W^{\pm}$ mass, one finds $v \simeq 246\, GeV$. Similar mass relations exist for all the fundamental fermions and the Z boson. The photon remains massless, consistent with experiment.

Another critical consequence of the Higgs mechanism in the Standard Model is the existence of a neutral, scalar particle, the so-called Higgs boson, denoted herein as H. This massive boson arises from the nonzero vev acquired by a member of the scalar doublet introduced to break electroweak symmetry. The mass of the Higgs itself is not predicted from the theory since it depends on the unknown Higgs self-coupling, λ:

$$M_{h_{SM}} = \sqrt{2\lambda v^2} \tag{2.3}$$

Not until this last piece of the puzzle is discovered can the absolute power of the Standard Model be known.

2.3 Consequences of the Higgs

The Standard Model is at best only an effective field theory. The Standard Model fails to incorporate massive neutrinos, whose signatures were first identified by the Super-K experiment in 1998 [5]. The model also does not incorporate the gravitational force, which, though feeble at the energy scales available to this generation of experiments, becomes comparable in strength to the other forces at an energy scale $\Lambda \approx M_{planck} = 10^{16}\ GeV$. Therefore the Standard Model must by definition be superseded by some more complete framework above the Planck scale. This means that at some energy scale below $10^{16}\ GeV$, the Standard Model must break down and new physics must come into play. The scale at which the Standard Model no longer adequately governs the fundamental particles and their interactions is sensitive to the mass of the Higgs. Thus a discovery of the Higgs not only offers insight into the robustness of the theory, but also sheds light on the ultimate reach of the Model's relevance.

The implications on the ultimate reach of the Standard Model are depicted in Figure 2.1 which shows the behavior of upper and lower bounds on M_H as a function of the Standard Model breaking scale Λ. Recall that the Higgs self-coupling is related to the vacuum expectation value of the Higgs and its mass by Eqn. 2.3. This coupling is required to be finite and positive for the theory to be consistent, since negative or infinite Higgs self-couplings would be unphysical. The Higgs self-coupling, λ, increases with energy; requiring the perturbation theory of the SM to remain viable at or below a certain energy scale, Λ, limits the coupling λ, and therefore M_H from above. One can also require that the vacuum expectation value acquired by the scalar Higgs is a true minimum of the electroweak theory. This is called the vacuum stability condition, and this requirement puts Λ-dependent constraints on M_H from below.

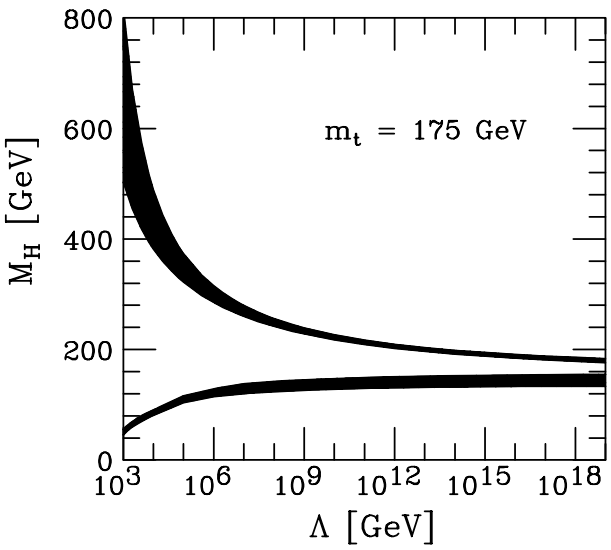


Figure 2.1: Lower and upper bounds on M_H as a function of energy scale Λ .

From Figure 2.1[7], one sees that if the Higgs mass is between 160 and 190 GeV/c^2, the SM is perturbative and therefore valid up to the Planck scale, where new physics

necessarily arises. However if the Higgs mass is outside this window, the Standard Model breaks down at some lower energy scale. For example, in the range 100 $GeV/c^2 < M_H < 130\ GeV/c^2$, Λ is of the order of a few hundred TeV. Several theories predict new phenomena in this regime; supersymmetry is a theory beyond the Standard Model that introduces new physics at the TeV scale.

Thus the discovery of the Higgs boson would have a significant impact on the present understanding of the fundamental particles and the forces through which they interact. That is why the search for the Higgs is of critical importance to the future of our understanding of the world around us.

2.4 Present Limits on the Higgs

An upper mass limit for M_H comes from SM unitarity constraints [6]. Unitarity in the SM prevents probabilities from being greater than one; Higgs contributions to certain SM processes could drive the predicted probabilities of these processes greater than one above some threshold M_H value. From unitarity constraints alone, $M_H < TeV/c^2$. Other more sensitive limits have been achieved via precision electroweak measurements and direct searches for the Higgs, each discussed below.

2.4.1 Electroweak precision measurements

An important goal of the past decade in experimental particle physics was the precision measurement of electroweak parameters predicted by the SM. Many of these parameters have a dependence on unknown quantities, such as the Higgs mass. For example, in electroweak theory, evaluation of the $W^{\pm}$ propagator yields a predicted value of the $W^{\pm}$ mass, M_W . The $W^{\pm}$ propagator involves Higgs loops at higher orders in perturbation theory; the calculation of the Feynman diagrams involving these Higgs loops thus introduces M_H as an unknown quantity. Theorists attempt to calculate

the perturbative expansion to high enough order to match experimental precision and shed some light on the unknowns in the calculation. The current experimental value of the $W^{\pm}$ boson mass is $80.41 \pm 0.04\ GeV/c^2$.

Important precision electroweak studies were conducted at the Large Electron-Positron Collider (LEP)[8], located at the CERN facility on the Swiss-French border. As its name suggests, the LEP machine collides beams of electrons and positrons; the first run of the LEP collider (LEP1) from 1989-1995 provided electron-positron collisions at a center-of-mass energy of $\sqrt{s} = M_Z$. The LEP machine was tuned precisely to this beam energy to produce millions of Z particles, giving experimenters the ability to precisely test electroweak couplings to the fundamental fermions. In LEP2 (1995-2000), accelerator design changes provided experimenters with higher beam energy and luminosity. These eleven years of precision electroweak data would prove valuable in the indirect search for the Higgs.

The four collider experiments at LEP [9]-[12], ALEPH, DELPHI, L3 and OPAL, were leaders in precision electroweak physics. The LEP Electroweak Working Group [13] was formed to organize and combine independent results from each of the LEP experiments, as well as incorporate electroweak measurements from experiments around the world. As of Winter 2003, after combining the electroweak data from the LEP experiments and other pertinent electroweak studies, the Working Group reported that based on the interdependence of M_W , m_t and M_H , a light mass Higgs would be favored [14]:

$$M_H < 211 GeV/c^2 \qquad \text{at } 95\% \text{ CL}$$

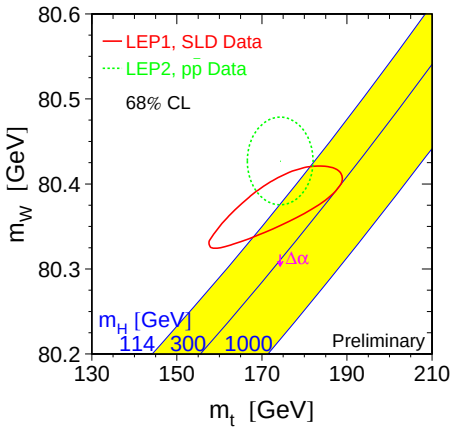


Figure 2.2: Winter 2003 M_W, m_t dependence for a variety of M_H values, overlayed with the measured values of M_W and m_t at 68% CL.

Figure 2.2 plots the theoretical prediction of M_W and m_t dependence for a range of Higgs masses, $M_H = 114 - 1000\ GeV/c^2$. The experimental measurements of the $W^{\pm}$ mass and the top mass are overlayed on the theory predictions. It should be noted that in Fig 2.2, the bounds on the measured M_W and m_t are only at the 68% confidence level (CL). The confidence level indicates the strength of belief in the statement being made.

Fig 2.2 illustrates an important behavior: corrections to M_W are proportional to the logarithm of the Higgs mass, *ie*, a large spectrum of possible Higgs masses ($M_H = 114 - 1000\ GeV/c^2$ in the combined result above) corresponds to a relatively small deviation in the predicted M_W . Thus, achieving the highest precision possible in the M_W measurement, as well as other electroweak observables, would be necessary to provide relevant bounds on the Higgs mass. There exists a similar strong dependence on m_t; according to [15] a 5 GeV shift in the measured value of m_t corresponds to a 35% deviation in the M_H prediction.

2.4.2 Direct Searches

The Higgs mass is not predicted by the SM, thus searches for very low mass Higgs bosons were conducted at various experiments over the past several decades. These attempts did not yield an observation; the mass region eliminated by these very low mass searches was completely superseded by the direct searches of the past ten years. Therefore a detailed discussion of the early searches is omitted here; however, the pioneering efforts should be noted[16].

Recent direct Higgs searches were conducted primarily at LEP in conjunction with the precision electroweak program. The four collider experiments searched millions of $e^+ - e^-$ collisions in search of Higgs bosons produced in association with a Z boson. This production mechanism is very similar to the channel that will be the focus of this search, as described in Section 2.5.

By mid-2000 the four collider experiments at LEP had not found any conclusive evidence for the Higgs. Time was running out for the Higgs searches at LEP; the facility was scheduled for shutdown near the end of 2000 in preparation for the construction of the next generation collider that would be built at CERN, the Large Hadron Collider (LHC). At the end of its lifetime, the LEP collider was pushed one final time past its design parameters in a last attempt to further the machine's reach in searching for the Higgs. A tantalizing signal was seen, an excess of events in a decay channel consistent with Higgs production with $M_H \simeq 115\ GeV/c^2$ [17]. Experimenters desperately tried to extend LEP2 a few more months to study this excess in detail. But the need to begin construction of the LHC prevailed and on November 2, 2000, the LEP collider was powered down for the last time. The four LEP collider experiments combined their efforts and set a final 95% CL lower mass bound of $M_H > 114.4\ GeV/c^2$ [18].

2.4.3 Implications on Present Analysis

From direct and indirect Higgs searches, it is clear that a low mass Higgs is favored in the range 114.4 $GeV/c^2 < M_H < 211\ GeV/c^2$. An analysis was completed that incorporated precision electroweak results along with the results of the direct searches at LEP [19]. This analysis calculated the probability distribution as a function of M_H; the data significantly favored a low mass Higgs in the region 110 $GeV/c^2 < M_H < 125\ GeV/c^2$. Given this information we choose to perform a search for the Higgs in the mass range 100 $GeV/c^2 < M_H < 150\ GeV/c^2$, properly covering an important region of interest.

2.5 Higgs Production at the Tevatron

Higgs bosons are produced through a variety of mechanisms at the Tevatron; Figure 2.3 shows the theoretical cross section for the three significant production mechanisms as a function of M_H.

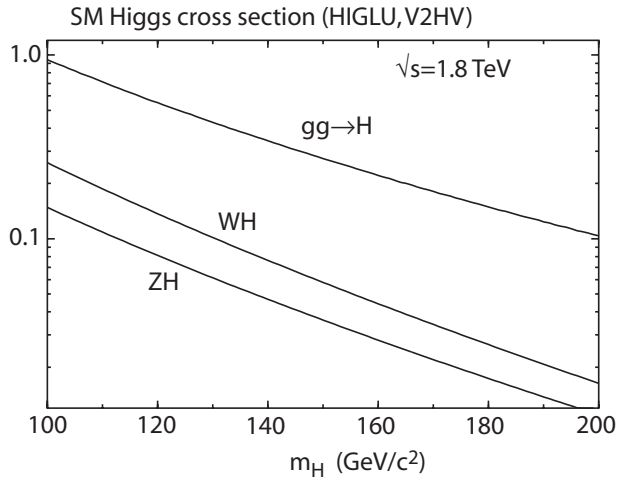


Figure 2.3: Theoretical prediction for WH production cross section in pb as a function of M_H.

The *cross section* can be thought of as the production rate for Higgs events in the p-$\overline{p}$ collisions. The dominant production mechanism at the Tevatron is via gluon-gluon fusion, represented diagramatically in Figure 2.4.

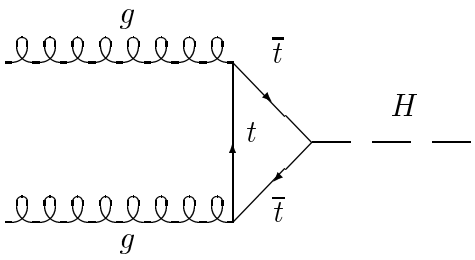


Figure 2.4: Gluon-gluon Higgs production through one top loop.

Gluon-gluon fusion necessarily contains a heavy fermion loop, since the Higgs boson does not couple directly to massless objects.

Higgs production at the Tevatron also proceeds via Higgs radiation off of a vector boson $V = W$ or Z. Copious numbers of $W^{\pm}$ and Z particles are created in the Tevatron environment; the resonances that are created can possess a mass on the high tail of their distribution if the momentum transfer from the hard scatter is significantly above the mass of the particle in question. These so-called off-shell resonances then can radiate a heavy object, such as a Higgs boson, and proceed to an on mass-shell state. This is precisely the mechanism behind the other two prominent Higgs production mechanisms at the Tevatron, known as *Higgstrahlung*. Higgs production in association with a $W^{\pm}$ boson is called WH production, and is called ZH production in the case of Higgs radiation off of a Z. It should be noted that the gluon-gluon fusion cross section is approximately a factor of four higher than that of the larger Higgstrahlung process, WH production.

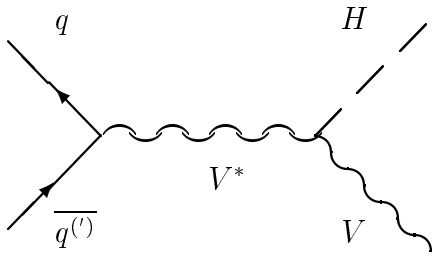


Figure 2.5: Higgs production in association with a vector boson at the Tevatron.

Understanding of the production mechanisms is not sufficient to design a Higgs search. The decay modes of the Higgs boson are critical, because it is the decay of the Higgs that provides the unique signature for which we will search. Figure 2.6 contains the decay modes and their branching fractions as a function of M_H. In the range 80 $GeV/c^2 < M_H < 130$ GeV/c^2, the Higgs decays predominantly to b-quark pairs. Above a mass of 130 GeV/c^2, the decay $H \rightarrow W^+W^-$ becomes dominant. This is because as the Higgs coupling is proportional to the mass of the objects it couples to; as the Higgs mass increases, more decay channels are available, and given sufficient mass in the resonance, it will decay to the heaviest pair of objects allowed. In the regime 80 $GeV/c^2 < M_H < 130$ GeV/c^2, the Higgs simply did not have enough mass in its initial state to decay to $W^{\pm}$ bosons; therefore the alternative is the next-most massive particle-antiparticle pair, a pair of b quarks.

The channel $gg \rightarrow H \rightarrow b\overline{b}$ produces 2 jets in the detector that will be evident in the tracking and calorimetry. *Jets* are sprays of particles which come from the decay of quarks and gluons created in the $p\bar{p}$ collision. The strong force prevents the existence of free quarks; when a quark is produced in a p–$\overline{p}$ collision, the quark must form a bound state with another quark, called a hadron. The vacuum can be thought of as a sea of quark-antiquark pairs; so the quark produced in the p–$\overline{p}$ interaction

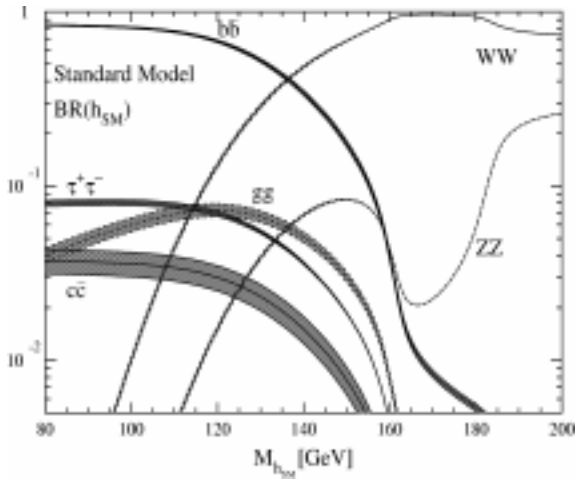


Figure 2.6: Higgs decay modes and their probabilities as a function of M_H.

rips one of these quarks from the vacuum in forming its bound state. This leaves one quarks from the vacuum now unpaired; this quarks subsequently also rips another quark from the vacuum, creating a bound state. This process (called *hadronization*) continues, creating many new hadrons which subsequently decay and appear in the detector as a jet of particles.

The gluon-gluon fusion channel is however tremendously difficult to identify at the Tevatron due to a formidable amount of other, less interesting interactions that share the same 2-jet final state. These so-called *background* processes have production rates orders of magnitude larger than the predicted Higgs production cross section. Management and understanding of the background is one of the most critical parts of any such search.

The more promising Higgs channel at the Tevatron is WH and ZH production. Despite a smaller cross section, these production mechanisms offer a signature whose backgrounds are considerably more manageable. The decay products of the vector boson offer additional handles for event identification.

In the analysis described herein, we focus on WH production. $W^{\pm}$ bosons decay to leptons 30% of the time, the remainder of the time to hadrons. The hadronic $W^{\pm}$ decay mode is however avoided when considering WH production in the M_H range of this analysis because it leads to a four jet final state, $WH \to q\overline{q'}b\overline{b}$. Like gluon-gluon fusion, this four jet final state also has formidable QCD backgrounds that we seek to avoid.

We instead rely on the leptonic decay of the $W^{\pm}$. The presence of a high p_T lepton from $W^{\pm}$ decay can provide a clean trigger for baseline event selection. The neutrino that is produced in the leptonic $W^{\pm}$ decay escapes the CDF detector without leaving telltale interaction remnants in their wake; however these neutrinos do carry away some of energy from the incident collision. The presence of neutrinos is inferred by an energy imbalance in the final state.

The $W^{\pm}$ boson decays to all three lepton families. We look for the decays $W \to e\nu$ and $W \to \mu\nu$ in this analysis; $W \to \tau\nu$ is not considered here because the large branching ratio ($\simeq$64%) of the τ lepton to hadronic final states makes τ identification considerably more difficult. However, instances of $W \to \tau\nu$ in which the τ decays to leptons ($\simeq$36%) are within the acceptance of the analysis.

In summary, the channel we investigate in this analysis is $WH \to \ell\nu b\overline{b}$ where $\ell = e$ or μ. We are not free however from the influence of background channels in this final state: prominent background channels include $W + b\bar{b}$ production , $t\overline{t}$ production, single top production and diboson channels, including WZ. Figure 2.7 shows the diagram for one of the contributing background process, $W + b\bar{b}$ production. The background contributions will be discussed in Chapter 5.

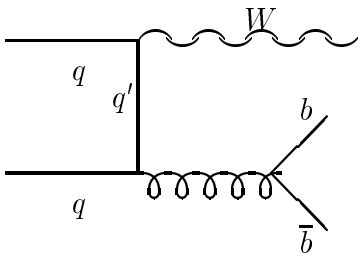


Figure 2.7: $W + b\bar{b}$ production, a background to WH for 100 $<M_H<$150 GeV/c^2.

CHAPTER 3

EXPERIMENTAL APPARATUS

The Higgs boson does not appear directly in our everyday world. Such new forms of matter are however accessible to mankind via their production and subsequent decay in highly energetic collisions of subatomic particles. Such collisions are manufactured at particle accelerator facilities like the Tevatron at Fermi National Accelerator Laboratory outside Chicago, IL (Fermilab for short). The search for the Higgs boson described herein utilizes collisions provided by the Tevatron accelerator, a machine that relies on the fields produced by superconducting magnets to steer beams of protons (p) and antiprotons ($\overline{p}$). The beams propagate in the Tevatron in 3.9 mi circumference circular orbits in a 5 cm diameter pipe a few stories underground.

The p and $\overline{p}$ beams are collided at two places along the circular track, one of which is the nominal center of the Collider Detector at Fermilab experiment (CDF), the detector with which we perform this search. This analysis focuses on data collected at CDF during Run 1 of the Tevatron collider (1992—1995). Protons and antiprotons are bound states of quarks; thus because quarks participate in the strong, weak and electromagnetic interactions, each of these forces can be probed in a p–$\overline{p}$ collision. High energy collisions allow for the production of such exotic particles as $W^{\pm}$ and Z bosons and the top quark, which was discovered at the Tevatron during Run 1.

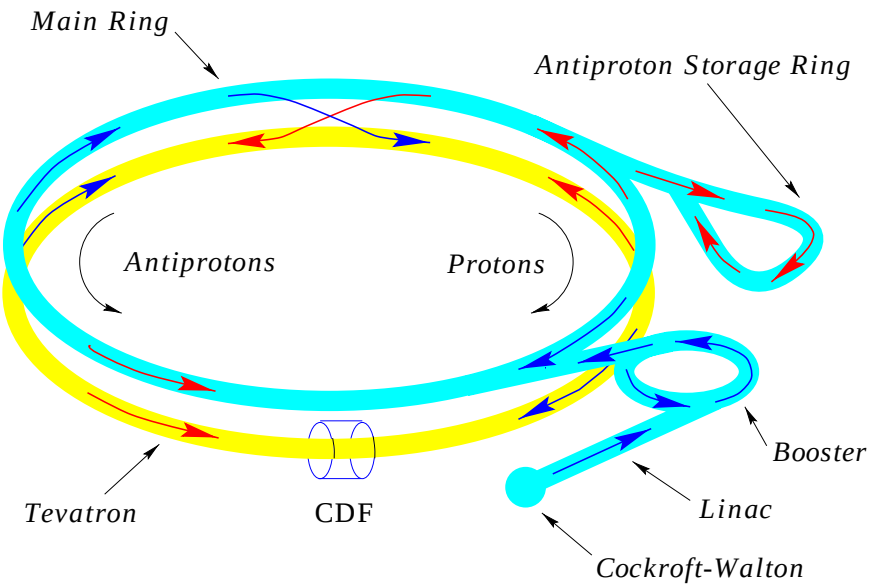


Figure 3.1: Diagram of the Tevatron accelerator complex.

Fermilab is the highest energy particle accelerator in the world, and is currently the only active facility capable of probing the Higgs sector.

3.1 The Tevatron Collider

Fermilab utilizes a series of accelerators to create the 900 GeV proton and antiproton beams collided in the Tevatron. The creation of the proton beam starts from a bottle of Hydrogen gas. The hydrogen atoms are ionized and then are accelerated by a Cockroft-Walton electrostatic pre-accelerator to an energy of 750 keV. The accelerated H^- ions then enter a 150 m linear accelerator, the Linac, where their energy is increased by the application of an oscillating electric field to 400 MeV. At the end of

the Linac the hydrogen ions are incident on a carbon foil where the electrons are removed from the hydrogen ions, leaving behind only the proton nuclei. These protons are then sent to the Booster, a 150 m diameter syncrotron. Like a linear accelerator, a synchrotron accelerator also exploits electric fields to impart energy to the particle beam. The synchrotron however uses magnetic fields to bend the trajectory of the protons into a circular path, allowing for the repeated application of the electric field during each revolution. The Booster synchrotron accelerates the protons to an energy of 8 GeV and operates in 12 cycles, thus providing 12 distinct bunches of protons that are transferred to another synchrotron, the Main Ring. This 6.3 km circumference synchrotron accelerates the proton bunches to an energy of 150 GeV; to keep the proton bunches in proper orbit during this step in the acceleration process, the Main Ring magnets are capable of generating magnetic fields up to 0.7 T. The twelve proton bunches are then coalesced into a single bunch of approximately 2×10^{11} protons and are injected into the Tevatron, the highest energy synchrotron accelerator in the world. The Tevatron is composed of superconducting magnets which reside directly below the Main Ring magnets in the same underground circular tunnel. The superconducting magnets of the Tevatron can generate magnetic fields up to 4.4 T. The Tevatron accelerates the protons to the colliding energy of 900 GeV. A total of six proton bunches are transferred to the Tevatron in the manner described above. The entire process described above takes approximately sixty seconds.

Production of the antiproton beam is slightly more complex. Bunches of 120 GeV protons from the Main Ring are extracted and focused onto a tungsten target. Interactions between the incident proton bunches and the nucleons in the tungsten create a spray of new particles, including — at some small rate — antiprotons with

an energy of 8 GeV, which are selected and focused with a lithium lens. The antiprotons are then transferred to a triangular-shaped synchrotron called the Debuncher; here the momentum spread and oscillations in the transverse plane of the antiproton bunches are reduced, a process known as "cooling". After the Debuncher the 8 GeV antiprotons are sent to another triangular-shaped synchrotron, called the Accumulator. Here the antiprotons are further cooled; once approximately 1×10^{12} antiprotons are "stacked", six antiproton bunches are transferred to the Main Ring, where the antiprotons are accelerated in the same fashion as the protons to 150 GeV. The six antiproton bunches are then transferred to the Tevatron and are accelerated to an energy of 900 GeV to match the energy of the proton beam. The transfer efficiency between the Accumulator and the Tevatron is low for antiprotons; consequently once the antiprotons reach the Tevatron a bunch typically contains only 30×10^9 particles. The antiproton bunches share the same beampipe as the protons, as well as the magnetic and RF fields used for acceleration.

The proton and antiproton bunches propagate in counter-circulating orbits in the Tevatron. The energy of each beam is 900 GeV; the center-of-mass energy of the colliding beams is 1.8 TeV. The beams travel in a helical path which allows for collisions at only two interaction regions. These two interaction regions are the locations of the two collider physics experiments at the Tevatron, CDF and DØ. The CDF and DØ detectors surround the interaction regions in an effort to record the remnants of the energetic p-$\overline{p}$ collisions. In a perfect world, the collisions would take place at the exact center of the detectors; however the actual collision point is gaussian distributed with a width along the z axis (z is along the beam axis) of 30 cm, and a width in the transverse $x - y$ plane of 35 μm.

The beams travel in the Tevatron at approximately the speed of light, which means that, given six bunches of p and $\overline{p}$, and the four mile circumference of the Tevatron, the collisions occur every 3.5 μs. During 1992—1995 collision runs in the Tevatron typically lasted $\sim$ 10 hours. More details on the operation of the Tevatron can be found in [20].

The instantaneous luminosity (L) is a measure of collision rate and is given by

$$L = \frac{N_p N_{\bar{p}} B f_0}{4\pi\sigma^2} \tag{3.1}$$

where N_p is the total number of protons per bunch, $N_{\bar{p}}$ is the total number of antiprotons per bunch (both given above), B is the number of bunches of each type, f_0 is the frequency of bunch revolution (47.4 kHz), and σ^2 is the cross-sectional area of the bunches ($\sigma^2 \sim 5 \times 10^{-5}\ cm^2$). The instantaneous luminosity of the accelerator falls exponentially with time due to spreading in the transverse plane of the beam (increase in σ) and losses of p and $\overline{p}$ in the transverse beam width tails and collisions (decrease in N_p,$N_{\bar{p}}$). In the 1992—1995 Run 1 of the Tevatron the peak instantaneous luminosity reached was $2.8 \times\ 10^{31}\ cm^{-2}\ s^{-1}$ while typical values were $0.54 \times\ 10^{31}\ cm^{-2}\ s^{-1}$ for Run 1a and $1.6 \times\ 10^{31}\ cm^{-2}\ s^{-1}$ for Run 1b.

3.2 The CDF Detector

CDF is located at one of the two colliding beam interaction regions of the Tevatron. The CDF detector is a multipurpose detector used to observe a wide range of physics processes produced from high-energy $p\bar{p}$ collisions. The CDF detector is designed to identify and measure the energy and momentum of electrons, muons, photons and jets.

A schematic drawing of the detector is shown in Figure 3.2; only a quarter of the detector is shown. The interaction point is in the lower righthand corner. CDF is

cylindrically and forward-backward symmetric about the transverse plane that passes through the interaction point.

The origin of the CDF coordinate system is the interaction point. The positive z-axis points along the beamline in the direction of the protons, the x-axis points radially outward from the center of the ring and the y-axis points upwards. In terms of angles, θ is the polar angle with respect to the z-axis, and ϕ is the azimuthal angle in the x-y plane. Typically, locations of particles are identified by the Lorentz invariant quantity *pseudorapidity*, η:

$$\eta = -\ln(\tan(\theta/2)) \tag{3.2}$$

Large η corresponds to regions very near the beam line (small θ); small η corresponds to the plane perpendicular to the z-axis at z=0. As an aside, a measure of the opening angle between two objects is given by ΔR:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.3}$$

Small values of ΔR indicate two objects that are highly collimated.

The CDF detector is divided into three η ranges: the central ($|\ \eta\ |<1.0$), the plug ($1.0 <|\ \eta\ |<2.4$) and forward regions ($|\ \eta\ |<2.4$). CDF is composed of a several smaller detector segments, each of which can be considered its own system. There are three main sections utilized in this analysis:

- charged particle tracking
- calorimetry
- muon detection

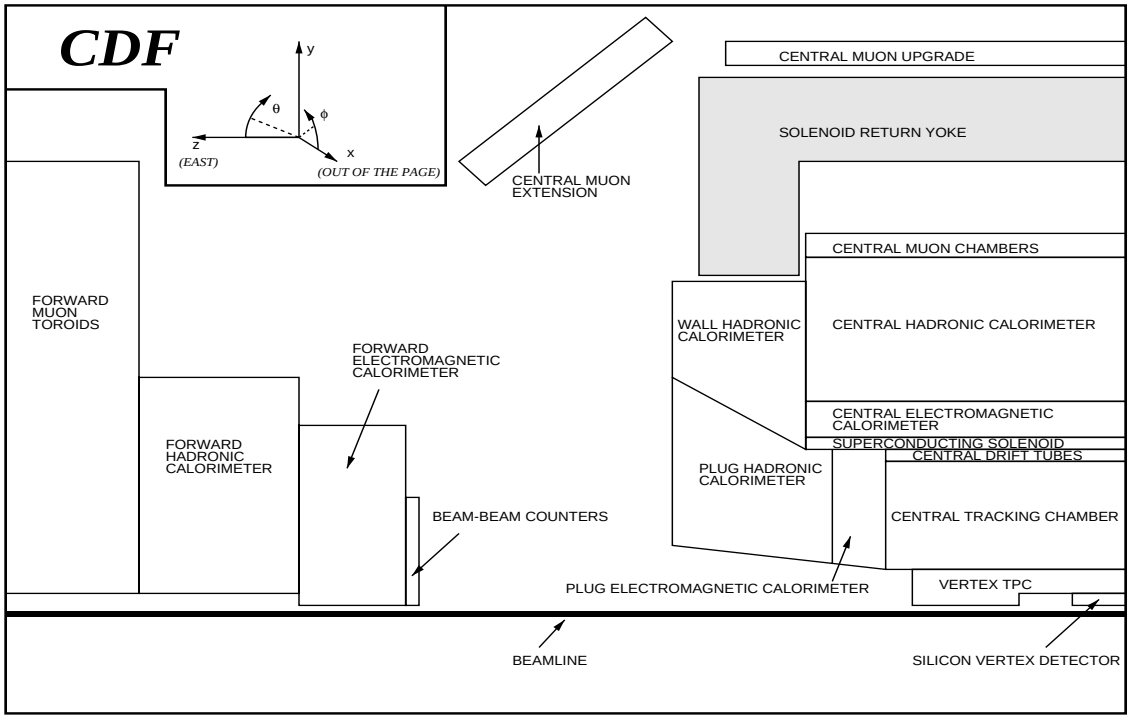


Figure 3.2: A one quarter side-view cross section of the CDF detector. The detector is forward-backward symmetric about the interaction region, located at the lower righthand corner of the figure.

We only focus here on the portions of the CDF detector most relevant for this analysis; a complete description can be found elsewhere [28]. Starting from the interaction point and moving radially outward, the tracking system is located inside a 1.5 m radius superconducting solenoid which produces a 1.4 T axial magnetic field. The axial magnetic field causes the trajectory of charged particles to curve in the transverse plane (recall, $\vec{F} = q\vec{v} \times \vec{B}$). From the curvature of the charged particle track, one can determine the particle's momentum. Outside the solenoid, the calorimetry system surrounds the tracking chambers. By employing two types of absorbing material, two types of calorimetry measurements are made: one that measures the amount of energy particles lose due to the electromagnetic interaction and one that gauges the particles' energy loss due to hadronic interactions. Electrons, photons and most hadrons are absorbed in the detector in the calorimetry. Muons interact minimally within the calorimetry, so they escape and are identified in the muon detectors at large radius. Each of these components are discussed in greater detail below.

3.2.1 The Tracking System

Charged particle tracking measurements are made using three components of the CDF detector: the Silicon Vertex Detector (SVX), the Vertex Time Projection Chamber (VTX), and the Central Tracking Chamber (CTC). Each of these detectors lies within the 1.4 T magnetic field, and each serves a specific purpose. Closest to the beampipe is the SVX which has the best track position resolution and is used to identify displaced vertices indicative of the decay of a B hadron. The VTX surrounds the SVX and is primarily used to identify the z position of an event's interaction point, or primary vertex. Surrounding the VTX is the CTC. The CTC was designed for the precise measurement of a charged particle track's momentum.

The SVX

The SVX provides accurate measurements in the transverse plane for the reconstruction of charged particle tracks. Details of the SVX detector and its performance can be found elsewhere [28]. The SVX consists of two end-to-end barrels that are aligned along the beampipe. There is a 2.15 cm separation between the two barrels at z=0. The active length of the SVX is 51 cm; this corresponds to a pseudorapidity coverage of $|\eta|$<1.9. Recall that $p\bar{p}$ collision positions are gaussian distributed with a spread of $\sigma \sim 30$ cm about z=0. Thus tracks from some events will be poorly measured by the SVX, which does not cover completely the interaction region. These features contribute to the track acceptance of the SVX, which was found to be $\sim$ 60%.

Each SVX barrel is divided into 12 wedges of 30° in azimuth. Figure 3.3 shows a diagram of an SVX barrel. Each wedge has four layers of silicon strip detectors. The innermost silicon layer resides at a radius of 2.86 cm and the outermost is at a radius of 7.87 cm, each radius measured from the beampipe. Each layer within a particular wedge is called a ladder. Each ladder has three 8.5 cm long silicon strip detectors. These three silicon segments contain a varying number of silicon strip channels, depending on layer: for example, the innermost ladder has 256 silicon strips while the outermost ladder has 768. The separation between silicon strips is 60 μm for the three inner layers and 55 μm for the outermost layer. This separation results in an $r-\phi$ position resolution of 13 μm. There are a total of 46080 SVX channels that are processed by 360 readout chips. These 46080 constitute roughly one-third of the total possible readout channels for the entire CDF detector. However only about 5% of the SVX channels are read out in a given event. Table 3.1 summarizes some of the more critical facts about the SVX.

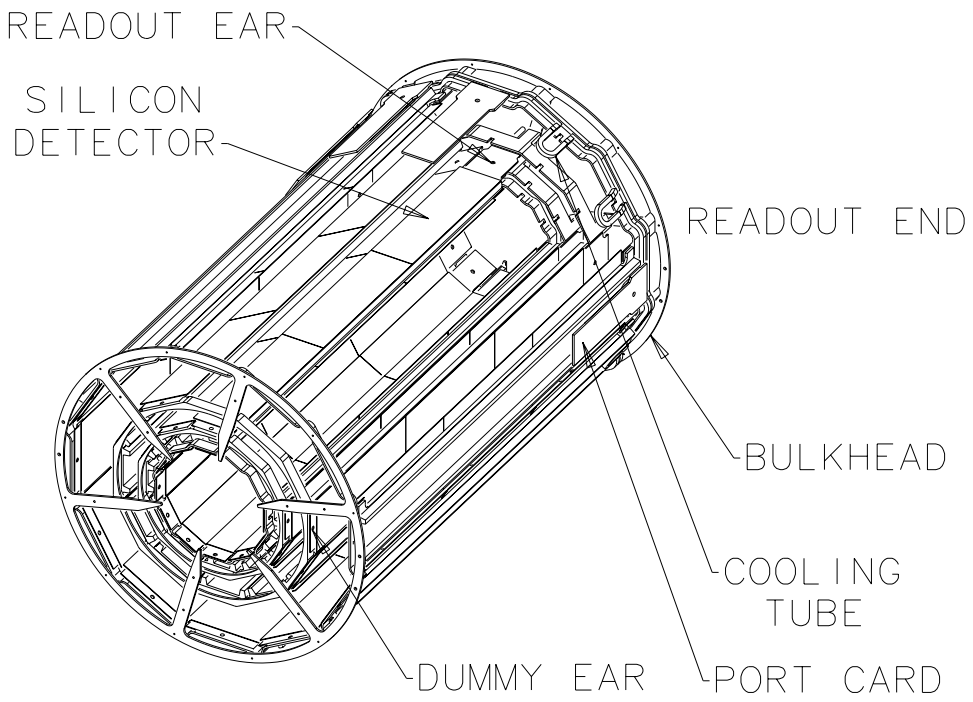


Figure 3.3: Portrait of one of the SVX barrels.

One of the primary uses for the SVX is the measurement of secondary vertices in the $r-\phi$ plane. Secondary vertices are indicative of the decay of b and c hadrons, which have constituent b or c quarks. The relatively long lifetime of these heavy quarks allow b and c hadrons to travel a small distance before they decay. This subsequent decay produces a secondary decay vertex that the SVX specializes in identifying.

The VTX

The collision rate, or luminosity, at the Tevatron is very large. It is frequently the case that more than one p-$\overline{p}$ interaction occurs per event. The VTX is used to associate charged particle tracks to their correct vertex along the beamline. The VTX is composed of 8 octagonal chambers that contain a 50% - 50% mix of argon and ethane gas. The complete VTX detector is 2.8 m long in z and extends from an inner radius of 8 cm (just outside the SVX) out to a radius of 22 cm from the beampipe. Each VTX chamber is divided into two 15.25 cm drift regions, separated by a high voltage grid.

The endcaps of the VTX are segmented azimuthally and consist of sense wires perpendicular to the beamline and the radial centerline of the wedges. Charged particles passing through the VTX ionize the gas and free electrons; the freed electrons then drift in the axial direction to the sense wires in the end caps. From the arrival time of the drift electrons, the VTX provides charged particle tracking in the $r-z$ plane. With coverage out to $|\eta|<3.5$, the VTX determines the z position of the event vertices to within 1 mm. Table 3.2.1 summarizes the physical characteristics of the VTX.

The CTC

The CTC is a large cylindrical open-wire drift chamber that measures charged particle momenta in the $r-\phi$ plane from the curvature of the particle's trajectory in the 1.4 T magnetic field. The CTC is 3.2 m long in z, with an inner radius of 0.3 m and an outer radius of 1.3 m, giving coverage over the pseudorapidity range of $|\eta|<1.0$. Sense wires run the length of the CTC chamber. There are 84 radial layers of sense wires in the CTC; these 84 layers are separated into nine superlayers, five of which are axial (wires run parallel to beam) and four of which are stereo (wires have $\pm 3^{\circ}$ azimuthal offset). Axial superlayers provide tracking information in $r-\phi$. These superlayers have twelve sense wires apiece. Stereo superlayers provide tracking information in the $r-z$ plane. The combination of the axial and stereo layers yield 3-dimensional tracking. The field wires of the CTC create a 1350 V/cm drift field. To compensate for the Lorentz angle produced by the coincident electric and magnetic fields, the wires in each superlayer are grouped into cells which are tilted by 45° with respect to the radial direction. This gives drift electrons trajectories which are azimuthal. Figure 3.4 shows a transverse view of the CTC endplate.

Tracks are reconstructed by fitting hits in the CTC to a helix. The curvature of the track is related to the transverse momentum of the particle. The momentum resolution of the CTC is

$$\delta P_T/P_T = 0.002c/GeV^{-1} \times p_T.$$

By combining tracking information from the CTC and the SVX, the momentum resolution improves to

$$\delta P_T/P_T = 0.001c/GeV^{-1} \times p_T.$$

The physical properties of the CTC are listed in Table 3.1.

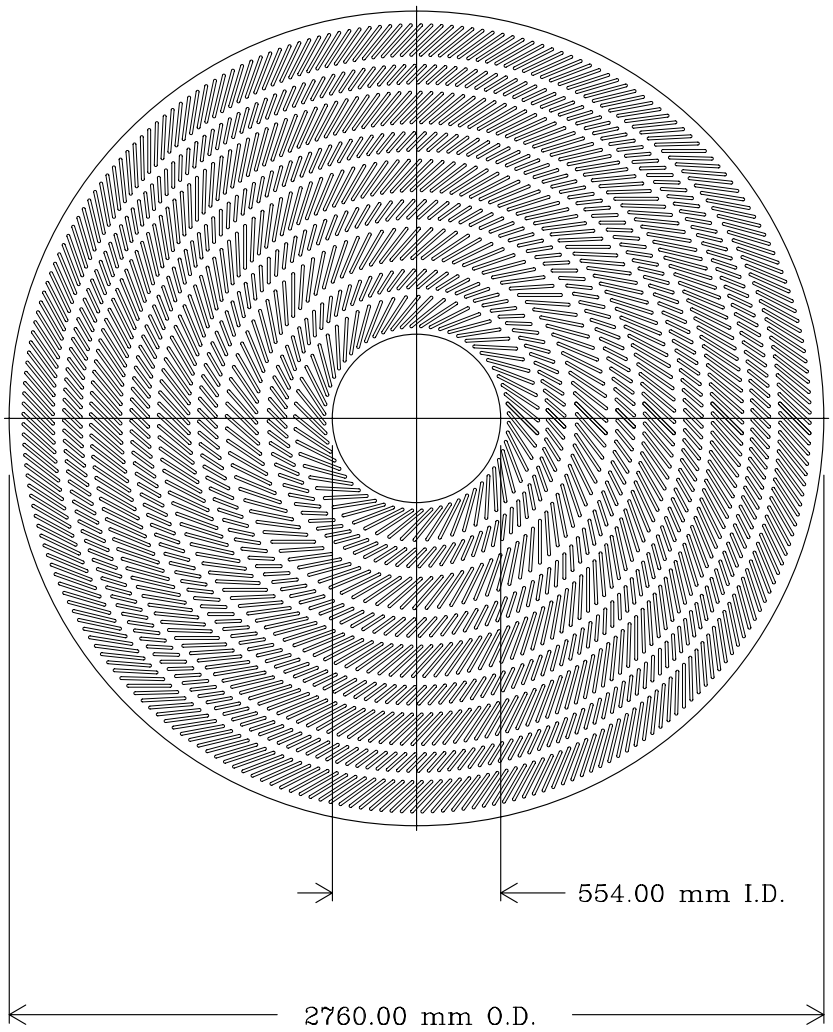


Figure 3.4: Transverse view of the CTC endplate illustrating the 9 superlayer geometry.

	Central tracking chamber (CTC)	Vertex tracking chamber(VTX)	Silicon vertex detector (SVX)
Pseudorapidity Coverage	$\|\eta\| < 1.5$	$\|\eta\| < 3.25$	$\|\eta\| < 1.2$
Inner, Outer Radii (cm)	30.9, 132.0	8, 22	2.7, 7.9
Length (cm)	320	280	26
Layers	60 axial, 24 stereo	24	4
Strip/Wire Spacing	10 mm	6.3 mm	60 μm (inner 3 layers) 55 μm (outer layer)
Spacial Resolution	$200\mu m$ $(r-\phi)$ 4 mm $(r-z)$	200-500 μm $(r-z)$	15 μm $(r-\phi)$
Momentum Resolution	$\delta P_T/P_T = 0.002 \times P_T$		$\delta P_T/P_T = 0.001 \times P_T$
Thickness	$\approx 0.015 X_0$	$\approx 0.0045 X_0$	$\approx 0.035 X_0$

Table 3.1: Description of the charged particle tracking chambers.

3.2.2 Calorimetry

Outside the tracking system resides the calorimetry. The 1.4 T axial magnetic field provided by the solenoid prevents particles with momentum in the transverse plane less than 350 MeV/c from escaping the tracking system. Such low momentum particles curl within the tracking system. However, particles with sufficient transverse momentum escape the tracking system and enter the calorimetry. In this analysis the calorimetry is used to determine the energy and direction of jets, to quantify the amount of missing energy ($\not\!\!E_T$) in the event, and to help in identifying electrons, photons and muons.

There are two types of calorimetry in the CDF detector: electromagnetic and hadronic. Both types of calorimeters consist of alternating layers of material. Incident particles interact with relatively dense absorbing material (lead in the electromagnetic calorimetry, iron in the hadronic calorimetry) and lose some fraction of their energy while creating cascades of secondary particles. Alternating with the absorbing material are an active medium, layers of scintillator. When encountering the scintillating material, the incident particles produce light; this light is collected by light guides on the end of each layer of scintillator. The amount of light is a measurement of the incident particle's energy as a function of depth.

Electromagnetic showers develop faster than hadronic showers, therefore the electromagnetic calorimeters are positioned closer to the interaction point than the hadronic calorimeters. The calorimeters are segmented into towers in $\eta - \phi$ space. Each tower points back to the geometric center of the detector. The calorimetry surrounds the solenoid and tracking chambers and cover a range of 2π in azimuth and a pseudorapidity $|\eta|<4.2$. The calorimetry consists of three subsystems: the central, plug, and forward.

The central electromagnetic (CEM), central hadronic (CHA), and wall hadronic (WHA) calorimeters cover approximately $|\eta| < 1.0$. The electromagnetic and hadronic portions of the central calorimetry are concentric cylindrical barrels that are divided into 15° wedges in azimuth. There are 24 such wedges on each side of z=0. Each wedge is further segmented into towers 0.1 units in pseudorapidity. The 48 wedges of the central calorimeters each contain 10 towers. The particle cascade caused by the particles incident on the calorimetry produce light in the scintillator. This light is collected by acrylic lightguides at the end of each tower and is then transmitted to photomultiplier tubes located at the back of each wedge. A cutaway cross-section view of a central calorimeter wedge is shown in Figure 3.5.

The CEM has 18 radiation lengths worth of material and an inner radius of 173 cm with a depth of 35 cm. The CHA contains 4.7 absorption lengths of material and extends beyond the CEM. The layout of the central calorimeter is shown in Figure 3.6. The measured energy resolution for electromagnetic showers is

$$(\sigma/E)^2 = (13.7\%/\sqrt{E_T})^2 + (2\%)^2.$$

The CEM was calibrated using a testbeam of electrons and is checked periodically using radioactive ^{137}Cs sources. The energy resolution for hadronic showers was measured from isolated pions; the resolution in the central hadronic calorimetry was determined to be

$$(\sigma/E)^2 = (50\%/\sqrt{E_T})^2 + (3\%)^2.$$

The region $0.6 < |\eta| < 1.1$ does not have complete CHA coverage. Additional hadronic calorimetry, the WHA, is in place to refine hadronic measurements in the region of the detector. The WHA makes hadronic energy measurements in a similar way as the CHA. Table 3.2.2 gives detailed information on each of the central calorimeters.

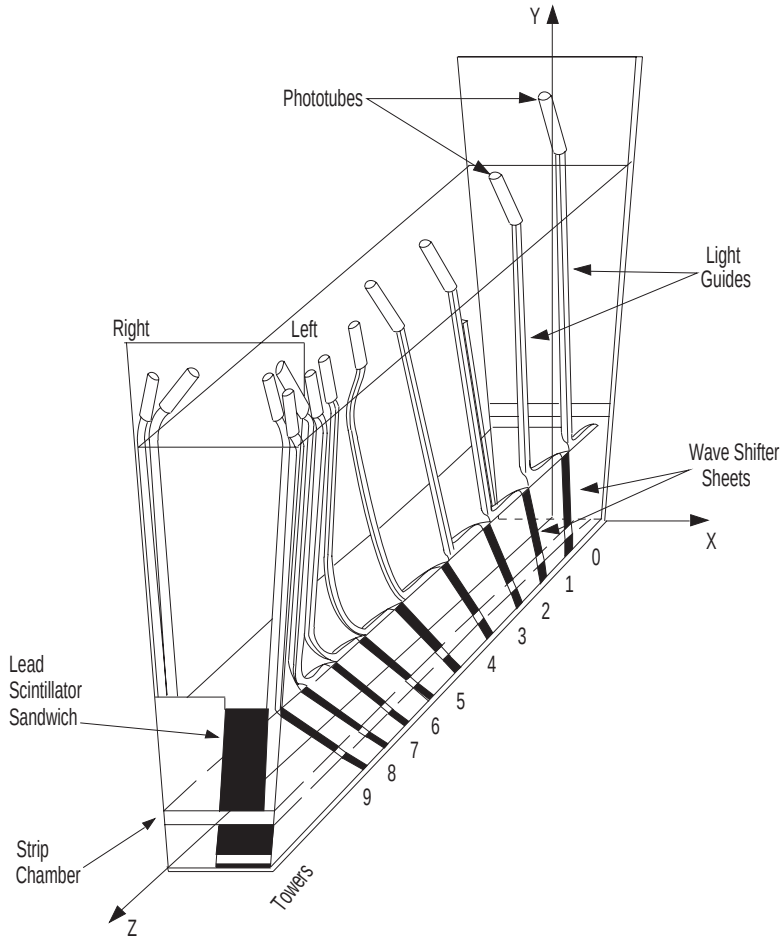


Figure 3.5: Diagram of a single central calorimetry wedge.

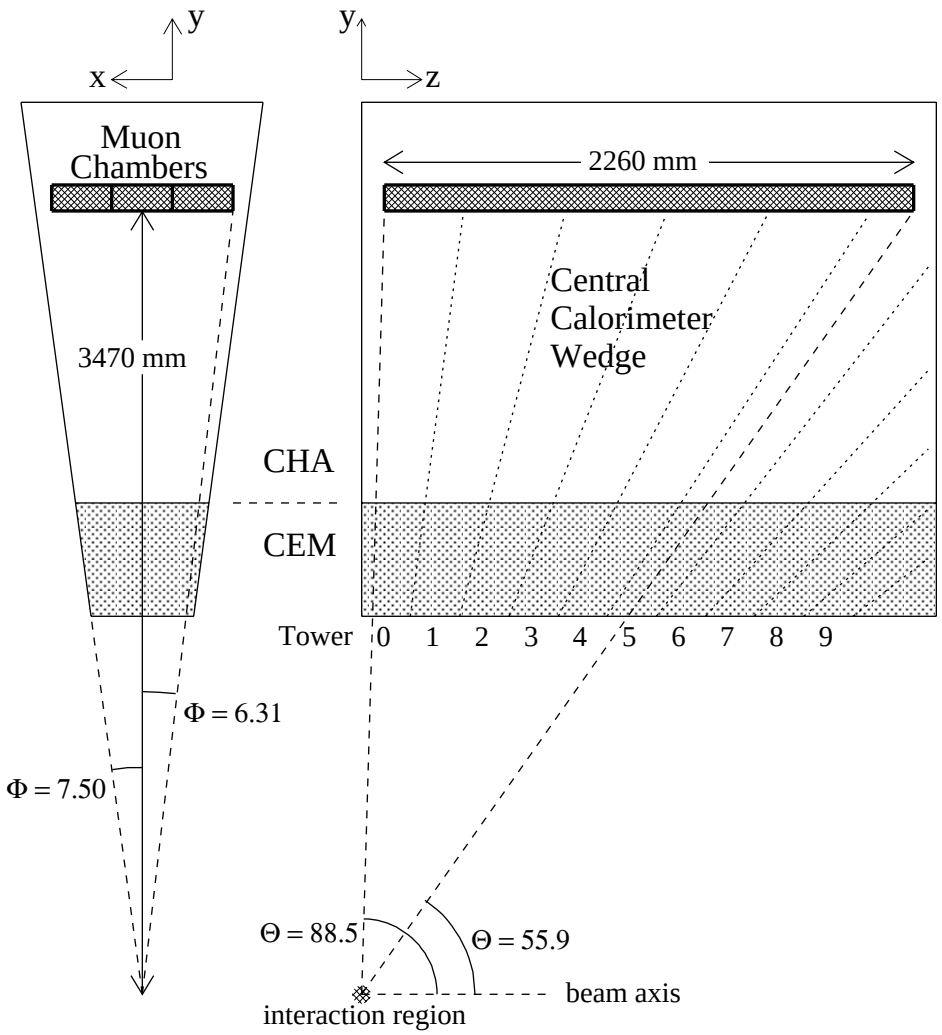


Figure 3.6: Geometry of a central calorimeter wedge and its towers.

	Central EM (CEM)	Central Hadron (CHA)	Endwall Hadron (WHA)
Coverage ($\|\eta\|$)	0 - 1.1	0 - 0.9	0.7 - 1.3
Tower Size ($\delta\eta \times \delta\phi$)	$0.1 \times 15^\circ$	$0.1 \times 15^\circ$	$0.1 \times 15^\circ$
Module Length	250 cm	250 cm	100 cm
Module Width	15°	15°	80 cm
Number of Modules	48	48	48
# Layers	31	32	15
Active Medium	polystyrene scintillator	acrylic scintillator	acrylic scintillator
Thickness	0.5 cm	1.0 cm	1.0 cm
Absorber	Pb	Fe	Fe
Thickness	0.32 cm	2.5 cm	5.1 cm

Table 3.2: Characteristics of the central and endwall calorimeters.

As depicted in Figure 3.5, within the CEM resides proportional strip chambers, called the CES. Located at 5.9 radiation lengths, the CES are positioned at the point in the CEM where maximal average electromagnetic shower development occurs. The CES have sense wires running parallel to the beamline that provide shower position in $r-\phi$ and perpendicular sense wires that provide position measurement in z. The CES is used primarily in this analysis in electron identification, which will be discussed in Section 5.1.1.

The plug ($|\ \eta\ |<2.4$) and forward ($|\ \eta\ |<4.2$) calorimeters are used to measure energies in the regions of the detector closer to the beamline. The energy measurements in these portions of the calorimetry are performed similarly as in the central calorimetry. One significant difference is the active material employed: in the plug

	Plug		Forward	
	EM (PEM)	Hadron (PHA)	EM (FEM)	Hadron (FHA)
Coverage($\|\eta\|$)	1.1 - 2.4	1.3 - 2.4	2.2 - 2.4	2.3 - 2.4
Tower Size ($\delta\eta \times \delta\phi$)	$0.09 \times 5^\circ$	$0.09 \times 5^\circ$	$0.1 \times 5^\circ$	$0.1 \times 5^\circ$
Active Medium	Proportional tube chambers with cathode pad readout			
Tube Size	0.7×0.7 cm^2	1.4×0.8 cm^2	1.0×0.7 cm^2	1.5×1.0 cm^2
# Layers	34	20	30	27
Absorber	Pb	Fe	94% Pb, 4% Sb	Fe
Thickness	0.27 cm	5.1 cm	0.48 cm	5.1 cm

Table 3.3: Characteristics of the plug and forward calorimeters.

and forward regions, proportional tube arrays containing argon-ethane gas are used to record the energy dissipated in the particle cascade caused by the incident particles. The plug and forward calorimeters are divided into electromagnetic (PEM,FEM) and hadronic (PHA,FHA) sections. Table 3.2.2 lists detailed characteristics of the plug and forward calorimeters.

3.2.3 The Muon Detectors

After escaping the tracking system, muons interact minimally within the calorimetry. But they do not escape detection in the CDF detector. A detector system specifically designed to identify muons is located at larger radii. CDF has three separate muon detectors: the central muon detector (CMU), the central muon upgrade (CMP),

and the central muon extension (CMX). All of the muon chambers are arrays of single wire, rectangular drift tubes.

The CMU is housed within the central calorimeter wedges, directly behind the CHA. The CMU covers the pseudorapidity range $|\eta| < 0.6$. It should be noted that the CMU has only 85% coverage in azimuth; this is due to 2.4° gaps between detector modules. From Figure 3.6 one can see the geometry of the CMU detectors beyond the central calorimeter. Each wedge contains three CMU chambers; each chamber contains four radial layers of four drift tubes. Each layer is offset by ($\sim$ 2 mm) to provide unambiguous ϕ measurement.

Four more layers of drift chambers, the CMP, reside beyond the CMU. A half meter of steel shielding resides between the CMU and the CMP chambers. This shielding is designed to reduce the number of hadrons which escape the hadronic calorimetry and proceed to the CMU and CMP. These so-called "punch-through" hadrons are a source of fake muons. The CMP operate in the same way as the CMU. CMP drift tubes are similarly staggered to remove position measurement ambiguity. In the region $|\eta|<0.6$, approximately 85% of the solid angle is covered by the CMU, 63% by the CMP, and 53% by both.

Muons in the range $0.6 < |\eta| < 1.0$ are detected by the CMX system. The CMX is comprised of four conical arches of drift tubes; layers of scintillator (the CSX) sandwich the CMX drift tubes and aid in the identification of real muons. The CMX covers 71% of the solid angle in the $0.6 < |\eta| < 1.0$ region.

3.2.4 Event Triggers

In Run 1 at CDF, bunch crossing occurred approximately every 3.5 μs with an average of one interaction per crossing during Run 1a, and three per crossing in Run

1b. That means between 280,000 and 840,000 collisions occurred at CDF in each second of operation. The Run 1 CDF detector had approximately 150000 channels available for readout on each and every event. Because of the large bandwidth and storage requirements, it is impossible to permanently record each event record: event size allows only ∼5-10 events to be written to magnetic tape per second. It would also be imprudent from a physics point of view, since not every collision has enough energy transfer to create an interesting physics process: think of a glancing blow between constituent quarks of the incident p and $\overline{p}$; such an event is fairly common and is not typically worthy of further scrutiny. It is clearly wise to judiciously select and record only those events that meet some minimal criteria that qualify them for further study after collection.

This is the task of the CDF event trigger, a three-level system that selects these minimally interesting events. The trigger system reduces the rate of events to be recorded to a much more manageable size. However, in the high rate Run 1 environment, this goal had to be tempered by the need to minimize the amount of "dead time", periods in which collisions occur but the trigger is incapable of processing the new data.

Each successive level of the trigger processes fewer events than the preceding level but with greater sophistication and more processing time. The Level 1 and Level 2 triggers are implemented in hardware, while Level 3 is implemented in software.

The Level 1 trigger is indeed deadtimeless, taking less than 3.5 μs to make its decision. The trigger is based on identification of raw energy clusters or an energy imbalance in the calorimeters or muon stubs in the muon chambers. These are the objects that can eventually be reconstructed to be electrons, neutrinos, photons, jets and muons. Level 1 selection reduces the event rate from 280 kHz down to

approximately 1 kHz. At Level 1, the energy clusters are required to exceed some region-specific threshold.

At Level 2 the trigger first makes use of tracking information. The central fast tracker (CFT) is a hardware processor that uses CTC hits to reconstruct high momentum tracks in $r-\phi$. The CFT has a momentum resolution of $\delta P_T/P_T^2 = 3.5\%$. Tracks found by the CFT are matched to clusters in the CEM to form electron candidates or to tracks in the muon chambers to form muon candidates. The Level 2 trigger also exploits more sophisticated calorimeter information; calorimeter clusters are formed by searching for a seed tower above a certain threshold and adding in neighboring towers which are over a lower threshold. The E_T, η , and ϕ are calculated for each energy cluster. Level 2 takes ~ 20 μs to make a decision during which time the detector ignores subsequent crossings. This is one source of deadtime, amounting to a few percent. A larger dead time fraction might be expected given the bunch spacing of 3.5 μs, but one must remember that the raw event rate is reduced by Level 1. The dead time calculation from Level 2 must also fold in effects of event pile-up from Level 1. The Level 2 trigger reduces the event rate to approximately 20–30 Hz.

The final portion of the multi-level trigger, Level 3, is implemented in software running on a PC farm. The Level 3 trigger software reconstructs events using a simplified version of the CDF offline code, which is a suite of software packages that takes raw detector data and reconstructs useful physics objects like jets, electrons, muons, etc. All events which pass the Level 3 trigger are written to 8 mm tape with a typical output rate being 3 - 5 Hz for Run IA and 8 Hz for Run IB.

CHAPTER 4

AN INTRODUCTION TO NEURAL NETWORKS

Each p-$\overline{p}$ collision is characterized by a number of kinematic and spacial variables pertaining to the final state particles we see in the detector such as the energy of a final state lepton, the opening angle between two jets, the reconstructed mass of a lepton and a neutrino, etc. The typical method of selecting events for a particular analysis imposes a requirement on several different event quantities. The specific requirement is determined from some known sample, either a simulated or control data set. Each requirement is typically independent. These criteria are called *rectangular cuts* because the sub-region of the entire phase space of possibilities from which events are ultimately selected is a well-defined multi-dimensional box having orthogonal boundaries defined by the independently-treated variables in the event selection.

Unlike rectangular cuts, there is a class of tools, called *multivariate techniques*, that have the virtue of being able to identify classes of events via the *correlations* between several variables in multiple dimensions. Such techniques exploit information from several variables simultaneously, while basic rectangular cuts only incorporate information from variables one at a time. The use of the correlations between the variables changes the region from which events are selected from a multi-dimensional box to some more complicated structure, such as, for example, a region with curved boundaries. Artificial neural networks are one such class of multivariate technique.

Neural networks are capable of, among other things, using information from a variety of variables in classification problems where events need to be cast into one of several bins.

Neural networks are also a form of *artificial intelligence*. A neural network is capable of classifying events only after presentation of numerous training vectors. During the presentation of this training set in which the correct classification answer is known, the functional form of the neural net adapts itself in order to better classify events. In this way the network *learns* to classify events, and therefore is related to tools in the realm of artificial intelligence.

The goal in many applications is to exploit information from a variety of variables simultaneously in an effort to identify a particular class of inputs. In principle one could do this on one's own, scouring through the events in search of a convenient region of phase space in which one can design a complex cut retaining most of one's signal class and rejecting as much background as possible. However, this approach quickly becomes too unwieldy for even a reasonably small number of input parameters. The neural network, with its ability to learn and adapt itself, is simply a tool that performs this task in an automatic and efficient way.

A comprehensive treatment of neural networks will not be presented here. Good resources exist that discuss neural networks in general (see for example [45]).

4.1 Neural Network Model

Artificial neural networks (NNs) are modeled after biological systems. The human brain is a biological neural network made of approximately 10^{11} nerve cells, or *neurons*. Each neuron consists of three main parts. The *cell body* receives an input

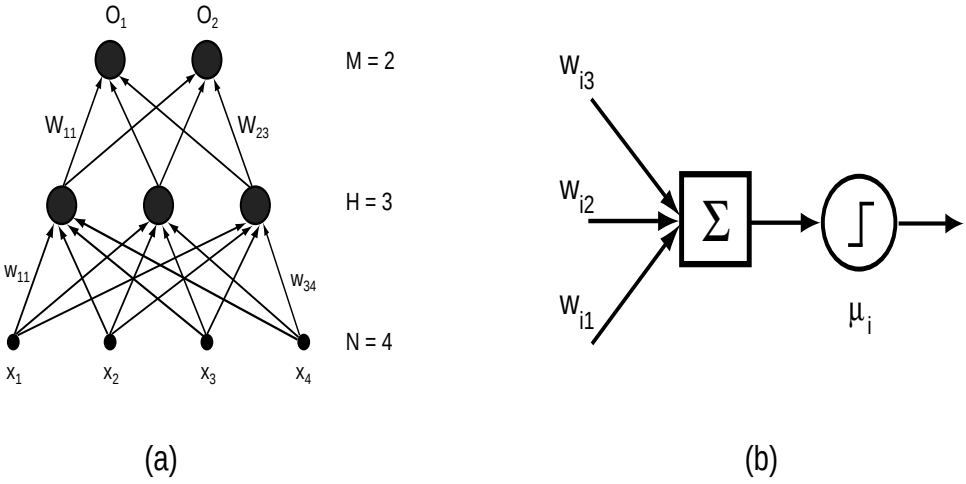


Figure 4.1: Visual aid of McCulloch - Pitts model(a), and schematic of a single node and its function(b).

from the surrounding cells by means of electric signals (ion flow). These electrical signals arrive at the cell body via *dendrites*. If the input signals exceed a threshold value then an electric discharge takes place. This outgoing message is transmitted via *axons* which are connected to the dendrites of other neurons. The first mathematical model of neural communication was developed by McCulloch and Pitts [47] in the 1940s. In this model, a neuron is seen as a binary threshold unit, and its connections to the surrounding neurons are represented by real numbers, called *weights*. Figure 4.1 [56] contains a drawing of the McCulloch - Pitts model. In this early model, the $j-th$ neuron performs the following calculation:

$$n_j = \Theta(\sum_i w_{ji} n_i - \mu_j), \qquad \text{where} \quad \Theta(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases} \tag{4.1}$$

In this equation, n_j is the output of neuron j and represents the state of the neuron: $1 = fired$, $0 = not\ fired$. The weight w_{ij} represents the strength of the connection between neuron j and neuron i. The output of neuron i is denoted by n_i, which is

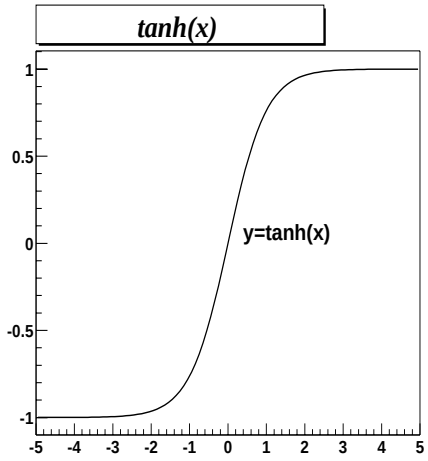


Figure 4.2: The sigmoid hyperbolic tangent function that we employ as our activation function.

in fact an *input* for neuron j. Finally, μ_j is the threshold value for neuron j; the weighted sum of the inputs must reach or exceed this threshold for the neuron to fire.

Models for NNs have advanced since the early pioneers, but the functional form of NNs has not changed dramatically. In modern implementations, sigmoid functions have replaced the step function utilized in Equation 4.1 as the so-called *activation function*. Sigmoids are a class of functions that can have a more gradual turn on than the step function. An example of a sigmoid that is a popular activation function is the hyperbolic tangent, shown in Figure 4.2.

NN architecture is represented by layers of nodes (the neurons in the biological example). An *input layer* of nodes represents the input information with which the NN will process events. The values of the N input nodes define an N-dimensional vector, denoted $\vec{I}$. In general there is some number of *hidden layers* of nodes; these layers are called hidden because they have no direct connection to information outside the NN. And finally there is the *output layer* which provides the event-by-event response

to the input information. The output layer has as many nodes as is necessary for the problem at hand; in the case of M output nodes, the output of the NN is denoted as an M-dimensional vector, denoted $\vec{M}$. *Feed-forward* NNs allow information flow in one direction, from input to output layer; other architectures allow for feedback loops, but this class of NN is not considered here.

4.2 Neural Network Learning

A NN can be thought of as a non-linear function that maps vectors of an input class to a user-defined target output vector. The NN takes as its input parameters certain quantities that characterize each class of events. One application of a NN in experimental particle physics is the classification of p-$\overline{p}$ collisions. We can consider different classes of collisions by the different final states that are produced. In such an application, one can use event wide kinematic variables, such as jet E_T, $\not\!\!E_T$, H_T, *etc.*, to differentiate between different production mechanisms.

One must however have example events of each class to present to the NN for the purposes of weight adjustment in order to differentiate between classes. One associates a target output vector for each class, which we will define to be $\vec{T}$.

Initially the weights and thresholds of the NN are random. Target NN output values for each class of event are supplied by the user, and via iterating over examples from each class, the form of this non-linear function evolves by minimization of a mean square error. The presentation of events and error minimization procedure is called *learning*; during the learning phase, the NN is being trained to properly identify each input class.

Each of the nodes above the input layer performs the sum

$$O_j = g(\sum_i W_{ji} \cdot Input_i - \theta_j) \tag{4.2}$$

where $g(x) = \tanh(x)$, O_j is the j-th output node, and W_{ji} and θ_j are the weights and threshold input to the node in question.

Each node performs this calculation; the result is then input to the next layer of nodes, and the process continues to the output layer. At the output layer, the resulting output vector is then compared to the target vector, and the deviation between desired and actual response is evaluated. The error measure utilized in the learning phase can be given by

$$E = \frac{1}{2}|\vec{T} - \vec{O}|^2 \tag{4.3}$$

where $\vec{T}$ and $\vec{O}$ are the target and actual output vectors. This error measure is then minimized; the weights and thresholds of the NN are updated via:

$$\omega_{t+1} = \omega_t + \Delta\omega_t \tag{4.4}$$

where

$$\Delta\omega_{t+1} = -\eta\frac{\partial E_t}{\partial\omega} + \alpha\Delta\omega_t \tag{4.5}$$

Here ω represents the complete set of weights and thresholds after update t in the NN, and E represents the error from Eqn. 4.3.

Two parameters in the above equations are in general unconstrained. η is known as the *learning rate*. Large η values indicate that the response of the weights in the NN during each update is scaled to large values; smaller values indicate a smaller response in the weights. α is known as the *momentum*; this parameter allows learning history information to influence the present weight update. Again, large α values indicate a

larger impact on previous updates; smaller α values remove the influence of previous weight updates.

Each node above the input layer also has a *threshold*, θ_j, associated with it. The weights and thresholds are the parameters of the NN that are adjusted in the learning phase when the mapping of input class to output vector is taught to the NN.

The weights and thresholds of a NN are its degrees of freedom. The number of degrees of freedom in a NN is given by:

$$N_{DoF} = N_H * (N_I + N_O + 1) + N_O \tag{4.6}$$

where N_I, N_H, and N_O are the number of input, hidden and output nodes of the NN. In general one needs a training sample containing a number of events which is 20-40 times the number of degrees of freedom of the NN for proper training.

By the iterative presentation of events of known origin, the NN adjusts its form in an attempt to match the desired output as best it can. An *epoch* is defined as one complete iteration through all available training events. In general, training proceeds through several thousands of epochs.

The NN is in general presented with two types of independent training samples:

- the learning sample, a sample of events whose origin are known and are used in NN training
- the generalization sample, which is a completely independent set of known events that is used as a monitor of network performance.

Training can be defined to be complete in several ways; typically the error distribution in the generalization sample plateaus after some number of event presentations, and this point is typically chosen to halt the training. One must take care though to

avoid over-training a NN. Over-training occurs when the learning process starts to focus on the specific qualities of the learning sample, that are not representative of the general class of events. Over-training is indicated by a gradual rise in the mean square error as the number of training event presentations increases.

4.3 Interpretation of Results

It has been shown [49] that under certain circumstances, the values of the output vector can be interpreted as Bayesian *a posteriori* probabilities. In order for this interpretation to be valid

- the learning must be accomplished with the minimization of a mean square error as described above, or some other similar error measure
- the training data must properly model the physical observable spectrum of events
- the output target vectors are orthogonal for each input class

The consequence provides a useful tool. Let us use an example relevant for our purposes. Consider a class of collision events, defined by some final state of particles. Let us assume that there are three mechanisms by which this final state can be produced. If one has consistent simulated events from each of the three classes of events, one can train a three output NN for event classification. The target output vectors should be (1,0,0), (0,1,0) and (0,0,1) for the three classes of event. After sufficient training, the value of each output node should correspond to the probability that an event belongs to one of the three classes.

CHAPTER 5

BASELINE EVENT SELECTION

Recall from Section 2.5 that the focus of this search is Higgs production in association with a $W^{\pm}$ boson in the range 100 $GeV/c^2 < M_H < 150$ GeV/c^2. We choose to look for events in which the associated $W^{\pm}$ decays leptonically; this helps reduce the number of background events in the sample that is ultimately selected.

In the range 100 $GeV/c^2 < M_H < 150$ GeV/c^2 the Higgs decays predominantly to a b-quark pair. These b quarks produce jets of particles that are detected in the CDF calorimetry. Large jet production rates necessitate efficient discrimination of b jets from light-quark jet background.

Therefore, the Higgs signature that is the focus of this search includes a highly energetic electron or muon, a neutrino, and two b jets. This chapter outlines how events possessing this signature are identified in the CDF Run 1 environment. Simulation studies were performed to determine the efficiency of our event selection on signal events; a discussion of these studies and the expected signal yields is below. Finally, we outline the expected background contribution.

5.1 Identification of Higgs Signature

Charged leptons from $W^{\pm}$ decay are critical in identifying $WH \rightarrow \ell\nu b\overline{b}$ events. Electrons and muons can be produced via a variety of mechanisms at the Tevatron,

including the semileptonic decays of B and C hadrons. Fortunately, the leptons from $W^{\pm}$ decay differ from those produced in the semileptonic decay of B- or C-hadrons. For instance, the leptonic $W^{\pm}$ decay products are typically isolated from jet activity. Also, leptons from $W^{\pm}$ decay are typically more energetic than leptons from the decay of hadrons. There are several criteria that exploit these differences, and we demand that candidate events satisfy these criteria in order to be considered in this analysis. The lepton identification variables are discussed in the following sections. Jet identification is also discussed below with special emphasis on the identification of jets coming from b quarks. There are also event topologies that are starkly irrelevant to this analysis that one can reject with minimal information. These event vetos are also presented in the following sections.

5.1.1 Electron Identification

Electrons are identified by a track in the CTC and energy deposition in the electromagnetic calorimetry. This analysis considers electron candidates only in the central region $|\eta|<1.0$. Electron candidates in the plug region suffer from limited tracking information, thus increasing the probability of electron mis-identification, or "fake" electrons. Fiduciality cuts [22] are applied that require that the shower from the incident electron in the calorimetry is sufficiently far away from wedge boundaries or other regions of poor energy measurement. The fiducial volume of the CEM covers approximately 84% of the solid angle in the region $|\eta|<1.0$. Electron candidates are required to be isolated from other energy depositions in the calorimetry. A three-dimensional reconstructed CTC track is also required to point to the CEM cluster.

This analysis requires that the electron candidate have a CEM cluster with $E_T>20$ GeV; this high E_T threshold assists in rejecting electrons from spurious sources like

semileptonic hadron decays. A large sample of Run 1 $Z \rightarrow e^+e^-$ events was utilized to establish the complete criteria for electron identification [22]. The following variables were investigated in the $Z \rightarrow e^+e^-$ sample and are used in this analysis for electron identification:

- E/P - the ratio of energy E deposited in the electromagnetic calorimeter to the track momentum P.

- E_{HAD}/E_{EM} - the ratio of the energy E_{HAD} deposited in the hadronic calorimeters to the energy E_{EM} deposited in the electromagnetic calorimeter (within the electron cluster).

- L_{shr} - the lateral shower profile for electrons, which compares the energy in adjacent CEM towers to the seed tower for the cluster. L_{shr} is defined as:

$$L_{shr} \equiv 0.14 \sum_i \frac{E_i^{obs} - E_i^{exp}}{\sqrt{(0.14\sqrt{E})^2 + \sigma^2_{E_i^{exp}}}} \tag{5.1}$$

  where the sum is over the towers adjacent to the seed tower, E_i^{obs} is the electromagnetic energy measured in tower i, E_i^{exp} is the energy expected from test beam electrons, $0.14\sqrt{E}$ is the CEM energy resolution, and $\sigma_{E_i^{exp}}$ is the uncertainty in E_i^{exp}.

- $|\Delta x|$ and $|\Delta z|$ - the separation between the extrapolated CTC track position and the CES chamber position in the $r-\phi$ plane, and $r-z$ plane respectively.

- χ^2_{strip} - the χ^2 is a goodness-of-fit measurement from the comparison of the CES shower profile in the z direction between the electron candidate and test beam electrons.

- z-vertex match - the distance along the beam axis between the primary vertex (interaction point) and the reconstructed track. If there is more than one vertex

in the event, the distance to the closest vertex is used. The closest vertex is required to have $|z| < 60\ cm$ measured from the $z = 0$ point.

- Iso - electron isolation, defined by $Iso = (E_T^{cone} - E_T)/E_T$, where E_T^{cone} is the sum of the transverse energies in towers within a cone of $R = 0.4$ around the track direction, and E_T is the electron cluster transverse energy.

The values for the requirements on each parameter are listed in Table 5.1. Electrons from $W^{\pm}$ decay are expected to resemble electrons from Z decay, so the application of identification criteria from the $Z \to e^+e^-$ analysis is reasonable here. This identification criteria was determined in the $Z \to e^+e^-$ analysis to be $\simeq$81% efficient in identifying electrons.

Variable	Requirement
E_T	>20 GeV
E/P	<1.8
E_{HAD}/E_{EM}	<0.05
L_{shr}	<0.2
$\|\Delta x\|$	<1.5 cm
$\|\Delta z\|$	<3.0 cm
χ^2_{strip}	<10
z-vertex match	<5.0 cm
Iso	<0.1

Table 5.1: Electron identification variables used in this analysis; fairly standard for similar analyses.

It should be noted that photons interacting with the material in the detector could convert to an e^+e^- pair, either of which could satisfy our electron selection. A *conversion* is indicated by the presence of an oppositely charged track that shares the same vertex as the candidate electron. These conversion electrons are excluded from

the selected sample via requirements on the e^+e^- invariant mass and the location of the vertex from which the candidate electron originated. This conversion removal vetos 91% of such events.

5.1.2 Muon Identification

In this analysis, muons are also restricted to the central regions of the detector, $|\eta|<1.0$. Candidate muons are required to have a CTC track matched to at least one muon *stub* in the CMU, CMP, or CMX systems (a stub is defined to be one fired muon chamber in one of the subsystems). Cosmic rays are a significant source of fake muons, as are hadronic showers that escape the calorimetry and are recorded in the muon system.

The requirement was imposed that each muon candidate have $p_T>20\ GeV/c$ to reduce the number of soft leptons from hadronic B and C decays. In the same spirit as in the determination of the electron identification variables, a $Z \rightarrow \mu^+\mu^-$ data sample was used to determine the muon selection criteria [22].

- E_{EM}, E_{HAD} - the energy deposition in the electromagnetic, and hadronic calorimeters, which should be small since muons traverse much of the calorimetry without depositing much energy.

- d_0 - the impact parameter. This is defined as the distance of closest approach between the reconstructed CTC muon track and the beam axis in the $r-\phi$ plane.

- $|\Delta x|$ - the CTC track-stub separation.

- z-vertex match - defined in a similar way as in the case of electrons.

- Iso - muon isolation. $Iso = (E_T^{cone} - E_T^{tower})/P_T$, where E_T^{cone} is the sum of the transverse energies in towers within a cone of $R = 0.4$ around the track direction, E_T^{tower} is the transverse energy measured in the tower associated with the muon track, and p_T is the transverse momentum of the muon track.

The requirements on each of these parameters are listed in Table 5.2. Fiducial cuts are also placed on the muon candidates; recall from Section 3.2.3 that the different muon subsystems do not have complete coverage in the region $|\eta|$<1.0. The combined muon identification efficiency was determined to be $\simeq$93%.

Variable	Requirement
p_T	>20 GeV/c
E_{EM}	<2 GeV
E_{HAD}	<6 GeV
$E_{EM} + E_{HAD}$	>0.1 GeV
d_0	<3 mm
$\|\Delta x\|_{CMU}$	<2 cm
$\|\Delta x\|_{CMP}$	<5 cm
$\|\Delta x\|_{CMX}$	<5 cm
Iso	<0.1
z-vertex match	<5 cm

Table 5.2: Muon identification variables used in this analysis; fairly standard for similar analyses.

5.1.3 Neutrinos and Missing Transverse Energy

As discussed in Section 2.5 the presence of neutrinos can be inferred by the energy imbalance in the detected final state particles, the so-called missing E_T or $\not\!\!E_T$. The total energy measured in the transverse plane of all final state particles should be zero, assuming that the energy in the transverse planes of the incident beams is zero.

A more appropriate quantity to use is momentum, given its vector nature. But since calorimeters measure energy, for these purposes we consider the energy measurements to have a direction as well. Note that at high energies, momentum and energy are equivalent up to a factor of c; for this reason, our employment of energy as a vector quantity is warranted. The raw $\not\!\!E_T$ is defined by

$$\not\!\!E_T^{raw} + \sum^{alltowers} E_T^{raw} = 0 \tag{5.2}$$

The sum is over all towers within $\mid \eta \mid$<3.6. The raw $\not\!\!E_T$ is corrected in the case that the primary lepton in the event is a muon. This is necessary because the energy deposition of the muon in the calorimeter is not indicative of the energy it carries off. In this case, the p_T of the muon is added component-wise to the raw E_T quantity, the muon's contribution to the CEM energy is removed and the resulting quantity defines the corrected $\not\!\!E_T$. Electron events have no need for such a correction, and so their final $\not\!\!E_T$ is equivalent to $\not\!\!E_T^{raw}$.

5.1.4 Jet Identification

Recall from the discussion in Section 2.5 that jets are sprays of particles from the hadronization of the quarks created in the p-$\overline{p}$ collisions that are the focus of this analysis. Jets will deposit most of their energy in a cluster of calorimetry towers. Cluster identification is an iterative process that starts from a seed tower possessing at least 3 GeV of transverse energy. Adjacent towers with E_T>1 GeV are added to the cluster; after each tower is included in the sum, the centroid of the cluster is re-calculated, weighting the different towers according to their energy. Only towers

within a cone of radius $R = 0.4$ are included in the calculation of the raw jet energy, E_T^{raw}. We demand that these jets have E_T>15 $GeV$and $\mid \eta \mid$<2.0.

The process $WH \rightarrow \ell\nu b\bar{b}$ contains at least two jets. We therefore require our selected events to have exactly two jets. These two jets may overlap such that they manifest themselves in the CDF detector as a single cluster; gluon radiation from the initial or final state partons may add additional jets. So although our signal events are not guaranteed to possess exactly two jets, one can reject a significant amount of background events if one restricts the search to events with only this jet multiplicity. For example, $t\bar{t}$ events are expected to have more final state hadrons than WH production; thus restricting the analysis to the 2 jet bin removes much of the $t\bar{t}$ background from consideration.

The jets from Higgs decay originate from b quarks. One of the major background sources in this analysis comes from $W^{\pm}$ boson production in the presence of light quark jets; therefore it is imperative to be able to differentiate b jets from light quark jets. This is accomplished by identifying secondary vertices in the SVX (discussed in Section 3.2.1). Recall that the long lifetime of the b quark leads to a macroscopic B-hadron decay length. These decay lengths depend on the specific hadron but in general are in the range $c\tau \approx 5mm$. The process of identifying secondary decay vertices is one form of *b-tagging*, a crucial tool in Tevatron searches.

In order to identify secondary vertices, one must first identify the primary event vertex. This is accomplished via a weighted fit of the z position from the VTX (see Section 3.2.1) and the SVX tracks in the event that are not consistent with originating from a secondary vertex. If multiple primary interactions exist in an event, then the primary vertex is defined to be the one with largest amount of track p_T associated with it.

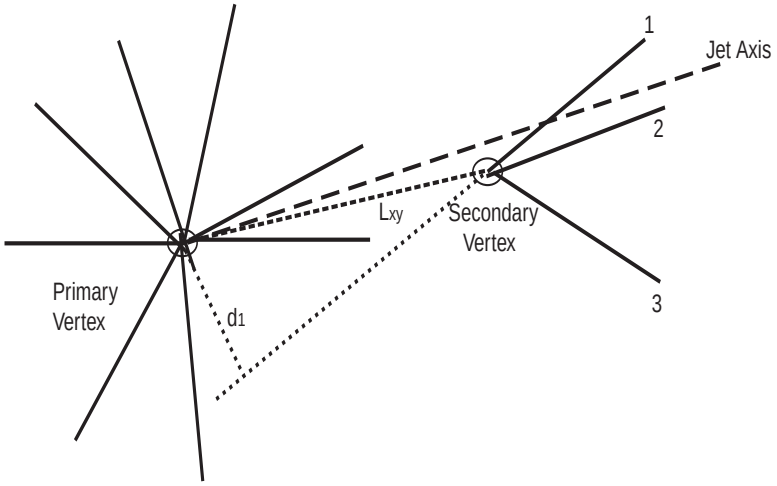


Figure 5.1: Diagram of the production of a secondary vertex from b quark production.

Jets are considered for b-tagging if they possess a raw E_T>15 GeV, are within $|\ \eta\ |$<2.0 and have at least two matching p_T>2.0 GeV/c SVX tracks. The matching criteria for SVX tracks requires that the opening angle between the jet and track is less than 35°. The SVX tracks are further required to have an impact parameter significance, $|d|/\sigma_d \geq 3$, where σ_d is the uncertainty on the impact parameter, d. A fit is performed in the x-y plane for the secondary vertex; the result is the signed decay length in the transverse plane, L_{xy}, between the primary and secondary vertex (shown in Figure 5.1). The sign of L_{xy} is positive (negative) if the secondary vertex and its associated jet are in the same (opposite) hemisphere. A jet is b-tagged if $|L_{xy}|/\sigma_{L_{xy}} \geq 3.0$ and $L_{xy} > 0$. In the jargon of CDF, this b-tagging algorithm is called SECVTX. We demand at least one of the 2 jets in our selected events to possess a SECVTX tag.

It should be noted that the per-jet tagging efficiency was measured in Run 1 in a sample enriched with b decays. The SECVTX algorithm described above was determined to be 37% efficient at identifying jets from b decays [38]. One should

Electron cuts	Muon cuts	
$E_T > 10\ GeV$	$P_T > 10\ GeV/c$	
$HAD/EM < 0.12$	If associated with a stub	If no stub
$E/P < 2.0$ (if in CEM)	$EM < 5\ GeV$, $HAD < 10\ GeV$	$EM < 2\ GeV$, $HAD < 6\ GeV$
$Iso < 0.2$	$\|\Delta x\|_{CMU,\ CMP,\ or\ CMX} < 5\ cm$	$\|\eta\| < 1.1$
	$Iso < 0.1$	$Iso < 0.2$

Table 5.3: Requirements for second lepton that qualify events for Z boson removal.

also be cautioned that, as with any identification algorithm, b-tagging is subject to spurious sources of secondary vertices not due to b-quarks. These sources include c, τ, and long-lived hadrons, such as K_s^0 and Λ.

5.1.5 Topology Vetos

There are classes of physics events that are clearly not relevant to this analysis. Some event topologies are clearly different; it is a straightforward exercise to eliminate these events from our selected sample. For example, events consistent with the decays $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ should not be considered part of our signal sample. We remove these Z events by identifying a primary lepton as described above; the event is further required to have a secondary lepton passing the cuts outlined in Table 5.3 [56]. Such events are excluded if the invariant mass of the lepton pair is around the Z mass: $75 < M_{\ell\ell} < 105\ GeV/c^2$.

A significant background source in this WH search is $t\bar{t}$ production. One decay channel available in $t\bar{t}$ production has both the $W^{\pm}$ from each t-quark decaying leptonically, the so-called dilepton mode. These events have two oppositely-charged electrons or muons, and because of these distinct decay products, such events can be easily removed. Rejection of these events is accomplished in two parts:

- events are rejected if they possess two oppositely-charged p_T>20 GeV/c leptons, $\not\!\!E_T$>25 $GeV$and at least two E_T>20 GeVjets

- events are rejected if they possess an isolated, p_T>15 GeV/c track with charge opposite the primary lepton.

5.1.6 Summary

The event selection chosen for this analysis attempts to isolate the $\ell\nu b\bar{b}$ channel. Events are required to:

- have at least one E_T (p_T) > 20 GeV central electron (muon)

- have $\not\!\!E_T$ > 20 GeV

- pass Z boson removal

- survive dilepton $t\bar{t}$ removal

- have exactly two E_T > 15 GeV, $\mid \eta \mid$ < 2.0 jets

- have one or more SECVTX tags among the two required jets

The lepton and $\not\!\!E_T$ requirements attempt to isolate the $W^{\pm}$ decay. Events are also required to have satisfied the high p_T lepton trigger and have a z vertex within 60 cm of the center of the detector.

The above selection removes a good deal of background processes. However formidable amounts remain, mostly from $W + b\bar{b}$ and $W + c\bar{c}$ production, W+jj production with a fake secondary vertex tag, $t\bar{t}$, single top production, and diboson production.

5.2 Expected Event Yields

Given the selection criteria that are utilized to identify the $\ell\nu b\bar{b}$ final state, one must now understand the expected event yields from signal and background in the Run 1 data set. We use Monte Carlo simulations to predict the yields for signal; the background prediction is taken directly from another analysis. Each of these prediction is discussed below.

5.2.1 Monte Carlo Simulations

An invaluable tool in performing searches of this type are Monte Carlo event simulations. These programs provide experimenters with a potentially unlimited simulated data set, with which one can study and optimize event selection criteria and determine the expected yield of signal and background events. QCD and Electroweak theory are used to simulate p-$\overline{p}$ interactions and the decays of the particles produced in the hard scatter. In a typical event, one has access to each particle's identity, mass, four-momentum, vertex position, and a list of parent and daughter particles. Complicated p-$\overline{p}$ events, like those from WH production and its backgrounds, typically contain a few hundred particles, many of which survive the showering process into the final state.

Because of the complicated nature of the QCD theory, Monte Carlo simulations are typically valid only to leading order; higher order contributions to processes have been included in some recent Monte Carlo event generators, but these have not been utilized here. B-hadron decays in this analysis are simulated using the QQ Monte Carlo package developed by the CLEO collaboration [29]. Other particle decays are determined by the branching fractions reported by the Particle Data Group [31].

Two Monte Carlo event generators are employed in this analysis. Pythia v6.129a [36] is a leading order Monte Carlo event generator and was used to model the signal process $WH \rightarrow \ell\nu b\bar{b}$ for six hypothetical M_H values in the range 100 $GeV/c^2 < M_H < 150$ GeV/c^2. Recall from Figure 2.5 that in WH production, two incident quarks annihilate to form an off-shell $W^{\pm}$ boson. These two quarks could be valence quark constituents of the proton and anti-proton (consider a $\overline{u}$ - d interaction). It is also possible for gluons and quarks from the "parton sea" within each proton to interact. The probability distributions of partons within the proton are not known exactly; in the Pythia samples, the GRV-94L parton distribution functions are utilized [37]. Pythia is also a useful generator because switches are provided with which one can turn on/off gluon radiation. The radiation of gluons off of initial and final state partons is not well understood, so the opportunity to switch this effect on and off allows one to gauge the effect of the particular radiation model employed. Pythia was used to model several of the background sources in addition to modeling our signal WH events.

ALPGEN [50], another leading order generator was used for the remaining background processes. ALPGEN calculates the hard interaction, *i.e.* the scattering of the incoming partons to the subsequent leptons and quarks; the output of this step is then interfaced to another Monte Carlo program, such as Herwig [51], where the leptons and quarks from the hard scatter are subjected to parton showering. The background samples generated with ALPGEN utilized the CTEQ5L [64] parton distribution functions. Although one has the capability to do so within ALPGEN, no minimum p_T, maximum $\mid \eta \mid$, or minimum ΔR requirements were placed on the outgoing partons in any of the samples generated with this program. ALPGEN also has the option of adjusting the Q^2 scale to which the hard scatter probes; the samples

used in this analysis had the Q^2 scale set to $Q^2 = M_V^2 + \sum(p_T^{jet})^2$, where $V = W, Z$ depending on the process desired.

This simulated particle data can then be input into a CDF detector simulation, called QFL [30]. This sophisticated detector simulation models the response of the CDF detector to the final state particles provided by the Monte Carlo event generators. The simulation approximates the tracking, b-tagging, calorimetry, and muon detection in the Run 1 CDF detector. The detector simulation is tuned according the response in electron test beam studies and actual p-$\overline{p}$ collision data.

5.2.2 Signal Acceptance

The WH event detection efficiency is calculated as follows:

$$\varepsilon_{WH} \equiv \varepsilon_{sel}^{WH} \cdot \varepsilon_{tagging} \cdot \varepsilon_{zprim} \cdot \varepsilon_{trig} \cdot \varepsilon_{lepID} \tag{5.3}$$

where

$\boldsymbol{\varepsilon_{sel}^{WH}} \equiv$ the kinematic event selection efficiency

$\boldsymbol{\varepsilon_{tagging}} \equiv$ the efficiency of requiring events to pass the tagging requirements

$\boldsymbol{\varepsilon_{zprim}} \equiv$ the efficiency of the cut $|z_{prim}| < 60cm$

$\boldsymbol{\varepsilon_{trig}} \equiv$ the high p_T lepton trigger efficiency

and

$\boldsymbol{\varepsilon_{lepID}} \equiv$ the efficiency of lepton ID cuts.

The tag efficiency, $\varepsilon_{tagging}$, is kept separate for clarity. Each of these factors is discussed below.

Kinematic Event Selection Efficiency for WH

Samples of 100k WH events were generated with Pythia in the mass range 100 $GeV/c^2 \leq M_H \leq 150$ GeV/c^2. The series of cuts was applied to the events in order

to construct a sample fulfilling the kinematic requirements of the selection. The efficiency of the event selection was measured by counting the fraction of events passing each successive cut.

The kinematic event selection efficiency, ε_{sel}^{WH} , is defined this way:

$$\varepsilon_{sel}^{WH} \equiv \prod_i \varepsilon_i \tag{5.4}$$

where the ε_i are defined by:

$\boldsymbol{\varepsilon_{geom}} \equiv$ the fraction of events with at least one electron or muon with $p_T > 15$ GeV/c and $\mid \eta \mid < 2.0$ in the event record

$\boldsymbol{\varepsilon_{match}} \equiv$ the fraction of events in which one of the high p_T leptons matches to an electron or muon candidate (NB: "match" is defined as $cos(\theta_{opening}) > 0.999$)

$\boldsymbol{\varepsilon_{ETpT}} \equiv$ the fraction of events in which the matched electron (muon) candidate has E_T (p_T) >20 $GeV(GeV/c)$

$\boldsymbol{\varepsilon_{fid}} \equiv$ the fraction of events in which the matched candidate is within the fiducial bounds of the Run 1 CDF detector

$\boldsymbol{\varepsilon_{Isol}} \equiv$ the fraction of events in which the matched candidate has Isol < 0.1

$\boldsymbol{\varepsilon_{MET}} \equiv$ the fraction of events with $\not\!\!E_T > 20$ GeV

$\boldsymbol{\varepsilon_{ZV}} \equiv$ the fraction of events which pass Z boson removal

$\boldsymbol{\varepsilon_{DV}} \equiv$ the fraction of events which pass the dilepton veto

and

$\boldsymbol{\varepsilon_{2jet}} \equiv$ the fraction of events in the 2 jet exclusive bin (see above for jet definition)

Each of the listed efficiencies is conditional upon the previous one. This sequential procedure was chosen because it is a good measure of the fraction of WH events with real high p_T leptons from $W^{\pm}$ decay.

Table 5.4 shows the value for each of these conditional efficiencies and the overall selection efficiency, ε_{sel}^{WH}, for various Higgs masses. The errors on each of the contributing efficiencies are from Monte Carlo statistics only. Systematic effects on the selection efficiency are discussed in Chapter 7.

One might remark that the match efficiency, ε_{match} , for these events is lower than expected, since the events are guaranteed to possess a high p_T lepton with $\mid \eta \mid < 2.0$ by virtue of passing the previous geometrical cut. The inefficiency incurred with this cut is mostly due to muons: from MC, $\varepsilon_{match}^{e} = 0.956$ and $\varepsilon_{match}^{\mu} = 0.552$. This muon inefficiency is due to the lack of muon detector coverage in the range $\mid \eta \mid > 1.1$. Plug electrons are removed later in the path, when fiduciality cuts are applied. Again from the MC electron and muon samples, $\varepsilon_{fid}^{e} = 0.623$ and $\varepsilon_{fid}^{\mu} = 0.831$.

The prior-to-tagging selection efficiency for WH events is 4-5.5% , depending on Higgs mass. For comparison, the $t\bar{t}$ selection efficiency measured in a similar way [33] was determined to be 10-12% . The main difference in WH events is that there is only one $W^{\pm}$ present that has the opportunity to decay leptonically and satisfy our lepton selection requirements. One should also note that in the Run 1 $t\bar{t}$ search, the analysis focused on the ≥ 3 jet bin, a cut which is slightly more efficient for $t\bar{t}$ events than our 2 jet requirement is for WH.

M_H (GeV/c^2)	100	110	120	130	140	150
ε_{geom}	0.294 ± 0.001	0.311 ± 0.001	0.333 ± 0.001	0.366 ± 0.002	0.407 ± 0.002	0.452 ± 0.002
ε_{match}	0.742 ± 0.003	0.749 ± 0.002	0.754 ± 0.002	0.766 ± 0.002	0.783 ± 0.002	0.798 ± 0.002
ε_{ETpT}	0.831 ± 0.003	0.826 ± 0.002	0.836 ± 0.002	0.846 ± 0.002	0.860 ± 0.002	0.874 ± 0.002
ε_{fid}	0.759 ± 0.003	0.770 ± 0.003	0.772 ± 0.003	0.778 ± 0.003	0.779 ± 0.003	0.783 ± 0.002
ε_{Iso}	0.778 ± 0.004	0.758 ± 0.004	0.771 ± 0.003	0.788 ± 0.003	0.810 ± 0.003	0.830 ± 0.002
ε_{MET}	0.824 ± 0.004	0.827 ± 0.004	0.833 ± 0.003	0.835 ± 0.003	0.839 ± 0.003	0.846 ± 0.003
ε_{ZV}	0.994 ± 0.001	0.993 ± 0.001	0.989 ± 0.001	0.982 ± 0.001	0.972 ± 0.001	0.971 ± 0.001
ε_{DV}	0.977 ± 0.002	0.971 ± 0.002	0.966 ± 0.002	0.948 ± 0.002	0.924 ± 0.002	0.907 ± 0.002
ε_{2jet}	0.469 ± 0.005	0.489 ± 0.005	0.496 ± 0.005	0.468 ± 0.005	0.412 ± 0.004	0.360 ± 0.004
ε_{sel}^{WH}	0.040 ± 0.001	0.044 ± 0.001	0.049 ± 0.001	0.053 ± 0.001	0.054 ± 0.001	0.055 ± 0.001

Table 5.4: ε_{sel}^{WH} and its contributing factors for various Higgs masses.

$M_H(GeV/c^2)$	100	110	120	130	140	150
$\epsilon_{\geq 1tag}$	0.37 ± 0.04	0.37 ± 0.04	0.33 ± 0.03	0.28 ± 0.03	0.19 ± 0.02	0.12 ± 0.01
Scale Factor	1.08 ± 0.11	1.08 ± 0.11	1.08 ± 0.11	1.07 ± 0.11	1.07 ± 0.11	1.06 ± 0.11
$\epsilon_{>1tag}$	0.40 ± 0.06	0.40 ± 0.06	0.36 ± 0.05	0.30 ± 0.04	0.20 ± 0.03	0.13 ± 0.02

Table 5.5: Fraction of WH events passing previous event selection satisfying the tag requirement.

Tagging Efficiency for WH

At least one SECVTX-tagged high-E_T central jet is demanded among the two required jets in each event. The tag efficiency for WH events for this tagging scenario is listed in Table 5.5 as a function of Higgs mass. The systematic error associated with these tagging algorithms was estimated in previous Run 1 lepton+jets analyses to be a $\pm$ 10% effect on the tag efficiency. We merely list the tag values and their systematic uncertainties in Table 5.5.

Also included in Table 5.5 are the values for the tag efficiency scale factor. Tagging scale factors attempt to reconcile the tag efficiency in the MC with that in the data. This scale factor was determined by examining the tag efficiency in MC $t\bar{t}$ events and scaling the efficiency to match the accepted value. A 10% uncertainty is inherent in the scale factor. We see though that the effect of this scale factor is only 6-8% in signal (compare to Run 1 $t\bar{t}$ tagging scale-up of 25%).

One can see that the tag efficiency is in the range 10-40% . The tagging efficiency decreases as M_H increases; the most significant factor influencing this effect is the steep decrease in the branching ratio of $H \to b\bar{b}$ for $M_H > 130\ GeV/c^2$.

Other Factors in WH Acceptance

The three factors ε_{zprim}, ε_{trig} and ε_{lepID} are taken directly from previous studies. The efficiency ε_{zprim} measures the effect of the requirement that the z coordinate of the primary event vertex be within 60 cm of the nominal interaction point. This efficiency was measured using $W \to \mu\nu$ data passing the standard lepton+jets selection criteria (prior to tagging) [33]; unlike the selection criteria for this search, that analysis considers the ≥ 3 jet bin, and this effect is incorporated in the systematic error. All other selection criteria are the same. This efficiency was found to be $\varepsilon_{zprim} = 0.949$

$\pm$ 0.023; the error is systematic and arises mostly from varying the jet multiplicity in the selection of the $W \to \mu\nu$ events [33].

Recall that the lepton identification criteria had their own efficiencies. These values need to be factored in when considering how many signal events are predicted to appear in the Run 1 data sample. With proper normalization of the contributions from the different subsystems in which the electrons and muons are recorded, the overall lepton identification efficiency is $\simeq$90%.

The high-p_T lepton trigger efficiency ε_{trig} was also measured. The same $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ samples utilized in determining the lepton identification criteria were used to measure the high p_T lepton trigger efficiency. Again, taking into account the different subsystems in which electrons and muons manifest themselves, the trigger was found to be $\simeq$85% efficient [40].

WH Acceptance Summary

Table 5.6 contains a summary of all the factors contributing to the overall detection efficiency of WH events as a function of Higgs mass as determined in our Pythia Monte Carlo samples. With this overall efficiency, and given the theoretical production cross section and total Run 1 integrated luminosity, one can calculate the number of WH events one expects to record in the Run 1 data set (Table 5.6, bottom line). Cross section numbers were provided by Hao and Willenbrock [42]. One can see that the expected number of Run 1 WH events is small.

Table 5.6 does not contain a complete list of uncertainties for all values. The impact of systematic errors was studied after advanced selection (described in Chapter 6.) was applied and is discussed in Chapter 7.

M_H (GeV/c^2)	100	110	120	130	140	150
ε_{sel}^{WH}	0.040 ± 0.001	0.044 ± 0.001	0.049 ± 0.001	0.053 ± 0.001	0.054 ± 0.001	0.055 ± 0.001
ε_{tags}^{WH}	0.40 ± 0.06	0.40 ± 0.06	0.35 ± 0.05	0.30 ± 0.04	0.21 ± 0.03	0.13 ± 0.02
ε_{zprim}	0.949 ± 0.023	0.949 ± 0.023	0.949 ± 0.023	0.949 ± 0.023	0.949 ± 0.023	0.949 ± 0.023
ε_{trig}	0.854 ± 0.014	0.854 ± 0.014	0.854 ± 0.014	0.854 ± 0.014	0.854 ± 0.014	0.854 ± 0.014
ε_{lepID}	0.903 ± 0.008	0.903 ± 0.008	0.903 ± 0.008	0.903 ± 0.008	0.903 ± 0.008	0.903 ± 0.008
ε_{WH}	0.011	0.012	0.012	0.011	0.008	0.005
Luminosity(pb^{-1})	105.1 ± 4	105.1 ± 4	105.1 ± 4	105.1 ± 4	105.1 ± 4	105.1 ± 4
Cross Section (pb)	0.26 ± 0.02	0.18 ± 0.01	0.14 ± 0.01	0.10 ± 0.01	0.08 ± 0.01	0.06 ± 0.01
Number Expected	0.32	0.24	0.19	0.12	0.07	0.03

Table 5.6: Total WH detection efficiency as a function of Higgs mass, and expected number of Run 1 events.

5.2.3 Backgrounds to $WH \to \ell\nu b\bar{b}$

The channel $WH \to \ell\nu b\bar{b}$ has several prominent background processes that share the same final state. These backgrounds include $t\bar{t}$ and single top production ($W*$ and W -*gluon*), channels that were the focus of several analyses in Run 1 and will of course continue to be of significant interest in the future. However in a WH search these processes are among the background.

A more significant type of background than these top-related channels comes from processes in which a leptonically decaying vector boson is accompanied by jet activity. These jets can come from gluon radiation from initial state quarks, or can arise from initial state gluons within the p or $\overline{p}$ that split to quarks. Real heavy-flavor jets can be produced in such cases, and in these instances the tag requirement is readily satisfied. One can also accidentally tag light quark jets, so one must also be sensitive to processes in which the accompanying jet is not necessarily from heavy flavor. We refer to these types of events as V+jets, where V=$W^{\pm}$ or Z.

Smaller contributions come from other sources. Diboson processes (mostly WZ, but also WW and ZZ) yield final states with leptons and jets; $\not\!\!E_T$ can appear naturally via $W^{\pm}$ decay, as in the case of WZ and WW, but one must also take care to consider spurious $\not\!\!E_T$ sources in ZZ events. The process $Z \to \tau\tau$ must also be considered, since a hadronically decaying tau can satisfy our jet requirement and a leptonically decaying tau can satisfy the $W^{\pm}$ selection.

The final state in $t\bar{t}$ production has the decay products of a second $W^{\pm}$ boson; by restricting ourselves to the 2 jet bin and vetoing events that pass the $t\bar{t}$ dilepton criteria, we exploit this distinguishing feature of $t\bar{t}$ production in an attempt to reject as much of this background as possible. However, single top, $W + b\bar{b}$ (a prominent V+jets channel) and WZ share the exact same final state ($\ell\nu b\bar{b}$) some fraction of the

time. There is no such distinguishing feature that we can easily use to eliminate the majority of these events. It will be the task of the NN to identify our signal events amidst these identical backgrounds.

These channels were the focus of much study in the context of the Run 1 CDF $t\bar{t}$ search. Much work was done to accurately predict the number of expected events from all these channels in the Run 1 data. By employing an equivalent event selection to previous analyses, we can incorporate the expected background yields directly. Table 5.7 has the expectations for each type of background [38].

Source	N_{exp}
$t\bar{t}$	5.4
W -$gluon$	1.5
$W*$	1.0
$W^{\pm}$+mistags	3.80
non-W	1.49
WZ	1.20
ZZ	0.16
WW	0.02
$W + c$	3.94
$W + c\bar{c}$	3.60
$W + b\bar{b}$	9.86
$Z \to \tau\tau$	0.70
$Z + c\bar{c}$	0.10
$Z + b\bar{b}$	0.54
Total	33.31

Table 5.7: Summary of background expectations for the process $WH \to \ell\nu b\bar{b}$ in $110pb^{-1}$ of Run 1 CDF data.

5.3 Summary

The baseline event selection chosen and tagging requirements corresponds to a WH event detection efficiency on the order of 0.5-1.1% . For a variety of SM Higgs masses, less than one WH event is expected to meet all the baseline selection criteria in the Run 1 data. By virtue of using the same baseline selection, predictions for the background channels were then taken directly from other analyses. It is evident from Table 5.8 that $t\bar{t}$ and W+HF events will dominate the selected dataset. The sensitivity with this basic selection for M_H=120 GeV/c^2 is 1/30.

	WH (M_H = 120)	$t\bar{t}$	SingleTop	Diboson	V +QCD
N_{exp}	0.19	5.4	2.5	1.38	24.06

Table 5.8: Summary of all background expectations and one representative signal scenario.

CHAPTER 6

ADVANCED EVENT SELECTION

Given the expected background contributions listed in Table 5.8 from the baseline selection criteria, it is clear that, if the WH production cross section is anything close to the prediction from theory, background events will overwhelm any signal. We seek now to increase our signal sensitivity by applying an advanced event selection. This advanced selection utilizes a neural network to help identify events consistent with signal while rejecting as much background as possible.

After selecting events with the network, we perform a counting experiment in windowed values of the reconstructed invariant mass of the 2 jets in the selected events. The di-jet system in signal events are daughters from Higgs decay; thus one expects the di-jet mass to be sensitive to the mass of the parent Higgs, M_H. This variable potentially has much information that can be used to better identify Higgs events and measure the mass of the resonance. More sophisticated techniques of exploiting this highly discriminating variable can be considered, such as a shape fit that returns a likelihood distribution for the number of signal events in a selected data sample. However, given the small statistics of the Run 1 data sample (recall that for the hypothesis M_H=120 GeV/c^2 we only expect 0.2 events prior to advanced selection) we chose in this analysis to forego more complex techniques for extracting a limit on the signal production rate.

Recall that we are performing a search for the Higgs at six different M_H hypotheses in the range 100 $GeV/c^2 < M_H < 150$ GeV/c^2. As will become evident, searches for each of these hypotheses is really a slightly different analysis. In the following sections we outline the advanced selection employed in this analysis for just one of the M_H hypotheses, $M_H = 120$ GeV/c^2. The extension to other mass values will be discussed in Section 7.5.

6.1 The NN Implementation

In this section we first describe the NN that is utilized in this analysis. This network was determined after a series of studies that considered many of the options one has when constructing a NN for such tasks. For instance,

- a spectrum of possible NN inputs was investigated.
- the number of background classes was varied and studied.
- the effect of the choice of the number of hidden nodes was considered.
- the effect of tuning several learning algorithm parameters was also studied.

By first describing the network that is utilized in this analysis, the impacts of the effects described above can more easily be put into context. The chosen NN implementation was determined using the *a priori* 95% CL limit on WH production achievable from simulation studies as the figure-of-merit, which we seek to minimize. The method for calculation of the *a priori* limit will be discussed in Chapter 7. The various optimization studies are described in Section 6.3 and can be skipped without loss of understanding.

In this analysis, the JETNET [46] software package was used for construction of the NN. JETNET was written specifically for physics applications, but it is general

enough to handle any type of classification problem. JETNET was chosen because of its wide use in the physics community. Available within JETNET are numerous tunable parameters the user has access to during optimization studies. An interface for JETNET to the ROOT analysis environment was also written and employed [48].

6.1.1 Training Samples

Before engaging in a description of the optimized network, a brief description of the training samples utilized in this analysis is warranted. As discussed in Chapter 4 the class of NN being utilized here is implemented in software. The training of the NN relies on simulated signal and background data samples that can be presented to the NN; the weights and thresholds of the NN are then adjusted as each simulated data sample is presented. For these purposes, Monte Carlo samples of each class of events were constructed using two different generators. Pythia v6.129a [36] was used for WH, $t\bar{t}$, single top, diboson, $W^{\pm}$+mistagged jets and $Z \to \tau\tau$. The signal WH samples were also used to determine our signal acceptance as described in Chapter 5.2.2. ALPGEN [50] was used to generate $W + b\bar{b}$, $W + c\bar{c}$, $W + c$, $Z + b\bar{b}$ and $Z + c\bar{c}$.

The non-W background is particularly difficult to model. These events are mostly QCD multi-jet production in which the high-p_T lepton requirement is satisfied by some hadron decay; this fake lepton along with a spurious $\not\!\!E_T$ source fake the presence of a $W^{\pm}$ boson. No generator efficiently produces these events; thus we have no shape information on this class of events. We choose here to model non-W events with distributions from $W + b\bar{b}$ MC as an approximation. One should keep in mind though that this background contributes less than 5% to the total background expectation prior to advanced selection.

6.1.2 The Chosen Network

The NN chosen for this analysis exploits information from eight event-level variables; the network then has eight input nodes. A single hidden layer with 12 hidden nodes was utilized, along with five output nodes. These five output nodes correspond to the five classes of events listed in Table 6.1. The 14 types of background processes were sorted into four classes; along with signal our problem is defined then as classifying events into five categories.

For the purposes of training, Monte Carlo data samples were constructed for each of the five classes of events. Each sample contained 4100 events, and each background contribution was normalized appropriately in relation to its fellow class members. According to Table 5.7 the largest background sources come from $W^{\pm}$ bosons produced in association with jets, and $t\bar{t}$ production. These types of backgrounds were grouped into separate classes to exploit inherent topological differences between the types of events.

	Class 1	Class 2	Class 3	Class 4	Class 5
Type of Events	WH (M_H=120GeV/c^2)	$t\bar{t}$	$W+b\bar{b}$ $W+c\bar{c}$ $W+c$ $W^{\pm}$+mistags non-W	Single top (W -*gluon* and $W*$)	WZ, WW,ZZ, $Z+b\bar{b}$, $Z+c\bar{c}$, $Z\rightarrow\tau\tau$
Target	(1,0,0,0,0)	(0,1,0,0,0)	(0,0,1,0,0)	(0,0,0,1,0)	(0,0,0,0,1)

Table 6.1: Five classes of events considered in this analysis. 14 types of background were separated into 4 classes.

The eight variables used as inputs to the NN were chosen from a superset of distributions that had some discriminating ability in 1D. The eight inputs are:

- $\not\!\!E_T$, which we expect to be larger in events in which the parent $W^{\pm}$ boson has more energy (such as a $t\bar{t}$ event)

- scalar H_T , the scalar sum of the p_T of the primary lepton, E_T of each jet and $\not\!\!E_T$, which again we would expect to be larger in the case of $t\bar{t}$ production due to the decay of the two heavy top quarks

- $\sum E_T^{Extra} = E_T^{j_1} + E_T^{j_2} + E_T^{\ell} - \not\!\!E_T$, a measure of the unclustered energy in each event, which we expect to be larger for $t\bar{t}$ events

- $M_{\ell\nu j_1}$, the invariant mass of the lepton, reconstructed neutrino and highest E_T jet; we expect this variable to peak at m_t for single-top production

- $(E_T^{j_1} - E_T^{j_2})$, the difference in E_T between the two jets in the event, which we expect to be smaller for $W + b\bar{b}$ production

- $M_{j_1 j_2 Obj_3}$, the mass of the j_1 - j_2 - Obj_3 system. The third object in these events — if one exists — is a $E_T > 8\ GeV$, $\mid \eta \mid < 5.0$ cluster that fails our E_T or η requirement.

- $M_{\ell\nu j_1 j_2}$, the invariant mass of the lepton, reconstructed neutrino and di-jet system, which we expect to be more isotropic in $t\bar{t}$ events.

- $p_T^{j_1 j_2}$, vector sum p_T of the j_1 - j_2 system, which we again expect to be isotropic in $t\bar{t}$ events.

These eight distributions are shown in Figures 6.1- 6.8.

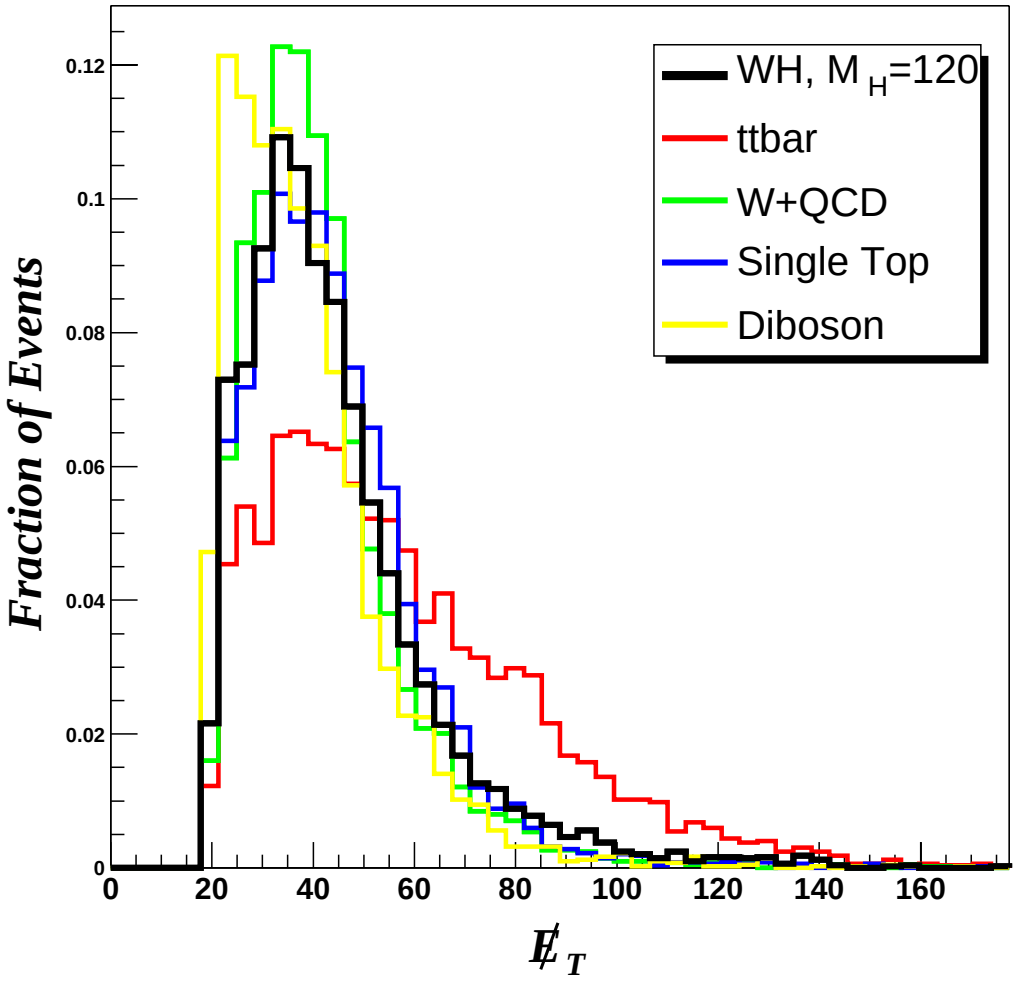


Figure 6.1: NN input variable $\not\!\!E_T$ for each of the five classes of events.

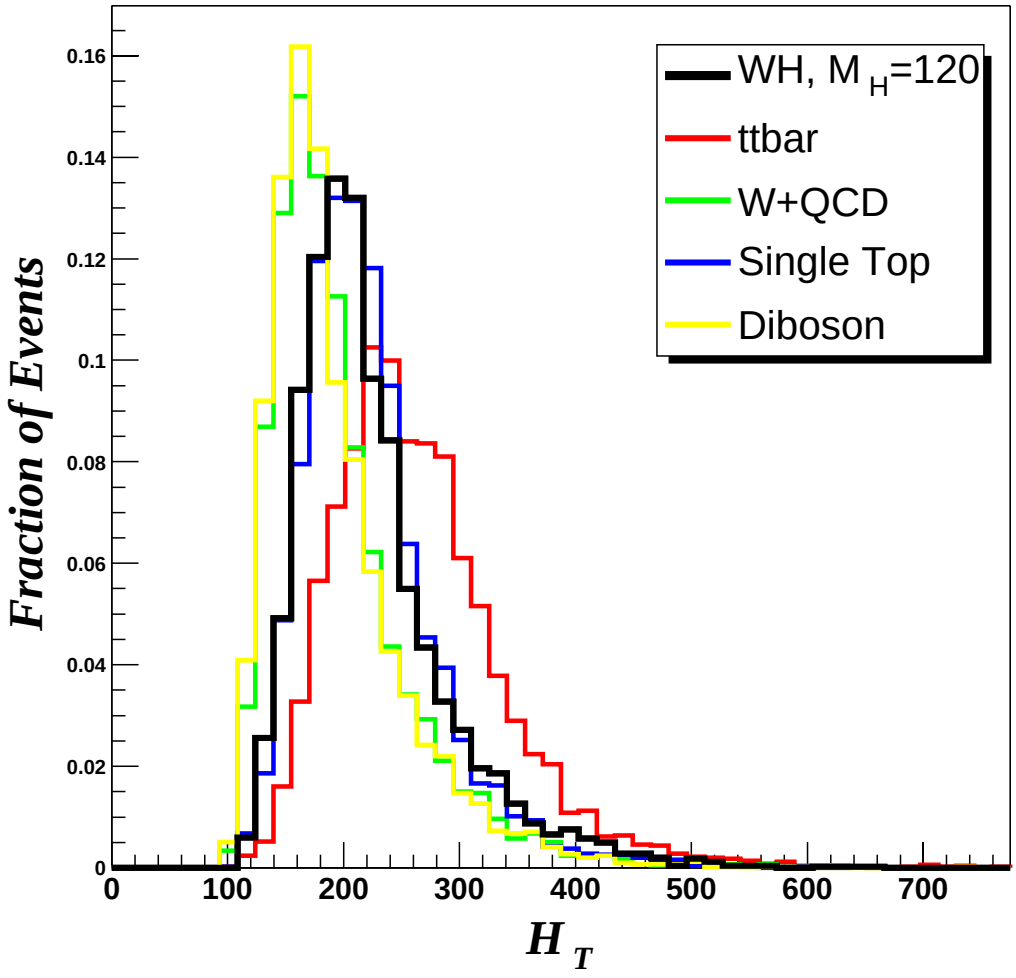


Figure 6.2: NN input variable scalar H_T for each of the five classes of events.

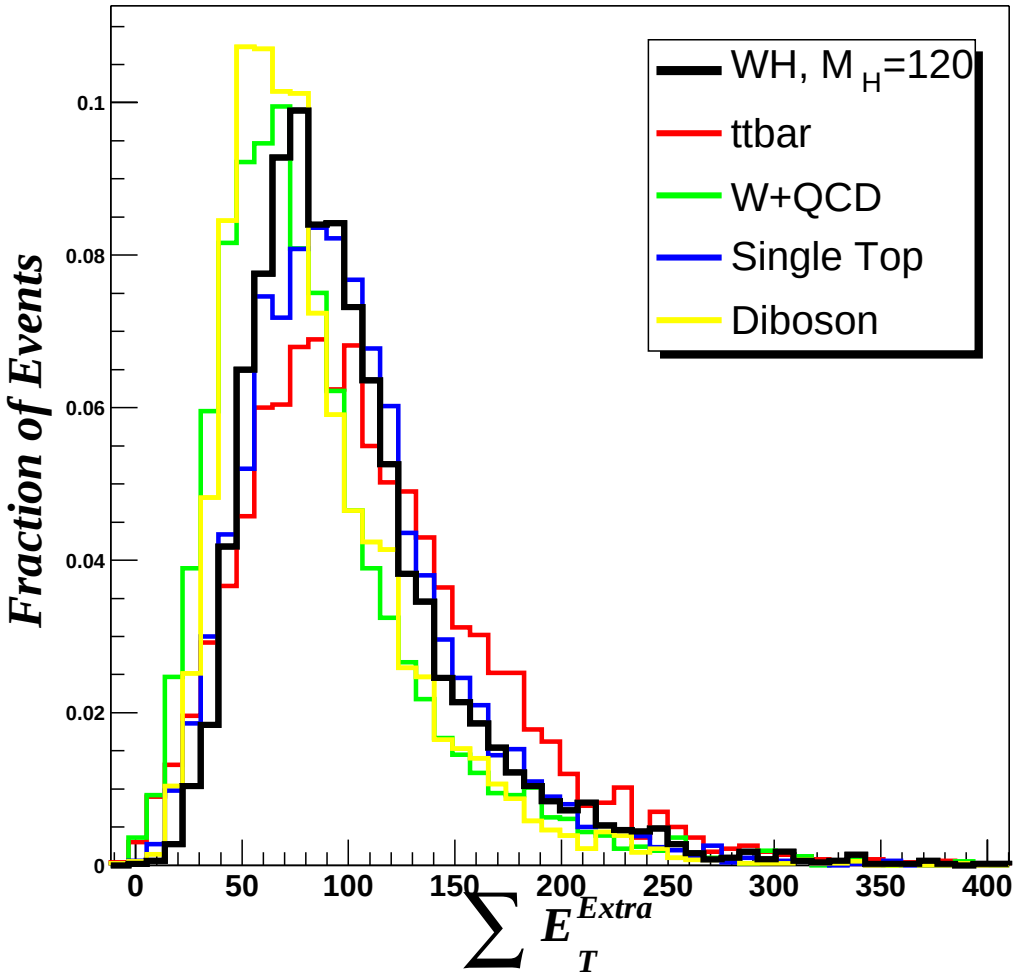


Figure 6.3: NN input variable $\sum E_T^{Extra}$ for each of the five classes of events.

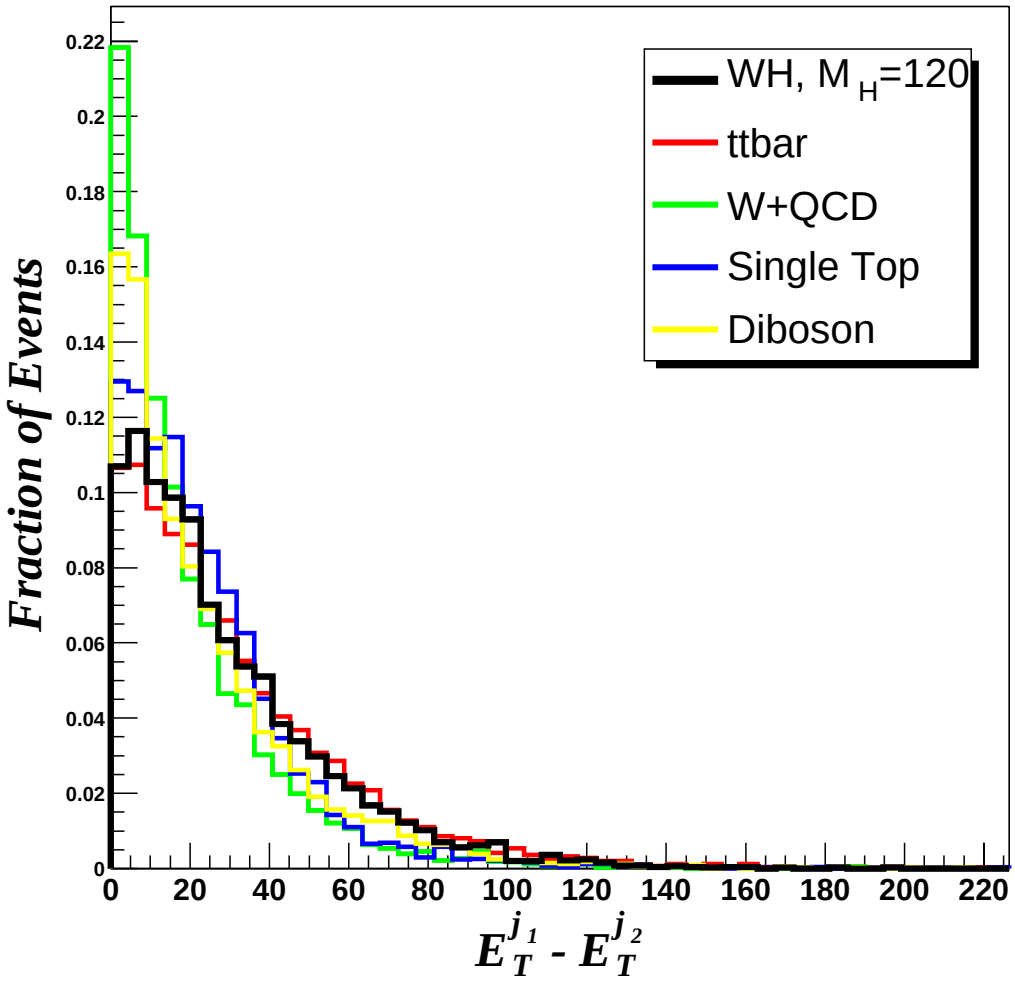


Figure 6.4: NN input variable (E_T^{j1} - E_T^{j2}) for each of the five classes of events.

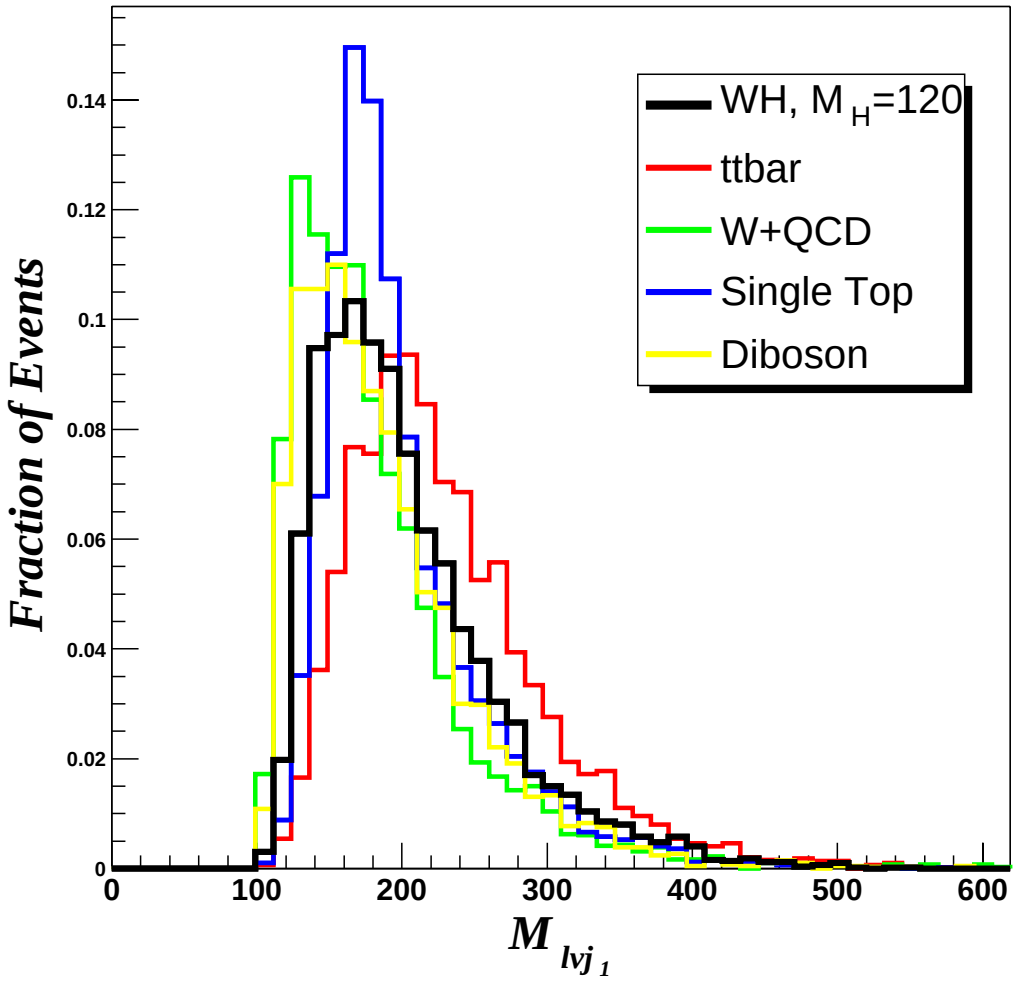


Figure 6.5: NN input variable $M_{\ell\nu j_1}$ for each of the five classes of events.

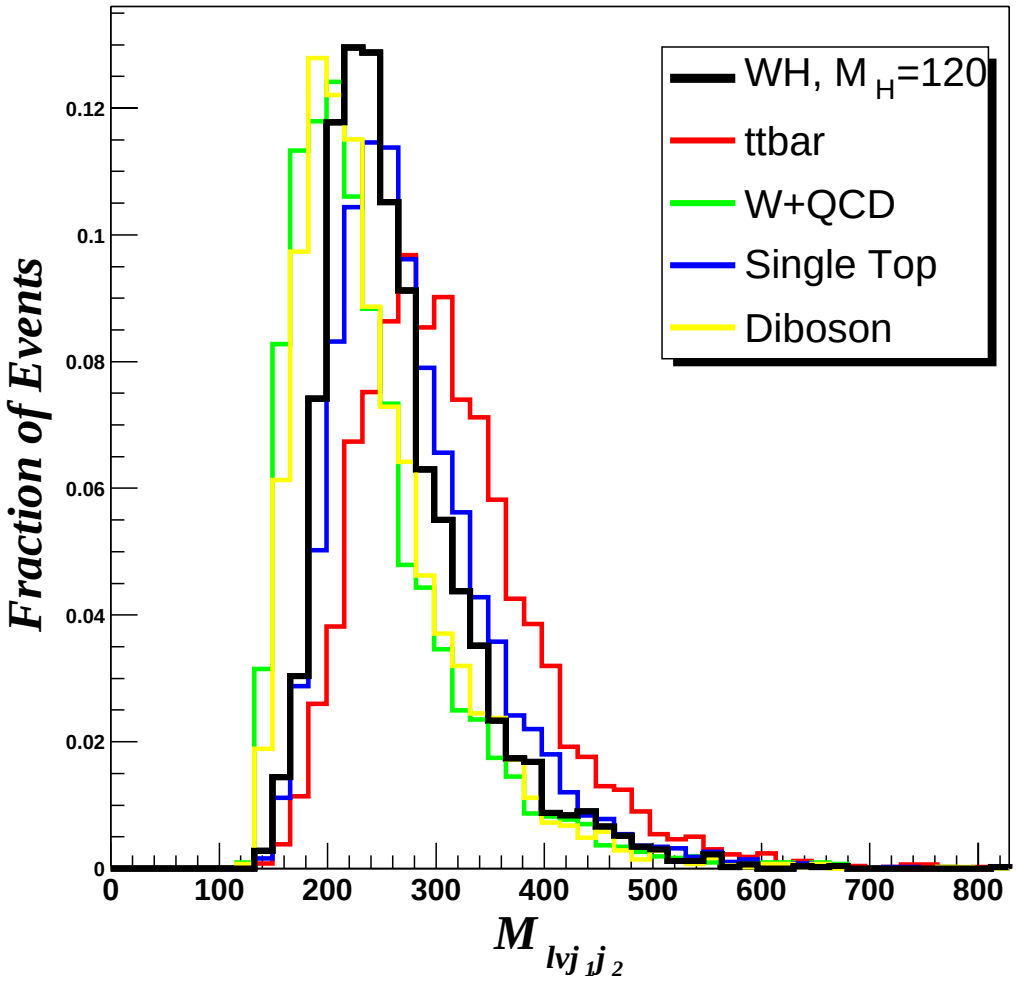


Figure 6.6: NN input variable $M_{\ell\nu j_1 j_2}$ for each of the five classes of events.

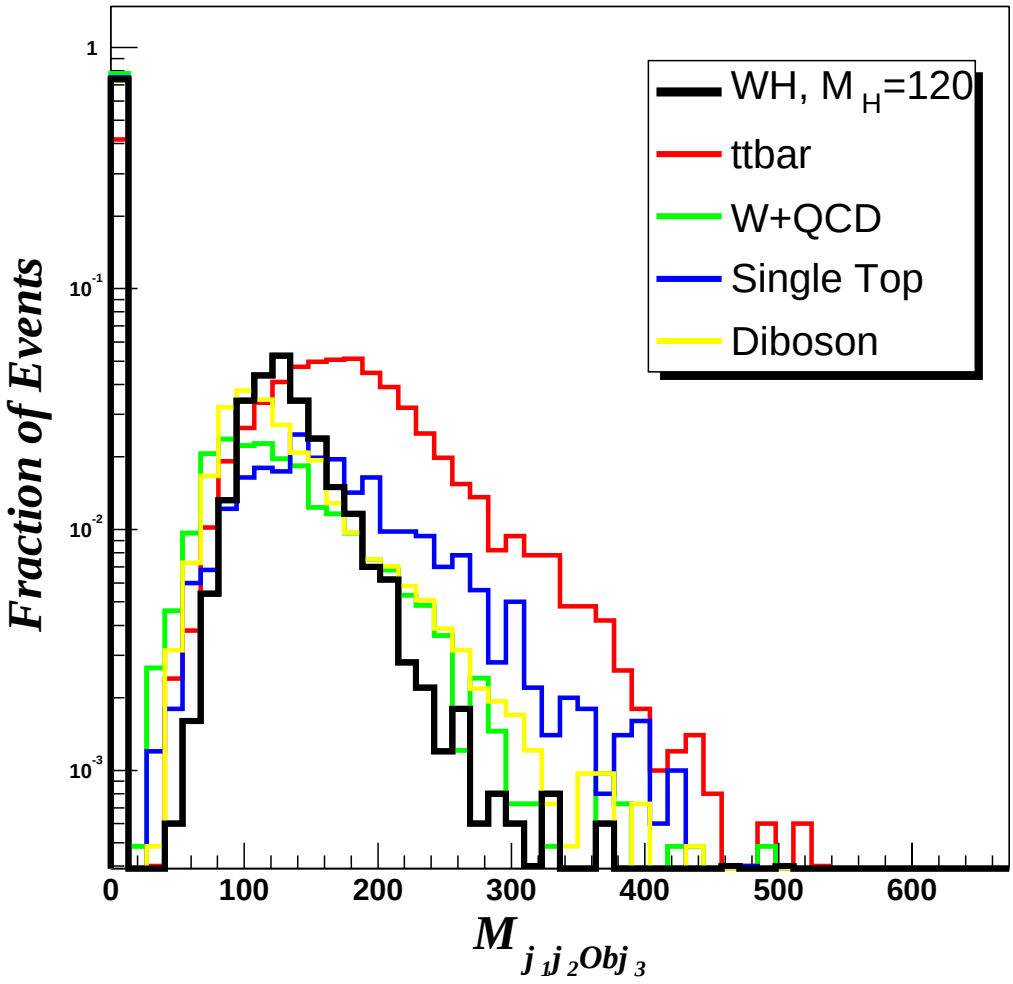


Figure 6.7: NN input variable $M_{j_1 j_2 Obj_3}$ for each of the five classes of events.

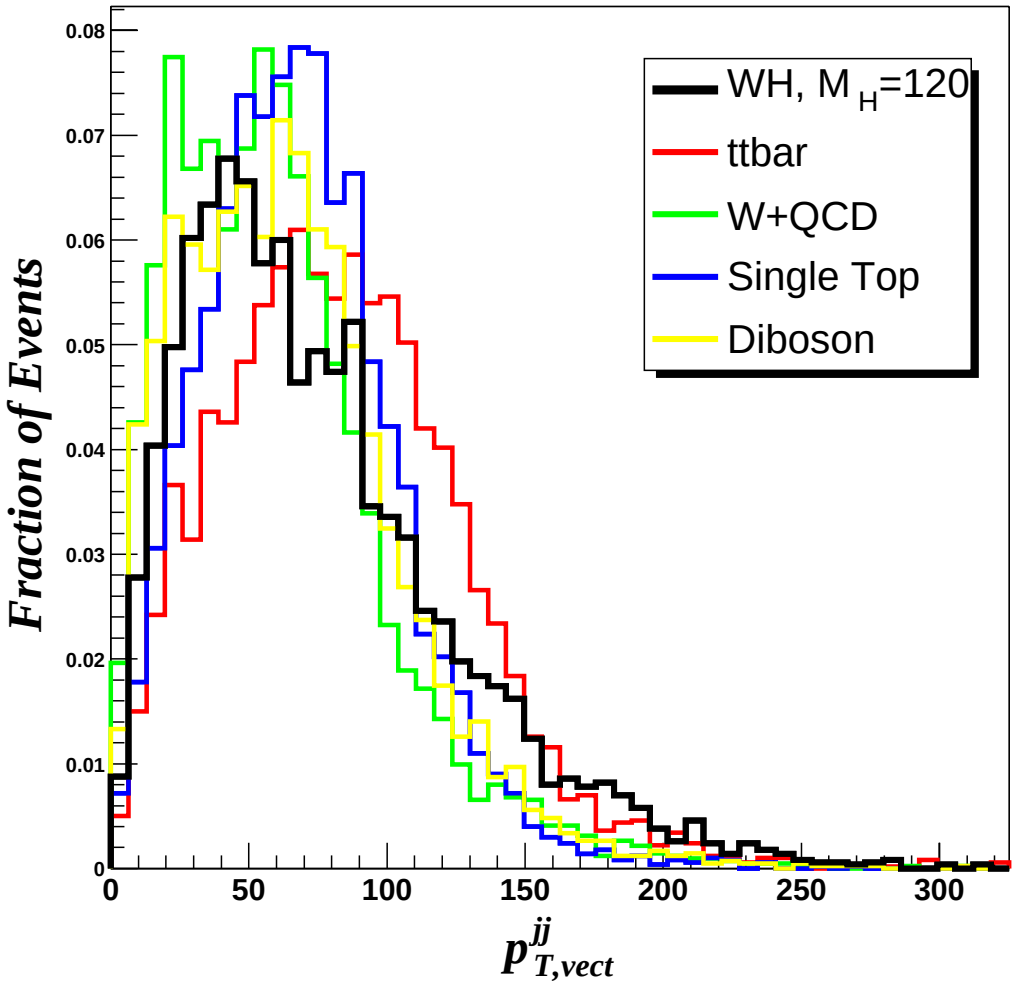


Figure 6.8: NN input variable $p_T^{j_1 j_2}$ for each of the five classes of events.

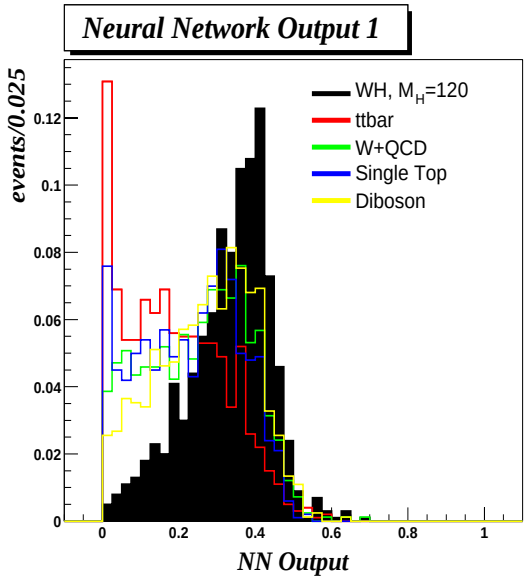


Figure 6.9: Node 1 output distributions for each of the five classes of events.

The network was trained for 10000 *epochs*; recall from Chapter 4 that an epoch is defined as the complete presentation of all available training vectors (events) for the purposes of weight updating. There were $\sim$ 4100 MC events in each of the MC training samples for the five classes of events. During training, the weights of the network were updated after every 100 presentations; the frequency of update is a tunable parameter, and the effect of adjusting this frequency was studied and is discussed later in Section 6.3.4. This means that over 10k epochs, the weights and thresholds of the NN were updated more than 2 million times.

The target output vectors for each class need to be supplied by the user. The target output vector for each class chosen for this analysis form the basis vectors for a 5D cube. For example, the target for Class 1 (WH events) was (1,0,0,0,0), the target for Class 2 ($t\bar{t}$) was (0,1,0,0,0), etc.

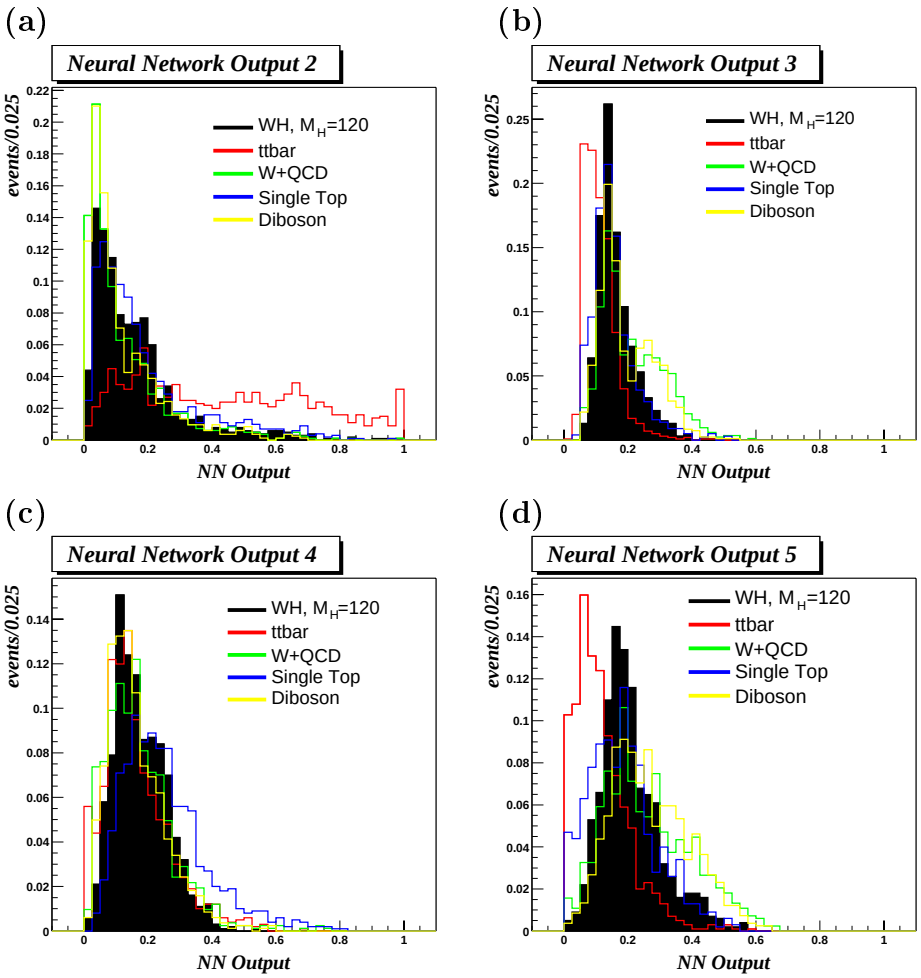


Figure 6.10: Nodes 2–5 output distributions for each of the five classes of events.

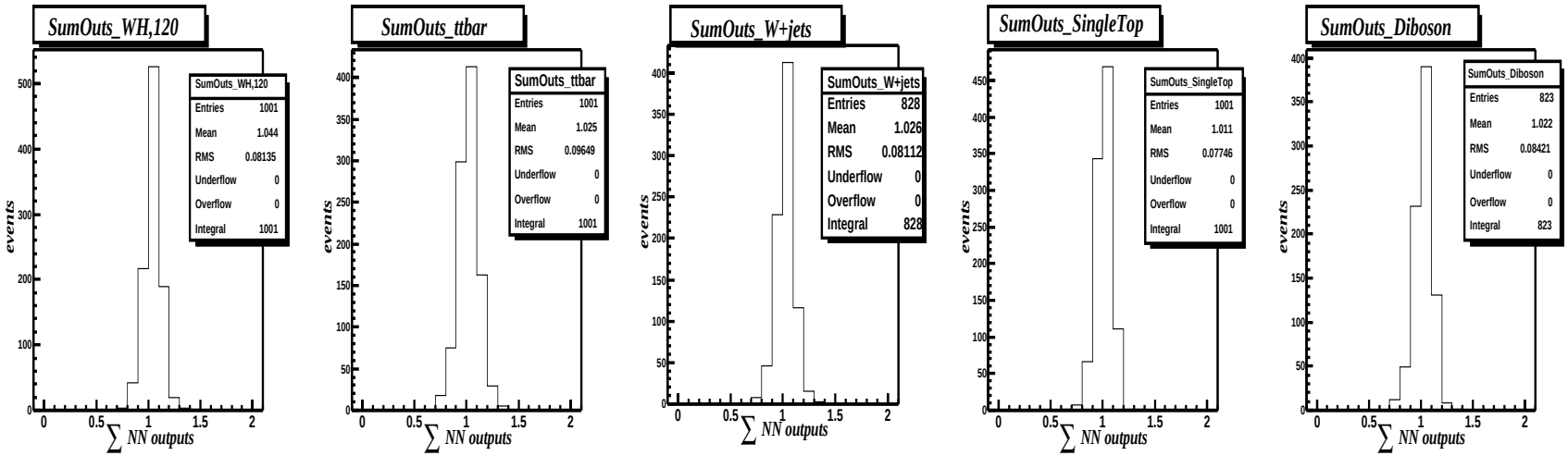


Figure 6.11: The sum of the NN output values for each of the five input classes.

The distributions for the five outputs is shown in Figure 6.9-6.10. We see that none of the classes produce a NN output value close to its target. It should be noted that although the NN output distributions for WH events do not match their target output (1,0,0,0,0), this is not indicative of poor training or a "broken" NN. In a perfect world in which each class of events is completely distinct, we would expect the NN output vector for events from each class to match its target value. However in reality, there is significant overlap between the samples. The NN output distributions merely indicate that there is no region in the 8D input phase space that is populated solely by signal events. What is critical, though, is that there is some separation in the NN output distributions, especially the distribution for NN output 1, the output node associated with signal events.

Recall one of the motivations for using a multiple-output neural network is that the values of each output node can be interpreted as Bayesian *a posteriori* class probabilities. We can check the validity of this claim by plotting the sum of the NN outputs for each input class. If these values can truly be interpreted as probabilities, then the sum for each sample should be 1. We see from Figure 6.11 that indeed the sum distribution is approximately centered on 1 for each input class with an RMS of 10% .

One can now use these NN output distributions and apply a final selection cut. A simple cut in the NN output distribution for output 1 was considered first, but an alternative technique proved to be more powerful. Recall that the NN output targets define a 5D cube. We consider the 1D distance (d) for events from each sample to the vertex of this cube that was the target for signal events, (1,0,0,0,0). This distance d is given by

$$d = \sqrt{(node_1 - 1)^2 + (node_2 - 0)^2 + (node_3 - 0)^2 + (node_4 - 0)^2 + (node_5 - 0)^2} \tag{6.1}$$

The distribution of d for each sample from the signal vertex is plotted in Figure 6.12. It is clear that this distribution for signal peaks at values less than 1.0, while for background the distribution has larger values. Thus one would expect a cut around $d^{thresh} \simeq 1.0$ to be optimal for retaining signal while rejecting background. Figure 6.13 shows the *a priori* limit achieved when implementing this NN selection as a function of d^{thresh}. In accordance with what was expected from our visual inspection of the 1D distance distributions, the optimal cut location is at $d^{thresh} = 0.95$. Events with $d < d^{thresh} = 0.95$ were retained. In designing this cut, one is just constructing a 5D sphere with radius d^{thresh} centered on the signal target (1,0,0,0,0) in the NN output space and measuring the efficiency for retaining events that lie within. For this reason we refer to the technique as the "hypersphere" cut.

6.2 Exploitation of the $M_{j_1 j_2}$ Distribution

It should be noted here that, in selecting input variables, attempts were made to avoid using variables that sculpted the $M_{j_1 j_2}$ distribution of the selected sample[1]. Figure 6.14(a) contains the $M_{j_1 j_2}$ distribution for events priori to the NN selection; Figure 6.14(b) contains the distribution for the post-NN-selection sample. In both Figure 6.14(a) and (b), the signal-only distribution is magnified by a factor of 10; in

[1]In WH events the $M_{j_1 j_2}$ distribution will peak near the hypothetical M_H being considered. In this advanced selection, we are selecting events that are most consistent with WH. Therefore if we make that decision based on $E_T^{j_1}$, $E_T^{j_2}$, and their separation, $\Delta R(j_1, j_2)$, then by definition we will be selecting events with similar $M_{j_1 j_2}$ distribution as WH, *i.e.*, sculpting the $M_{j_1 j_2}$ distribution.

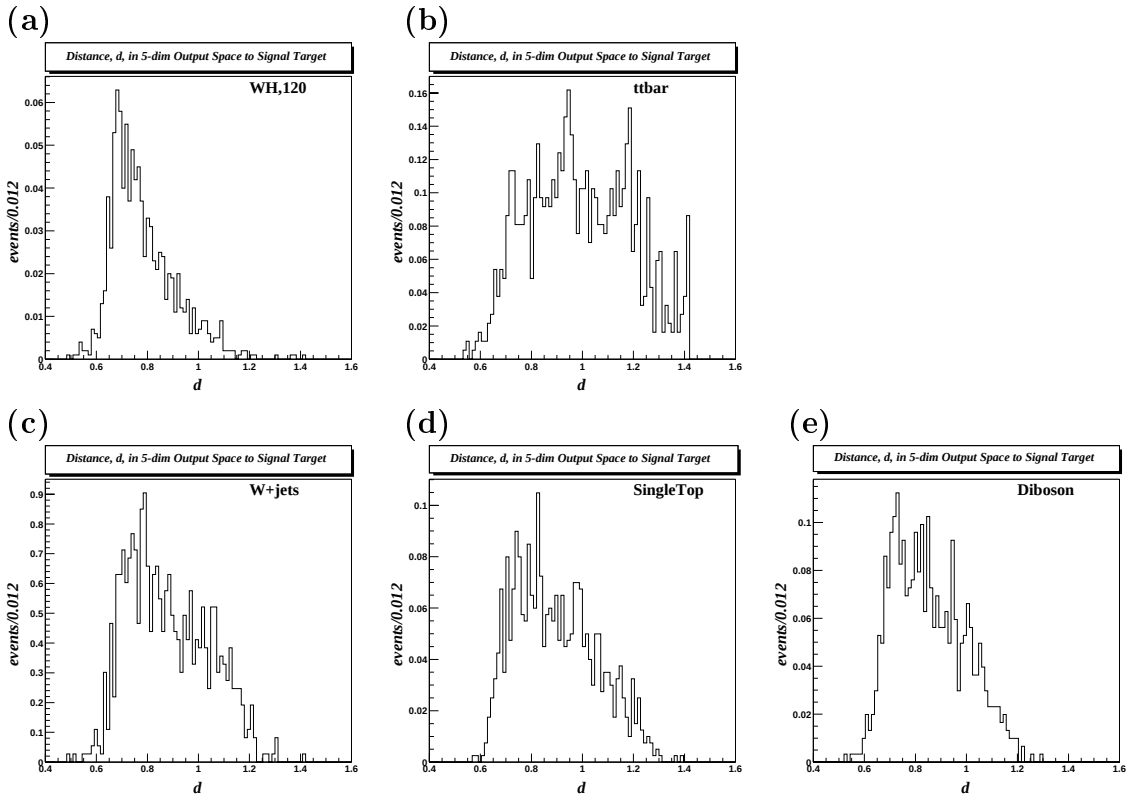


Figure 6.12: The scalar distances in the 5D output space of the events from each of the five input classes to the target vertex associated with signal. Note that the maximum distance between events is equivalent to the diagonal length of the face of the unit cube; in any dimensionality this value is $\sqrt{2} \simeq 1.4$.

the S+B distributions, however, the signal contribution is normalized properly to its Run 1 expectation. One can see very little evidence for any M_H mass bump in the S+B $M_{j_1 j_2}$ distribution.

Ideally, one would examine the $M_{j_1 j_2}$ distribution for the events passing the advanced selection criteria and perform a likelihood fit that yields the contribution from signal. However, due to the small statistics of the Run 1 data sample, it was not clear that performing a fit for the signal contribution would have any benefit over more crude signal extraction methods. In lieu of fitting for the signal contribution in the

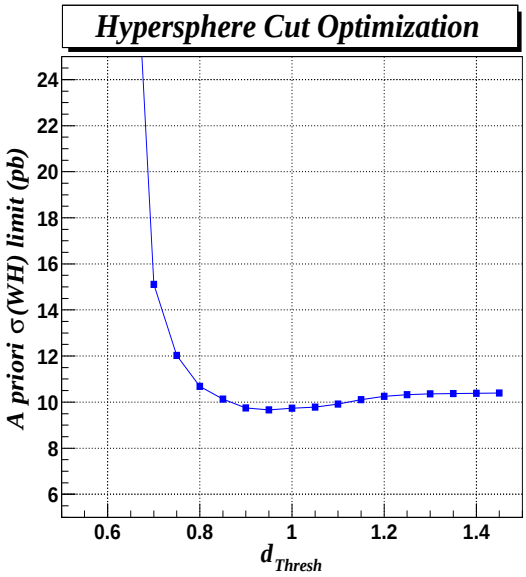


Figure 6.13: The *a priori* limit as a function of hypersphere radius for final advanced selection.

final selected sample, in this analysis a counting experiment was performed in a window of $M_{j_1 j_2}$. The requirement 80 $GeV/c^2 < M_{j_1 j_2} <$ 150 GeV/c^2 provided the final step in our advanced event selection for the hypothesis M_H=120 GeV/c^2. Table 6.2 contains the efficiencies for each input class of events for both advanced selection criteria, the "hypershpere" NN cut and di-jet mass window requirement. From Table 6.2 one can calculate that with this advanced selection, the S/$\sqrt{B}$ increased from 1/30 to 1/22 for M_H = 120 GeV/c^2.

Note that only the results of the advanced selection in simulation studies for scenario M_H = 120 GeV/c^2 has been presented thus far. The extension of the analysis to other M_H values will be discussed in Section 7.5.

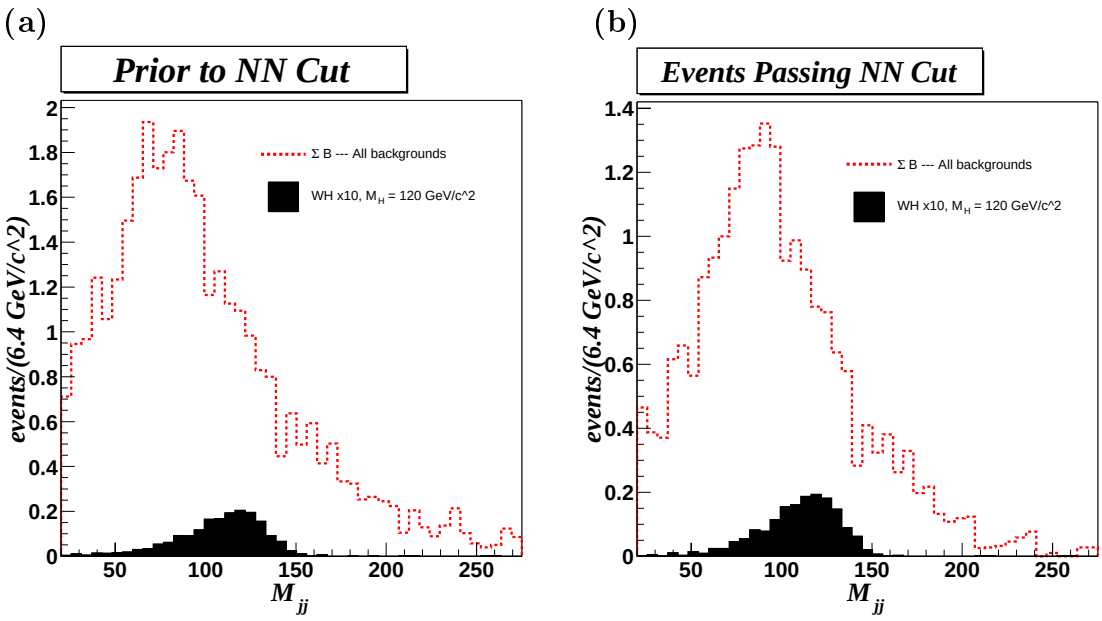


Figure 6.14: The $M_{j_1 j_2}$ distribution for WH signal (x10) and summed backgrounds before (a) and after (b) the NN cut.

6.3 Tests of Choices

With the above expected yields for background and signal with M_H= 120 GeV/c^2, one can calculate the *a priori* limit on WH production cross section, the method for which will be discussed in Chapter 7. The *a priori* limit was determined to be 8.4 pb for M_H= 120 GeV/c^2.

Each tunable choice in the advanced selection was studied independently. While studying the effect of one variable, choices had to be made regarding the other unrelated parameters. Consequently one could argue that the results of these studies are dependent on the settings chosen for these other parameters. An exhaustive study would be ideal, in which one tuned all variables in all combinations; but given that there are several tunable parameters with an unlimited number of possible settings, such an exhaustive study is intractable. Here we attempt to understand the impact

	Class 1	Class 2	Class 3	Class 4	Class 5
Type of Events	WH (M_H=120GeV/c^2)	$t\bar{t}$	$W+b\bar{b}$ $W+c\bar{c}$ $W+c$ $W^{\pm}$+mistags non-W	Single top (W -*gluon* and $W*$)	WZ, WW,ZZ, $Z+b\bar{b}$, $Z+c\bar{c}$, $Z\rightarrow\tau\tau$
Prior N_{exp}	0.19	5.4	22.69	2.5	2.72
ϵ_{hyp}	0.886	0.451	0.661	0.630	0.740
$\epsilon_{M_{jj}}$	0.881	0.552	0.483	0.563	0.593
ϵ_{AdvSel}	0.781	0.249	0.319	0.355	0.439
Post N_{exp}	0.15	1.3	7.2	0.89	1.2

Table 6.2: Expectations before and after advanced selection for each of the five classes of events. Cut: $d < d^{thresh} = 0.95$. Note that $\epsilon_{AdvSel} = \epsilon_{hyp} \times \epsilon_{M_{jj}}$.

of each of these choices the best we can, while acknowledging the neglect of myriad other possibilities.

6.3.1 Determining N_{Output}

We have 14 types of background processes in addition to our signal, which implies we are dealing with a 15 class problem. In Chapter 4 it was noted that a N-class problem motivates a NN with N output nodes. Clearly then a 15-output NN is called for here. However, such a NN is somewhat complex.

In general, the number of degrees of freedom in a 3-layer NN is given by:

$$N_{DoF} = N_H * (N_I + N_O + 1) + N_O \tag{6.2}$$

where N_I,N_H, and N_O are the number of input, hidden and output nodes of the NN. The number of degrees of freedom is just the total number of weights and thresholds (w_{ji} and θ_j in Eqn.4.2) for the entire network. For adequate NN training, it is recommended that one have training samples for each input class that contain a number of events at least 10 times the number of degrees of freedom for the net [52];

we seek to have 20 times the number of degrees of freedom in order to be confident the weights and thresholds of our network are well-constrained during the learning phase. A 15-output NN increases the required number of MC events for adequate learning significantly; it is therefore in our interest to reduce our 15-class problem to something more manageable. To this end, the 14 types of backgrounds were searched for subsets that shared common distribution shapes.

Because of their unique topologies, four of the background channels — $t\bar{t}$, $W + b\bar{b}$, single top and WZ — were chosen as good candidates for baseline background categories. Each of these four background channels produces the $\ell\nu b\bar{b}$ final state in a completely different way, and so it is natural to base the learning of the network on these four classes. From Table 5.7 it also clear that $t\bar{t}$ and $W + b\bar{b}$ are two of the most prominent background sources; it is therefore reasonable to at the very least separate these two channels into different classes.

The two single top production mechanisms were the focus of several Run 1 searches [62], [54]. These analyses looked for single top without regard to identifying each individual production mechanism. This is warranted because the kinematic distributions for the two single-top mechanisms are very similar. It is reasonable to adopt the same convention here and combine these two single top sources.

The distributions of the remaining background processes were scanned and compared to the four baseline categories. By inspection it was determined that $W^{\pm}$+mistags, $W + c\bar{c}$ and $W + c$ events could be grouped together with $W + b\bar{b}$. Herein we will refer to this input class generically as "$W^{\pm}$+jets". We also choose to include the non-W background with this $W^{\pm}$+jets class, modeling the shape of these events with $W + b\bar{b}$ events. WW and ZZ events were naturally paired with the major diboson process WZ . Because of their small size, the remaining backgrounds ($Z + b\bar{b}$,

$Z + c\bar{c}$, $Z \to \tau\tau$) are grouped also with WZ . This class of events will be somewhat inappropriately referred to as "diboson".

Our 15 class problem was thus reduced to 5 classes (Table 6.1). This five class problem motivates a 5-output NN; the NN output target values for each of the five classes of events is also listed in Table 6.1. The requirements for training such a NN – given a reasonable number of input and hidden nodes – do not exceed the amount of MC events we have in hand, so we are satisfied with this reduction in complexity.

Other implementations were tested that had fewer number of background classes, and therefore a smaller number of output nodes. It was shown [60] that the performance of the NN was insensitive to the number of output nodes; an implementation with just one background class performed similarly to the five background class NN.

Because of our desire to reduce the number of degrees of freedom of the NN, one might ask why we chose the more complex five output NN. There are several reasons. First, recall that we get the normalization of the baseline predicted background contributions from other, similar analyses. By employing several background classes instead of just one, the analysis is less sensitive to problems in the overall expected background yield. Also, having a five output NN potentially gives one more handles for understanding the final selected sample. It was possible that such a trained NN could, via well-separated output distributions, reliably sort events into each of the backgrounds classes. As evident from Figure **??** this was not exactly the case.

6.3.2 The NN Inputs

Thirty-one possible NN inputs were considered. Out of this set of 31 variables, a baseline subset were chosen according to following guidelines:

- The variables explored should seek to exploit the differences between the production mechanisms for the $\ell\nu b\bar{b}$ final state in signal and background.

- The input variables should avoid sculpting the di-jet mass, $M_{j_1 j_2}$.

A compatibility test was utilized that compared signal distributions to all four background classes in an effort to identify a subset of variables that were most discriminating (Table 6.3) and which satisfied the criteria listed above. The compatibility test considers each distribution bin-to-bin and records the maximum bin-height difference among the classes of events. Distributions that differ significantly by definition have at least one bin that has different normalized height. This test offers a quantitative measure of incompatibility between signal and background among the different input variables considered. In Table 6.3 small values indicate that there are no bins with significantly larger normalized contents, and thus two compatible distributions.

The statistics in Table 6.3 were used to determine a subset of 17 possible NN inputs; this cream-of-the-crop was used as the basis for the NN input studies. The variables chosen for study were:

- $\not\!\!E_T$

- $M_{\ell\nu}^{Trans}$

- scalar H_T , the scalar sum of the p_T of the primary lepton, E_T of each jet and $\not\!\!E_T$.

- N_{Extra}^{Objs}, the number of $E_T > 8\ GeV$, $\mid \eta \mid < 5.0$ clusters not satisfying default jet requirements.

- $\sum E_T^{Extra} = E_T^{j_1} + E_T^{j_2} + E_T^{\ell}$ - $\not\!\!E_T$

	WH to WH	WH to $t\bar{t}$	WH to $W^{\pm}$+jets	WH to Single Top	WH to Diboson
$E_T^{j_1}$	0.000	0.197	0.292	0.018	0.229
$E_T^{j_2}$	0.000	0.240	0.234	0.113	0.222
$\Delta R(j_1, j_2)$	0.000	0.058	0.150	0.093	0.060
$M_{j_1 j_2}$	0.000	0.361	0.373	0.326	0.354
E_T^{ℓ}	0.000	0.024	0.050	0.074	0.154
η^{ℓ}	0.000	0.013	0.033	0.015	0.029
$\Delta R(\ell, j_1)$	0.000	0.097	0.041	0.043	0.043
$\Delta R(\ell, j_2)$	0.000	0.016	0.058	0.059	0.059
$\not{E}_T$	0.000	0.248	0.072	0.034	0.130
$M_{\ell\nu}^{Trans}$	0.000	0.155	0.110	0.042	0.141
H_T	0.000	0.296	0.244	0.026	0.282
N_{Extra}^{Objs}	0.000	0.321	0.045	0.017	0.006
$\sum E_T^{Extra}$	0.000	0.115	0.194	0.031	0.179
$\langle \sum E_T^{Extra} \rangle$	0.000	0.316	0.062	0.019	0.027
$\mid \eta \mid_j^{Max}$	0.000	0.051	0.039	0.170	0.051
$M_{\ell\nu j_1}$	0.000	0.210	0.177	0.066	0.126
$M_{\ell\nu j_2}$	0.000	0.249	0.087	0.137	0.108
$\lvert M_{\ell\nu j} - m_t \rvert_{Min}$	0.000	0.062	0.070	0.163	0.078
$\lvert M_{\ell\nu j} - m_t \rvert_{Max}$	0.000	0.123	0.079	0.158	0.057
$(E_T^{j_1}$ - $E_T^{j_2})$	0.000	0.032	0.188	0.097	0.108
$M_{j_1 j_2 Obj_3}$	0.000	0.336	0.086	0.068	0.037
$M_{\ell\nu Obj_3}$	0.000	0.321	0.045	0.017	0.013
$M_{\ell\nu j_1 j_2}$	0.000	0.294	0.231	0.104	0.189
$cos(\theta(\ell^{fakeWRF}, H_{12}^{labRF}))$	0.000	0.034	0.093	0.066	0.024
$cos(\theta(j_{tag}^{H12RF}, H12^{labRF}))$	0.000	0.040	0.052	0.154	0.040
$cos(\theta(j_{tag}^{H12RF}, \ell^{H12RF}))$	0.000	0.046	0.060	0.128	0.040
$cos(\theta(\ell^{WRF}, W^{labRF}))$	0.000	0.138	0.038	0.081	0.112
$p_T^{j_1 j_2}$	0.000	0.144	0.142	0.097	0.093
$p_T^{\ell\nu j_1 j_2}$	0.000	0.181	0.088	0.046	0.036
Sphericity	0.000	0.050	0.053	0.009	0.075
Aplanarity	0.000	0.035	0.078	0.022	0.100

Table 6.3: Bin height fractional difference for 31 interesting variables, compared for all input classes to signal, WH.

- $\langle \sum E_T^{Extra} \rangle = \frac{\sum E_T^{Extra}}{N_{Extra}^{Objs}}$
- $M_{\ell\nu j_1}$
- $M_{\ell\nu j_2}$
- $|M_{\ell\nu j} - m_t|_{Max}$
- $(E_T^{j_1}$ - $E_T^{j_2})$
- $M_{j_1 j_2 Obj_3}$, the mass of the j_1 - j_2 - Obj_3 system. The third object in these events — if one exists — is a $E_T > 8\ GeV$, $|\ \eta\ | < 5.0$ cluster that fails our E_T or η requirement.
- $M_{\ell\nu Obj_3}$
- $M_{\ell\nu j_1 j_2}$
- $cos(\theta(j_{tag}^{H12RF}, \ell^{H12RF}))$, the cosine of the opening angle between the leading tagged jet boosted to the rest frame of the j_1 - j_2 system (called H12) and H12 in the lab frame.
- $cos(\theta(\ell^{WRF}, W^{labRF}))$, the cosine of the opening angle between the primary lepton boosted to the rest frame of the ℓ - ν system (denoted W) and W in the lab frame.
- $p_T^{j_1 j_2}$, vector sum p_T of the j_1 - j_2 system.
- $p_T^{\ell\nu j_1 j_2}$, vector sum p_T of the system defined by the primary lepton, $\not\!\!E_T$ and jets.

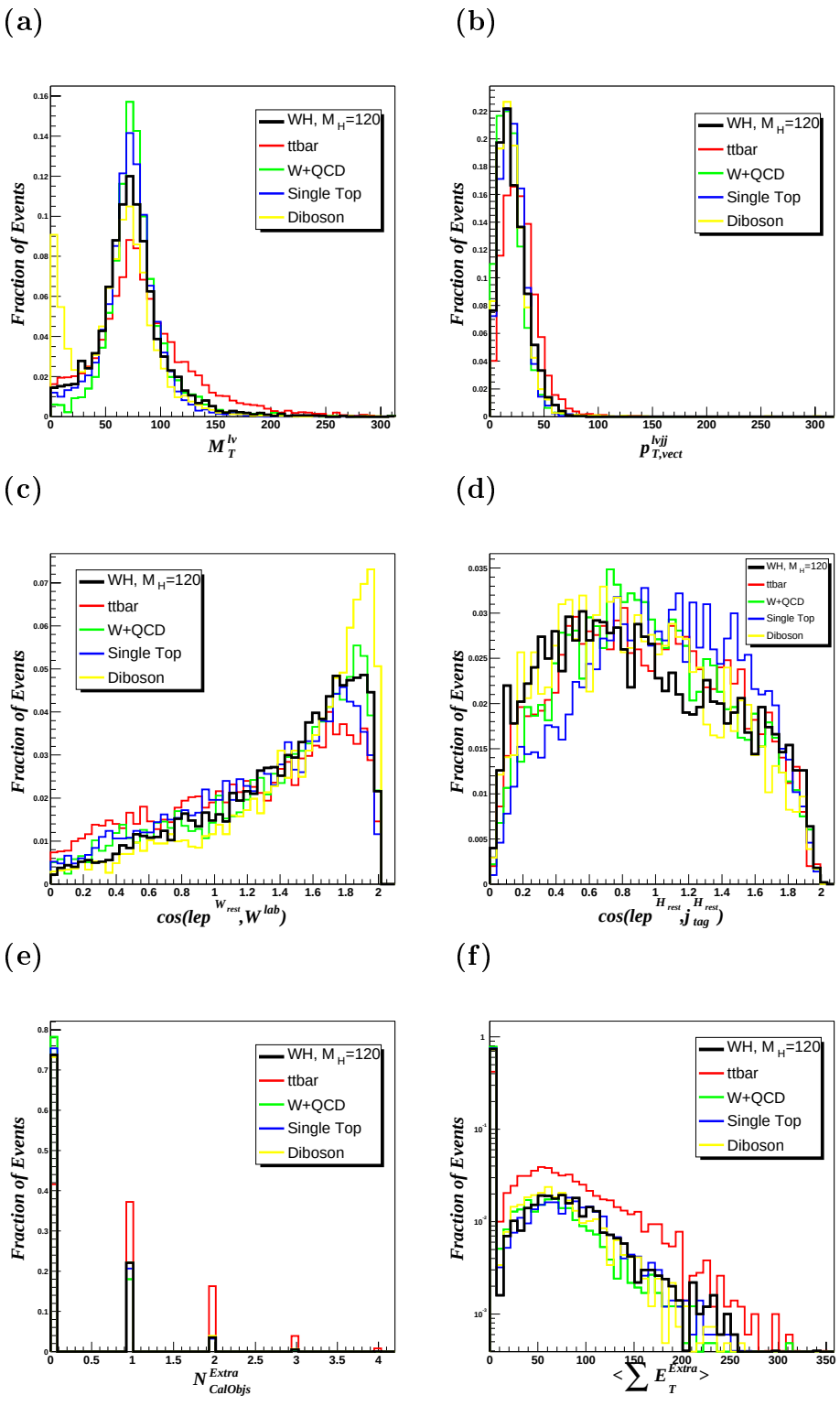
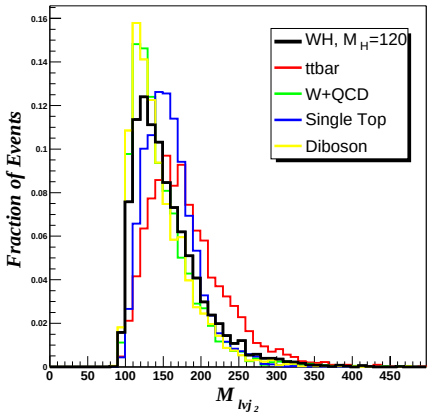
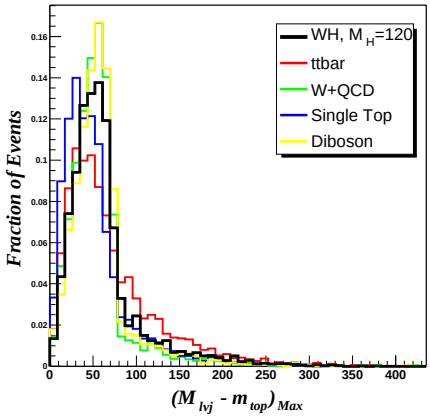


Figure 6.15: More studied NN inputs: (a) $M_{\ell\nu}^{Trans}$; (b) $p_T^{\ell\nu j_1 j_2}$; (c) $cos(\theta(\ell^{WRF}, W^{labRF}))$; (d) $cos(\theta(j_{tag}^{H12RF}, \ell^{H12RF}))$; (e) N_{Extra}^{Objs}; and (f) $\langle \sum E_T^{Extra} \rangle$.

(a)



(a)



(a)

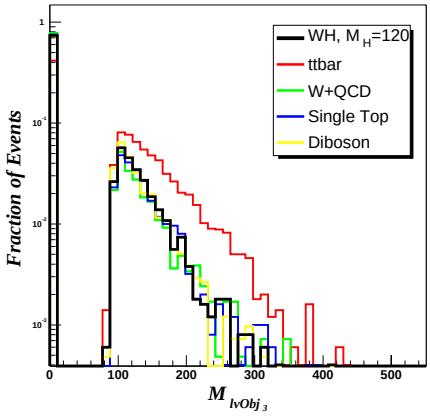


Figure 6.16: Studied NN inputs: (a) $M_{\ell\nu j_2}$; (b) $|M_{\ell\nu j} - m_t|_{Max}$; (c) $M_{\ell\nu Obj_3}$.

Inputs	Hypersphere+$M_{j_1 j_2}$ Window *A priori* limit (pb)
All 31	8.11
27 of 31 (All except $E_T^{j_1}$, $E_T^{j_2}$, $\Delta R(j_1, j_2)$, and $M_{j_1 j_2}$)	8.21
Selected 17	8.33
Best 13: $\not\!\!E_T$, H_T,N_{Extra}^{Objs}, $\sum E_T^{Extra}$, $\langle \sum E_T^{Extra} \rangle$,$M_{\ell\nu j_1}$,$M_{\ell\nu j_2}$ ($E_T^{j_1}$ - $E_T^{j_2}$),$M_{j_1 j_2 Obj_3}$,$M_{\ell\nu Obj_3}$,$M_{\ell\nu j_1 j_2}$, $cos(\theta(j_{tag}^{H12RF}, H12^{labRF}))$,$cos(\theta(\ell^{WRF}, W^{labRF}))$, $p_T^{j_1 j_2}$,$p_T^{\ell\nu j_1 j_2}$	8.60
Best 8: $\not\!\!E_T$, H_T, $\sum E_T^{Extra}$, $M_{\ell\nu j_1}$,($E_T^{j_1}$ - $E_T^{j_2}$),$M_{j_1 j_2 Obj_3}$, $M_{\ell\nu j_1 j_2}$,$p_T^{j_1 j_2}$	8.40
Best 7: $\not\!\!E_T$, H_T, $\sum E_T^{Extra}$, $M_{\ell\nu j_1}$,($E_T^{j_1}$ - $E_T^{j_2}$),$M_{j_1 j_2 Obj_3}$, $M_{\ell\nu j_1 j_2}$	8.48
Best 6: $\not\!\!E_T$, H_T, $M_{\ell\nu j_1}$, ($E_T^{j_1}$ - $E_T^{j_2}$),$M_{j_1 j_2 Obj_3}$,$p_T^{j_1 j_2}$	8.52
Best 5: $\not\!\!E_T$, H_T, $M_{\ell\nu j_1}$, ($E_T^{j_1}$ - $E_T^{j_2}$),$M_{j_1 j_2 Obj_3}$	8.68

Table 6.4: *A priori* limits for different input configurations utilizing the NN+$M_{j_1 j_2}$ window selection.

Eight of these variables were determined to be the optimal set and were used in the NN utilized in this analysis. They are plotted in Figures 6.1- 6.8. The remaining variables that were considered are plotted in Figures 6.15- 6.16; only the $M_H = 120$ GeV/c^2 scenario is plotted for signal.

A 17-22-5 NN was trained using these seventeen variables as inputs. The *a priori* limit achievable with such a network, using the hypershpere method and $M_{j_1 j_2}$ window discussed in Section 6.1.2, was found to be 8.33 pb.

As was discussed earlier, our finite amount of MC statistics motivates that we utilize a NN with as few degrees of freedom as possible that achieves a reasonable *a priori* limit. Our training MC samples for each input class have approximately 4100 events apiece; this allows for NNs with up to around 200 degrees of freedom, if one considers a reasonable number of hidden nodes in a single hidden layer.

Our 17-22-5 NN discussed above exceeds 200 degrees of freedom (it has nearly 500). To reduce the complexity of the NNs in this study we choose to reduce the number of inputs to $N_{Input} < 9$; this allows us to later freely vary the number of hidden nodes while remaining well beneath the limits of our MC samples.

Table 6.4 lists the *a priori* limit achievable from MC for a variety of input choices. The inputs were ranked according to their maximum bin-height differences from Table 6.3, and the top subsets were chosen. Some of the input configurations in Table 6.4 (*All 31*, *27of31*, *Selected 17*, and *Best 13*) have >200 degrees of freedom and are therefore not considered for further study. One should also note that the scenarios utilizing 31 and 27 inputs incorporate $E_T^{j_1}$, $E_T^{j_2}$, $\Delta R(j_1, j_2)$, and $M_{j_1 j_2}$ in the NN, which we seek to avoid. We seek to retain as much information on $M_{j_1 j_2}$ as possible, and employing these variables would sculpt our $M_{j_1 j_2}$ distribution.

From Table 6.4 one sees that the *a priori* limit achievable for the tested scenarios for which our MC is sufficient have very similar values, ≈8.4–8.7 *pb*. The eight input implementation was chosen because in this implementation the effect of the Hypersphere NN cut is greatest. Note that this is by no means an exhaustive study, but that reasonable choices were made that led to the subset of inputs considered.

6.3.3 The Hidden Nodes

We are free to choose the number of hidden nodes in the NN. Recall that we are restricting this analysis to NNs with only one hidden layer of nodes. Using the same inputs discussed in 6.1.2, we trained an 8-input, 5-output NN, varying the number of hidden nodes in the single hidden layer. Table 6.5 contains the *a priori* limit as a function of N_H.

N_H	*A priori* Limit (*pb*)
6	8.43
7	8.41
8	8.58
9	8.52
10	8.41
11	8.40
12	8.40
13	8.52
14	8.46
15	8.52
16	8.54

Table 6.5: Results of N_H impact on *a priori* limit for a 8-N_H-5 NN trained with standard back-propagation using update frequency=10, learning rate=1.0, alpha=0.5.

Previous experience has shown us that the variation of N_H has a very weak effect on the *a priori* limit achievable in such studies [54]. From Table 6.5 one can see that the same conclusion can be drawn for this analysis.

6.3.4 Learning Algorithm Parameters

Recall from Section 4.2 that there are a few tunable parameters within the back-propagation algorithm, including the learning rate, η, and the momentum α. One can also adjust the frequency of weight update, *i.e.*, updating weights after 10 events,

100 events, 1000 events, *etc.*. Optimization of these parameters was performed [60]; the NN was shown to be fairly insensitive to reasonable adjustments in each case.

The learning algorithm described in Section 4.2 is called standard back-propagation. There are variants to this algorithm, and two other members of the back-propagation family were also tested, Langevin and Manhattan learning [46]. Table 6.6 shows the results of training 8-12-5 NNs with these new learning algorithms. There was no significant improvement in the *a priori* limit when employing other back-propagation variants.

Learning Algorithm	*A priori limit* (*pb*)
Standard Back-Propagation	8.40
Manhattan	8.47
Langevin	8.55

Table 6.6: *A priori* limit for a few variations of the back-propagation algorithm.

JETNET has a built-in ability to eliminate (drive to small values) unnecessary or superfluous weights in one's NN. This capability is called *pruning*; one has the freedom to turn this feature on and off. The default NN described in Section 6.1.2 had pruning active. Table 6.7 shows a comparison between equivalent NNs, except that one was trained with pruning active, the other with pruning inactive. We see that pruning had little impact on the ultimate *a priori* limit achieved with this NN.

6.3.5 Determination of $M_{j_1 j_2}$ Window

Our post-NN event selection imposes that the di-jet mass lie within the window 80 GeV/c^2<$M_{j_1 j_2}$<150 GeV/c^2. This window was the optimal choice from a study in

Pruning	*A priori* limit (pb)
Active	8.40
Inactive	8.38

Table 6.7: *A priori* limit for two JETNET scenarios: pruning ON, and pruning OFF.

which $M_{j_1 j_2}^{Min}$ was varied between 50 and 95 GeV/c^2, and $M_{j_1 j_2}^{Max}$ was varied between 135 and 180 GeV/c^2. The possible $M_{j_1 j_2}^{Min}$, $M_{j_1 j_2}^{Max}$ values were varied in 15 GeV/c^2 increments. Using a variation any smaller would have us searching for effects that were significantly below the Run 1 dijet mass resolution of 15% .

The optimization study was performed using post-NN selection samples, and the results are described in Table 6.8. The NN employed is the optimal 8-12-5 described in Section 6.1.2. It should be noted that the Higgs mass hypothesis utilized here is M_H=120 GeV/c^2, and that the optimal choice of $M_{j_1 j_2}$ window is sensitive to the Higgs mass being considered. This study will be repeated for each hypothesis considered in the final analysis.

6.3.6 Summary of Tests

We have studied several different advanced selection scenarios for this $WH \to \ell\nu b\bar{b}$ search. A reasonable input space was chosen for our NN whose training requirements do not exceed the size of our MC samples. We have shown that the *a priori* limit on the WH cross section for this analysis is not particularly sensitive to the tunable parameters of the NN algorithm. Different NN architectures were also studied; the *a priori* limit dependence on the number of hidden and output nodes was not significant. We conclude that our 8-12-5 NN is a reasonable choice. There is somewhat more

$M_{j_1 j_2}$ Window (GeV/c^2)	*A priori limit* (pb)
$50<M_{j_1 j_2}<180$	9.13
$65<M_{j_1 j_2}<180$	8.91
$80<M_{j_1 j_2}<180$	8.77
$95<M_{j_1 j_2}<180$	8.99
$50<M_{j_1 j_2}<165$	8.98
$65<M_{j_1 j_2}<165$	8.73
$80<M_{j_1 j_2}<165$	8.57
$95<M_{j_1 j_2}<165$	8.71
$50<M_{j_1 j_2}<150$	8.84
$65<M_{j_1 j_2}<150$	8.60
$80<M_{j_1 j_2}<150$	8.40
$95<M_{j_1 j_2}<150$	8.42
$50<M_{j_1 j_2}<135$	9.12
$65<M_{j_1 j_2}<135$	9.14
$80<M_{j_1 j_2}<135$	8.91
$95<M_{j_1 j_2}<135$	8.97

Table 6.8: Varying the $M_{j_1 j_2}$ window in post-NN selection. The requirement 80 $GeV/c^2<M_{j_1 j_2}<150$ GeV/c^2 was optimal for $M_H = 120$ GeV/c^2.

stronger dependence in *a priori* limit on the $M_{j_1 j_2}$ window, and this cut needs to be re-optimized for other M_H values.

It should be noted that while much effort was put into investigating different architectures and input configurations, the effect of any of these changes is dwarfed by the inclusion of systematic errors on the expected number of signal and backgrounds events (as will become evident in Chapter 7). So one should understand that as long as reasonable choices are made (and from the above it appears that our choices are reasonable), the chosen NN implementation will not be the dominant source of uncertainty in the ultimate limit measurement.

6.4 Summary

An advanced selection has been designed to aid in the identification of $WH \rightarrow \ell\nu b\overline{b}$ events amidst large amounts of background. We have demonstrated that the employment of this advanced selection, which utilizes a NN and a window requirement in $M_{j_1 j_2}$, reduces the *a priori* limit over a standard technique. These studies have been described for one Higgs mass hypothesis, $M_H = 120\ GeV/c^2$; extensions to other M_H values will be described in Section 7.5. Also, the method of translating predicted signal and background yields into the *a priori* limit will be discuss in the following Chapter.

CHAPTER 7

LIMITS ON THE WH CROSS SECTION FROM SIMULATION STUDIES

In our search for WH production in the $\ell\nu b\bar{b}$ final state, we have applied a two-stage event selection. From simulation studies,we have a prediction for the number of signal and background events we expect to see in the Run 1 data sample. In this document we have described in detail only the expected yields for the hypothesis $M_H = 120\ GeV/c^2$, but a variety of low mass Higgs hypotheses will be considered. Outfitted with the expected yields for $M_H = 120\ GeV/c^2$, we are now capable of measuring an upper limit on the WH production rate.

7.1 Calculating a Limit

From Table 6.2 we see that for $M_H = 120\ GeV/c^2$, $N_{exp}^{S,theory} = 0.15$ and $N_{exp}^{B} = 10.6$ events. Given these numbers, we seek to calculate the *a priori* 95% confidence level (CL) limit on WH production for the mass hypothesis $M_H = 120\ GeV/c^2$.

We first assume that the number of observed events, N_{Obs}, is Poisson distributed:

$$P(x,\mu) = \frac{\mu^x e^{-\mu}}{x!} \tag{7.1}$$

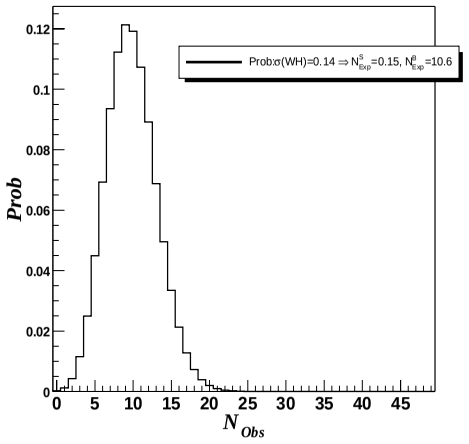


Figure 7.1: The Poisson distribution in N_{Obs} , with mean and variance equal to $N_{exp}^{S,theory}+N_{exp}^{B}$=10.75 events using σ_{WH}^{theory}=0.14 *pb*.

where x=N_{Obs}, and μ=the expected number of observed events.

A feature of the Poisson distribution is that its mean and variance are equal; in the above equation, the mean and variance are denoted by μ. The equivalence of mean and variance in the Poisson probability implies that as the mean increases, the impact of the statistical uncertainty (width of Poisson distribution) becomes less critical with respect to the mean value. For example, if we considered a Poisson with mean μ=9 events, then its statistical uncertainty, σ_{stat}, would be 3 events, and the resolution from statistics (σ_{stat}/μ) would be 3/9=0.33, a 33% uncertainty. If instead we consider a Poisson with a mean μ=100, then it would have statistical uncertainty σ_{stat}=10 and resolution of σ/μ = 0.10, a 10% uncertainty. Depending on the mean of the Poisson, the statistical uncertainty for a specific scenario may be significantly greater or significantly smaller than the systematic uncertainty. This is a feature to keep in mind at all times.

In this analysis, the mean of the Poisson distribution is equal to the sum of the number of signal and background events our selection studies predict. Figure 7.1 shows what this Poisson distribution looks like in N_{Obs} for $\mu = N_{exp}^{S,theory} + N_{exp}^{B} = 10.75$ events.

In this analysis, we seek to measure a limit on the WH cross section, so let us ignore for the time being the theory predictions and consider other possible WH cross section values. Figure 7.2 shows the Poisson distribution in N_{Obs} for σ_{WH}=5,12,20 pb (note that $\sigma_{WH}^{theory} = 0.14$ pb).

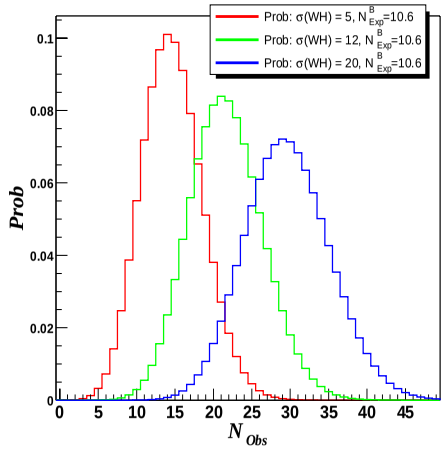


Figure 7.2: Poisson distributions for σ_{WH}=5,12,20 pb.

By constructing these Poisson distributions for various values of σ_{WH}, one can map the entire σ_{WH} versus N_{Obs} plane, pictured in Figure 7.3. The three curves in Figure 7.2 represent the three horizontal bins at σ_{WH}=5,12,20 pb in Figure 7.3. One can see that for $\sigma_{WH} = \sigma_{WH}^{theory} = 0.14$ pb, the largest probability scenario predicts N_{Obs}=9 – 12 events, which is a direct result of our expectation $N_{exp}^{S,theory} + N_{exp}^{B} = 10.75$ events.

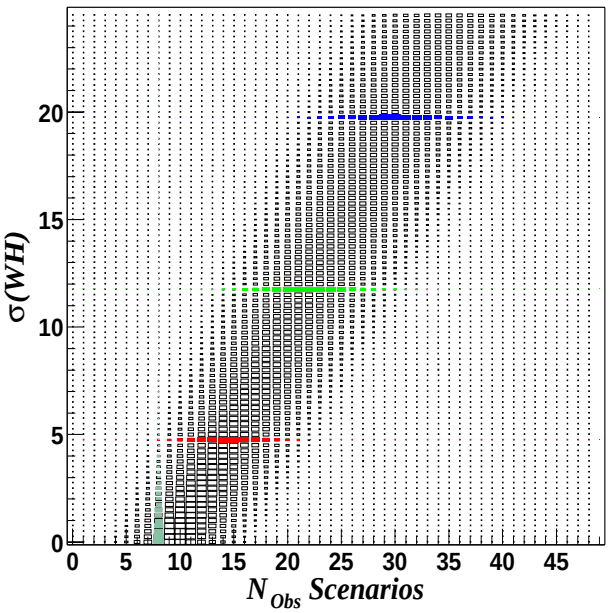


Figure 7.3: Poisson probabilities of hypothetical σ_{WH} values as a function of N_{Obs}.

The possible outcomes of this analysis applied to the Run 1 data sample are the different N_{Obs} scenarios plotted along the horizontal axis of Figure 7.3. Consider a single bin of N_{Obs}, say N_{Obs}=8 events; this scenario is a single vertical bin of Figure 7.3. We can look at the probability distribution for σ_{WH} in this bin, which is pictured in Figure 7.4. Since we expect $N_{exp}^{S,theory}+N_{exp}^{B}$=10.75 events with σ_{WH}^{theory}=0.14 pb, and since N_{Obs}=8 events corresponds to a deficit with respect to expectation, we see that in the N_{Obs}=8 events bin, small values of σ_{WH} are favored. If we would look at a bin of N_{Obs} that corresponds to an excess over expectation, larger σ_{WH} would be favored.

We refer to the σ_{WH} probability distribution pictured in Figure 7.4(a) as the *likelihood* distribution, $\mathcal{L}(\sigma_{WH}|N_{Obs} = 8)$, for σ_{WH} given N_{Obs}= 8 events. In any single N_{Obs} bin, the 95% CL limit on the WH cross section is defined as:

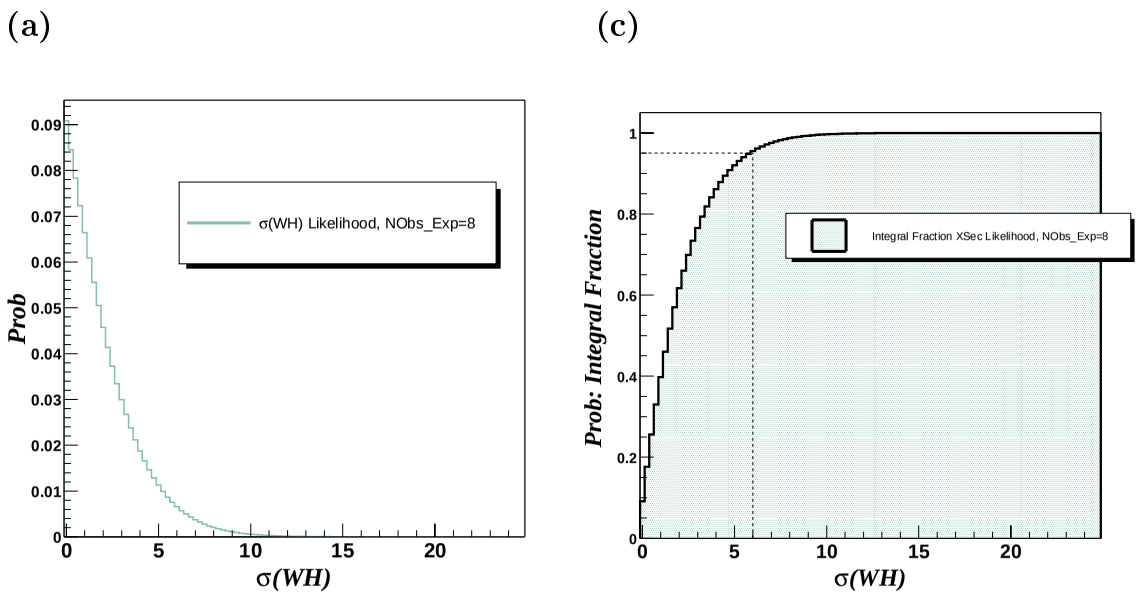


Figure 7.4: (a): The likelihood distribution for N_{Obs}= 8 events; (b): Fractional inverse integral plot for the N_{Obs}= 8 events likelihood.

$$0.95 = \frac{\int_0^{\sigma_{WH}^{95}} d\sigma_{WH} \mathcal{L}(\sigma_{WH}|N_{Obs})}{\int_0^{\infty} d\sigma_{WH} \mathcal{L}(\sigma_{WH}|N_{Obs})} \tag{7.2}$$

Equation 7.2 just determines the value of σ_{WH} at which 95% of the integrated likelihood distribution is to the left in Figure 7.4. Figure 7.4(b) is the fractional integral plot of the distribution in Figure 7.4(a); from Figure 7.4(b) we see that for N_{Obs}= 8 events, the 95% CL limit on σ_{WH} is $\approx 6.0pb$. The 95% CL limit on σ_{WH} can be calculated for each N_{Obs} value; Table 7.1 contains these limits for several values of N_{Obs}. Recall that σ_{WH}^{theory}= 0.14 pb, for which the total number of expected signal and background was $N_{exp}^{S,theory}$+N_{exp}^{B}=10.75 events.

N_{Obs}	σ_{WH} Limit (pb)
2	3.4
3	3.7
4	4.0
5	4.4
6	4.9
7	5.4
8	6.0
9	6.7
10	7.4
11	8.2
12	9.1
13	10.1
14	11.1
15	12.1
16	13.2
17	14.3
18	15.4
19	16.6
20	17.7
21	18.8
22	20.0

Table 7.1: The 95% CL limit on σ_{WH} for a few values of N_{Obs}. These limit values do not incorporate the effect of systematic errors.

7.2 The *A Priori* Limit

The method described above provides one with the 95% CL limit on the WH production cross section as a function of the different possible outcomes of the analysis. We seek now a measure by which one can incorporate information from all the possible outcomes of the experiment. We call this figure-of-merit the *a priori* limit, and it is calculated by weighting the limits from each N_{Obs} scenario by their Poisson probability given $\sigma_{WH}=\sigma_{WH}^{theory}=0.14$ pb. Thus the *a priori* limit is given by:

$$\sigma_{WH}^{95,apriori} = \sum_{N_{obs}} \sigma_{WH}^{95} * P(N_{obs}|N_{exp}^{S,theory} + N_{exp}^{B}) \tag{7.3}$$

where $P(x|\mu)$ is the Poisson distribution in x with mean μ. The *a priori* limit is useful in comparing different choices within the analysis; in this search we used the *a priori* limit as the benchmark in optimization studies. One can also use the *a priori* limit in comparing completely different analyses.

With the prescription described above,

$$\sigma_{WH}^{95,apriori} = 8.4pb \qquad \text{(no systematics) for } M_H = 120\ GeV/c^2$$

The procedure described for calculating the 95% CL limit on the WH production cross-section is a Bayesian calculation. There is some controversy in the HEP community regarding Bayesian and Frequentist methods which we choose not to address here. Studies were performed using the Frequentist approach of Feldman and Cousins [58] for the calculation of limits. The Bayesian treatment was chosen because of the consistent prescription for the incorporation of systematic errors, which is the focus of the remainder of this note. It should be noted that for some N_{Obs} values, the Feldman-Cousins method produced smaller 95% CL limits, and that this Frequentist method yielded a slightly smaller *a priori* limit (see Table 7.2 for a small number of N_{Obs} examples). But without a consistent technique for incorporating systematic errors, the Frequentist approach is disregarded.

N_{Obs}	σ_{WH} Limit (pb)	
	Bayesian	Frequentist
5	4.4	2.5
10	7.4	6.9
15	12.1	13.2
20	17.7	18.9
A priori limit (pb)	8.4	8.1

Table 7.2: A comparison of Bayesian and Frequentist 95% CL limits for a few N_{Obs} scenarios.

7.3 Impact of Systematic Errors

The measurement of a limit on the WH production cross-section — or any limit, for that matter — is not as neat and tidy as the prescription described in Section 7.1. In the treatment above, we measured the efficiency for selecting signal and background events from samples of simulated events generated by Monte Carlo programs. These simulated collisions were then input to the CDF Run 1 detector simulation, and our selection criteria was applied to the measurable quantities that approximate the detector's response to the outgoing particles from these collisions. Overall efficiencies were calculated for each signal and background channel, and, along with cross section and luminosity values, expectation values for each channel were calculated.

However we know there are uncertainties on the event expectations we have calculated. We know that these Monte Carlo programs occasionally make approximations at particular points in their generation of physics collisions; what impact do these choices have on our efficiency measurements, which rely on these Monte Carlo events for input? We have a final state that includes a lepton, a neutrino (which manifests itself as $\not\!\!E_T$) and two jet objects; therefore any shortcomings of the simulation of the

detector to leptons and jets and their energy measurements will correspond to mismeasured efficiencies for such final states. The Run 1 luminosity was not measured exactly (105.1 pb^{-1}); how do we incorporate a ∼4% uncertainty on this value (105.1 $\pm 4pb^{-1}$)?

The answers to these questions will be pursued in the following sections. A few caveats should be noted before we embark on the in-depth discussion of the inclusion of systematic errors. First, the effect of systematics will only be studied with regard to signal and the two largest background sources: $t\bar{t}$ and $W^{\pm}$+QCD (Class 3), which constitute ∼85% of the predicted background sample passing all selection criteria. Systematics for all $W^{\pm}$+QCD (Class 3) backgrounds will be approximated by $W + b\bar{b}$ production only; because of the shape similarities between each of the Class 3 backgrounds (see [60]), this approximation is not without reason. The effect of systematic errors in background Classes 4 and 5 (Class 4: single top; Class 5: diboson, Z+X) will not be measured directly. Instead we will use the average of the cumulative systematic uncertainties on the expected number of $t\bar{t}$ and $W + b\bar{b}$ events to define the uncertainty on the expectations from Classes 4 and 5. This is a reasonable approximation since the Class 4 and Class 5 backgrounds are not wholly different topologically from the more dominant backgrounds; also, the systematic effects are not strongly mechanism-dependent, as will be apparent below.

7.3.1 Jet Energy Scale

There are several uncertainties in the correction factors that are applied to jet energies. The basis of these jet corrections include:

- relative energy scale – correcting for any η dependence in the jet energy measurement

- absolute energy scale – corrections from attempts to determine incoming particles' energies from calorimeter response in testbeam and actual Run 1 data
- extraneous event activity – more than one p-$\overline{p}$ interaction within the same event will affect how well jet energies are measured
- energy spillage from considerably wide jets – we define a jet to lie within a cone of radius 0.4; this correction takes into account jet energy beyond this cone.

The effect of these energy scale uncertainties was measured by adjusting each of the contributing parameters by ± 1 σ and then re-measuring the efficiencies for WH, $t\bar{t}$, and $W + b\bar{b}$. The portion of the event selection that is most affected by this systematic effect is the 2-jet requirement. Figure 7.5 shows the distribution of the number of E_T>15 GeV, $\mid \eta \mid$<2.0 jets for each sample being considered here (we call this number $N_{Jets}^{HighET,cent}$). We see in Figure 7.5(b), for example, that the $t\bar{t}$ efficiency for this requirement decreases ∼8% when the energy scale uncertainties are shifted $+1\sigma$, and the $t\bar{t}$ jet-selection efficiency increases by ∼8% when the energy scale uncertainties are shifted -1σ. This is a consequence of the nature of the $N_{Jets}^{HighET,cent}$ distribution for $t\bar{t}$, which is peaked in the 3- and 4-jet bin. As the jet energy scale is decreased by 1σ, some fraction of ≥3-jet events are demoted to the 2-jet bin when the energies of one or more of their jets are shifted below our E_T>15 GeV threshold. Similarly, when the jet energy scale is increased by 1σ, 1-jet events with other, weaker jet objects that nominally fail our jet selection, are promoted to the 2-jet bin. The converse of this argument is evident through the effect in the $W + b\bar{b}$ sample, whose $N_{Jets}^{HighET,cent}$ distribution peaks in the 0- and 1-jet bin. The $N_{Jets}^{HighET,cent}$ distribution for WH peaks in the 2-jet bin, so one would expect to see as much promotion and

demotion for both of the $\pm\sigma$ scenarios, essentially washing out the effect, which we do see some evidence for.

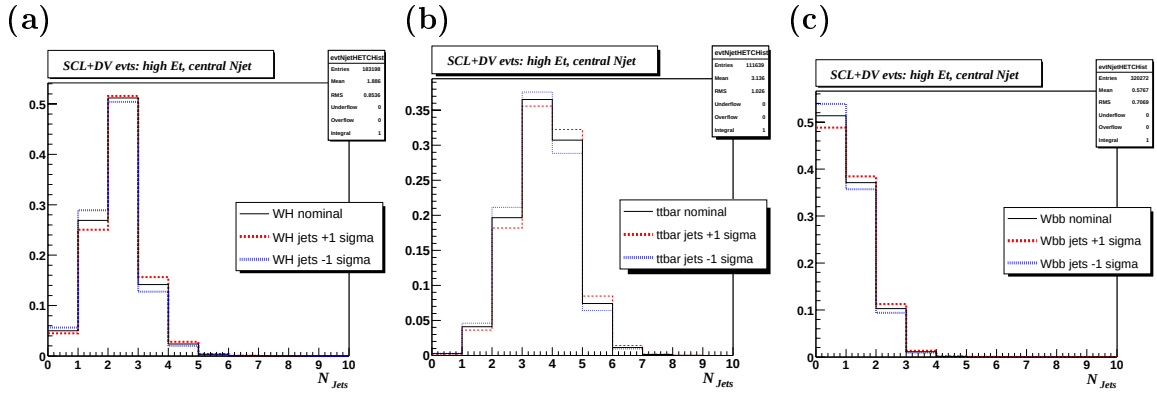


Figure 7.5: Effect of varying the jet energy scale uncertainties $\pm 1\ \sigma$ on WH, $t\bar{t}$, and $W + b\bar{b}$ samples on the number of E_T>15 GeV, $\mid \eta \mid$<2.0 jets.

	Efficiency Change (%)		
	WH	$t\bar{t}$	$W + b\bar{b}$
Shift Correction Factors -1σ	-2.7	13.7	-6.1
Shift Correction Factors +1σ	0.6	-11.3	6.6

Table 7.3: Efficiency changes in % when shifting the factors contributing to the jet energy corrections by $\pm 1\ \sigma$.

While we have focused on the effect of this systematic error on the jet selection, it should be noted that the jet energy scale systematic effect also affects other portions of the event selection, most prominently the dijet mass requirement, 80 GeV/c^2<$M_{j_1 j_2}$<150 GeV/c^2. The change in jet energies can also affect the neural network selection, but this effect is expected to be small since we were careful not to

rely too heavily on jet-E_T influenced variables as our inputs. Although the details are not reported here, the effect in these requirements was studied and was shown to be significantly smaller ($\sim$3%) than in the jet selection ($\sim$8%).

Later in this note we will discuss our implementation of pseudoexperiments in the construction of smeared likelihood distributions. Care was taken to preserve the correlations among the efficiency changes in Table 7.3 between each channel. This was accomplished by approximating the uncertainties in the jet energy correction factors with a Gaussian with mean zero and unit variance, selecting the shift in correction factor randomly from this Gaussian in each pseudoexperiment, and then mapping out the response in the efficiency change according to a linear scaling to the values in Table 7.3. Such a method ensures that the unphysical scenario in which signal and both background channels fluctuated up or down simultaneously is avoided. Figure 7.6 shows the selection efficiencies for WH, $t\bar{t}$, and $W+b\bar{b}$ over 10000 pseudoexperiments.

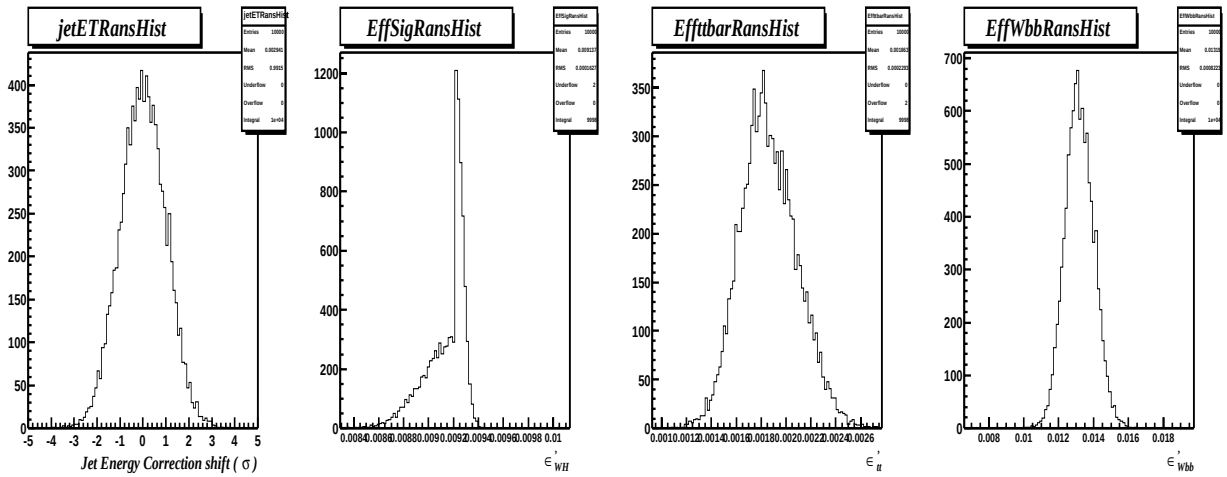


Figure 7.6: Distribution of efficiency changes in pseudoexperiments due to jet energy scale effects. Correlations between WH, $t\bar{t}$ and $W+b\bar{b}$ channels are preserved.

7.3.2 Radiation Modeling

Monte Carlo generators do a fine job of providing experimenters with simulated collisions showered to final states with which to understand detector response and the impact of selection criteria. However, there are underlying models in these programs that have some inherent uncertainty. For example, the modeling of initial state radiation (ISR) — radiated gluons off p- or $\overline{p}$-constituent partons — is not completely understood, and in order to achieve reasonable results with these programs some assumptions were made.

The amount of initial state radiation can affect the number of jets in each event. This is a sensitive issue, considering that we have a somewhat strict N_{Jets}=2 requirement. To evaluate the impact that ISR has on our selection efficiency, we generate WH, $t\bar{t}$, and $W + b\bar{b}$ samples using Pythia for the parton shower and request all ISR be OFF, and then re-measure the cumulative efficiency changes.

We quantify the effect of ISR as half the absolute difference between the nominal efficiency and the no-ISR efficiency:

$$\delta\epsilon^{ISR} \equiv \frac{|\epsilon^{noISR} - \epsilon^{0}|}{2\epsilon^{0}} \tag{7.4}$$

With this prescription we measure the effect of over-estimating the amount of ISR in our samples by looking at the extreme case in which all ISR is removed. It is clear that considering such a scenario in which there is no ISR is unphysical. Thus we approximate the impact of ISR as half the difference between the nominal and no-ISR scenarios. We also assume that an underestimate of the ISR in our samples will have an equivalent efficiency change as measured with the no-ISR samples.

ISR effects were shown to have a 10-11% effect on the WH and $t\bar{t}$ selection efficiency. The effect in the $W + b\bar{b}$ samples however was much smaller. This is due to the nature of a $W + b\bar{b}$ event; recall from Figure 2.7 that the b quarks produced in such events come from the splitting of a gluon that radiated off an initial state parton. So in our $W + b\bar{b}$ Monte Carlo sample, when we turn off ISR, we are really turning off second-order initial state radiation effects while ensuring that we still have gluon radiation to provide the b jets.

In a similar vein, there are some uncertainties encountered in the models of final state radiation (FSR). To evaluate the impact of FSR, one does not have the luxury of simply turning off FSR in Pythia[2]. Instead one must look at the subset of no-ISR events that have all their jets matched to partons from the hard scatter; such events are assumed to have had no hard gluon emissions and can then be considered to be essentially a no-ISR+no-FSR sample. We then define quantify the effect of FSR with respect to the no-ISR sample:

$$\delta\epsilon^{FSR} \equiv \frac{|\epsilon^{noISR,noFSR} - \epsilon^{noISR}|}{2\epsilon^{noISR}} \tag{7.5}$$

FSR effects were determined to have a 10-15% effect on the backgrounds, and a 6% effect on signal.

7.3.3 Parton Distribution Functions

When generating p-$\overline{p}$ collisions, some model of the structure of the proton must be employed. There is some uncertainty associated with these parton distribution

[2]This is not true, actually. One can turn off FSR in Pythia, but this also turns off the soft gluon emissions that give jets their shapes, which is a quality that we absolutely must retain.

functions(PDFs), and these uncertainties could have an effect on our signal and background efficiencies.

Pythia v6.129a [36] was used to generate the WH and $t\bar{t}$ samples. The default PDF choice in that version of Pythia was GRV94L [63], a leading order calculation. The $W + b\bar{b}$ samples were generated with ALPGEN [50]; unfortunately, ALPGEN does not have GRV94L as part of its linked PDF library. So the $W + b\bar{b}$ samples was generated with CTEQ5L [64], another leading order calculation from the CTEQ group.

To gauge the effect of these PDF choices, new Pythia samples were generated using CTEQ5L instead of GRV94L, and new ALPGEN $W+b\bar{b}$ samples were generated using MRST01 [65]. The efficiencies for WH, $t\bar{t}$, and $W + b\bar{b}$ were re-evaluated using these new samples. The effect on the selection efficiency was defined:

$$\delta\epsilon^{PDF} \equiv \frac{|\epsilon^{newPDF} - \epsilon^{0}|}{\epsilon^{0}} \tag{7.6}$$

The PDF choice changed the WH and $t\bar{t}$ efficiencies by only a few percent. The effect in $W + b\bar{b}$ was slightly larger, $\delta\epsilon^{PDF}_{Wbb} = 8\%$. These values are not significantly different from other, similar Run 1 $\ell\nu b\bar{b}$ final state analyses.

7.3.4 Signal and Background Generators

Besides the PDF choice, there could be other sources of systematic error arising from the Monte Carlo generator used for constructing our signal and background samples. Within the generator, models are employed that describe, for example, the hadronization process. The models employed are determined by choices made by the authors. We seek to study the impact of these choices by comparing the signal

and background acceptances for different Monte Carlo generators. To estimate the systematic error due to using Pythia for the WH and $t\bar{t}$ samples, we generate WH and $t\bar{t}$ samples using ALPGEN and re-measure the selection efficiencies. To estimate the systematic error due to generating $W+b\bar{b}$ using ALPGEN, we use a Pythia $W+b\bar{b}$ sample generated for another analysis.

We define the effect on the selection efficiency this way:

$$\delta\epsilon^{Gen} \equiv \frac{|\epsilon^{newGen} - \epsilon^{0}|}{\epsilon^{0}} \tag{7.7}$$

It should be noted that our ALPGEN $W+b\bar{b}$ sample was generated with loose parton-level demands: we asked for two outgoing b-partons from the hard scatter satisfying $p_T > 0.0$, $|\eta| < 15.0$, and $\Delta R(b_1, b_2) > 0.0$. Because of the time required to generate $W + b\bar{b}$ events, we utilized the Pythia $W + b\bar{b}$ sample from the latest Run 1 single top search [54]. However in this sample we did not have the freedom to make equivalent parton-level demands as were imposed in the ALPGEN sample. Dissimilar parton-level cuts on the MC samples could yield misleading efficiency effects that are not generator-induced *per se*. For this reason, we assume that uncertainties associated with the generator choice will have only a minor effect on the basic event selection. Other Run 1 analyses have estimated the effect due to generator choice to be $\sim$4% for the basic selection we are utilizing in this analysis. We then consider only generator-induced effects on the advanced selection efficiencies, the neural network and $M_{j_1 j_2}$ cuts. We find that the generator choice imposes a 12-13% shift on the $t\bar{t}$ and $W + b\bar{b}$ selection efficiencies, while the effect is very small in WH.

7.3.5 The Top Quark Mass

The top quark mass was measured in Run 1 to be approximately 175 $\pm 5\ GeV/c^2$. For our Monte Carlo $t\bar{t}$ samples, we demanded however that m_t=175 GeV/c^2. We seek to quantify the effect of the $\pm 5\ GeV/c^2$ uncertainty on the measured top mass on our $t\bar{t}$ selection.

To evaluate this effect, we generate with Pythia new $t\bar{t}$ samples: one sample with m_t= 170 GeV/c^2 and another sample with m_t= 180 GeV/c^2. We define the effect on the $t\bar{t}$ efficiency to be the largest of these two scenarios. We find that for m_t= 170 GeV/c^2, the $t\bar{t}$ efficiency increases by 2% , but in the case m_t= 180 GeV/c^2, the $t\bar{t}$ efficiency decreases by 5.6%. We assume that the uncertainty on m_t has no impact on the WH and $W + b\bar{b}$ samples.

7.3.6 Other Systematic Effects

There are a few more systematic uncertainties that we must account for. The uncertainty in the b-tagging, the lepton identification, the Run 1 high p_T lepton trigger, and the Run 1 luminosity measurement have been studied previously, allowing us to use their results directly. Table 7.4 has a summary of the b-tagging systematic, which varies in each channel according to the tag efficiency (the uncertainty from tagging was determined to be 10% of its value; we also insert a 10% uncertainty on our tagging scale factor [44]). The lepton identification and high p_T lepton trigger uncertainties were measured in tandem; the uncertainty on the product of these efficiencies was determined to be 10% as well. Given that ϵ_{lepID}=0.854 and ϵ_{trig}=0.903 [44], we find that these effects change the efficiency in each channel by 7.7% . Finally the uncertainty on the Run 1 luminosity measurement affects the expected number of events in

each channel in the same way as any change in efficiency. Thus we must account for this 4% effect because the Run 1 luminosity was determined to be $105.1 \pm 4\ pb^{-1}$ [38].

	WH	$t\bar{t}$	$W + b\bar{b}$
Prior-to-scaling ϵ_{tag} (%)	32.5 ± 3.3	37.3 ± 3.7	33.9 ± 3.4
Scale Factor	1.08 ± 0.11	1.08 ± 0.11	1.08 ± 0.11
Post-scaling ϵ_{tag} (%)	35.1 ± 4.9	40.4 ± 5.7	37.3 ± 5.2
Uncertainty (% of ϵ_{tag})	14.0	14.0	14.0

Table 7.4: Raw tagging efficiency, scale factor and subsequent 14% uncertainty on ϵ_{tag} for WH, $t\bar{t}$, and $W + b\bar{b}$.

7.3.7 Summary of Impact of Systematic Effects

We have investigated several sources of systematic error in this analysis, a search for $WH \rightarrow \ell\nu b\bar{b}$ in the Run 1 data. We attempted to be sensitive to effects that affected measurement of the efficiency for identifying an $\ell\nu b\bar{b}$ final state. Table 7.5 contains a summary of all the systematic effects considered in this analysis. For the purposes of calculating the cumulative effects, the larger of the $\pm 1\ \sigma$ uncertainties for the jet energy scale was used. We see that we will know the expected number of signal WH events to 21% ; we do a bit worse in the background channels, only being able to determine the expectations to within ∼30% . We use these numbers for the signal and backgrounds to approximate the systematic effect in Classes 4 and 5 as possessing a 25% uncertainty as well.

Table 8.2 lists a summary of the expected number of events in each Class including these systematic effects (Table 8.2 is just an updated version of Table 6.2 but with

uncertainty on the expected yields). These results are consistent with other Run 1 analyses in the lepton+jets channel [34], [38].

Effect	Efficiency Change (%)		
	WH	$t\bar{t}$	$W + b\bar{b}$
jet energy corrections -1σ	-2.7	13.7	-5.9
jet energy corrections +1σ	0.6	-11.3	6.7
ISR	10.4	10.9	1.7
FSR	5.7	10.8	14.9
PDF	3.8	3.2	8.3
Generator	0.2	12.1	13.1
m_t	0.0	5.6	0.0
LepID+Trigger	7.7	7.7	7.7
Tagging	14.0	14.0	14.0
Luminosity	3.8	3.8	3.8
Cumulative	20.7	29.7	27.9

Table 7.5: A summary of the systematic effects influencing this analysis for signal WH and the two largest backgrounds, $t\bar{t}$ and $W + b\bar{b}$.

	Class 1	Class 2	Class 3	Class 4	Class 5
Type of Events	WH (M_H= 120 GeV/c^2)	$t\bar{t}$	$W + b\bar{b}$ $W + c\bar{c}$ $W + c$ $W^{\pm}$+mistags non-W	Single top (W -$gluon$ and $W*$)	WZ, WW,ZZ, $Z + b\bar{b}$, $Z + c\bar{c}$, $Z \to \tau\tau$
Luminosity	105.1	105.1	105.1	105.1	105.1
Cross Section(pb)	0.14	6.5		1.7, 0.73	
Basic Efficiency	1.2E-2	8.0E-3			
Basic N_{exp}	0.19	5.4	22.7	2.5	2.7
Advanced Efficiency	0.781	0.249	0.319	0.355	0.439
Final N_{exp}	0.15 $\pm$ 0.03	1.3 $\pm$ 0.4	7.2 $\pm$ 2.0	0.9 $\pm$ 0.2	1.2 $\pm$ 0.3

Table 7.6: Run 1 event expectations with systematic uncertainties on signal and backgrounds.

7.4 Limit Calculation Including Systematic Effects

The spirit of the calculation does not change very much from what was described in Section 7.1. However, where previously we used a single-valued efficiency for WH, $t\bar{t}$, and $W + b\bar{b}$ events, now one must replace these efficiencies with functions that incorporate the uncertainty that is evident from the studies outlined above. This technique is called *smearing* the likelihood and is a somewhat standard practice when dealing with the effects of systematic error.

We must perform integrals similar to those in Equation 7.2 but now there are a few more terms in the integrands. Equation 7.2 is replaced by:

$$0.95 = \frac{\int_{\sigma_{WH}=0}^{\sigma_{WH}^{95}} \int_{\epsilon=0}^{\infty} \begin{array}{c}[d\epsilon d\sigma_{WH} \mathcal{L}(Lum, \epsilon_{WH}, \epsilon_{t\bar{t}}, \epsilon_{Wbb}, \sigma_{WH}|N_{Obs}) \\ *\mathcal{G}(\epsilon_{WH}, \delta\epsilon_{WH}) * \mathcal{G}(\epsilon_{t\bar{t}}, \delta\epsilon_{t\bar{t}}) * \mathcal{G}(\epsilon_{Wbb}, \delta\epsilon_{Wbb})]\end{array}}{\int_{\sigma_{WH}=0}^{\infty} \int_{\epsilon=0}^{\infty} \begin{array}{c}[d\epsilon d\sigma_{WH} \mathcal{L}(Lum, \epsilon_{WH}, \epsilon_{t\bar{t}}, \epsilon_{Wbb}, \sigma_{WH}|N_{Obs}) \\ *\mathcal{G}(\epsilon_{WH}, \delta\epsilon_{WH}) * \mathcal{G}(\epsilon_{t\bar{t}}, \delta\epsilon_{t\bar{t}}) * \mathcal{G}(\epsilon_{Wbb}, \delta\epsilon_{Wbb})]\end{array}} \tag{7.8}$$

In the above expression, ϵ_i represents the nominal efficiency for channel i, and $\delta\epsilon_i$ represents the uncertainty for each channel reported in Table 7.5. This expression will be evaluated for each value of N_{Obs} as before.

It is potentially quite a chore to integrate these expressions by hand. Instead we perform a large number of *pseudoexperiments* in order to construct the smeared likelihoods properly, which can then be integrated numerically, saving us time. In our pseudoexperiments we perform the following procedure a large number of times for each N_{Obs} scenario:

1. Gaussian fluctuate the efficiency for each channel according to each systematic effect.

2. This yields new WH, $t\bar{t}$, and $W + b\bar{b}$ efficiencies $(\epsilon'_{WH}, \epsilon'_{t\bar{t}}, \epsilon'_{Wbb})$, and a new luminosity$(Lum')$.

3. Construct the pseudoexperiment likelihood, $\mathcal{L}_{pse}(Lum', \epsilon'_{WH}, \epsilon'_{t\bar{t}}, \epsilon'_{Wbb}, \sigma_{WH}|N_{Obs})$

4. Add $\mathcal{L}_{pse}$ to an ensemble likelihood $\mathcal{L}^{\star}$

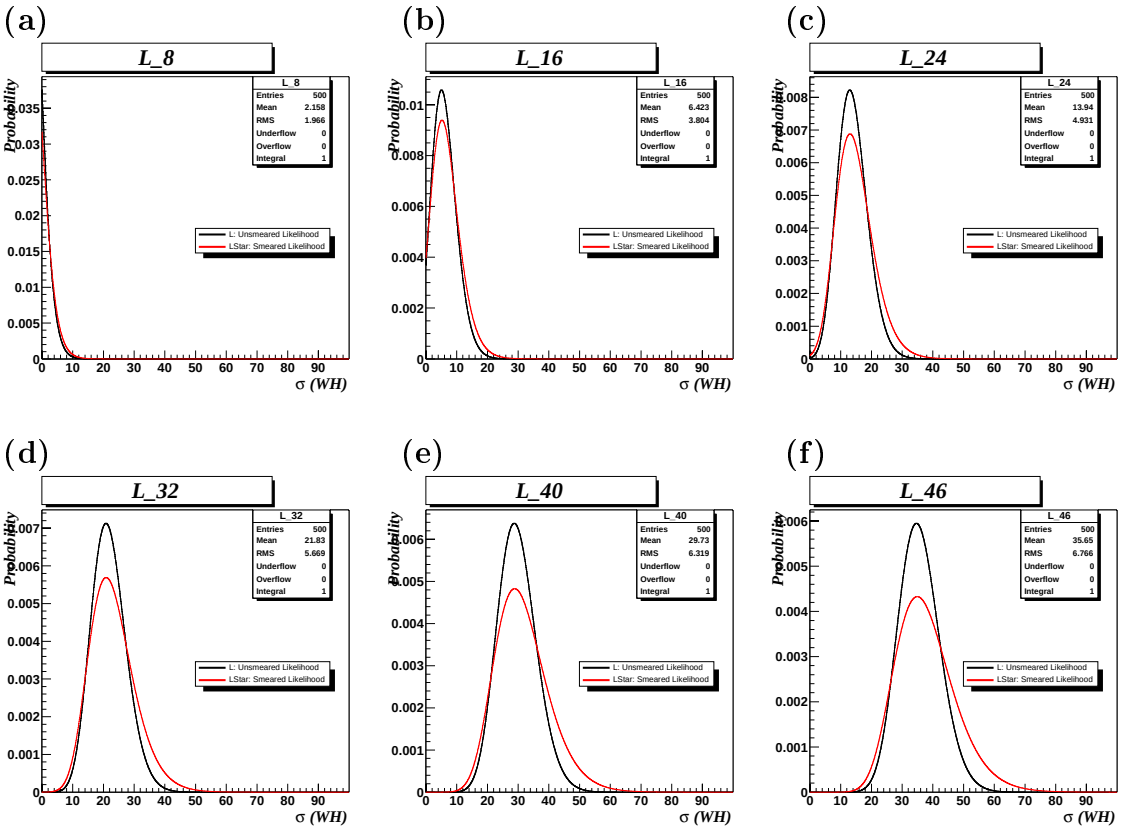


Figure 7.7: A comparison of unsmeared and smeared σ_{WH} likelihoods for N_{Obs}= 8, 16, 24, *etc.* values.

We note again that in these pseudoexperiments, the correlations in the WH, $t\bar{t}$, and $W + b\bar{b}$ samples induced by fluctuating the parameters that contribute to the jet

energy scale uncertainty (one of the dominant systematics) are preserved. (See Figure 7.6.) The Gaussians we use to fluctuate the efficiency in each pseudoexperiment have means equal to the efficiency before fluctuation and variance equal to $\delta\epsilon$ for each effect. After a large number of pseudoexperiments, the normalized version of $\mathcal{L}^{\star}$ is the smeared likelihood we seek. This distribution can then be integrated numerically to determine the 95% CL limit in the N_{Obs} bin in question:

$$0.95 = \frac{\int_0^{\sigma_{WH}^{95}} d\sigma_{WH} \mathcal{L}^{\star}(\sigma_{WH}|N_{Obs})}{\int_0^{\infty} d\sigma_{WH} \mathcal{L}^{\star}(\sigma_{WH}|N_{Obs})} \tag{7.9}$$

Figure 7.7(a)-(f) contains the $\mathcal{L}^{\star}$ distribution for a variety of N_{Obs} values. We see that in fact these smeared likelihoods have means and widths deviated from the nominal cases. We then follow this procedure to calculate the 95% CL limit in each of the N_{Obs} bins. Table 7.7 contains the limit values for a number of N_{Obs} scenarios, represented graphically in Figure 7.8 overlayed with the baseline Poisson. We see that the limit on WH in the most likely outcomes of the Run 1 is $\sim$8 pb. We can then use Equation 7.3 to calculate the *a priori* limit including systematic effects. We find:

$$\sigma_{WH}^{95,apriori} = 10.0pb \qquad \text{(with systematics) for } M_H = 120\ GeV/c^2$$

7.5 Extension to Other M_H Values

The description in this document has up to now considered only the M_H=120 GeV/c^2 hypothesis. We now seek to measure the WH production cross section limit

N_{Obs}	σ_{WH} Limit (pb)
2	3.9
3	4.3
4	4.8
5	5.3
6	5.8
7	6.5
8	7.2
9	8.0
10	8.9
11	9.8
12	10.8
13	11.9
14	13.1
15	14.2
16	15.3
17	16.6
18	17.8
19	19.2
20	20.4
21	21.8
22	23.1

Table 7.7: The 95% CL limit on σ_{WH} for a few values of N_{Obs} including the effect of systematic errors.

for six other M_H values in the range 100 $GeV/c^2 < M_H < 150$ GeV/c^2 at 10 GeV/c^2 intervals. The extension of the analysis to other M_H is values is straightforward.

The NN described in Section 6.1 was trained using simulated signal data with $M_H = 120$ GeV/c^2. It is natural to expect that for each M_H value considered that a new NN would be trained using signal samples for the Higgs mass hypothesis in question. However, there is little M_H dependence on the performance of the NN when used with other M_H hypotheses. For example, Table 7.8 shows the *a priori* limit for two NNs considering the $M_H = 140$ GeV/c^2: one utilizes a new NN trained with the proper signal sample; and one simply uses the $M_H = 120$ GeV/c^2 NN. One can see

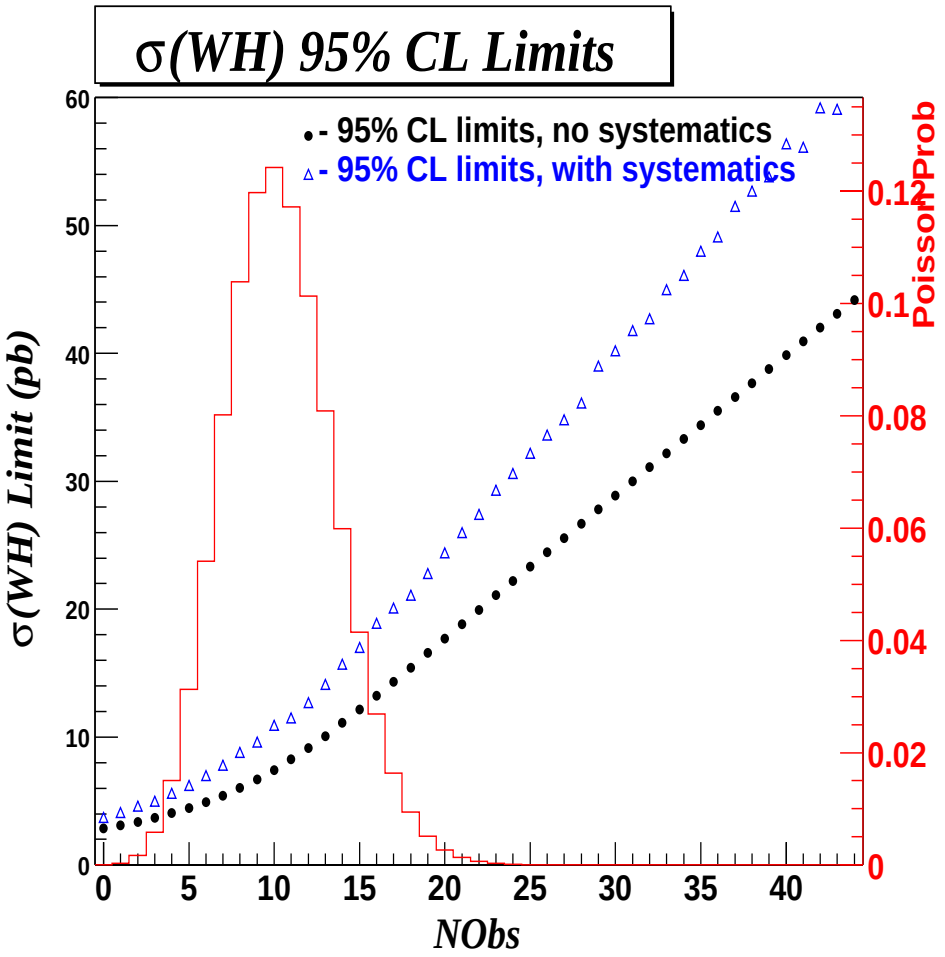


Figure 7.8: Poisson probability distribution with mean, variance $N_{exp}^{S,theory}+N_{exp}^{B}$=10.53 (right axis) overlayed with a graph of the 95% CL limit (left axis), both plots as a function of N_{Obs}, including impact of systematic errors.

that there is no significant difference between these two implementations. Therefore, we conclude that it is safe for the hypothesis $M_H = 140\ GeV/c^2$ to use the NN trained with $M_H = 120\ GeV/c^2$ samples. We further assume that the $M_H = 120\ GeV/c^2$ NN is suitable for each Higgs mass being considered. This is a reasonable assumption since by construction we avoided input variables that were sensitive to the di-jet mass $M_{j_1 j_2}$ which is strongly dependent on M_H.

Neural Network	*A priori* limit
Trained with M_H= 140 GeV/c^2 signal	13.48
Trained with M_H= 120 GeV/c^2 signal	13.50

Table 7.8: Comparison of two NN implementations for M_H= 140 GeV/c^2 analysis.

The di-jet mass requirement did however have to be re-optimized for other M_H values. Recall that in our description of the M_H=120 GeV/c^2 analysis, as our final advanced selection requirement we demanded 80 $GeV/c^2 < M_{j_1 j_2} <$ 150 GeV/c^2. Clearly this $M_{j_1 j_2}$ requirement is not a good choice for $M_H >$ 140 GeV/c^2, as our requirement will cut out a large amount of signal while letting unnecessary background levels at low $M_{j_1 j_2}$. This cut was re-optimized for each M_H value, and the final $M_{j_1 j_2}$ requirement for each M_H hypothesis is listed in Table 7.9.

M_H (GeV/c^2)	$M_{j_1 j_2}$ Window (GeV/c^2)
100	$60 < M_{j_1 j_2} < 130$
110	$70 < M_{j_1 j_2} < 140$
120	$80 < M_{j_1 j_2} < 150$
130	$90 < M_{j_1 j_2} < 160$
140	$100 < M_{j_1 j_2} < 170$
150	$110 < M_{j_1 j_2} < 180$

Table 7.9: Optimized $M_{j_1 j_2}$ requirements for each M_H value considered in this analysis.

With the final portion of the selection defined in Table 7.9 we essentially have a slightly different analysis for each M_H value. The effect of systematic errors was addressed in the same way described in Section 7.3 for the hypotheses M_H= 110

and 140 GeV/c^2. The systematic effects for M_H= 100, 130, and 150 GeV/c^2 were approximated via linear fits from the three M_H values for which the effects were measured (Figure 7.9). Table 7.10 contains a summary of the systematic effects as a function of M_H . The numbers of expected events passing all selection for signal and background are listed in Tables 7.11- 7.15.

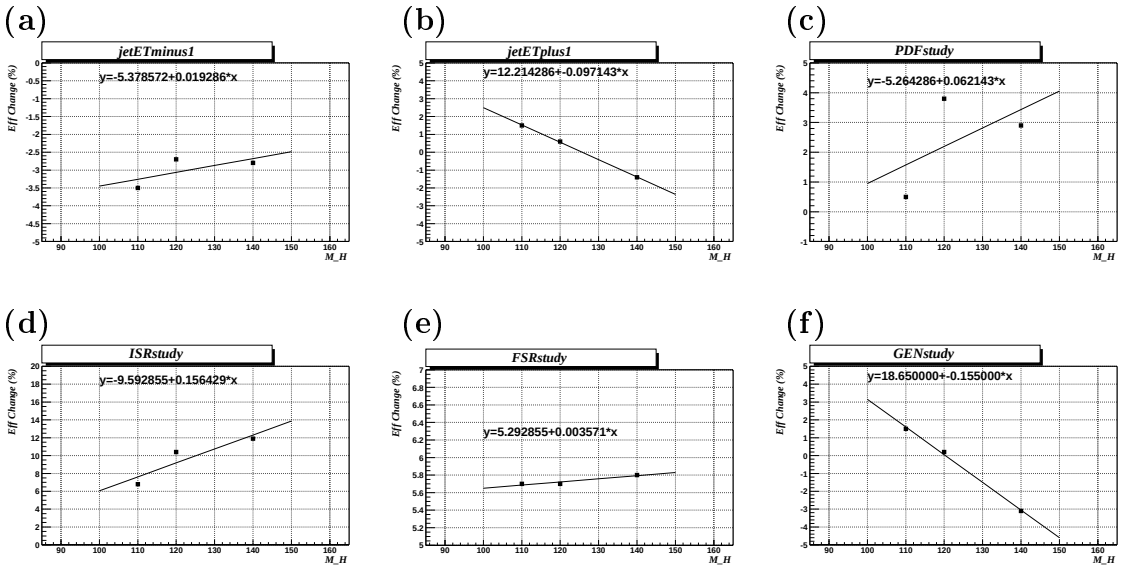


Figure 7.9: Effect of absence of initial state and initial state+final state radiation on WH, $t\bar{t}$, and $W + b\bar{b}$ samples on the number of E_T>15 GeV, $\mid \eta \mid$<2.0 jets.

Effect	WH Efficiency Change (%) M_H=100 GeV/c^2	110	120	130	140	150
jet corrections -1σ	-3.5	-3.5	-2.7	-2.9	-2.8	-2.5
jet corrections +1σ	2.5	1.5	0.6	-0.4	-1.4	-2.4
ISR	6.1	6.8	10.4	10.7	11.9	13.9
FSR	5.7	5.7	5.7	5.8	5.8	5.8
PDF	1.0	0.5	3.8	2.8	2.9	4.1
Generator	3.2	1.5	0.2	1.5	3.1	4.6
m_t	0.0	0.0	0.0	0.0	0.0	0.0
LepID+Trigger	7.7	7.7	7.7	7.7	7.7	7.7
Tagging	14.0	14.0	14.0	14.0	14.0	14.0
Luminosity	3.8	3.8	3.8	3.8	3.8	3.8
Cumulative	19.1	19.1	20.7	20.9	21.7	23.3

Table 7.10: A summary of the systematic effects influencing this analysis for signal WH at a variety of M_H values. The impact of the systematic effects were measured directly in the M_H=110,120, and 140 GeV/c^2 samples; the effects were then fit to get the impact for M_H=100,130,150 GeV/c^2.

	Class 1	Class 2	Class 3	Class 4	Class 5
Type of Events	WH (M_H= 100 GeV/c^2)	$t\bar{t}$	$W+b\bar{b}$ $W+c\bar{c}$ $W+c$ $W^{\pm}$+mistags non-W	Single top (W -*gluon* and $W*$)	WZ, WW,ZZ, $Z+b\bar{b}$, $Z+c\bar{c}$, $Z\rightarrow\tau\tau$
Luminosity	105.1	105.1	105.1	105.1	105.1
Cross Section(pb)	0.26	6.5		1.7, 0.73	
Basic Efficiency	1.10E-2	8.00E-3			
Basic N_{exp}	0.32	5.4	22.7	2.5	2.7
Advanced Efficiency	0.796	0.260	0.386	0.349	0.567
Final N_{exp}	0.25 $\pm$ 0.05	1.4 $\pm$ 0.4	8.8 $\pm$ 2.5	0.9 $\pm$ 0.2	1.5 $\pm$ 0.4

Table 7.11: Run 1 event expectations with systematic uncertainties on signal WH and its backgrounds for hypothesis M_H= 100 GeV/c^2.

	Class 1	Class 2	Class 3	Class 4	Class 5
Type of Events	WH (M_H= 110 GeV/c^2)	$t\bar{t}$	$W + b\bar{b}$ $W + c\bar{c}$ $W + c$ $W^{\pm}$+mistags non-W	Single top (W -*gluon* and $W*$)	WZ, WW,ZZ, $Z + b\bar{b}$, $Z + c\bar{c}$, $Z \to \tau\tau$
Luminosity	105.1	105.1	105.1	105.1	105.1
Cross Section(pb)	0.18	6.5		1.7, 0.73	
Basic Efficiency	1.20E-2	8.00E-3			
Basic N_{exp}	0.24	5.4	22.7	2.5	2.7
Advanced Efficiency	0.813	0.268	0.367	0.364	0.529
Final N_{exp}	0.20 ± 0.04	1.4 ± 0.4	8.3 ± 2.3	0.9 ± 0.2	1.4 ± 0.4

Table 7.12: Run 1 event expectations with systematic uncertainties on signal WH and its backgrounds for hypothesis M_H= 110 GeV/c^2.

	Class 1	Class 2	Class 3	Class 4	Class 5
Type of Events	WH (M_H= 130 GeV/c^2)	$t\bar{t}$	$W + b\bar{b}$ $W + c\bar{c}$ $W + c$ $W^{\pm}$+mistags non-W	Single top (W -*gluon* and $W*$)	WZ, WW,ZZ, $Z + b\bar{b}$, $Z + c\bar{c}$, $Z \to \tau\tau$
Luminosity	105.1	105.1	105.1	105.1	105.1
Cross Section(pb)	0.10	6.5		1.7, 0.73	
Basic Efficiency	1.10E-2	8.00E-3			
Basic N_{exp}	0.12	5.4	22.7	2.5	2.7
Advanced Efficiency	0.730	0.233	0.267	0.340	0.339
Final N_{exp}	0.09 ± 0.02	1.3 ± 0.4	6.1 ± 1.7	0.9 ± 0.2	0.9 ± 0.2

Table 7.13: Run 1 event expectations with systematic uncertainties on signal WH and its backgrounds for hypothesis M_H= 130 GeV/c^2.

	Class 1	Class 2	Class 3	Class 4	Class 5
Type of Events	WH (M_H= 140 GeV/c^2)	$t\bar{t}$	$W+b\bar{b}$ $W+c\bar{c}$ $W+c$ $W^{\pm}$+mistags non-W	Single top (W -*gluon* and $W*$)	WZ, WW,ZZ, $Z+b\bar{b}$, $Z+c\bar{c}$, $Z\rightarrow\tau\tau$
Luminosity	105.1	105.1	105.1	105.1	105.1
Cross Section(pb)	0.08	6.5		1.7, 0.73	
Basic Efficiency	8.0E-3	8.00E-3			
Basic N_{exp}	0.07	5.4	22.7	2.5	2.7
Advanced Efficiency	0.654	0.215	0.203	0.303	0.220
Final N_{exp}	0.04 $\pm$ 0.01	1.2 $\pm$ 0.4	4.6 $\pm$ 1.3	0.8 $\pm$ 0.2	0.6 $\pm$ 0.2

Table 7.14: Run 1 event expectations with systematic uncertainties on signal WH and its backgrounds for hypothesis M_H= 140 GeV/c^2.

	Class 1	Class 2	Class 3	Class 4	Class 5
Type of Events	WH (M_H= 150 GeV/c^2)	$t\bar{t}$	$W+b\bar{b}$ $W+c\bar{c}$ $W+c$ $W^{\pm}$+mistags non-W	Single top (W -*gluon* and $W*$)	WZ, WW,ZZ, $Z+b\bar{b}$, $Z+c\bar{c}$, $Z\rightarrow\tau\tau$
Luminosity	105.1	105.1	105.1	105.1	105.1
Cross Section(pb)	0.06	6.5		1.7, 0.73	
Basic Efficiency	5.0E-3	8.00E-3			
Basic N_{exp}	0.03	5.4	22.7	2.5	2.7
Advanced Efficiency	0.526	0.182	0.173	0.263	0.147
Final N_{exp}	0.02 $\pm$ 0.005	1.0 $\pm$ 0.3	3.9 $\pm$ 1.1	0.7 $\pm$ 0.2	0.4 $\pm$ 0.1

Table 7.15: Run 1 event expectations with systematic uncertainties on signal WH and its backgrounds for hypothesis M_H= 150 GeV/c^2.

	σ_{WH} Limit (pb)					
N_{Obs}	M_H=100 GeV/c^2	110	120	130	140	150
0	3.3	3.0	3.2	3.8	6.0	11.7
1	3.6	3.2	3.5	4.2	6.8	13.5
2	3.9	3.6	3.9	4.7	7.7	15.5
3	4.3	3.9	4.3	5.2	8.8	18.0
4	4.7	4.3	4.8	5.8	10.0	20.8
5	5.2	4.7	5.3	6.5	11.4	24.0
6	5.7	5.2	5.8	7.3	13.0	27.7
7	6.3	5.7	6.5	8.2	14.7	31.5
8	6.9	6.3	7.2	9.2	16.6	35.8
9	7.6	7.0	8.0	10.3	18.8	40.5
10	8.4	7.7	8.9	11.5	21.0	45.3
11	9.2	8.4	9.8	12.7	23.4	50.4
12	10.1	9.2	10.8	14.1	25.8	55.2
13	11.0	10.1	11.9	15.4	28.2	60.2
14	12.1	11.1	13.1	16.9	30.9	65.1
15	13.2	12.1	14.2	18.4	33.3	70.1
16	14.3	13.1	15.3	19.8	35.6	74.4
17	15.4	14.2	16.6	21.4	38.4	78.4
18	16.6	15.2	17.8	22.9	40.9	81.9
19	17.8	16.3	19.2	24.5	43.4	85.1
20	19.0	17.5	20.4	26.0	46.1	87.5
21	20.4	18.7	21.8	27.6	48.6	90.0
22	21.7	19.9	23.1	29.2	51.3	91.9
23	22.9	21.0	24.4	30.7	53.7	93.2
24	24.2	22.1	25.9	32.5	56.7	94.3
25	25.4	23.2	27.0	33.9	58.9	95.2
26	27.0	24.6	28.3	35.4	61.3	96.0
27	28.1	25.7	29.6	37.0	63.9	96.5
28	29.5	26.9	30.9	38.5	66.3	97.0
29	30.9	28.1	32.2	40.0	68.8	97.4
30	32.1	29.2	33.4	41.4	70.9	97.7
31	33.5	30.5	35.0	43.3	73.9	98.0
32	34.8	31.6	36.2	44.8	75.9	98.2
33	36.1	32.8	37.5	46.4	78.2	98.4

Table 7.16: The 95% CL limit on σ_{WH} for values of N_{Obs} including the effect of systematic errors for six values of M_H.

M_H	*A priori* Limit (pb)	
	No Systematics	With Systematics
100	9.5	11.3
110	8.4	9.9
120	8.4	10.0
130	9.3	10.9
140	13.5	15.6
150	24.3	28.5

Table 7.17: *A priori* limit as a function of M_H before and after the inclusion of systematic effects.

The 95% CL limit as a function of observation scenarios is listed in Table 7.16 for each of the M_H hypotheses being considered in this analysis. Table 7.17 contains a comparison of the *a priori* limit as a function of M_H both before and after the inclusion of systematic effects. We see that the *a priori* limit is in the range 10-11 pb for $M_H < 130\ GeV/c^2$; however the limit increases significantly for $M_H > 130$ GeV/c^2. This is due to the small branching ratio of $H \rightarrow b\bar{b}$ for larger M_H values. As M_H grows the number of $\ell\nu b\bar{b}$ final state WH events decreases; thus our signal expectation gets smaller and smaller, making it more and more difficult to establish a limit at the same level as in the scenarios $M_H < 130\ GeV/c^2$.

CHAPTER 8

RESULTS

Equipped with the expected number of signal and background events, the limit on WH production cross section for each possible outcome of the analysis (in number of observed events, N_{Obs}) was determined. We can now apply the method to the actual Run 1 data sample and see what the actual outcome will be.

8.1 Results from the CDF Run 1 Data Sample

The Run 1 data sample contains 1057 events passing the minimal $W^{\pm}$ selection and 2 jet requirement. Of these 1057, only 42 events possess the required one or more b-tagged jets. The Run and Event Number are listed for these 42 events are Table 8.1.

The 42 events listed in Table 8.1 that pass our baseline event selection were first sent through the 5-output NN. The output distributions for data and the four background classes from Monte Carlo are shown in Figures 8.1- 8.2. One of the motivations for using a multiple output NN is that by having an output node for each class of events, then the values of each output node represent event-by-event Bayesian *a posteriori* class probabilities. Thus one should verify that, in the data, the output values behave like probabilities.

	Run	Event	Trigger	Number of SECVTX tags	Pass NN	Pass $M_{j_1 j_2}$ (for $M_H = 120$)
Run 1a	40190	98182	e	1	Y	N
	41627	87219	e	1	Y	Y
	45776	386857	e	1	N	N
	46357	511399	μ	1	Y	Y
	46935	266805	e	1	Y	Y
	47439	128290	e	1	Y	Y
	47689	80060	μ	2	N	N
Run 1b	60705	93795	μ	1	Y	Y
	60766	299452	e	1	Y	N
	61167	368226	e	2	N	N
	61377	114526	μ	1	N	N
	63603	4029	μ	1	Y	Y
	63883	935	e	1	Y	Y
	63946	43019	μ	1	N	N
	64126	52063	e	1	Y	Y
	64916	499208	e	2	Y	N
	64997	46557	e	1	Y	Y
	65022	34157	e	1	N	N
	65298	907072	e	1	Y	Y
	65384	266051	e	1	N	N
	65470	4390	e	1	N	N
	65741	654870	μ	2	Y	Y
	65750	106257	μ	1	Y	Y
	66103	190513	e	1	Y	N
	66103	563542	e	1	Y	Y
	66103	743101	μ	1	Y	N
	66412	121506	e	1	Y	Y
	66518	203555	μ	1	N	N
	67692	420568	e	1	Y	Y
	68044	53510	e	1	Y	N
	68231	157759	μ	1	Y	N
	68374	364586	e	1	Y	N
	68423	3326	μ	1	N	N
	68464	275644	e	1	Y	N
	68593	37659	μ	1	Y	Y
	68637	225974	μ	1	Y	Y
	68774	150313	μ	1	Y	Y
	69498	36574	μ	1	Y	Y
	69520	136405	e	1	N	N
	69683	21986	e	1	Y	Y
	69709	173294	e	1	Y	Y
	69761	157205	μ	2	Y	N

Table 8.1: List of 42 Run 1 events that pass the baseline event selection. Also listed for each event is whether the event passed the neural network and $M_H = 120$ $M_{j_1 j_2}$ selection.

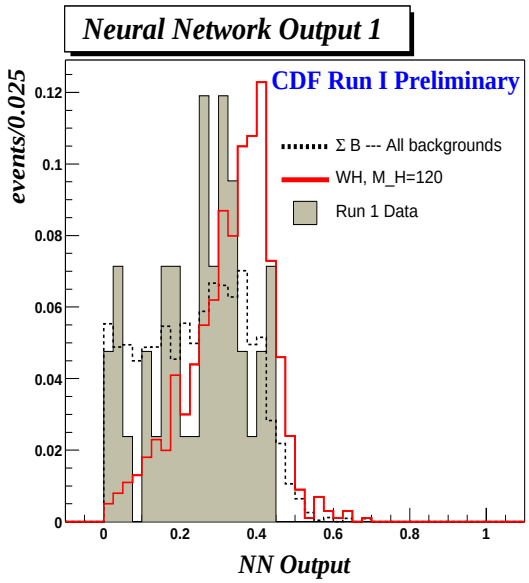


Figure 8.1: Data distribution for values from NN output node 1, the node associated with signal. Overlayed are the expected shapes from signal and summed background. All distributions are normalized to equal area.

Figure 8.3 shows the sum of the output values for each data event; we see that the data in general exhibit the property

$$\sum_{i=1}^{5} NNOut_i(data) = 1$$

as one would expect if the Bayesian *a posteriori* class probabilities argument is valid. Recall from Section 6.1 that the NN selection was not imposed directly in the NN output value for any of the output nodes; instead, we cut in the distance, d, in the 5D output space for each event from the target output vector associated with signal ($d < 0.95$ was shown to be optimal). Figure 8.4 contains the distribution of d for the 42 $W^{\pm}$+2jet events. The cut $d < 0.95$ is 74% efficient on our data sample, and thus eliminates 11 events.

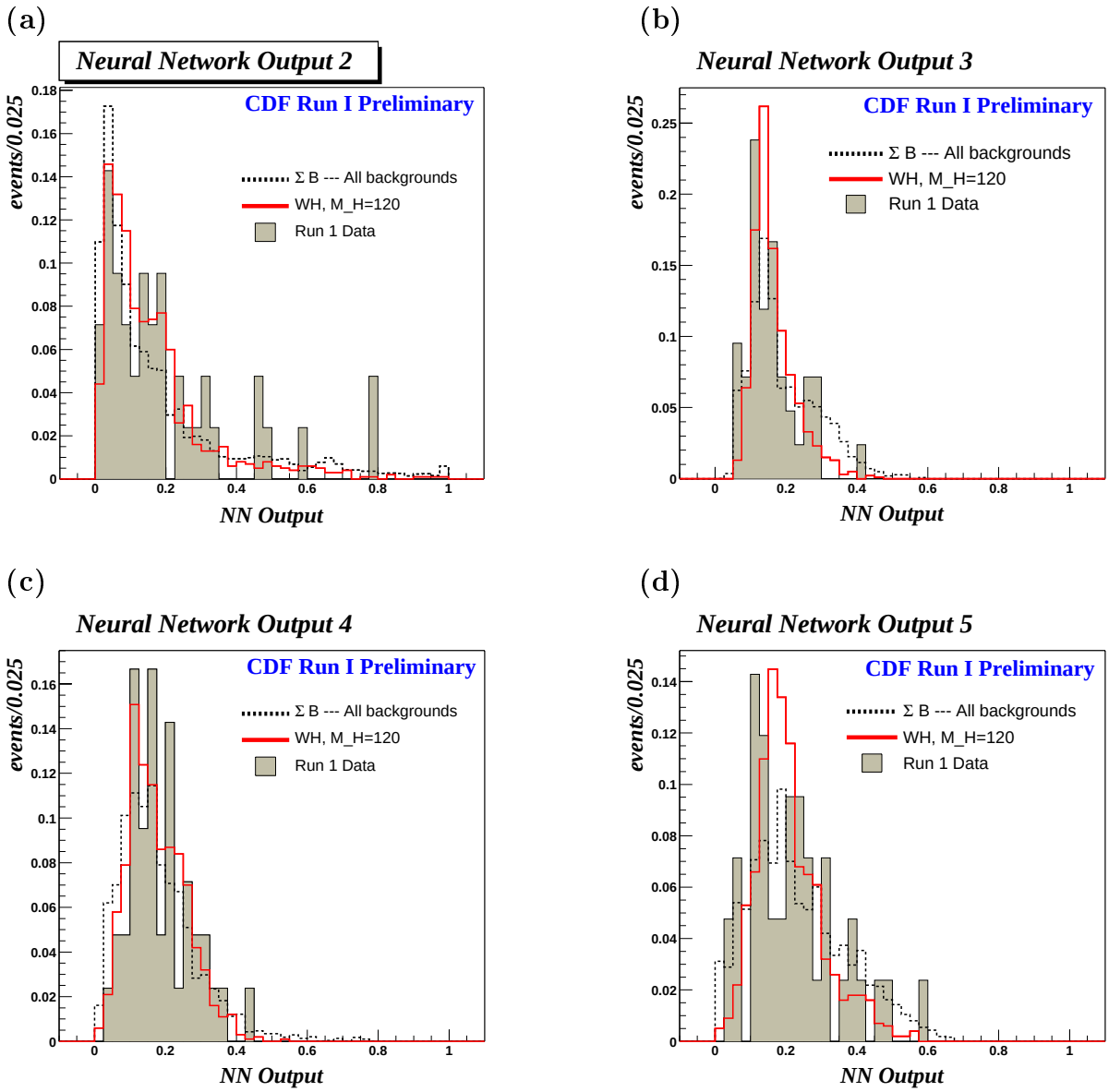


Figure 8.2: Data distribution for values from NN output nodes 2–5, the nodes associated with background. Overlayed are the expected shapes from signal and summed background. All distributions are normalized to equal area.

The advanced selection is completed by the M_H-dependent cut in the reconstructed di-jet mass, $M_{j_1 j_2}$. Care was taken in the construction of the NN to avoid

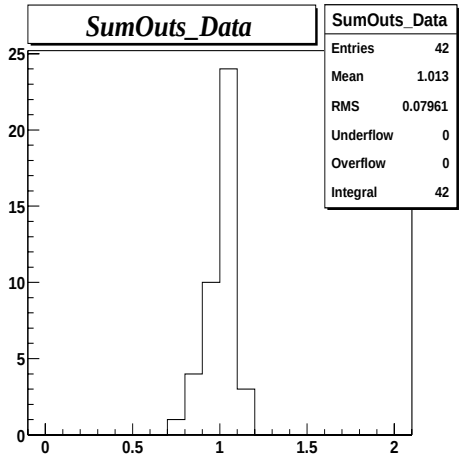


Figure 8.3: Sum of the values returned by the 5 NN output nodes for 41 data events. It is encouraging that the sum is centered about one, lending credence to the outputs-as-probabilities argument.

choosing input variables that would sculpt the $M_{j_1 j_2}$ distribution of the selected sample. Recall that the $M_{j_1 j_2}$ cut varies for different values of M_H (as listed in Table 7.9). Figure 8.5 shows the data $M_{j_1 j_2}$ distribution (overlayed with the background expectation) for events prior to the NN cut (a), for those events passing the NN cut (b), and for events failing the NN selection (c).

The results of our counting experiment are listed in Table 8.2. We see, for example, that for the hypothesis M_H=120 GeV/c^2, we expect 10.6 $\pm$ 2.1 events from background, and we see 21, a $\sim$3 σ excess considering the statistical uncertainty on the expected 10.6 events. A 2–3 σ excess is evident for each mass; this is to be expected since there is significant overlap between the $M_{j_1 j_2}$ window for each M_H hypothesis. Given these observation outcomes, we can use the values in Table 7.16 to determine the 95% CL limit on the WH production cross section as a function of M_H, which is contained in Table 8.3. It should be noted that several other Run 1 analyses investigating the same final state also saw such an excess of events over

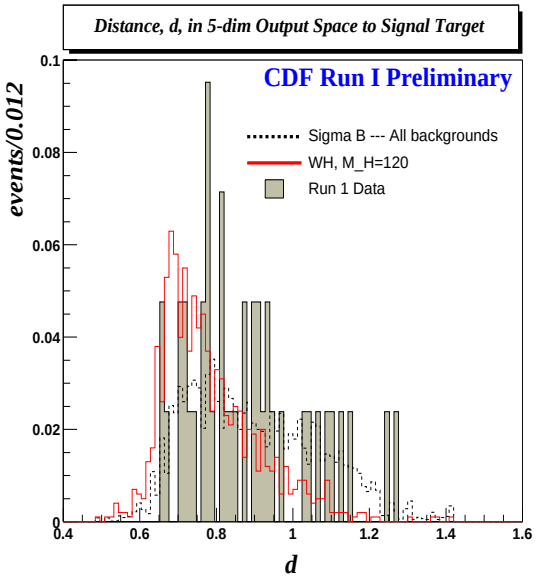


Figure 8.4: Data distribution for d, the parameter in which the NN selection is imposed. The cut $d < 0.95$ is 73% efficient in the 41 event Run 1 data sample. Overlayed are the distributions for signal and summed background, normalized to equal area.

the background prediction; a comparison to two of these analyses is discussed in Appendix A.

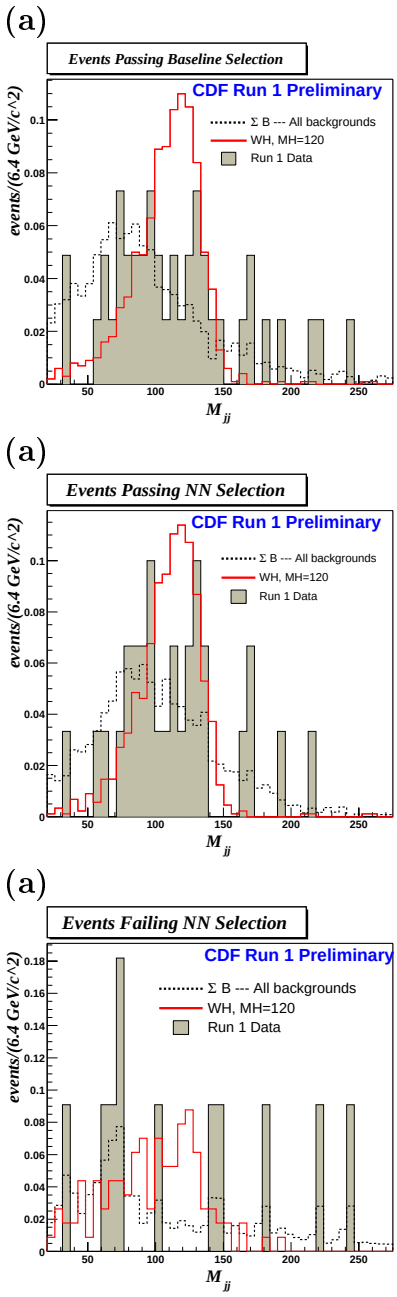


Figure 8.5: $M_{j_1 j_2}$ distribution for 42 data events before NN cut, the surviving 31 events after the NN cut, and the $M_{j_1 j_2}$ for the 11 events that do not satisfy our NN selection.

CDF Run I Preliminary

	Signal	Background					Data	
M_H (GeV/c^2)	(theoretical)	$t\bar{t}$	$W+b\bar{b}$, $W+c\bar{c}$, $W+c$,non-W, $W^{\pm}$+mistags	Single top (W -*gluon* and $W*$)	WZ,WW,ZZ $Z+b\bar{b}$, $Z+c\bar{c}$, $Z\rightarrow\tau\tau$	Bkgd Total	Events	Prob
100	0.25 ± 0.05	1.4 ± 0.4	8.8 ± 2.5	0.9 ± 0.2	1.5 ± 0.4	12.6 ± 2.6	19	0.0899
110	0.20 ± 0.04	1.4 ± 0.4	8.3 ± 2.3	0.9 ± 0.2	1.4 ± 0.4	12.0 ± 2.4	22	0.0183
120	0.15 ± 0.03	1.3 ± 0.4	7.2 ± 2.0	0.9 ± 0.2	1.2 ± 0.3	10.6 ± 2.1	21	0.0092
130	0.09 ± 0.02	1.3 ± 0.4	6.1 ± 1.7	0.9 ± 0.2	0.9 ± 0.2	9.2 ± 1.8	17	0.0255
140	0.04 ± 0.01	1.2 ± 0.4	4.6 ± 1.3	0.8 ± 0.2	0.6 ± 0.2	7.2 ± 1.4	15	0.0141
150	0.02 ± 0.005	1.0 ± 0.3	3.9 ± 1.1	0.7 ± 0.2	0.4 ± 0.1	6.0 ± 1.2	12	0.0306

Table 8.2: Counting experiment results for M_H values considered in this analysis.

CDF Run I Preliminary

	Cross Section 95% CL Limits (pb)			
	A priori Limit		Data Results	
M_H (GeV/c^2)	No Systematics	With Systematics Effects	σ_{WH}	$\sigma_{VH}*$ BR($H \to b\bar{b}$)
100	9.5	11.3	17.8	22.5
110	8.4	9.9	19.9	24.1
120	8.4	10.0	21.8	23.4
130	9.3	10.9	21.4	18.0
140	13.5	15.6	33.3	18.5
150	24.3	28.5	55.2	17.9

Table 8.3: The 95% CL limit on the WH production cross section as a function of M_H with *a priori* results for comparison.

In previous CDF Run 1 $WH \to \ell\nu b\bar{b}$ analyses, the result was quoted not in terms of a limit on the WH cross section, but as a limit on the product $\sigma_{VH}*$ BR($H \to b\bar{b}$), where $V = W$ or Z. This facilitated the combination of results with searches in the ZH channels. We can translate our WH result into a result on $\sigma_{VH}*$ BR($H \to b\bar{b}$); the limit values are listed in Table 8.3, and this new result is plotted in Figure 8.6 with a comparison to the former CDF Run 1 $\ell\nu b\bar{b}$ search. Both analyses saw an excess of events, and therefore both analyses arrive at similar results.

8.2 Higgs Prospects in Run 2

Recall that this search was performed in the Run 1 data sample collected by the CDF collaboration operating at Fermilab. Run 2 at the Tevatron, already underway, exploits a p-$\overline{p}$ beam that is collided at slightly higher energy than that of Run 1. In addition, higher luminosity and improved detectors offer promising opportunities to search for new phenomena over the next several years.

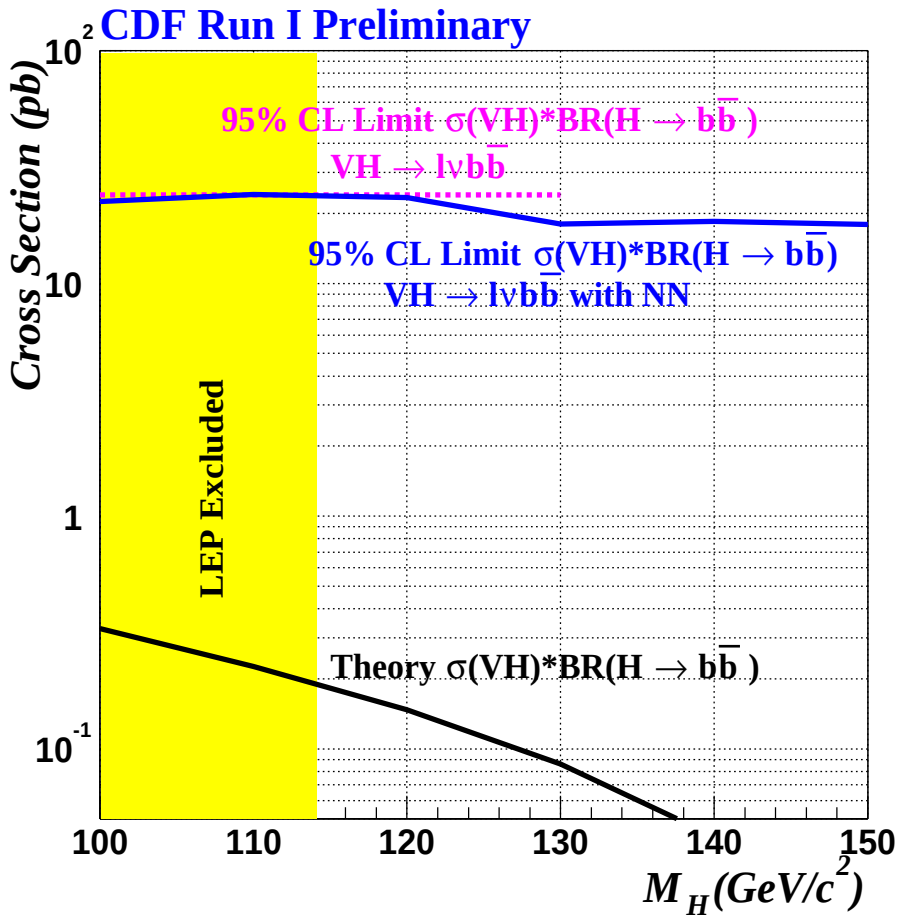


Figure 8.6: The 95% CL limit on the σ_{VH} times $BR(H \to b\bar{b})$ as a function of M_H.

In this analysis, it has been demonstrated that the employment of the advanced selection in this analysis increases our signal sensitivity. For example in the case of $M_H = 120\ GeV/c^2$, we see a 34% increase in sensitivity when comparing the expected signal and background yields before ($S_0/\sqrt{B_0} = 1/30$) and after ($S_{final}/\sqrt{B_{final}} = 1/22$) advanced selection. Because of its ability to reduce background levels, a neural network will be exploited in Run 2 Higgs searches.

One can ask how much luminosity is needed in this analysis to identify WH events with statistical significance. By "statistical significance" it is meant that the signal

must be some factor larger than the statistical uncertainty in the background, thus strengthening one's confidence that any signal-like excess that is seen is not likely due to a background fluctuation. The uncertainty on the background is taken to be $\sqrt{B}$. For "observation" of a signal, the significance is usually defined as $S/\sqrt{B} = 3$, known as 3-σ observation condition. The luminosity requirement for 3-σ observation for this analysis is described below.

We seek:

$$\frac{S}{\sqrt{B}} = 3 \tag{8.1}$$

We can scale our final signal and background numbers by some value f:

$$\begin{aligned} \frac{S}{\sqrt{B}} &= \frac{f \times S_{final}}{\sqrt{f \times B_{final}}} \\ &= \sqrt{f} \times \frac{S_{final}}{\sqrt{B_{final}}} \end{aligned} \tag{8.2}$$

In this analysis, $\frac{S_{final}}{\sqrt{B_{final}}} = 1/22$ and with this value one finds that $f = 4356$. This means that one would need >4000 times as much Run 1 data to make a 3-σ WH observation in this channel with this method.

Run 2 Tevatron WH searches will contend with similar signal and background levels. The daunting luminosity requirement will be ameliorated by exploiting a more appropriate baseline event selection that reduces background levels. For example, contrary to what is done in this analysis, one can require that both of the 2 jets in the selected events be tagged. By making this requirement, the signal efficiency is of course reduced, but background contributions decline dramatically. For these reasons, Run 2 WH searches will exploit this double-tag requirement. The double tag requirement was not utilized here because signal statistics were predicted to be extremely small to begin with, and so the baseline event selection was designed to retain as much signal as possible at the expense of also increasing background levels.

However this issue is less critical in Run 2, where it is expected that each collider experiment will accumulate a dataset at least 20 times larger than that of Run 1.

Run 2 WH searches will also benefit from extended lepton identification and b tagging. Lepton identification in Run 2 will extend into the plug and forward region of the detector, unlike Run 1 in which only central leptons were considered. Likewise, in Run 2 CDF will have the ability to b tag jets out to larger $\mid \eta \mid$ due to a new silicon tracker that is nearly twice as long as the Run 1 version. This and other forms of extended coverage will help boost the overall signal yield; it is hoped that tools like neural networks will be able to help reduce the backgrounds that are also introduced in these regions. A recent study [55] indicates that with 5-10 fb^{-1} of Run 2 data, a Higgs discovery is possible up to $M_H = 120\ GeV/c^2$ and exclusion at the 95% CL is possible up to $M_H = 135\ GeV/c^2$.

CHAPTER 9

CONCLUSIONS

Despite the phenomenal success of the Standard Model, the mechanism in the framework for the origin of mass in the universe has no experimental evidence. The Higgs boson is a yet-undiscovered particle whose existence is linked to the dynamics of the imposition of mass to the spectrum of fundamental particles within the Model. This new form of matter has yet to be discovered, and elucidation of the Higgs sector is one of the prominent goals in contemporary experimental particle physics.

A new search has been performed that looks for the Higgs signature in the remnants of proton-antiproton collisions provided by the Tevatron accelerator at Fermi National Accelerator Laboratory. This analysis utilizes data collected by the Collider Detector at Fermilab (CDF) experiment during Run 1 of the Tevatron (1992-1995). The search was designed to focus on a production mechanism in which the Higgs is produced in association with a leptonically decaying $W^{\pm}$ boson. A neural network was used to select events consistent with Higgs production, and was shown in simulation studies to provide significant increase in signal sensitivity over conventional methods.

In the Run 1 CDF dataset, the technique identified an excess of selected events over the predicted background yield. This excess is consistent with previous CDF Run 1 searches in similar channels; the nature of the excess needs to be identified, and will

be the focus of much scrutiny in Run 2 at CDF, which is underway. An upper bound on the WH production cross section is measured for Higgs mass hypotheses in the range 100 $GeV/c^2 < M_H < 150$ GeV/c^2. The upper bound on the WH cross section is a factor of two larger than the *a priori* result in simulation studies. The new measured upper bound is 100 times larger than the prediction from theory; thus no portion of the Higgs mass window considered could be excluded. This limit is however a slight improvement on the upper limit reported by previous CDF WH searches.

APPENDIX A

CONNECTION TO OTHER ANALYSES

It is of interest to see how the advanced selection employed in this analysis treats events from other lepton+jet analyses that found similar excesses in the $W^{\pm}$+2 jet bin. One such analysis is the latest $t\bar{t}$ cross section measurement [38], which saw an excess in both the single- and double-SECVTX-tagged Run 1 samples (herein we denote "ST" as meaning exactly one SECVTX tag and "DT" meaning exactly two SECVTX tags). Table A.1 compares the background expectation and result from the WH search described above. In the $W^{\pm}$+2 jet bin, the $t\bar{t}$ cross section analysis sees 35 single tag events with an expectation of 29.6 $\pm$ 2.7 and 5 DT events with an expectation of 2.4 $\pm$ 0.6. We see a similar excess in the ST bin, but limited statistics make it difficult to say much of anything about the DT bin.

Another analysis that saw an excess in the lepton+jets final state is the superjets analysis [61], which studied the soft-lepton tag rate in the $W^{\pm}$+2,3 jet bin. A superjet is defined as a jet which possesses both a SECVTX and SLT tag. The superjets analysis found an excess of events containing a superjet, identifying 13 such events despite a SM prediction of 4.4 $\pm$0.6. Eight of these 13 superjet events are in the $W^{\pm}$+2 jet bin that is included in this analysis. Since we employ here an advanced

M_H (GeV/c^2)	N^B_{exp}(Total)	Data	N^B_{exp}(ST)	Data	N^B_{exp}(DT)	Data
100	12.6 ± 2.6	19	11.7 ± 2.4	17	0.92 ± 0.19	2
110	12.0 ± 2.4	22	11.1 ± 2.2	20	0.88 ± 0.18	2
120	10.6 ± 2.1	21	9.8 ± 1.9	20	0.79 ± 0.16	1
130	9.2 ± 1.8	17	8.5 ± 1.7	16	0.71 ± 0.14	1
140	7.2 ± 1.4	15	6.6 ± 1.3	14	0.58 ± 0.11	1
150	6.0 ± 1.2	12	5.5 ± 1.1	11	0.48 ± 0.10	1

Table A.1: Breakdown of the Run 1 events passing our advanced selection into the 1 tag (ST) and 2 tag (DT) exclusive bins.

selection designed to be sensitive to $WH \rightarrow \ell\nu b\overline{b}$, it is interesting to see if this analysis considers these events to be consistent with WH signal or background.

Table A.2 contains a description of the fate of the superjet events when confronted with our advanced selection. We see that one of the superjet events fails our dilepton veto. The dilepton veto is imposed via two criteria:

- Standard dilepton veto: an event is vetoed if it possesses 2 oppositely charged high p_T leptons, $\not\!\!E_T > 25$ and two $E_T > 20$ jets
- Extended dilepton veto: an event is vetoed if it possesses a primary lepton and a $p_T > 15$, isolated oppositely charged track

In [61] a similar – though less stringent – extended dilepton veto is imposed for the superjets analysis. There they demand only that the isolated track have $p_T > 10\ GeV$. The extended dilepton veto employed here is the standard implementation from the Run 1 top group [62]; other similar analyses vetoed this event [54], [62].

Beyond this difference, one sees that the NN selection employed here cannot supply a clear verdict on these superjet events. In the seven such events that pass our baseline criteria, four are retained by the NN selection while three are rejected. The di-jet

mass cut differs for each M_H hypothesis considered, so we simply list each superjet event's $M_{j_1 j_2}$ value in Table A.2 for reference.

Run	Event	Pass Basic Selection	Pass NN Selection	$M_{j_1 j_2}$
46935	266805	Y	Y	100.3
41540	127085	N	—	—
41627	87219	Y	Y	98.1
46357	511399	Y	Y	97.8
61167	338226	Y	N	183.8
65384	266051	Y	N	142.0
69520	136405	Y	N	75.6
65741	654870	Y	Y	129.7

Table A.2: The eight Run 1 $W^{\pm}$+2 jet superjet events, and how they fare in the selection employed in this analysis. One event fails the baseline selection (the dilepton veto) in agreement with [62].

Another Run 1 lepton+jets analysis that identified an excess in the $W^{\pm}$+2 jet bin is the latest Run 1 single-top search [54]. In this analysis, 41 $W^{\pm}$+2 jet events were identified with an expectation of ~ 30. Because of equivalent selection, the $W^{\pm}$+2 jet data sample constructed for this single-top analysis should exactly match the $W^{\pm}$+2 jet sample in this analysis, prior to any advanced selection. Yet there is a one event discrepancy between these two analyses, the origin of which is to be determined. It should be noted however that both the superjets analysis and the previous Run 1 single top search each exploit similar baseline event selections and each arrive at a different Run 1 $W^{\pm}$+2 jet data sample.

In [54] a 3-output NN was used to distinguish single top ($W*$ and W -*gluon* combined) from the dominant backgrounds, $t\bar{t}$ and $W^{\pm}$+QCD. A subset of the $W^{\pm}$+2

jet events was found to be more consistent with single top than background; the events in which the single-top probability was $> 50\%$, along with their signal probabilities from the NN single-top search are listed in Table A.3, along with their NN output values from this analysis. Recall that here we employ a 5-output NN; the value for each output node can be considered the probability that each event belongs to signal (Output Node 1) or the four background classes (Nodes 2-5). From Table A.3 we see that only a few candidate signal events from NN single-top search are also determined to be most likely from single-top here in this analysis. It should be noted that in the NN single top analysis, WH events were not considered to be a significant portion of the background; thus one would not expect the NN used in that analysis to be able to discriminate between WH and single top, two very similar final states.

There are several Run 1 lepton+jets analyses that identified an excess of events. One can argue that these events are single top, scalar quarks, WH production or plain, old $W^{\pm}$+QCD background. It is evident that something is not well-understood in this final state, and such analyses deserve further investigation in Run 2.

Run	Event	CDF 5742 Single Top Prob	Five-class Probabilities					CDF 6468 Most Likely
			Signal	$t\bar{t}$	$W + b\bar{b}$, $W + c\bar{c}$, etc.	Single-top	Diboson, Z+X	
40190	98182	0.642	0.159	0.092	0.169	0.212	0.272	Diboson
41627	87219	0.667	0.291	0.173	0.121	0.356	0.123	Single Top
60766	299452	0.591	0.370	0.029	0.209	0.184	0.211	WH
61167	368226	0.563	0.064	0.472	0.056	0.252	0.110	$t\bar{t}$
61377	114526	0.534	0.127	0.026	0.296	0.155	0.421	Diboson
63883	935	0.582	0.312	0.315	0.102	0.116	0.123	$t\bar{t}$
64126	52063	0.628	0.305	0.144	0.114	0.203	0.230	WH
64997	46557	0.593	0.169	0.183	0.087	0.116	0.273	Diboson
65384	266051	0.614	0.195	0.470	0.103	0.213	0.073	$t\bar{t}$
65470	4390	0.544	0.022	0.056	0.424	0.210	0.302	$W + b\bar{b}$,$W + c\bar{c}$...
65750	106257	0.636	0.342	0.089	0.169	0.326	0.183	WH
66103	743101	0.799	0.259	0.225	0.086	0.430	0.118	Single Top
67692	420568	0.766	0.308	0.148	0.107	0.282	0.141	WH
68044	53510	0.746	0.409	0.051	0.135	0.275	0.225	WH
68423	3326	0.658	0.030	0.581	0.063	0.117	0.125	$t\bar{t}$
68593	37659	0.698	0.441	0.060	0.190	0.223	0.174	WH
68774	150313	0.614	0.350	0.115	0.146	0.318	0.153	WH
69683	21986	0.815	0.165	0.241	0.106	0.323	0.059	Single Top

Table A.3: Events from [54] that were determined to be consistent with single top, and how they are classified in this analysis.

APPENDIX B

THE POWER OF THE NEURAL NETWORK

We seek to quantify the power of the employment of a NN in this analysis. The *a priori* limit has been used throughout this document to illustrate the rationale behind each choice that was made in the advanced selection implementation. Here we will again use the *a priori* limit as our figure of merit.

We have chosen to employ 8 inputs in our NN. A fair comparison that gauges the power of the NN itself would compare such a NN implementation to an advanced selection in which rectangular cuts are placed in these same eight variables; after all, the NN is just carving out a slightly more complicated region of this 8D space than rectangular cuts would. Such a study was performed in 105.1 pb^{-1}:

$$\sigma_{WH}^{95,apriori} = 9.7pb \qquad \text{(8D rectangular cuts) for } M_H = 120\ GeV/c^2$$

We can then consider what Run 1 luminosity would have been required to establish an *a priori* limit of 8.4 *pb* using the 8D rectangular cuts method. Table B.1 contains the *a priori* limits for several hypothetical Run 1 luminosity numbers. We see that in order to have an *a priori* limit of 8.4 *pb* using the rectangular cuts, one would have needed to have 135 pb^{-1} of Run 1 data. So one can conclude that the incorporation

of the NN alone exploiting equivalent variables is the equivalent of having an extra 30 pb^{-1} of Run 1 data, an effective increase in luminosity of $\sim$30% .

Luminosity(pb^{-1})	WH *a priori* limit (pb)
90	10.7
105.1	9.7
120	9.0
135	8.4
150	7.8
165	7.4

Table B.1: *A priori* limit achievable with rectangular cuts analysis in the eight variables used as NN inputs. Note that the actual total integrated luminosity in Run 1 was 105.1 ± 4 pb^{-1}.

This effect is important for Run 2, where we seek to extract as much information on the Higgs as possible given limited luminosity. There is nothing in this Run 1 analysis that cannot be applied to a similar WH analysis in Run 2; background and signal rates will of course need to be re-measured and the details hammered out. But there is no reason to not expect similar gains in sensitivity (and therefore decreases in required luminosity for discovery) from the NN in Run 2. It is therefore worthwhile to exploit a NN in the Higgs Run 2 searches.

BIBLIOGRAPHY

[1] F. Abe *et al.* [CDF Collaboration], Phys. Rev. Lett. **73**, 225 (1994) [hep-ex/9405005].

[2] K. Kodama *et al.* [DONUT Collaboration], hep-ex/0012035.

[3] D. E. Groom *et al.*, Eur. Phys. J. **C15**, 1 (2000).

[4] P.W. Higgs, Phys. Rev. Lett. **12**, 132 (1964)

[5] Super-K experiment: http://neutrino.kek.jp/news/980604.N98SK/

[6] B. W. Lee, *et al.*, *Phys. Rev. Lett*, **38** (1977) 883

[7] K. Riesselmann, hep-ph/9711456.

[8] LEP collider: http://cern.web.cern.ch/CERN/Divisions/SL/lep2page.html

[9] ALEPH Detector: http://alephwww.cern.ch/HANDBOOK/

[10] DELPHI Detector: http://delphiwww.cern.ch/ offline/physics/delphi-detector.html

[11] L3 Detector: http://l3www.cern.ch/

[12] OPAL Detector: http://opal.web.cern.ch/Opal/tour/detector.html

[13] LEP Electroweak Working Group http://lepewwg.web.cern.ch/LEPEWWG/

[14] LEP Electroweak Working Group Winter 2003 Proceedings http://lepewwg.web.cern.ch/LEPEWWG/plots/winter2003/

[15] Talk, "Electroweak Precision Data Global Higgs Analysis," DESY Zeuthen, February 28, 2003.

[16] J. F. Gunion, H. E. Haber, G. L. Kane and S. Dawson, SCIPP-89/13.

[17] LEP Higgs Working Group, Nov. 3, 2000, Submitted to LEP Committee and CERN Research Board

[18] "Search for the Standard Model Higgs Boson at LEP", LEP Higgs Working Group CERN-EP/2003–011

[19] J. Erler, Phys. Rev. D **63**, 071301 (2001) [arXiv:hep-ph/0010153].

[20] See

`http://www.fnal.gov`

for a significant amount of information on the daily operation of the Tevatron.

[21] F. Abe *et al.*, Nucl. Inst. Meth. Phys. Res. Sect. A **271**, 387 (1988)

[22] M. Kruse, Ph.D. Thesis, Purdue University(1996).

[23] M. Carena *et al.*, hep-ph/0010338.

[24] M. Spira, hep-ph/9810289.

[25] Physics at Run II Workshop - Higgs Bosons http://fnth37.fnal.gov/higgs/higgs.html

[26] "Higgs Searches in Run 2 at the Tevatron" FERMILAB-CONF-99/156-E. CDF/PUB/EXOTIC/PUBLIC/5019

[27] J. Goldstein, C. S. Hill, J. Incandela, S. Parke, D. Rainwater and D. Stuart, hep-ph/0006311.

[28] F. Abe *et al.*, Nucl. Inst. Meth. Phys. Res. Sect. A **271**, 387 (1988)

[29] J. Lewis,*et al.*,"CLEOMC: The CDF Interface to the CLEO Monte Carlo(QQ)", CDF/ANAL/MONTECARLO/PUBLIC/2724, 1994.

[30] A. Caner,*et al.*,"CDFSIM+QFL Simulation of the CDF Detector", CDF/ANAL/MONTECARLO/PUBLIC/2177, 1993.

[31] K. Hagiwara *et al.* [Particle Data Group Collaboration], Phys. Rev. D **66**, 010001 (2002).

[32] CDF 6309, "Neural Network Implementation for a $WH \to \ell\nu b\bar{b}$ Search"

[33] CDF 3403, "$t\bar{t}$ Acceptance Calculation for the Run 1B SVX B-Tag Analysis"

[34] CDF 4112, "Search for New Particles $X \to b\bar{b}$ Produced in Association with 6 ..."

[35] CDF 4939, "Godparent Report on the new $t\bar{t}$ Cross Section..."

[36] T. Sjostrand, Computer Physics Commun. 82 (1994) 74.

[37] M. Gluck, E. Reya, and A. Vogt, Z. Phys. C **67** 433 (1995)

[38] T. Affolder *et al.* [CDF Collaboration], Phys. Rev. D **64**, 032002 (2001) [arXiv:hep-ex/0101036].

[39] See, for example, CDF 3461, "Observation of top dileptons with hadronically decaying tau leptons."

[40] CDF 2954, "COMBINED LEVEL 1 AND LEVEL 2 HIGH PT MUON TRIGGER EFFICIENCIES IN RUN"

[41] CDF 2883, "LEPTON IDENTIFICATION EFFICIENCIES FOR TOP DILEPTON ANALYSIS"

[42] Email correspondence, July 1, 2002.

[43] CDF 6414, "Limit Calculation and Incorporation of Systematic Errors in a Search for $WH \to \ell\nu b\overline{b}$ "

[44] CDF 6304, "Acceptance Calculation for $WH \to \ell\nu b\overline{b}$ Search"

[45] See, for example, C.M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon, Oxford, 1998.

[46] L. Lonnblad, C. Peterson, H. Pi, T. Rognvaldsson, Comput. Phys. Commun. 81, 185 (1994)

[47] W. McCulloch and W. Pitts, Bull. Math. Biophys. **5**, 115-133 (1943)

[48] CDF 5434, "A ROOT Interface to JETNET"

[49] M.D. Richard, R.P. Lippmann,"Neural Network Classifiers Estimate Bayesian a posteriori Probabilities", *Neural Comput.* **3**,461 (1991).

[50]

[50] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, arXiv:hep-ph/0206293.

[51] G. Marchesini *et al.*, Comput. Phys. Commun. **67**, 465 (1992).

[52] C. Peterson and T. Rognvaldsson, "An Introduction to Artificial Neural Networks", *Proc. 1991 CERN Summer School of Computing, CERN Yellow Report* **92-02**, 113-170 (1992)

[53] S. Wolinski and P. Savard, "An upper limit on Standard Model single top production in CDF Run 1 data using fitting techniques", CDF 5199

[54] C. Ciobanu, "Using A Neural Network to Identify Single Top in Run 1 Data", CDF 5742

[55] B. Winer, *et al.*, "Results of the CDF Higgs Sensitivity Study", CDF 6353

[56] C. Ciobanu, "A Neural Network Search for Single Top Quark Production in CDF Run I Data", Ohio State University Ph.D. Thesis, 2002.

[57] M. Carena *et al.*, arXiv:hep-ph/0010338.

[58] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998) [arXiv:physics/9711021].

[59] CDF 6414, "Systematic Errors in the Run 1 Neural Network $WH \to \ell\nu b\bar{b}$ Search"

[60] CDF 6309, "Neural Network Implementation for a $WH \to \ell\nu b\bar{b}$ Search"

[61] P. Giromini, *et al.*, "Evidence for a non-Standard Model signal...", CDF 4322

[62] S. Wolinski, P. Savard, "An Upper Limit on Standard-Model Single Top...", CDF 5199

[63] M. Gluck, E. Reya and A. Vogt, Z. Phys. C **67**, 433 (1995).

[64] H. L. Lai *et al.*, Phys. Rev. D **51**, 4763 (1995) [arXiv:hep-ph/9410404].

[65] Need citation for this PDF

[66] CDF 6468, "Results from a New Run 1 $WH \to \ell\nu b\bar{b}$ Search"