Eric Thomas Moore

Candidate

Department of Physics and Astronomy

Department

This dissertation is approved, and it is acceptable in quality and form for publication on microfilm:

Approved by the Dissertation Committee:

, Chairperson

Accepted:

Dean, Graduate School

Date

A Neural Network Measurement of the $t\bar{t}$ Pair Production Cross Section in the Lepton+jets Channel

by

Eric Thomas Moore

B.S., Physics, University of Texas at Dallas, 1993

M.S., Physics, University of Texas at Dallas, 1994

Dissertation

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Physics

The University of New Mexico
Albuquerque, New Mexico

May 2002

Dedication

To my wife, Judy, for her love, support, and sacrifice. And to my children for helping me to always maintain my perspective.

"After ecstasy, the laundry." – Zen saying

Acknowledgments

Working in an environment like that which exists at CDF is difficult and I would like to thank the people who made my work easier, or at times just made it tolerable.

I thank my adviser, Michael Gold, for his assistance, advise, and criticism; I would also like to thank Michael for the contributions which he made to the many papers that we co-authored. I would like to thank Steve Worm for his assistance in the navigation of CDF's *shark-filled waters*, and for finding locations which provided adequate sustenance in suburban Chicago.

Here in Albuquerque, I would like to thank Martin Hoeferkamp for his material support without whom I would not have had: the computational facilities needed to continue this work, the quiet space for my labors, or the almost constant supply of candy necessary to continue my efforts. In addition to Martin, my fellow graduate students, Nicki Bruner & Alex Brandl, have been encouraging, and have provided assistance in the formatting of this document.

At an experiment like CDF literally thousands of people over many years play a role in the work accomplished. I would like to acknowledge the contributions of the people (past and present) who have dedicated so much to the acquisition of the data which made this analysis possible.

And to my parents, I won't say they encouraged me along this chosen path, but I do believe they deserve some of the blame for my obstinate aggression toward wind-mills.

Finally on the page, but first in my heart, I would like to give special thanks to those whom have sacrifice the most to this work, my family: Judy, Tiffany, Caitlin, and Keenan.

A Neural Network Measurement of the $t\bar{t}$ Pair Production Cross Section in the Lepton+jets Channel

by

Eric Thomas Moore

Abstract of Dissertation

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Physics

The University of New Mexico
Albuquerque, New Mexico

May 2002

A Neural Network Measurement of the $t\bar{t}$ Pair Production Cross Section in the Lepton+jets Channel

by

Eric Thomas Moore
B.S., Physics, University of Texas at Dallas, 1993
M.S., Physics, University of Texas at Dallas, 1994
Ph.D., Physics, University of New Mexico, 2002

ABSTRACT

We have analyzed the 90 pb^{-1} of data from Run1B at *CDF* in order to extract the top/anti-top quark($t\bar{t}$) production cross section. This analysis uses a set of kinematic variables to discriminate the $t\bar{t}$ signal from the W-boson + jets($W+jets$) background. Unlike *CDF*'s previous top cross section measurements, tagging of b quarks is not required here. Using a Artificial Neural Network(*ANN*) the kinematic variables were combined into a single discrimination parameter. With this parameter we performed a binned maximum likelihood fit of our data sample to Monte Carlo generated background ($W+jets$) and signal ($t\bar{t}$) distributions. Taking the minimum of the negative log-likelihood value, we determine the most probable number of signal events in our sample, do a background subtraction and use this number ($N_{t\bar{t}}$) to extract the $t\bar{t}$ cross section, with a resulting total $t\bar{t}$ cross section of $\sigma_{t\bar{t}} = 5.1^{+1.5}_{-1.6}$ pb.

Table of Contents

List of Figures

List of Tables

Preface

The analysis presented in this thesis began as an extension of a topological measurement utilizing H_T to determine the $t\bar{t}$ cross section (H_T being the sum of the energy transverse to the beamline). This work has been summarized in Appendix C, and was presented at the Centennial meeting of the *American Physical Society* in 1999. During the investigation of this early analysis it soon became clear that there were many other kinematic variables which were potentially of interest in separating our primary background (W+$jets$) from $t\bar{t}$ signal.

Once several different kinematic variables had been incorporated into the analysis it became evident that a conventional analysis would require a prohibitive amount of work to be done in determining correlations of various kinematics. In order to handle this issue it was decided to use an *Artificial Neural Network* (*ANN*). The use of the *ANN* eliminated the necessity of examining the correlations between variables, as this was now done by the *ANN* itself.

The *ANN* combines all the kinematic values feed into it into a single output value providing the maximum possible separation between signal and background events. Although the *ANN* played a crucial role in this analysis my continuing suspicion of their veracity resulted in a very simple structure to the neural network and limited its application. Where other analyses have been very ambitious in applying *ANN*s to the separation of all the significant backgrounds to their respective signals we have limited the direct application of the *ANN* to the separation of the signal from our primary background (W+$jets$).

Once the network was trained the selected data was past through the trained network. The data output distribution was then compared to the output distributions for background and signal events. This comparison of data to Monte Carlos (MC) was done using a likelihood fit. The resulting fit yields a number of signal like events (N_S). Some of the events which where identified as signal like are actually from secondary backgrounds. This demanded the development of a new method for the removal of secondary backgrounds.

In order to determine the proportion of secondary background events which are identified as signal like it was necessary to utilize MC samples of each of the secondary backgrounds. The samples of secondary backgrounds where put through our *ANN* to get output distributions for each of these backgrounds. The outputs for these backgrounds are then fit against the same MC distributions which the data was fit against, this then yields the proportion of background events which are signal like. Along with an estimate for the total amount of secondary backgrounds in our data sample, we can use the proportion of signal like events to determine the amount of background contamination in the previously measured N_S. Subtracting the secondary background contamination from N_S yields the actual number of signal events ($N_{t\bar{t}}$).

With the $N_{t\bar{t}}$ the $t\bar{t}$ cross section ($\sigma_{t\bar{t}}$) can be calculated.

Chapter 1

Introduction

> *"My studies in Speculative philosophy, metaphysics, and science are all summed up in the image of a mouse called man running in and out of every hole in the Cosmos hunting for the Absolute Cheese."*
>
> *–Benjamin De Casseres*

1.1 The Standard Model

Of the four known fundamental forces of nature, all excepting gravity are described by the interaction of matter particles, fermions, through the exchange of force particles known as gauge bosons. The theory which describe these processes is called the standard model of particle physics. At the time of writing, no significant deviations from the standard model have been observed. The possible exception to this is oscillations in the neutrino sector and the non-zero mass of the neutrinos this oscillation implies.

Matter & Force

The matter particles of the standard model are fermions, particles that obey Fermi-Dirac statistics and having half-integer spins. Fermions come in two varieties:

quarks, and leptons. There are six known quarks and six known leptons. The quarks and leptons are divided into three separate groupings or ‘generations’. This structure is summarized in Tables 1.1 , 1.2, 1.3. The particle of central interest in this analysis is the top quark, a third generation quark. In its left-hand helicity state it is the weak isospin partner of the bottom quark. This sixth quark was discovered at Fermi National Accelerator Laboratory (FNAL) by the *CDF* and DØ experiments in 1995 [1] [2].

$$\begin{array}{cccc} & \textbf{Quarks} & & \textbf{Q} \\ \begin{pmatrix} u \\ d \end{pmatrix}_L, u_R, d_R & \begin{pmatrix} c \\ s \end{pmatrix}_L, c_R, s_R & \begin{pmatrix} t \\ b \end{pmatrix}_L, t_R, b_R & \begin{pmatrix} +2/3 \\ -1/3 \end{pmatrix} \\ & \textbf{Leptons} & & \\ \begin{pmatrix} e \\ \nu_e \end{pmatrix}_L, e_R & \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_L, \mu_R & \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}_L, \tau_R & \begin{pmatrix} -1 \\ 0 \end{pmatrix} \end{array}$$

Table 1.1: The fermions of the Standard Model. The left-handed particles reside in weak-isospin doublets. The right-handed particles reside in weak-isospin singlets.

The force particles of the standard model are bosons, particles that obey Bose-Einstein statistics having integer spins. These are called gauge bosons because of the gauge field nature of the theories describing the forces of the standard model. The electromagnetic force carrier is the photon (γ). The weak force is mediated by three vector bosons (W^+, W^-, Z^o). The strong force, color, is mediated by eight gauge bosons called gluons(g). The $SU(N)$ group theories have N^2-1 associated gauge bosons. Therefore, $SU(2)$ has 3 gauge bosons (weak force) and $SU(3)$ has 8 gauge

Force	**Gauge Bosons**	**Strength(α)**	**Range(m)**	**Time Scale(s)**
Electromagnetic	Photon (γ)	$\alpha \approx 10^{-2}$	∞	$10^{-20} - 10^{-16}$
Weak	W^+, W^-, Z^0	$G_F \approx 10^{-13}$	$< 10^{-16} cm$	$\geq 10^{-12}$
Strong	Gluon (g)	$\alpha_S \approx 1$	$< 10^{-13} cm$	$\geq 10^{-23}$
Gravity	Graviton ?	$G \approx 6 \times 10^{-39}$	∞	∞

Table 1.2: The ranges, strengths, and typical lifetimes associated with the three Standard Model interactions, and with gravity. (α is dimensionless coupling constant)

bosons (strong force). The coupling constant of a given force is a measure of the strength of that force. The gauge bosons are summarized in Tables 1.2 & 1.3. The Higgs boson will be discussed in Section 1.1.3.

Generation	**Particle**	**Type**	**Mass [32]** (GeV/c^2)	**Couplings**
1*st*	u	quark	0.0015 to 0.005	α, G_u, α_S, G
	d	quark	0.003 to 0.009	α, G_d, α_S, G
	ν_e	lepton	$< 10^{-8}$	G_{ν_e}, G
	e	lepton	5.11×10^{-4}	α, G_e, G
2*cd*	c	quark	1.1 to 1.4	α, G_c, α_S, G
	s	quark	0.06 to 0.17	α, G_s, α_S, G
	ν_μ	lepton	< 0.0002	G_{ν_μ}, G
	μ	lepton	0.106	α, G_μ, G
3*rd*	t	quark	175 ± 5.2	α, G_t, α_S, G
	b	quark	4.1 to 4.4	α, G_b, α_S, G
	ν_τ	lepton	< 0.018	G_{ν_τ}, G
	τ	lepton	1.78×10^{-3}	α, G_τ, G
	γ	gauge boson	0	G_F, G
	W^+, W^-	gauge boson	80.4	α, G_F, G
	Z^0	gauge boson	91.187	G_F, G
	g	gauge boson	0	G_F, α_S, G
	H(Higgs)	boson	> 95.3	G_F, G

Table 1.3: The masses of the fermions & bosons of the Standard Model. The couplings indicate the forces to which the forces which they couple to (the charges they carry). The stated masses for the ν_μ & ν_τ assume ν_3 is the dominate mass eigenstate.

Particle Interactions

Fermions interact with one another through mediating fields as in the classical theory of electromagnetism. However in relativistic quantum theory these fields are quantum fields, and the forces arise as the virtual exchange of their quanta (the gauge bosons described above). For example the interaction of two electrons via the exchange of photons (γ) as in Figure 1.1 is equivalent to the classical view of the electron acting on an EM field which then acts on the second electron. The photon is simply the quantum of that EM field. Such reactions are described by the well known Quantum Electrodynamics(QED) and is associated with the gauge group $U(1)_{EM}$.

Every gauge group has an associated gauge symmetry, this symmetry (if unbroken) gives rise to a conserved quantity. In $U(1)_{EM}$ that conserved quantity is electric charge.

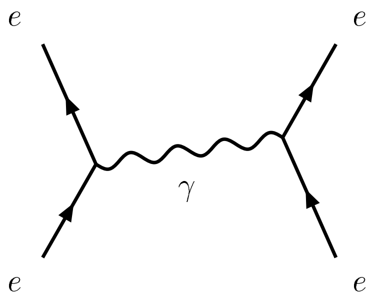
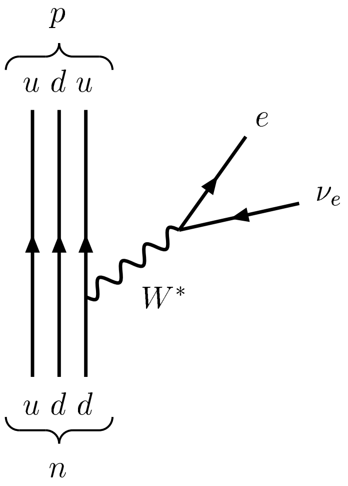


Figure 1.1: Electron-electron scattering via the exchange of a photon(γ).

A typical weak interaction is the decay of a neutron into a proton. A particle may decay weakly as in beta-decay where the initial particle the neutron decays into a proton, or more specifically the down quark within the neutron decays into an up quark, Figure 1.2.



$$n \rightarrow p + e + \nu_e$$
$$d \rightarrow u + e + \nu_e$$

Figure 1.2: Beta decay.

The W^+ boson in this reaction is not a real W because the neutron does not have the

80 GeV necessary to create a real W boson. The large mass of the W-boson suppresses the decay, therefore the decay time of the neutron is long (approximately 12 minutes). Due to the high mass of the weak bosons the weak force is usually slow as shown in Table 1.2. However if you start with a particle more massive than the W-boson then a real W-boson may be produced by the weakly decaying massive particle. This is precisely the case with a top quark, because the top is more massive than the W it can decay weakly to a real W. This allows the reaction to take place very quickly. The weak force in this case acts more quickly than the strong force- a top quark does not live long enough to hadronize. This will be discussed further in Section 1.2. The weak force is chiral- only particles which are in their left-handed helicity state and anti-particles in their right-handed helicity state interact weakly. The electromagnetic and weak forces together are best described by the electroweak theory first proposed by Glashow, Salam, and Weinberg. This theory which incorporates both electromagnetic and weak interactions is associated with the group $SU(2)_L \times U(1)_Y$. The L indicates that it is the left-handed weak isospin doublet which participates in the interaction and Y stands for *hypercharge.* They are related to the electric charge Q by,

$$Q = T_3 + \frac{1}{2}Y, \tag{1.1}$$

where T_3 is the third component of the weak isospin. The quantum numbers of the fermions in the first generation are,

	T	T_3	$\frac{1}{2}Y$	Q
u_L	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{2}{3}$
d_L	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{6}$	$-\frac{1}{3}$
ν_{eL}	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	0
e_L	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	-1
u_R	0	0	$\frac{2}{3}$	$\frac{2}{3}$
d_R	0	0	$-\frac{1}{3}$	$-\frac{1}{3}$
e_R	0	0	-1	-1

these are the same for each of the next two generations of fermions (the weak bosons also have electric charge). It is unknown whether the neutrino occurs in a right-handed helicity state.

The strong force is well described by quantum chromodynamics (QCD). This (like the weak theory) is a non-abelian gauge theory. This means that rotations of the symmetry group, in this case $SU(3)_C$, are not commutative. This aspect of the theory is a manifestation of the fact that gluons, the gauge bosons of the strong force, not only mediate the force but also are carriers of the charge of the strong force ('color'). This is very different from photons, γ, which mediate the electromagnetic force but are not themselves charged particles.

Together these theories, along with the Higgs mechanism (Section 1.1.3), form the Standard Model and characterize our best understanding of the three forces which they attempt to describe.

1.1.1 Hadrons

Hadrons are composite states of quarks bound together by the strong force. Since quarks and not leptons are carriers of strong charge, 'color', only quarks appear in hadrons. This explains why atomic electrons are not bound inside the nucleus of atoms as are protons and neutrons.[1]

Hadrons are classified into two groups: baryons, and mesons. Baryons are hadrons with half-integer spins, and mesons are integer spin hadrons. Each quark in the hadron has a spin of one-half therefore if the hadron contains an even number of quarks the hadron will have integer spin and be a meson, otherwise it will have half integer spin and be a baryon. The best known examples of baryons are the nucleons, protons and neutrons. The proton contains two up quarks and one down quark. The neutron contains one up quark and two down quarks.

---

[1] Though protons and neutrons (like all hadrons) are colorless bound states, their constituent quarks and gluons are not colorless. Color interactions occur between hadrons via π exchange, binding them within the nucleus.

$$p = (uud), \qquad n = (udd) \tag{1.2}$$

Likewise, superpositions of up and down quark states make pions the most common type of mesons,

$$\pi^+ = u\bar{d} \;, \qquad \pi^- = d\bar{u} \;, \qquad \pi^o = \frac{u\bar{u} - d\bar{d}}{\sqrt{2}} \;. \tag{1.3}$$

1.1.2 Confinement & Jets

It was not until the 1970^s, with evidence from deep inelastic scattering experiments, that quarks began to be thought of as true physical particles and not just theoretical constructs. This was due in large part to the fact that no isolated quark (or gluon) has ever been observed. This phenomenon is known as confinement. Quarks and gluons are believed to be bound into 'colorless' hadrons.

Unlike the well-known inverse square law of electromagnetism, where the force goes as $1/r^2$, the strong force goes as r. As the distance between two quarks increases the force between them increases. In this way the force becomes great enough, as the quarks move apart, to pair produce quarks out of the vacuum which then bind to the two initial quarks. This is illustrated in the Figure 1.3. Note that energy is clearly conserved here so long as the energy required to 'stretch' the quarks apart is greater than the mass of the two particles created out of the vacuum.

In this way many hadrons may be produced from a single initial quark pair. This process, known as hadronization, results in a shower of hadrons. This shower of hadrons is known as a 'jet'. It is these jets that are observed in particle detectors, and through these observations the originating quarks and gluons are inferred. The momentums of these hadrons may be used to reconstruct the momentum of the originating parton.

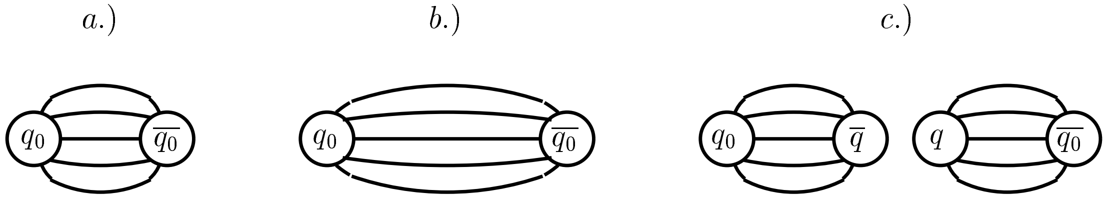


Figure 1.3: Schematic of the color lines of force between two partons, as described in the Lund string model. a.) quark pair which is close to asymptotically free. b.) quark pair which is being stretched as in a high energy collision. c.) quarks after a two more quarks have been pair produced from the vacuum. [4]

1.1.3 Higgs Mechanism

The Higgs field couples to all elementary particles with mass. The mass of a particle can be considered a measure of the strength of this coupling. The standard model has at its base the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$. It accommodates electroweak symmetry breaking by the introduction of a weak-isospin doublet of scalar fields Φ with hypercharge, $Y = +1$. The introduction of these fields is the key to the Higgs mechanism which is a vital part of the standard model, as it is currently constructed. Experimental evidence for the Higgs does not exist but the search for it will be of great experimental interest in the near future including Run II of *CDF*.

Without the Higgs mechanism the standard model has no means of accommodating the masses of the fermions or bosons which are experimentally seen. The weak-isospin doublet of scalar fields is,

$$\mathbf{\Phi} = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix} , \tag{1.4}$$

where φ^+ and φ^0 are complex fields with the quantum numbers

A scalar potential is added by hand,

$$V = \mu^2 |\mathbf{\Phi}|^2 + \lambda |\mathbf{\Phi}|^4 \qquad |\mathbf{\Phi}|^2 = \mathbf{\Phi}^\dagger \mathbf{\Phi} , \tag{1.5}$$

	T	T_3	$\frac{1}{2}Y$	Q
φ^+	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
φ^0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	0

where λ is the self-coupling of the scalar field. When μ^2 is chosen to be negative ($\mu^2 < 0$) a non-vanishing expectation value of $|\mathbf{\Phi}|^2$ appears in the physical vacuum state. The appearance of a preferred direction in weak-isospin space spontaneously breaks the symmetry of the theory [4]. Expanding the scalar field about its zero expectation value yields,

$$v = \sqrt{\frac{-\mu^2}{\lambda}} = (G_F\sqrt{2})^{-\frac{1}{2}} \approx 245\ GeV. \tag{1.6}$$

This is known as the electroweak scale. The fact that the top mass ($m_t \approx 175\ GeV$) is within a factor of two of this scale raises great interest in whether top physics might cast light on the process of spontaneous symmetry breaking.[5] The spontaneous breaking of electroweak symmetry gives mass (leading order) to the weak bosons,

$${M_W}^2 = \frac{\pi\alpha}{G_F\sqrt{2}sin^2\theta_W} \tag{1.7}$$

and

$${M_Z}^2 = \frac{{M_W}^2}{cos^2\theta_W}. \tag{1.8}$$

Where α is the fine structure constant, G_F is the Fermi weak coupling constant, and θ_W is the weak mixing angle. When the symmetry is broken ($SU(2)_L \times U(1)_Y \rightarrow SU(1)_Q$) three generators are broken, the three Goldstone bosons are then interpreted as longitudinal $W_L^{\pm}$, Z_L fields. The remaining unbroken generator is electric charge (Q). In addition to the weak boson masses a physical scalar (spin-0) particle, the Higgs boson, appears in the theory. Every fermion has its own Yukawa coupling(G_f) to the Higgs field, and thereby acquires a mass,

$$m_f = \frac{G_f v}{\sqrt{2}} \ . \tag{1.9}$$

The search for the Higgs boson will be of primary importance in the next run of the Tevatron at both CDF and DØ. How $t\bar{t}$ production impacts these searches will be discussed in Section 1.2.6.

1.2 The Top Quark

1.2.1 Why should the Top Quark Exist?

There was a short period of time prior to the discovery of the τ lepton(by Martin Perl) when nothing was known of a third generation of particles. However since this discovery and certainly since the discovery of the first third generational quark, the bottom quark [10], the search for the top quark was of considerable concern in the particle physics community. It should be noted that before the discovery of the charm quark M. Kobayashi and T. Maskawa pointed out that a third generation of quarks would allow for CP violation.[11] But why were these discoveries viewed by physicists as such compelling evidence for the top quark?

Chiral Anomalies

The answer to this question is several fold. One early concern about a model without a top quark had to do with the renormalizability of such a theory. There are certain anomalies ('chiral' anomalies) which can arise when three gauge bosons interact via a closed loop of fermions, even finite anomalies are forbidden. This is illustrated in Figure 1.4 and is known as a triangle diagram.

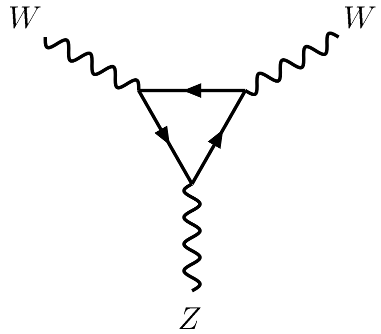


Figure 1.4: An example of a standard model triangle diagram. These types of diagrams can give rise to chiral anomalies. Where at least one of the three must be left-handed.

It turns out that if the charges of the left-handed fermions in each generation sum to zero,[6]

$$\sum Q_f = \sum Q_l \; + \; 3 \cdot \sum Q_q = 0, \qquad (1.10)$$

this anomaly is eliminated and the model is chiral free. Prior to the discovery of any of the third generation fermions this was the situation. However given the b-quarks charge another charge +2/3 fermion,

$$\sum Q_f = -1 \; + \; 3 \cdot \left[\left(\frac{2}{3} \right) + \left(-\frac{1}{3} \right) \right] = 0 \qquad (1.11)$$

the top quark, is necessary to make this sum equal to zero. Therefore in order to eliminate chiral anomalies and get a renormalizable theory it is necessary to introduce the top. This is also circumstantial evidence for some higher symmetry (Grand Unification).[12]

Forward-Backward Asymmetry

Another critical clue to the existence of the t-quark is the angular distribution in $b\bar{b}$ production from e^+e^- collisions. There are two first order diagrams which contribute to this Figure 1.5.

A symmetric distribution is obtained from the γ portion of this process. However when the diagram with the massive Z-boson is included a forward-backward asymmetry arises. When the center of mass energy, $\sqrt{s}$, of the e^+e^- collisions is at or near the Z resonance this forward-background asymmetry in $b\bar{b}$ production becomes significant.

By carefully measuring the asymmetry occurring at the $Zb\bar{b}$ vertex [4] a measurement of the b-quark isospin can be made. The Z-boson couples to the $b\bar{b}$ through

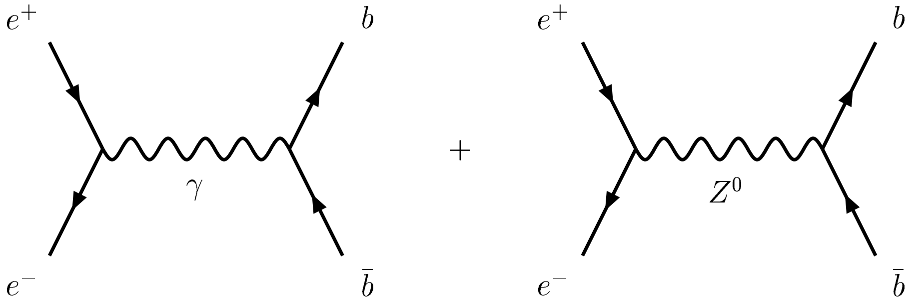


Figure 1.5: Lowest order production of $b\bar{b}$ pairs in e^+e^- collisions.

vector and axial vector couplings. The strength of these couplings (v_b and a_b) are given by,

$$\sqrt{\frac{G_F {M_Z}^2}{2\sqrt{2}}}\gamma_\mu(v_b - a_b\gamma_5) \tag{1.12}$$

where v_b & a_b are,

$$v_b = 2[I_3^L(b) + I_3^R(b)] - 4e_b sin^2\theta_W \qquad and, \qquad a_b = 2[I_3^L(b) + I_3^R(b)] \ , \tag{1.13}$$

the $I_3^L(b)$ and $I_3^R(b)$ are the third components of the isospin of the b-quark [6]. The forward-backward asymmetry is sensitive to the relative size of these two couplings of the $Vb\bar{b}$ vertex.

$$A_{FB} = \frac{3a_e v_e a_b v_b}{({v_e}^2 + {a_e}^2)({v_e}^2 + {a_e}^2)} \tag{1.14}$$

Measurements indicate that the third components of isospin for the b-quark are:

$$I_3^L(b) = -0.490^{+0.015}_{-0.012} \qquad I_3^R(b) = -0.028 \pm 0.056. \ \ [7,\ 8] \tag{1.15}$$

This indicates that the right-handed b-quark is a singlet state, but with $I_3(b) \approx 1/2$ the left-handed b-quark should be part of a isospin doublet. The partner in this doublet is by definition the top quark.

b-quark: singlet or doublet

A bottom quark might decay weakly by a flavor changing current to a lighter quark. The CKM matrix gives an indication, in elements V_{ub} & V_{cb}, of the likelihood of such decays to occur. The $b \rightarrow cW$ reaction is seen frequently enough that reasonable measurements of the coupling factors have been made [9]. This flavor changing charged current (FCCC) is shown in Figure 1.6 The process, $b \rightarrow cW$, shown here is a standard model process known to occur. This process can only happen if the b-quark is part of an isospin doublet, or if the b-state mixes with the charm state. However if this FCCC, $(b \leftrightarrow s) \rightarrow cW$, happens then it would have to be possible for the flavor changing neutral current (FCNC), $(b \leftrightarrow s) \rightarrow sZ$, process to occur Figure 1.7.

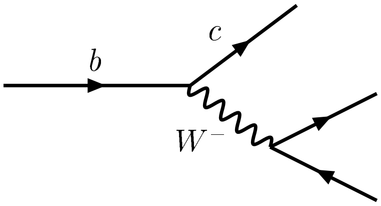


Figure 1.6: An allowed standard model decay, if the b-quark is part of an isospin doublet.

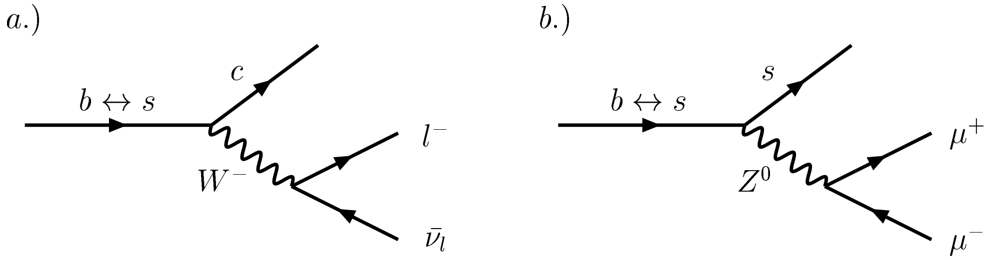


Figure 1.7: How a b-quark singlet state might mix with lighter quarks is shown. Where the W-boson decays into leptons. a.) flavor changing charged current. b.) flavor changing neutral current.

Both the W & Z bosons may decay into quarks or leptons. When they decay leptonically the W-bosons decay to a single detectable lepton and the Z-boson decays to two detectable leptons. If both these processes were possible than it would be expected that the ratio of the weak couplings would be,

$$\frac{\Gamma(b \to l^+ l^- X)}{\Gamma(b \to l^- \nu_l X)} \approx 0.11 \ [9] \tag{1.16}$$

Despite this expectation no l^+l^- pairs have been observed in b-decays, as would be expected if the b-quark were a isospin singlet, and therefore decayed to a Z-boson part of the time.

For all these reasons the existence of the top quark was anticipated long before any experimental evidence of its existence was found. And the large mass of the top was indicated by B^o-$\bar{B}^o$ mixing.

1.2.2 Top Discovery

In the 1970s the search for the top quark began in earnest at e^+e^- colliders, the first important experiment to join the search was PETRA at DESY. PETRA set the first confidence limit on the mass of the top, $m_t > 23\ GeV$ [10]. A few years later the TRISTAN experiment improved this limit to 30 GeV [13]. Then in 1984 the UA1 experiment at CERN SppS thought they had found the top quark [14]. This signal was proved to be false within a year by the UA1 group itself. The backgrounds, mostly the $W + jets$ backgrounds, had been underestimated [15]. These experiments eventually increased the mass limit to $m_t > 45.8\ GeV$ [16]. However it would have to wait for the development of $p\bar{p}$ colliders and the significantly higher energy realms which they lay open to allow further work to be done on the top mass.

The UA1 and UA2 experiments ran through 1989 and produced mass limits of $m_t > 60\ GeV$ and $m_t > 69\ GeV$ respectively [17]. In its first year of running, 1988-89, CDF raised the mass limit to $m_t > 91\ GeV$ [18]. DØ soon moved this limit up to $m_t > 128\ GeV$ [19, 20]. In April 1994 the CDF collaboration claimed evidence for $t\bar{t}$

production [21]. With 19.3 pb^{-1} of data and 12 candidates CDF set the production cross section at $\sigma_{t\bar{t}} = 13.9^{+6.1}_{-4.8} pb$ and the mass at $m_t = 174 \pm 10^{+13}_{-12}\ GeV$ [1, 22]. DØ found nine candidate events setting $\sigma_{t\bar{t}} = 8.2 \pm 5.1 pb$ and $m_t \approx 180\ GeV$ [2, 19].

Before the summer of 1995 both CDF & DØ experiments had announced discovery of top [1, 2]. Twenty years after the PETRA's earliest mass limit. Current mass and $\sigma_{t\bar{t}}$ measurements are shown in Figures 1.9 , 1.10. Figure 1.8 shows one of the first CDF candidate events.

1.2.3 Top Quark Production

Top quarks can be produced as single top quarks or in quark-antiquark pairs. Single top production is more easily swamped by backgrounds is not expected to be seen in the Run I data set. The theoretical single top cross section is $\sigma_{tb} = 2.43\ pb$.[30] [77]

Single top Production

Single top production happens via two processes. Firstly by the creation of an off mass shell W which decays to a quark-antiquark' ($q\bar{q}$') pair Figure 1.11. The prime signifies that the second quark is of a different type than the first. Secondly a quark and gluon can interact via a W and a secondary b-quark Figure 1.12. Producing a top quark and an anti-bottom quark. This measurement has not been made but will be made at FermiLab during Run II. This will provide a measurement of the V_{tb} element of the CKM matrix,

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.17}$$

The values of the CKM matrix determine the amount of mixing that takes place between the different quarks. This measurement could confirm three generational unitarity, which would contra-indicate a 4^{th} generation of quarks, proving at least that no mixing between a 4^{th} generation and the known three generations is possible. [32]

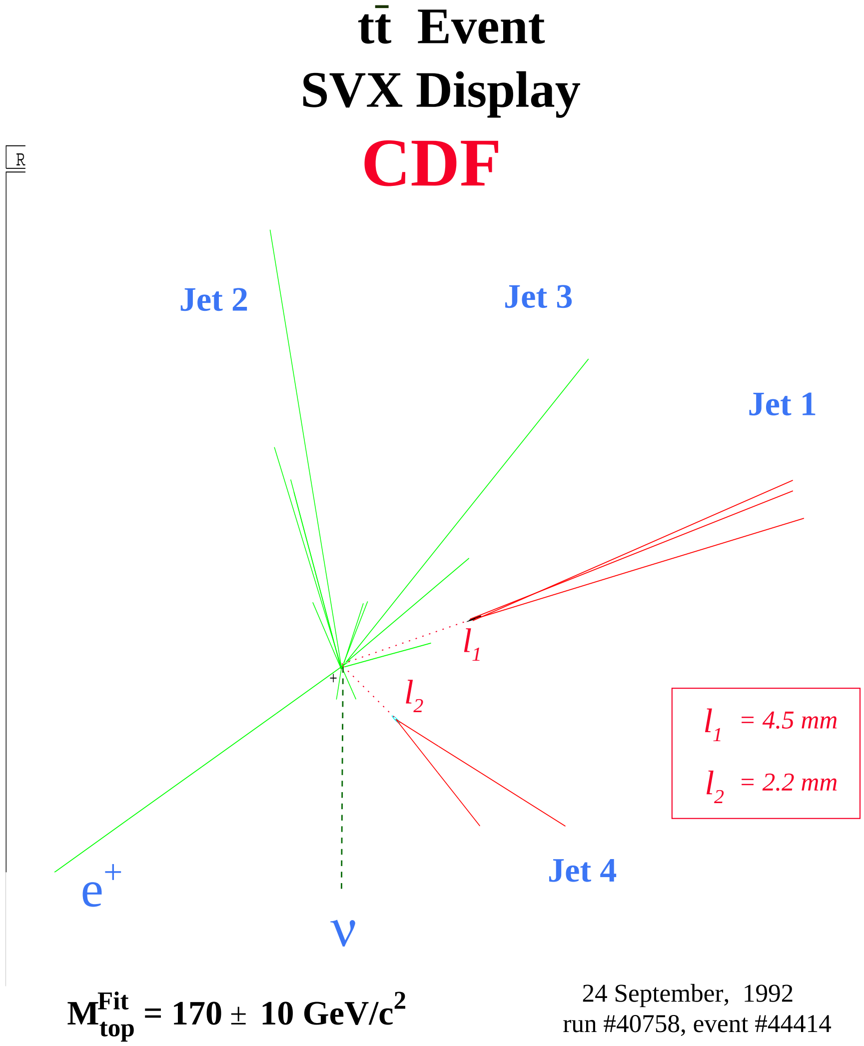


Figure 1.8: Very early $t\bar{t}$ candidate $l + jets$ event with SVX b-tags. The two jets with displaced vertices, jets 1 & 4, indicate b-quarks. Such a displacement occurs because the lifetime of the b-quark is relatively long. A typical Lorentz factor here is $\gamma = E/m_b \approx 10$. Therefore, the distance traveled by the B-meson is $< d >= \gamma \cdot c \cdot \tau \cong 10 \cdot (3 \times 10^{10}\ cm/s) \cdot 10^{-12}\ s = 0.3\ cm$.

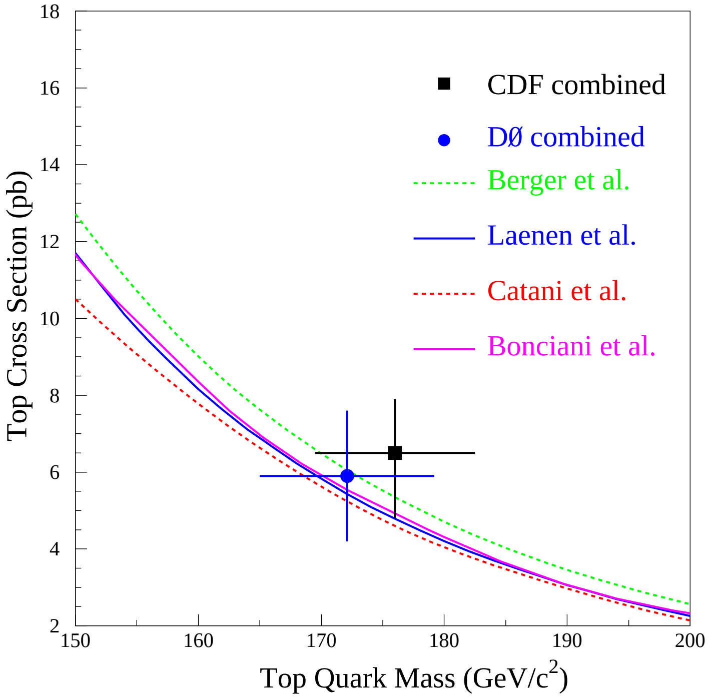


Figure 1.9: Mass of top and cross section of $t\bar{t}$ as compared with theory [29].

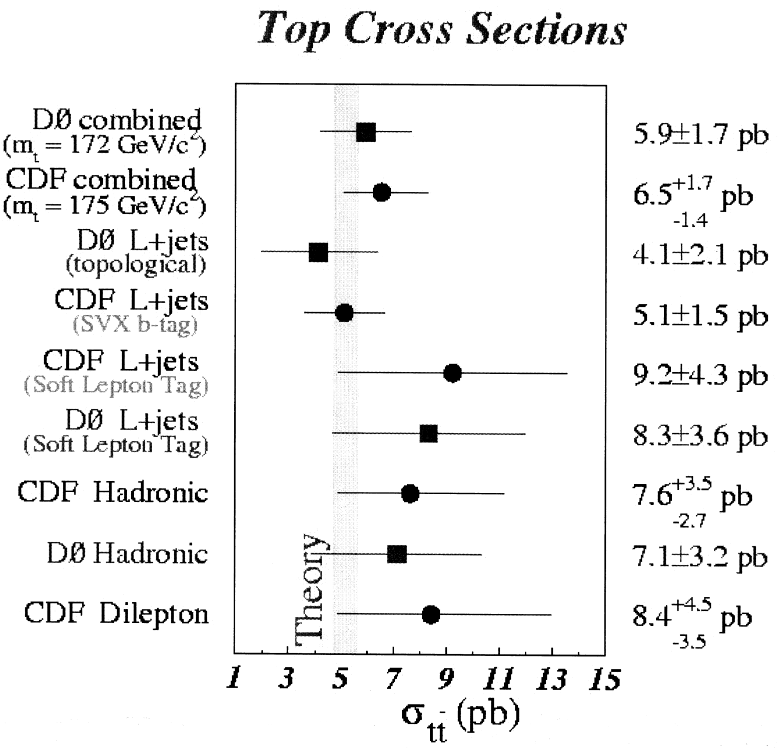


Figure 1.10: Current combined CDF and DØ $t\bar{t}$ production cross section numbers.

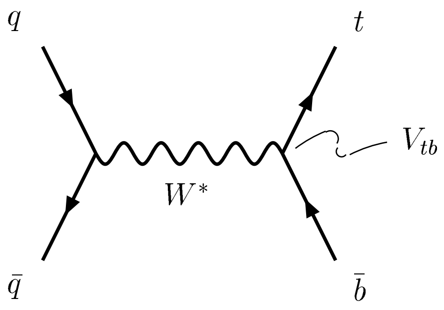


Figure 1.11: Lowest order processes for QCD single *top* production from $q\bar{q}$ annihilation.

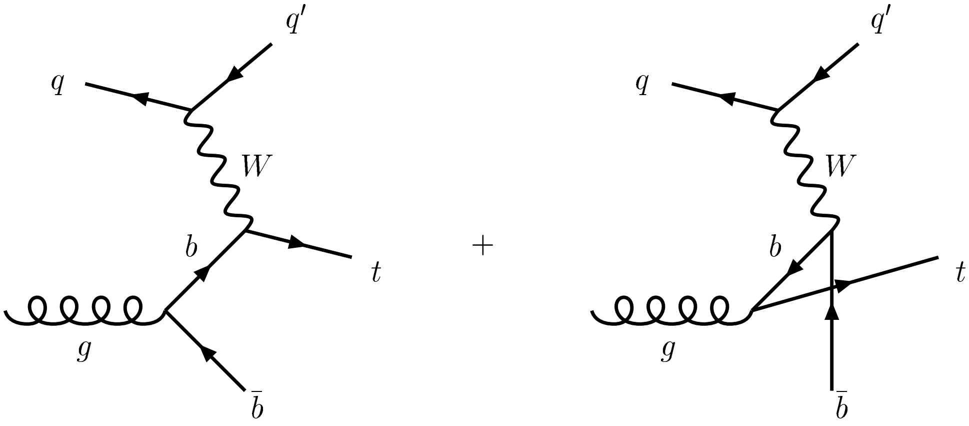


Figure 1.12: Lowest order processes for QCD single *top* production from qg scattering.

$t\bar{t}$ Production

This analysis is concerned with $t\bar{t}$ production. Pair production also occurs via two processes. One production mechanism is gluon fusion. In this process two gluons fuse into one which then strongly decays into a $t\bar{t}$ pair. There are three production diagrams for this process Figure 1.14. The other production mechanism is quark annihilation. A quark anti-quark annihilate with each other producing a gluon which then pair produces a $t\bar{t}$ pair Figure 1.16. The relative probability of each of these mechanisms is dependent on the center of mass energy ($\sqrt{s}$) of the initial interaction.

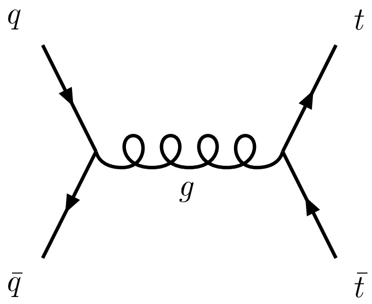


Figure 1.13: Lowest order process for $t\bar{t}$ production from $q\bar{q}$ annihilation.

At the Tevatron $\sqrt{s}$ the $q\bar{q}$ process is favored, 90% of the $t\bar{t}$ pairs are produced this way.

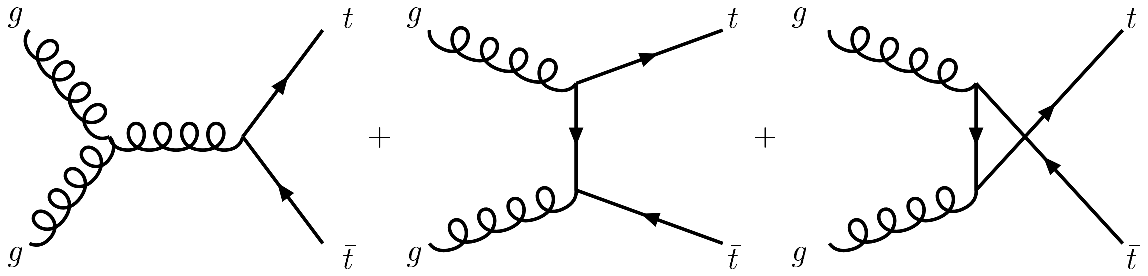


Figure 1.14: Lowest order processes for QCD single top production from gluon fusion.

1.2.4 Top Decay

The top quark decays weakly into a bottom quark and a real W-boson, Figure 1.15. Although weak decays are usually rather slow (Table 1.2) this is not the case for the top quark. In this case the weak force acts very quickly before the top has time to hadronize. For the measured top mass, $m_t = 175\ GeV$, the decay theoretical width is $\Gamma_{top} \approx 1.8\ GeV$ and its lifetime is $\hbar/\Gamma_{top} \approx 10^{-25} seconds$.

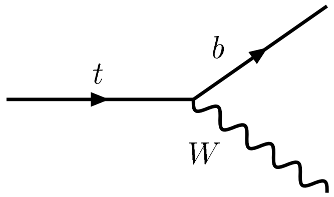


Figure 1.15: The top quark decays into a W boson and a bottom quark.

Neglecting terms of the order of $\frac{m_b^2}{m_t^2}$, α_S^2, and $\left(\frac{\alpha_S}{\pi}\right)\frac{M_W^2}{m_t^2}$ the decay width is given by,[28]

$$\Gamma_{top} = \frac{G_F\ m_t^3}{8\pi\sqrt{2}} \left(1 - \frac{M_W{}^2}{m_t{}^2}\right)^2 \left(1 + 2\frac{M_W{}^2}{m_t{}^2}\right) \left[1 - \frac{2\alpha_S}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right] . \tag{1.18}$$

This compares with hadronization timescales of $\approx 10^{-24}$ seconds. Therefore it is the high top mass which is responsible for the top's short lifetime.

1.2.5 Signature & Branching Ratios

The decay channels of $t\bar{t}$ pairs are classified by the varying decay modes of the W-bosons produced in the process. The W can either decay into quarks or leptons, Figure 1.16. If it decays into leptons that pair of leptons may be any one of the three lepton generations: (e, ν_e), (μ, ν_μ), or (τ, ν_τ). If it decays into quarks those quarks will be one of two different flavor pairs, (u, d) & (c, s), which conserve charge and are kinematically allowed. Each flavor pair is triply degenerate coming in three different color combinations. Therefore each quark flavor pair is three times as likely as a given lepton pair Table 1.4.

One channel is both the W-bosons decaying to leptons, known as the dilepton channel. If both Ws decay to quarks and thereby produce jets it is the all-hadronic channel. And finally the channel of interest in this analysis the *lepton + jets* channel is when one W decays leptonically and the other hadronically. The branching ratios of each of these channels can easily be seen from the Table 1.4. Since τ leptons present special problems for detection most analyses do not attempt to include them and sometimes have to contend with them as backgrounds. Therefore the effective branching ratios are smaller if events with τ leptons are not included, as is the case in most analyses, this one included,

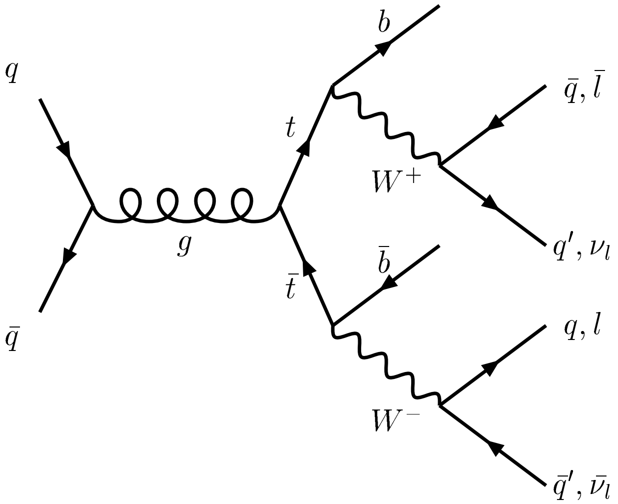


Figure 1.16: Lowest order process for QCD $t\bar{t}$ production from $q\bar{q}$ annihilation. And the tops decay, $t \rightarrow Wb$, with the final W decay.

$$
\begin{aligned}
Br(WW \rightarrow l l \nu \nu) &= 4/81 \qquad &\text{(dileptons)}\\
Br(WW \rightarrow l \nu q q') &= 24/81 \qquad &\text{(lepton+jets)}\\
Br(WW \rightarrow q q' q q') &= 36/81 \qquad &\text{(all hadronic)} \; .
\end{aligned}
$$

The all-hadronic channel has the largest branching ratio, however it also has the largest backgrounds. The channel with the smallest branching ratio, the dilepton channel, is also the cleanest having the fewest backgrounds with which to contend. The lepton + jets channel ($l + jets$), used in this analysis, is a compromise between branching ratios and backgrounds.

1.2.6 Why is Top Important?

Currently the foremost experimental goal in particle physics is to find the Higgs boson. If this discovery is made all the standard model particles will have been measured, Table 1.3. The top quark plays an important role in this search in at least two ways. First the mass of the top has already, along with the W boson mass, given us a idea of the mass range over which to expect to find the Higgs boson if it exists, Figure 1.17.

The dominant production mechanism for standard model Higgs production at

W^+
decay

	$e\nu_e$	$\mu\nu_\mu$	$\tau\nu_\tau$	ud	$c\bar{s}$
$c\bar{s}$	3/81	3/81	3/81	9/81	9/81
$u\bar{d}$	3/81	3/81	3/81	9/81	9/81
$\tau\nu_\tau$	1/81	1/81	1/81	3/81	3/81
$\mu\nu_\mu$	1/81	1/81	1/81	3/81	3/81
$e\nu_e$	1/81	1/81	1/81	3/81	3/81

W^+
decay

Table 1.4: Standard Model branching ratios for $t\bar{t}$ W decays.

the Tevatron is by gluon fusion($gg \rightarrow H$). Because of the high QCD backgrounds in this mode, it is believed that $q\bar{q} \rightarrow V^* \rightarrow VH \rightarrow Vb\bar{b}$, where V is either a W or Z, provides one of the best opportunity for discovery at *CDF*, Figure 1.18. [12]

Clearly for this process $t\bar{t}$ will pose a major background problem. The signature for this process has the potential to appear top like.

Standard model processes are by no means the only processes for which $t\bar{t}$ background will be a major concern. We have already discussed the importance of the Higgs mechanism in the standard model as well as the way in which it is incorporated as a scalar doublet with a single physical Higgs boson emerging in the theory. However this picture is more complicated when one looks at the Minimal Supersymmetric Standard Model (MSSM). In the MSSM five Higgs particles are predicted: two CP-even scalars (h & H), one CP-odd pseudoscalar (A), and two charged bosons ($H^\pm$). It is worth noting here that in the MSSM the bottom-Higgs Yukawa couplings are enhanced making the b-sector of great interest in supersymmetric Higgs searches [39].

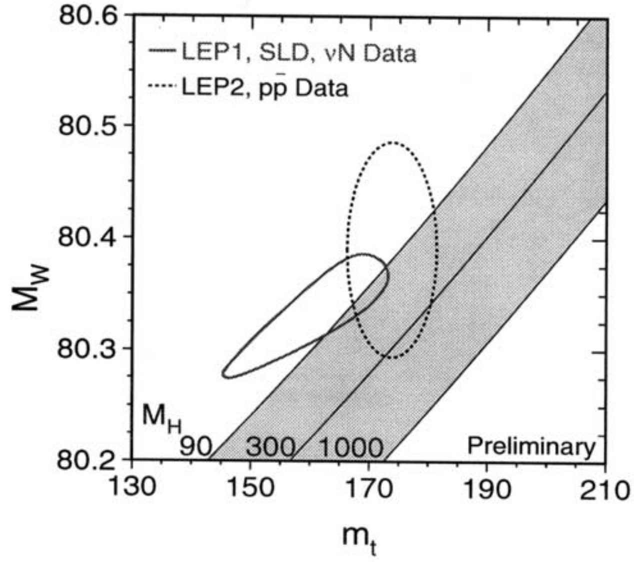


Figure 1.17: The dependence of the Higgs mass on the top and W masses. Band represent Higgs boson masses from 90 to 1000 GeV. The contours shown represent one σ confidence level. This figure taken from reference [33].

One example of such a search is $p\bar{p} \to b\bar{b}\varphi \to b\bar{b}b\bar{b}$ where $\varphi = h, H, A$ [40]. To lowest order these processes are shown in Figure 1.19 These four jet events would be strongly masked by $t\bar{t}$ production.

Perhaps the best example of a SUSY search with high $t\bar{t}$ background is the search for the supersymmetic top (stop($\tilde{t}$)) quark. A commonly predicted $\tilde{t}$ decay is $\tilde{t} \to b + \tilde{\chi}_1^+$ and $\tilde{\chi}_1^+ \to Wb + \tilde{\chi}_1^0$ or $\tilde{\chi}_1^+ \to bl + \tilde{\chi}_1^0$ [41]. Where $\tilde{\chi}_1^0$ is the Lightest Supersymmetric Particle (LSP). This decay is compared to the top decay in Figure 1.20.

The similarity of these events is self evident. They are especially similar given that leptonic W decay modes make high missing energy normal in top decays. The $\tilde{\chi}_1^0$, being the LSP, is stable and therefore only observed in the detector as missing energy.

There are even a number of non-$SUSY$ searches which are substantially im-

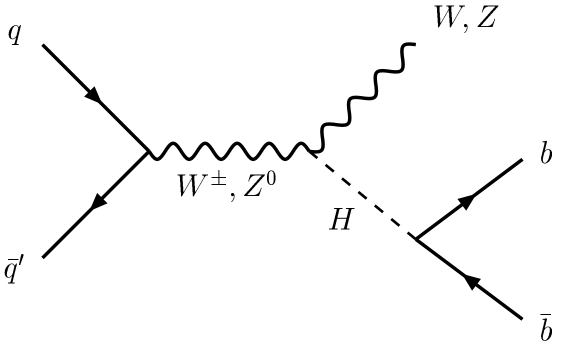


Figure 1.18: Possible Higgs production at the Tevatron. Three weak bosons are shown in the final state.

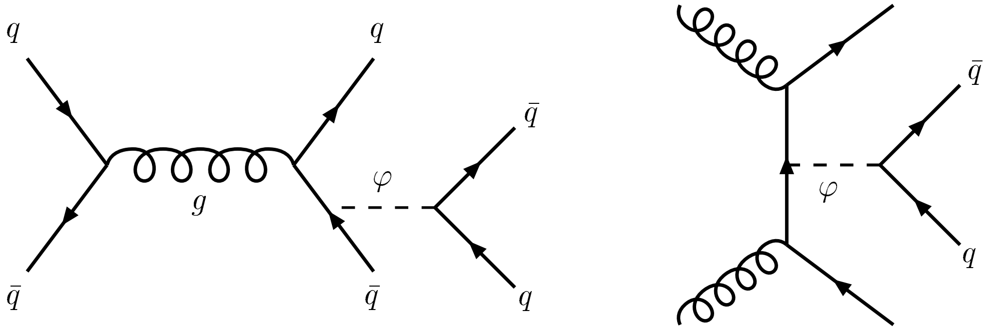


Figure 1.19: Lowest order process for QCD $t\bar{t}$ production from $q\bar{q}$ annihilation.

pacted by top backgrounds. For example in technicolor there are technimesons $\rho_T \to W\pi_T \to Wb\bar{b}$ [36].

These are just a sampling of some of the searches for which $t\bar{t}$ production will be a significant background. Clearly this is reason enough to study $\sigma_{t\bar{t}}$, however it should be remembered that top is of substantial interest on its own accord. One reason for this, how single top impacts three generational unitarity, has been discussed in Section 1.2.3. The best reason for studying top is, simply the examination of an interesting area of physics, in this case physics that may greatly effect the electroweak scale.[5]

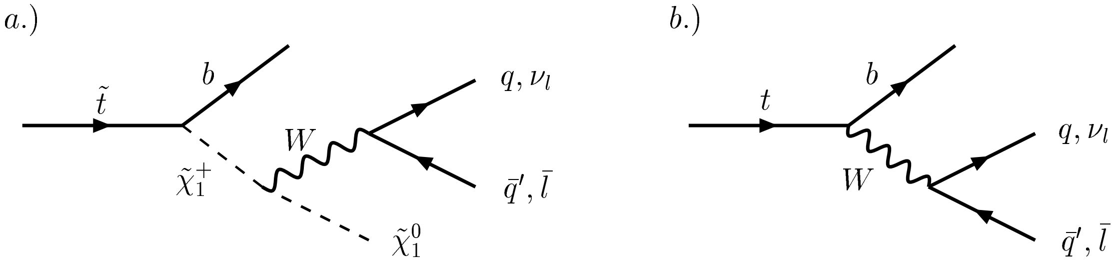


Figure 1.20: A comparison of top and stop quark decays. a.) stop decay. b.) top decay.

Chapter 2

Detector

2.1 The Tevatron

The Tevatron is a superconducting accelerator with a main ring radius of one kilometer. It is a proton anti-proton collider with 6 bunches each of p's and $\bar{p}$'s each at 900 GeV, for a center of mass energy ($\sqrt{s}$) of $1.8 TeV$. Each of the p bunches contains approximately 2×10^{11} protons and each $\bar{p}$ bunch contains approximately 5.5×10^{10} anti-protons.

The production of protons for the Tevatron begins with a bottle of H_2 gas. Electrons are added to the hydrogen to create H^- ions. These ions are then accelerated to 750 keV using a Cockcroft-Walton electrostatic accelerator and sent into a 150 m linear accelerator (linac). The linac accelerates the ions to 400 MeV/c^2. At the end of the linac the ions are passed through a copper foil which strips them of both of their electrons. The protons which emerge from the foil are then injected into the Booster. The *Booster* is a 23 m radius synchrotron; it accelerates the protons to 8 GeV before injecting them into the Main Ring. The *Main Ring* is a 1 km radius accelerator which generates magnetic fields of 0.7 $Tesla$. The protons are accelerated by 0.5 MeV/c^2 per turn until they reach 150 GeV, they are coalesced into bunches and the bunches are injected into the Tevatron. Superconducting magnets in the Tevatron, which is directly below the Main Ring, generate fields of 0.66 T to 4.4 T.

The p bunches are finally accelerated to 900 GeV in the Tevatron. This whole process takes approximately one minute.

The process for anti-protons is the same up to the Main Ring, then 120 GeV protons are shunted out of the Main Ring and fired at a tungsten target. From the resultant shower residue $\bar{p}$'s are selected and focused with a lithium lens. From there they are sent into the *Debuncher* where they are "stocastically cooled", resulting in a phase space reduction of the beam (i.e., the energy spread of the beam is reduced). From the debuncher the cooled $\bar{p}$'s are sent into a concentric ring, the accumulator, where they are "stacked" (stored and accumulated) until there are approximately 100×10^{10} $\bar{p}$'s. Six bunches of $\bar{p}$'s are then injected into the Tevatron and accelerated to full energy (900 GeV). Because the $\bar{p}$'s stack at a rate of about 4×10^{10}per hour the whole process for $\bar{p}$'s takes about one day [42].

Once the p and $\bar{p}$ bunches are in the Tevatron and circulating in opposite directions around the ring the stacking procedure is continual. Because of this continued operation of the Main Ring it is necessary for the Main Ring to be diverted above the *CDF* detector to minimize contamination from particles coming off the ring.

Counter rotating beams are made to collide at two interaction points around the Tevatron: D0 and B0. B0 is the point were the Collider Detector at Fermilab (*CDF*) resides. Circulating around the ring requires $21\mu s$, with 6 bunches collisions take place at a rate of one crossing per $3.5\mu s$. Collisions continue until the luminosity degrades too badly due to collisions, beam losses and beam-beam interactions.

2.2 The *CDF* Detector

The Tevatron is currently the highest energy collider in the world at a center of mass energy of $\sqrt{s} = 1.8\ TeV$. The *CDF* detector is one of two detectors designed to exploit this high interaction energy. It is designed to measure the various high energy particles which emerge from the $p\bar{p}$ interaction region. *CDF* has the capacity to directly measure: electrons, muons, photons and jets. A jet is a shower of collimated hadrons. Jets are defined in the *CDF* detector to be the energy deposited inside a "cone" (a predetermined solid angle) and which track back to a single interaction point. There are three primary hardware components to the *CDF* detector: tracking chambers, calorimetry and muon chambers. Each of the hardware components in addition to triggers and offline reconstruction will be described in the following sections; for a more complete description, refer to the *CDF* "blue book" [43].

2.2.1 Tracking Systems

There are three distinct tracking systems which work in conjunction to make a precision measurement of a particle's path through the detector. The three tracking components in the *CDF* detector are the Silicon Vertex detector (SVX), the Vertex Time Projection Chamber (VTX) and the Central Tracking Chamber (CTC).

Silicon Vertex Detector

The SVX is a silicon microstrip detector which lies at the center of the *CDF* , immediately outside the beamline. The SVX is divided into two barrels each 51 cm in length, adjoining at $z = 0$. A schematic of one barrel is shown in Figure 2.4.

There are four layers of silicon varying from a radius of 2.9 cm to 7.9 cm in each barrel. Each layer consists of varying numbers mechanical/electrical units (ladders) depending on the radius of the layer around the beamline and forming a cylinder of microstrip sensors. A ladder has three single sided microstrip sensors bonded together end to end running the length of the barrel. The strips on the sensors are $60 \mu m$ pitch running axially.

One of primary values of the SVX detector has been its ability to find displaced vertices. This has been of great interest for tagging b quarks from $t\bar{t}$ events in Run I at the Tevatron, and was instrumental in the discovery of the top quark. Although this part of the *CDF* detector has been central to past *CDF* top analyses it is of somewhat less concern in this analysis because b-tagging is not used to identify top quarks. The efficiency on b-tagging is limited, and the independence from b-tagging of this analysis is one of its primary benefits.

Vertex Detector

The VTX is a time projection chamber with sense wires running through the chamber. The chamber is segmented azimuthly into eight pieces. The gas in the chambers are approximately a 50/50 mix of argon and ethane. Sense wires run outward from the beamline. Tracks through the gas yield tracking information allowing the primary vertex in the event to be determined. The drift times to the sense wires provides time of arrival information which is used to locate the event vertex's z position to within 1 mm. The VTX extends from the outside of the SVX out to a radius of 22 cm.

Central Tracking Chamber

The primary purpose of the CTC is to measure charged particle momenta. Lepton identification comes from coincidences between the CTC and the EM calorimetry or muon chambers. The particle's momentum is determined by it's path in the 1.5 T magnetic field. There are sense wires running the length of the CTC (along the beamline). There are nine "super-layers" with alternating "axial" and "stereo" wires, Figure 2.5. The axial layers have wires running parallel to the beamline and the stereo layers have wires running $\pm 3\deg$ from beamline parallel. The combination of axial and stereo information give a full 3-dimensional track reconstruction. To correct for the Lorentz angle produced by crossed $\vec{E}$ and $\vec{B}$ fields, the wires create a 1350 V/cm drift field. The magnetic field in the CTC is provided by a solenoid magnet just outside the CTC. The superconducting solenoid is liquid He cooled. Momentum resolution of the CTC is $\delta P_T/{P_T}^2 = 0.002(GeV/c)^{-1}$, and the inclusion of SVX information improves momentum resolution by a factor of two [43].

2.2.2 Calorimetry

Particles with transverse momentum of greater the 350 MeV have the potential to escape the solenoid which surrounds the CTC. Outside the tracking region of the detector calorimeters are located. Precision energy resolution is obtained from two categories of calorimetry: hadronic and electromagnetic. Hadronic calorimetry for the measurement of strongly interacting particle energies and electromagnetic calorimetry for the measurement of electromagnetically interacting particle energies, γ, e.

The central region of the detector out to a pseudorapidity ($|\eta|$),

$$\eta \equiv -ln\left[tan\left(\frac{\theta}{2}\right)\right] \ ,$$

of 1.1 is covered by the central hadronic(CHA) and the central electromagnetic(CEM) calorimeters. They are divided azimuthly into 24 equal wedges ("projective towers"). Every wedge is divided along η into towers, where each tower covers $\Delta\eta = 0.1$. The electromagnetic calorimeters lie directly inside a corresponding hadronic calorimeter in any given tower, this allows the electromagnetic and hadronic energies to be compared. The region from $(1.1 \leq |\eta| \leq 2.4)$ is considered to be the end plug region and higher $|\eta|$ is considered the forward region. The coverage of these towers is shown in Figures 2.2, 2.3.

Electromagnetic Calorimeters

In the central region the electromagnetic shower counters use lead interspersed with scintillator as the active medium. Outside the central region proportional tube chambers with cathode pad readout are used. Proportional strip and wire chambers, central shower counters(CES), at a shower maximum depth of six radiation lengths provide spatial resolution of $\approx 2\ mm$, Figure 2.6. In the end plug region three different radiation depths are sampled (3.0, 3.8 and 14.2). The forward region samples at two depths. The energy resolution for the electromagnetic calorimetry varies from $\sigma/E = 2\%$ at 50 GeV in the central region to $\sigma/E = 4\%$ at 50 GeV in the end and forward regions.

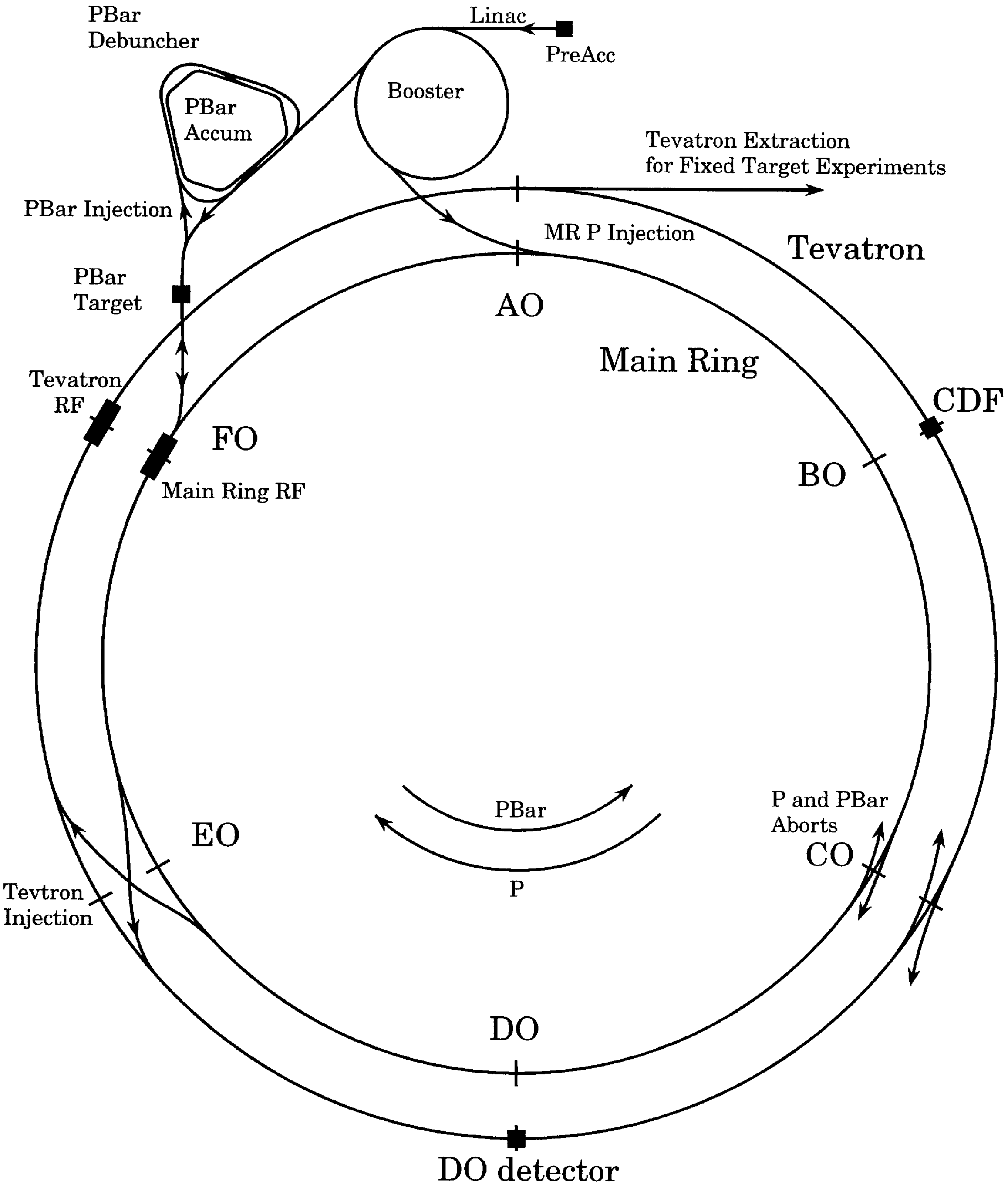


Figure 2.1: Schematic view of the accelerator assembly.

Figure 2.2: A cross section of one quadrant of the *CDF* detector. It is symmetric in the other quadrants.

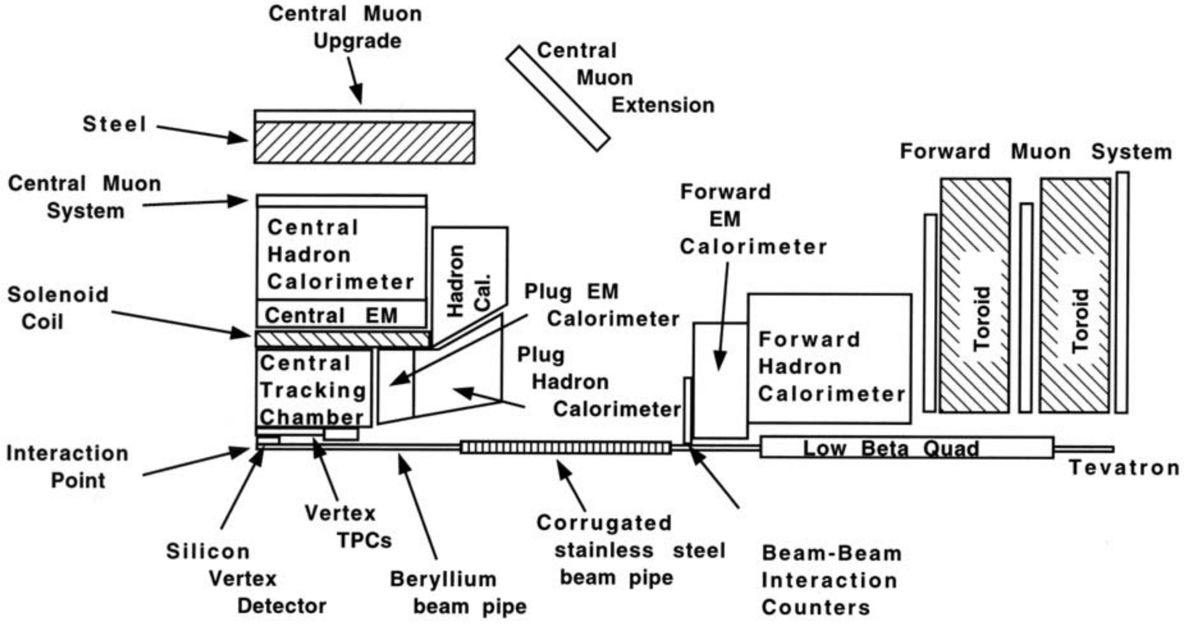


Figure 2.3: Another cross section of one quadrant of the *CDF* detector. It is symmetric in the other quadrants. (not to scale)

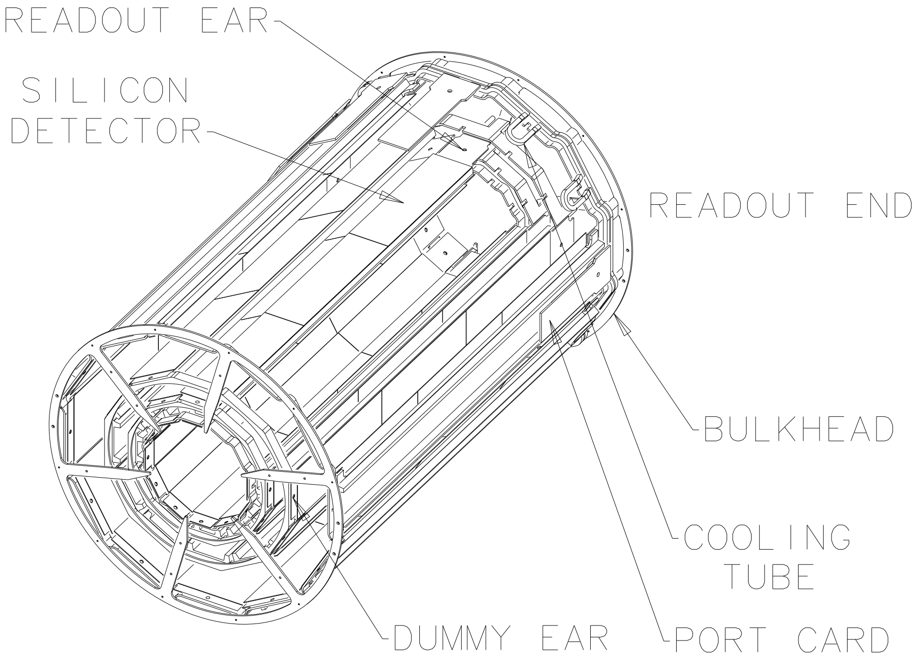


Figure 2.4: A schematic of one barrel of the SVX.

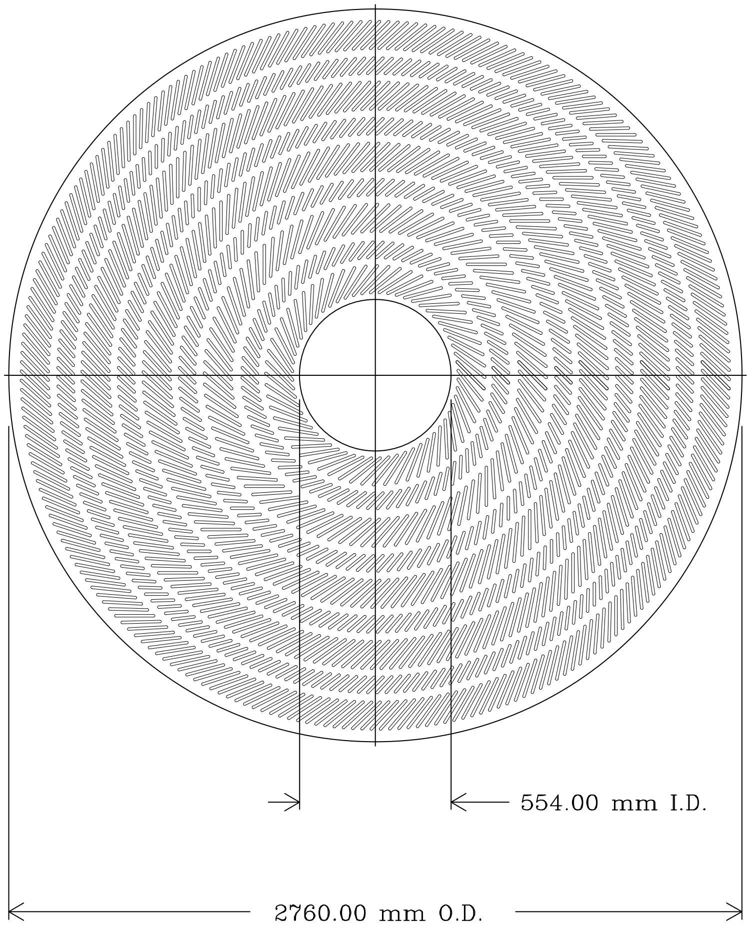


Figure 2.5: A schematic view of the end plane of the CTC.

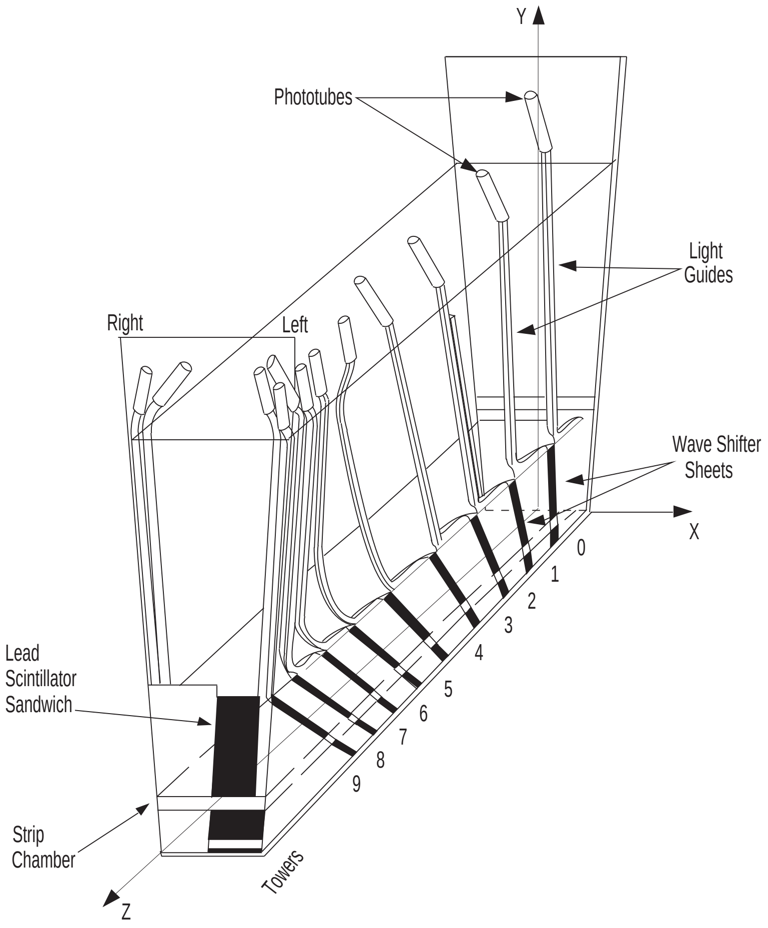


Figure 2.6: Schematic view of one module of the central calorimeter [43].

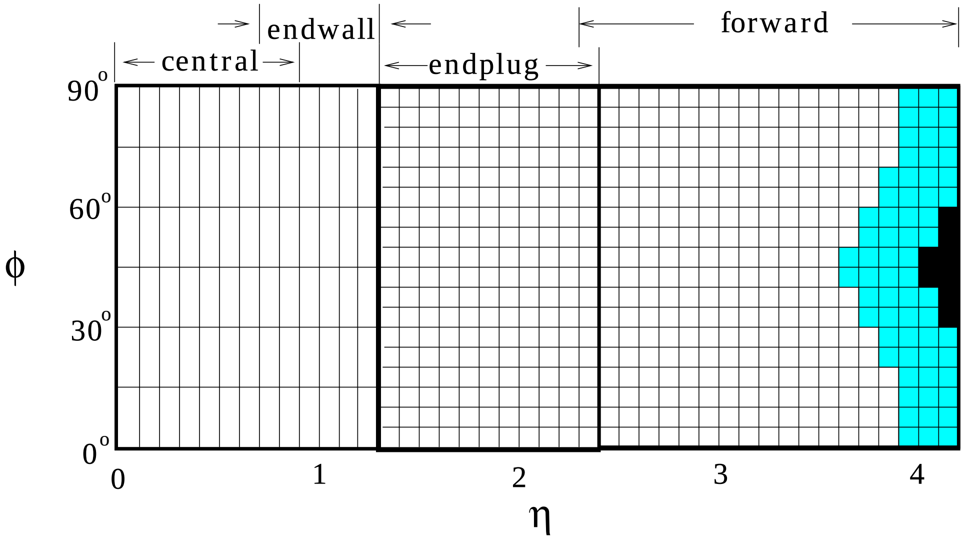


Figure 2.7: The segmentation in $\eta - \phi$ of the calorimeters. The shaded area in the forward region is only covered by electromagnetic calorimeters. The blackened area has no coverage [43].

Hadronic Calorimeters

The hadronic calorimeters in the central region consist of alternating steel plates and plastic scintillator. The end and forward regions are proportional tube chambers with cathode pad readouts. The central region of the hadronic calorimeter is defined slightly differently from the electromagnetic with $|\eta| \leq 1.3$. The end & forward regions are defined the same as for the electromagnetic calorimeters. The energy resolution for the hadronic calorimetry is $\sigma/E = 11\%$ at 50 *GeV* in the central region to $\sigma/E =$(14-20)% at 50 *GeV* in the end and forward regions.

2.2.3 Muon Detection

Due to their high mass ($m_\mu >> m_e$) muons are essentially minimum ionizing in the calorimeters. *CDF* has two systems in the central region for detecting muons which

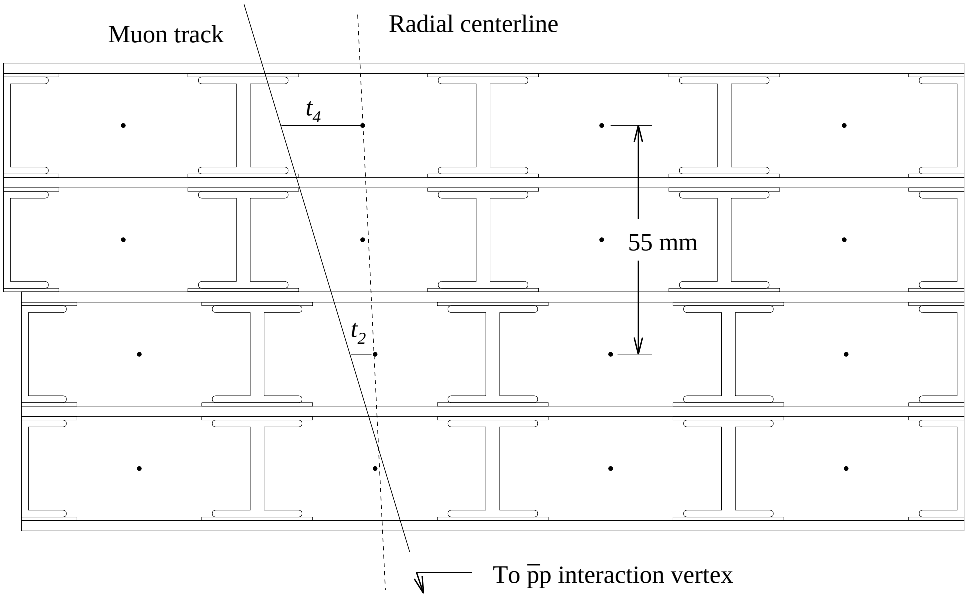


Figure 2.8: Schematic view of the CMU [43].

penetrate the calorimetry. Directly outside the central calorimeters is the central muon detector (CMU). The CMU covers $|\eta| \leq 0.65$ and consists of four layers of drift chambers, Figure 2.8. Momentum is determined by comparison of drift times in the different layers. It is possible to measure the muons ϕ and z by time to distance relation and charge division. The momentum resolution of the CMU is $\sigma P_T/{P_T}^2 = 0.002(GeV/c)^{-1}$. If the muon has a $P_T \geq 1.4\ GeV/c$ it will penetrate the CMU.

Outside the CMU is a steel absorber 0.6 m thick and outside this is the central muon detector upgrade(CMP). The CMP also has four layers. The position resolution in ϕ is 250 μm and in z is 1.2 mm. The system is essentially 100% efficient for fiducial muons [43]. This dual muon system helps veto pions which might punch through the CHA. CMU muons with confirmation in the CMP are $\approx$ 95% pure.

The CMX provides additional coverage in the region of $0.65 \leq |\eta| \leq 1.0$. All three components(CMU,CMP and CMX) are shown in Figure 2.3.

2.2.4 Beam-Beam Counters

A plane of scintillator counters are on the face of each of the forward and backward shower counters, Figure 2.3. These scintillators are called beam-beam counters(BBC) and provide a “minimum bias” trigger for the detector in addition to a luminosity measure, Figure 2.10. The timing on these counters is $\sigma < 200\ ps$, which gives the best time of interaction measurement. The timing difference also gives a $\pm 4cm$ vertex measure.

It is crucial to know the total integrated luminosity of the data taken. In $p\bar{p}$ reactions instantaneous luminosity is given by,

$$\mathcal{L} = \frac{N_p N_{\bar{p}} N_B f}{4\pi\sigma^2}, \tag{2.1}$$

where N_p & $N_{\bar{p}}$ are the number of protons & anti-protons per bunch. In this formula N_B is the number of bunches of each type, f is the revolution frequency of the collider, and the $4\pi\sigma^2$ is the cross sectional area per bunch. The integrated luminosity over Run I is shown in Figure 2.10. The total integrated luminosity in Run IB was $86.3\, pb^{-1}$ [93]. This is the data sample used in this analysis.

2.2.5 Data Taking

The *CDF* detector has approximately 150,000 electronic channels to be read out. Beam crossings occur every 3.5 μs with one to three collisions per crossing, for an event rate of ≈ 50-75 kHz at the original design luminosity of $10^{30} cm^{-2} s^{-1}$. The luminosity increased throughout Run 1 to $10^{31} cm^{-2} s^{-1}$. Not all these events can be written to tape, most events in any case are likely not to be of great interest. Because of this several layers of selection (trigger levels) are applied prior to events being written. *CDF* utilizes a three level trigger system to bring the number of events down to a manageable size. The sooner events are discarded the less deadtime there is, and the more events which can be readout.

Level 1

The level 1 trigger relies on fast output from preamplifiers to quickly make the initial decision to keep or throw away an event. This decision is made before the next beam crossing, in less than $3.5\mu s$, therefore operation of level 1 triggers are dead timeless. Dead time is incurred if the trigger takes more time to analyze an event than it has between beam crossings. Triggers at level 1 which are relevant to top quark searches are calorimeter and central muon triggers. The calorimeter trigger relies on the E_T, it is summed from all calorimeter towers which individually pass a low(1 GeV) threshold cut were the sum must be over some high threshold. The three muon triggers use fast outputs from the CMU,CMP, and CMX. At a luminosity of $10^{31} cm^{-2} s^{-1}$ the trigger rate at level 1 is $\approx 1\ kHz$.

Level 2

Events passing level 1 triggers are sent to the level 2 triggers. During this stage information is collected from the tracking, calorimetry, and muon systems. The central fast tracker(CFT) using fast timing information from the CTC determines if tracks are present in the CTC with an efficiency of $\approx 94\%$ for high P_T tracks ($P_T > 10\ GeV/c$). Dedicated hardware gathers information from the calorimeter towers, this information is examined for clusters of towers containing events above some threshold of energy which would tend to imply the presence of a jet, electron, or photon. The CMU, CMP, and CMX provide information which can be matched with CFT tracks to provide indication of the presence of muon candidates. How long this level 2 process takes varies according to the complexity of the event. Simple events typically take 20 to 30 μs but more complex events, especially those with a large number of tracks, can take hundreds of microseconds. At a luminosity of $10^{30} cm^{-2} s^{-1}$ the trigger rate of level 2 was 20 Hz with approximately 4% dead time.

Level 3

The final trigger level, level 3, is entirely software based. Eight Silicon Graphics multi-cpu PowerServers are used to run this software trigger. At this level a full event reconstruction has been done on each event. At level 3 electrons and muons have been fully reconstructed, the jets have been clustered, and the missing transverse energy($\not{E}_T$) has been calculated. Once again the majority of the time required at this level is the reconstruction of tracks. At the final operating luminosity of $10^{31} cm^{-2} s^{-1}$ the trigger rate of 5 Hz with approximately 10% dead time. These events are written to tape.

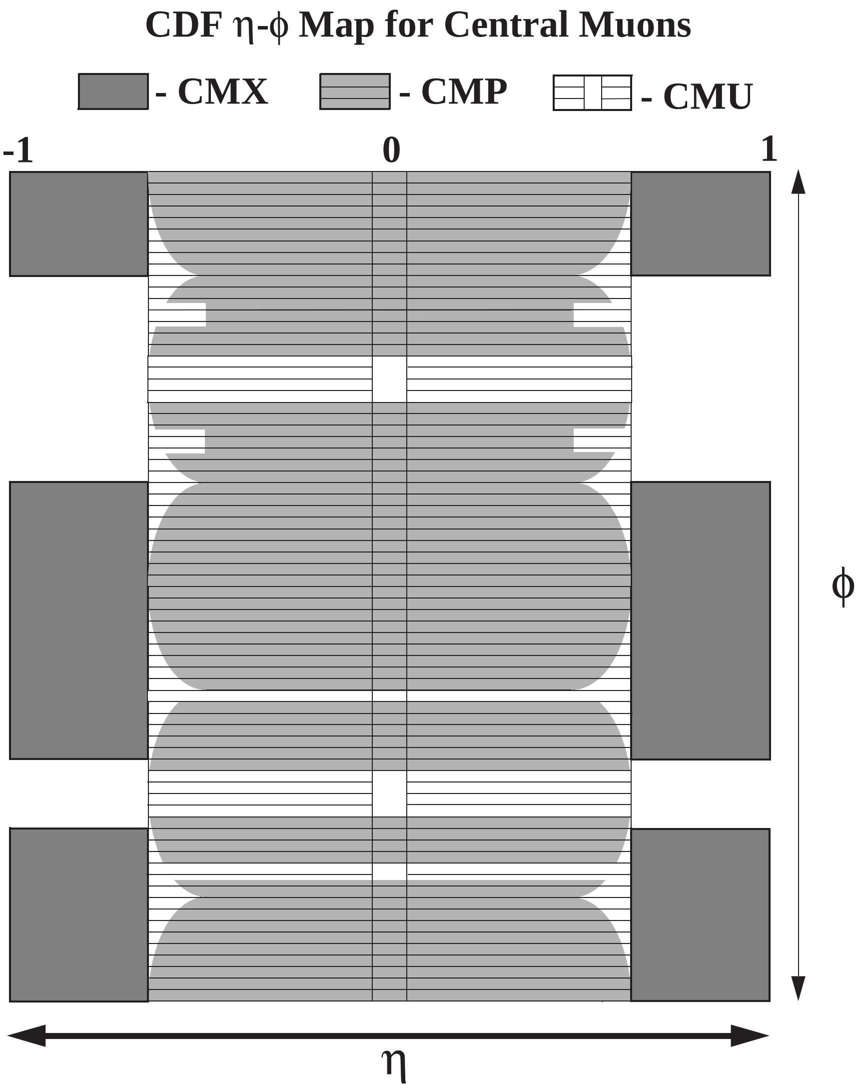


Figure 2.9: The segmentation in $\eta - \phi$ of the central muon systems [43].

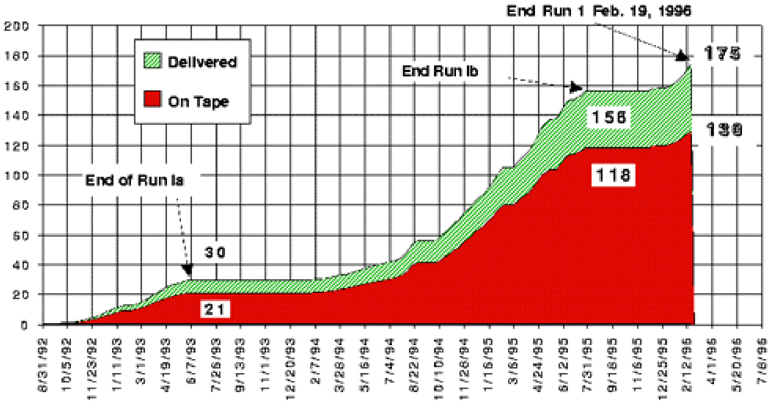


Figure 2.10: Integrated luminosity over the course of Run I [43].

Chapter 3

Neural Networks

3.1 The Brain

Until a century ago very little was known of the internal structure of the human brain, although the central nervous system has been studied for centuries. In fact the electrical nature of nervous impulses was first studied by Luigi Galvani in the 1780s. The brain was first shown to be composed of individual “neural” cells (*neurons*) in 1888 by Santiago Ramon y Cajal, using a technique developed by Camillo Golgi. The two shared the 1906 Nobel prize in medicine for this work. Over the last half century a great deal of work has been done on the structure of these *neurons* so that today quite a bit is known about their structure and behavior.

A single *neuron* has a fairly well known physical structure. These neurons have a central cell body known as the *soma*. From the *soma* many hair-like structures project radially, these are the *dendrites*. Also coming from the *soma* is a single larger fiber known as the *axon*, this tubular structure divides into branches at the other end which terminate in end bulbs. These bulbs are in close proximity, typically 200 *nm*, to the *dendrites* of other neurons. The gap between the *dendrites* and the bulbs is called the *synapse*, Figure 3.1 shows the the basic structure of a neuron.

A neuron in the “resting state” has certain concentrations of ions both inside and outside the cellular membrane. The most common ions involved in a cell are K^+,

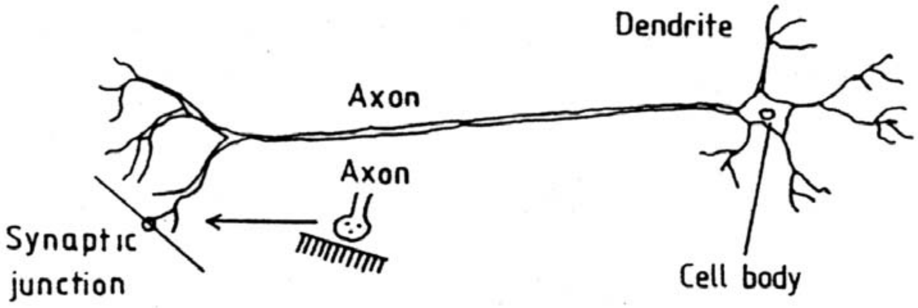


Figure 3.1: Schematic veiw of a neural cell.

Na^+, and Cl^-. In its rest state the axon of the neuron is neutral along its entire length. Although the fluid, inside and outside the axon, is electrically neutral the concentrations of various ions may be substantially different inside and outside the axon. The Na^+ concentration inside the axon is typically 15 mol/m^3, the concentration outside is typically 140 mol/m^3. This means that there exists great potential for a diffusion response across the cellular membrane. In the rest state the cell membrane is largely impermeable to the ions until the axon is stimulated. At the point of stimulation the cell membrane of the axon becomes permeable and charged ions flow across the cell membrane from neighboring parts of the axon. In this way the neighboring point on the axon is stimulated and that region becomes permeable, in this process a signal flows down the length of the cell body. This signal moves along the axon at a speed of 30 m/s to 150 m/s.

Once a signal has reached the nerve endings it triggers a neuro-chemical response at the synapses. These signals trigger a response in the dendrites of a neighboring neuron, the post-synaptic cell. The signal is transmitted along the dendrites to the soma. It is here at the soma that all the signals from the dendrites are combined and if the signals coming in exceed some threshold value the neuron “fires” and a signal is sent down the axon of this post-synaptic cell. Exactly how this combining of the signals from different dendrites takes place in the brain is not well understood, especially how the timing of signals from various dendrites affect the process. Fortunately the process for an *Artificial Neural Network (ANN)* will be much simpler to describe.

The complexity of the human brain is not due to the intrinsic complexity of the individual neurons in the brain but in the large numbers of neurons present. A rodent the size of a human being would have a brain 1/30 the size of a human brain [45]. This tends to indicate that it is the over-abundance of neurons that leads to the greatly increased capacity of the human brain over that of the rodent.

The human brain has $\approx 10^{11}$ neurons. Each neuron in the brain and spinal cord has a few thousand synapses. Therefore the total number of synapses in the human brain are of the order of 10^{14}. It is neuro-chemical changes that affect the activity of synapses. These properties of the neurons reflect, at the cellular level, memory. It is therefore synaptic responses in the brain which are the physiological key to learning.

The efficiency of a particular synapse changes over time according to its activity. Repeated stimulation of the pre-synaptic neuron will result in an increased magnitude of the response in the post-synaptic neuron. This strengthens that pathway and weakens surrounding pathways. This mechanism is known as *Hebb's rule* and is thought to play the dominate role in the physiology of learning.[46]

3.2 Artificial Neural Networks

The expression *Artificial Neural Network* refers to a broad class of mathematical models which utilize a distributed architecture. Such architectures have multiple parallel processing centers, nodes, instead of a single linear processing stream. These models are similar, and inspired by, the neural model of the brain just described. The term neural network (w/o the artificial) is usually reserved for an actual brain. However unless specified the terms *ANN*, neural network, and network will all be used interchangeable with and in reference to an Artificial Neural Network.

Although there are many different types of *ANNs* they may be placed into two broad categories, "supervised" and "unsupervised". The first category, supervised learning, takes place when relationships in data are determined by comparison to some established "parameter domain". This parameter domain is typically a data set where the desired relationships are already known, using this information a network may be "trained" to find these relationships in other similar data sets. By contrast, unsupervised learning is when a network discovers "natural" groupings in the data without any constraints. Both supervised and unsupervised learning takes place through intrinsic relationships in the data and not by a predetermined set of rules.

One example of supervised learning is given by a *feed forward multi-layered perceptron* (MLP) network. The word perceptron is historical having to do with the workings of the nodes in a network. Each grouping of nodes is referred to as a layer. The term feed-forward indicates that each node in the network connects only to nodes in other layers and not neighboring nodes in the same layer. The most important consequence of this is to make recurring loops impossible. Although there are many other types of supervised networks this is one of the most important types of networks for *high-energy particle physics* and is the one used in this analysis.

3.3 Why use a Neural Network ?

A hardware *ANN* in a neural computer could contain billions of elements; but hundreds of billions (10^{11}) are in a human brain. The complexity of neural interconnections in the brain is beyond our current understanding, a single neuron in the human brain may have well in excess of 1000 synaptic connections. The selection of synaptic strengths is very crude in artificial neural models and much work on this learning process is needed before it comes close to the subtlety of even the simplest brain. Despite the crudeness of artificial neural networks they can be of considerable value in simple pattern recognition.

3.4 Neural Network Applications

Artificial Intelligence & Artificial Neural Networks

Artificial Intelligence (*AI*) also known as expert systems are sometimes applied to similar situation as *ANN*s. Expert systems are very different from artificial neural networks. Although expert systems have become chic especially in commercial applications they are substantially different from *ANNs* in the way they deal with problem solving. Expert systems essentially solve problems through a logical deconstruction of the problem to be solved. In essence an expert system must know the "rules" by which a problem is solved, this is certainly not the case for *ANNs*. A neural network does not have a set of logical instructions on how to solve the problem at hand, instead it learns how to recognize a broad class of inputs that would be associated with a given output.

3.4.1 Biomedical

Although expert systems have been applied to medical diagnostics with limited success, many problems in medicine appear to be more appropriate to a neural network approach. Medicine and biology in general are not based on well established "simple" rules which can be applied to a wide variety of problems, as is common in physics, but on the observers ability to classify a complex system where the details of the underlying dynamics may be poorly understood. In medicine the underlying biological dynamics for a system may be far too complex to be qualified and even if those dynamics can be well understood the complexity may lead to chaotic behavior which renders this knowledge useless to a medical clinician. This situation would seem to be ideally suited to the application of an artificial neural network.

Pap Smears

An interesting application to medicine is to the diagnostic testing for cervical cancer, the Pap smear. A method of diagnosis involving microscopic screening of mucus smears (developed by G. N. Papanicolaou), this valuable diagnostic tool involves the examination of a sample of $\approx 10^5$ cells which may have only a small percentage of abnormal cells. A neural network (Papnet) has been developed by Neuromedical Systems Inc to automatically examine cells for possible malignancies. Papnet is trained on a sample of inputs for healthy and malignant cells. Possibly malignant cells are tagged for examination by a pathologist. This changes the problem for the human examiner from a search for a few hundred abnormal cells in $100,000$ to a few hundred abnormal cells in perhaps $1,000$ thereby greatly reducing the probability of producing a false negative result.

Nuclear Magnetic Resonance Imaging

One of the most important applications of *ANNs* in medicine is in one of its most traditional roles, simple pattern recognition. Medical image processing is an area in which neural networks have excelled. Nuclear Magnetic Resonance Imaging (MRI) has become an invaluable tool in medical diagnostics. Interpretation of the images created during a MRI have always been an issue of some difficulty. Neural networks provide a relatively quick and accurate way to identify various abnormalities in such images, in addition to simple classification of varying types of tissue. It has been demonstrated that a neural network approach provides better delineation between regions of different tissue than similar Bayesian approaches [45]. Neural networks can also be applied to other types of image and signal processing situations like, x-ray radiography, position emission tomography (PET), ultrasound, electro-cardiogram (ECG), etc.

Neural networks have also been applied, with varying degrees of success, to general medical decision making and diagnosis. Although most such tools currently utilize a linear computing model, expert system, some attempts to apply *ANNs* have been made [47, 48]. An *ANN* has the advantage over the *AI* approach in that medical diagnostic decisions are frequently made with poor or little information and a physician may have to suggest a course of treatment based on partial or vague symptoms. This is the type of situation in which a neural network can demonstrate its true strength, the ability to make good educated guesses. The neural network is not deterred from making an assessment of probability based on partial information, the operator must of course always take this into account. The diagnostic *ANN* has at least one major advantage over a physician, an indifference to likelihoods. The *ANN* will not discount a symptoms importance simply because flu is more likely than Hanta virus.

3.4.2 Economics

There have been many attempts to apply neural networks and expert systems to financial problems. Subjects as diverse as predicting when a company might go bankrupt [49] to forecasting the GDP of Singapore [50] have been investigated using neural networks. Unfortunately an in-depth examination of the results of these endeavors as a whole is made very difficult due to the proprietary nature of most of this research. As is frequently the case with privately funded research, results of that research is often not shared. The result of this is very much the same as happens so often in biology; information that may be critical to the progress of other researchers is unavailable, thereby stifling progress.

Despite the frigid intellectual environment which tends to permeate economic research on such topics, it is clear that some substantial individual efforts have been made to investigate the usefulness of neural networks in economic prediction. A good example of this would be prediction of stock prices. One such study has been done by Y. Yoon and G. Swales [51], utilizing information such as market valuation and stock price appreciation in addition to qualitative information gathered from company presidents annual letters to stockholders. This qualitative information involves examining the letters content for themes like confidence, growth, anticipated losses/gains, new products, etc.. The data was then set by the researchers to reflect these themes. This kind of content analysis method is not uncommon in traditional social science research. The network constructed is essentially a four layer feed forward network with a buy/sell output. Their performance was 77.5 percent accurate as compared with 65 percent for a multiple discriminant analyses.

3.4.3 Pattern Recognition

Perhaps the most traditional application of neural networks is in pattern recognition. Many other problems addressed by *ANNs* are in fact only a certain type of pattern recognition problem. It could be argued that every neural network problem is simply a pattern recognition problem. This may in fact apply to most human thought. The analysis described in this thesis is one good example of a pattern recognition problem. The prototypical problem in pattern recognition is automated reading, or handwriting recognition. A symbol which might be easily recognized by a person like a script '3' may be very difficult for a machine to read.

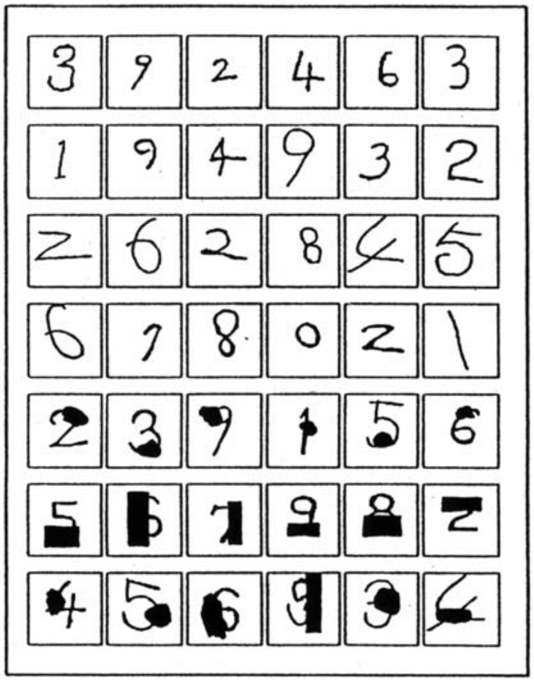


Figure 3.2: Various examples of test patterns used in network training for handwriting recognition. Figure taken from reference [52].

As a result of the huge amount of information processed by machine a great amount of effort has been put into automated reading technologies. *ANNs* have been of great use in these systems (often hybrid) *ANN* & *AI* systems. The postal services of almost every developed country now uses automated sorting and pattern recognition systems to facilitate quick delivery. A friend of mine on recently moving back to Austria found that a third of his boxes had been sent to Australia, this he attributed to the stupidity of someone working for the postal office. He was probably not quite right. It is more likely due to the stupidity of *something* working at the post office and not *someone*.

3.4.4 HEP Applications

Neural networks have been applied to a number of problems in High Energy Physics (HEP). Artificial neural networks have been employed in HEP for applications as diverse as simple triggering systems and complete offline double beta decay analyses.

Triggering

Artificial neural networks have been employed in preprocessing of data as a trigger system for jets in a number of different experiments. In at least some of theses cases efficiencies were improved [53]. A third level trigger has been investigated to identify electrons & photons at *CDF*.[54]

Mass Reconstruction

A feed-forward network was employed to reconstruct the invariant mass of the W-boson. Twenty-four variables are used as inputs to the *ANN* with substantial preprocessing of data (“intelligent” variables). The performance of this network is shown to be superior to conventional methods [55].

B-tagging

B-tagging using neural networks in Z^o decays has been examined at DELPHI [56]. In this analysis very simple variables are used as inputs, as contrasted with the use of intelligent variables in the mass reconstruction analysis. This analysis also examined the multivariate methods so that that result may be compared with the neural network approach. In order to compare the two approaches two quantities, purity and efficiency are defined.

$$b - efficiency(\epsilon_b) \equiv \frac{\#\ of\ correctly\ tagged\ b - quarks}{total\ \ of\ b - quarks}$$

$$b - purity(p_b) \equiv \frac{\#\ of\ correctly\ tagged\ b - quarks}{total\ \ of\ tagged\ (correct\ \&\ incorrect)\ b - quarks}$$

[45] The neural network approach give consistently better results than the various multivariate methods, Figure 3.3.

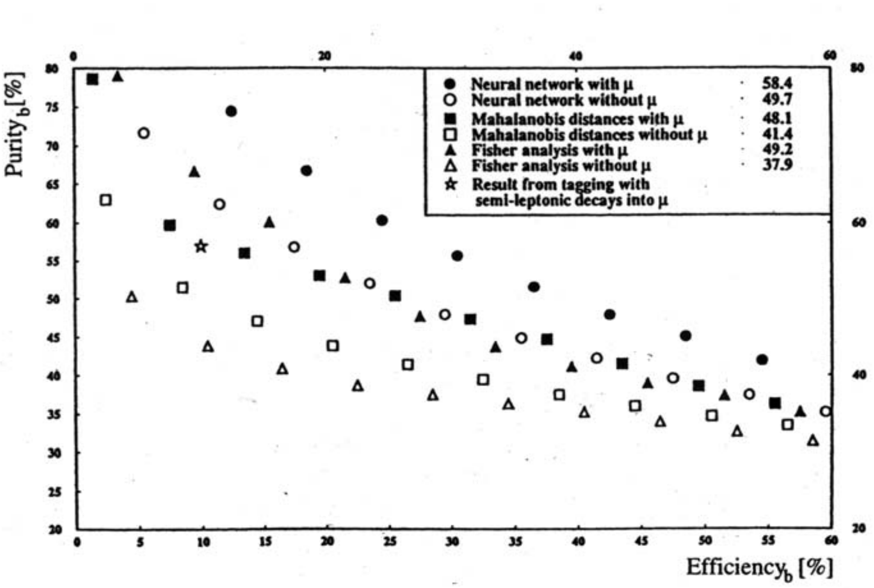


Figure 3.3: Comparison of neural network methods to various classical methods of event tagging. Figure taken from reference [45].

There has even been a $t\bar{t}$ production cross section measurement made in the DØ Collaboration at FermiLab.[33] This measurement used few intelligent variables in the $e\mu$ dilepton channel. A simple multivariate comparison indicated that the neural network cut used in that analysis was more efficient. Another measurement at *CDF* was examined in the all-hadronic channel.[60]

Chapter 4

Kinematics & Topology

4.1 Detector Definitions

In this chapter we describe the kinematic variables used as input parameters to our *ANN* . These variables shall make up our input pattern and represent all the information which the neural network will have about the events presented to it. Many of the variables examined where selected for possible inclusion due to the studies of Andy Beretvas and Morris Binkley [57] and/or due to DØ studies [82]. A complete list of the variables used is presented at the end of the chapter, Table 4.1.

Before we begin describing the variables, we first examine some basic definitions relevant specifically to *CDF* and some of a more general nature.

- The coordinate system used at *CDF*is cylindrical. The 'z' coordinate is along the beamline where positive 'z' is is the direction traveled by the proton bunches. The 'θ' is the usual polar angle and the 'ϕ' is the azimuthal angle.

- Due to the partonic structure of the proton it is convenient to define a quantity pseudo-rapidity (η),

$$\eta \equiv -ln\left[tan\left(\frac{\theta}{2}\right)\right] \ .$$

- A cone radius or general distance in η-ϕ space is

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

The coordinates above make it easy to define transverse quantities. These are important because in hadronic collisions the center of mass of the underlying partonic hard scattering process is randomly boosted along z, ΔR is invariant with regard to boosts along z.

- The transverse momentum is the momentum perpendicular to the beam-line,

  $$P_T = P sin\theta \ ,$$

  where P is the total particle momenta. ($\theta = 0$ would correspond to the beamline in the proton direction.)

- The transverse energy is

  $$E_T = E sin\theta \ .$$

  In practice the P_T and E_T are very close to being equivalent when the energy of the particle is much greater than the rest mass.

- The transverse missing energy, $\not{E}_T$, is simply the transverse energy not accounted for when all known energies perpendicular to the beam are balanced.

- The longitudinal missing energy, $\not{E}_z$, will also be used but is not directly measured and is complicated to determine, that discussion is left to Section 4.4.

4.2 "Simple" Variables

The use of *Artificial Neural Networks* for the study of heavy quark identification has been extensively studied [59, 56]. Some of these analyses employed traditional "intelligent" variables (*Ex:* Aplanarity) and others utilized large numbers of "simple" variables (*Ex:* Jet P_T). The top quark itself has been searched for using intelligent variables as input pattern to an *ANN* by the DØ Collaboration [82]. Many of the intelligent kinematic variables which may be of value in $t\bar{t}$ searches have been carefully examined for their ability to separate $t\bar{t}$ from its primary background of $W + jets$ [57]. The detector definitions above are more than enough to determine many of the variables used in this analysis. All the "simple" variables such as,

$$P_T(lepton),\ P_T(jet),\ \not{E}_T,\ \ < \eta^2 >, \ldots,$$

can be understood with no further explanation.

4.3 "Intelligent" Variables

Variables like aplanarity ($\mathcal{A}$), sphericity ($\mathcal{S}$) and thrust describe the shape of an event. These variables were developed for e^+e^- collisions to study jet events. Much of the information contained in these variables can be summarized in terms of a normalized momentum tensor (T_{ab}),

$$T_{ab} = \frac{\sum_i^{Njets} P_a(i) P_b(i)}{\sum_i^{Njets} |P(i)|^2} \ ,$$

where a,b run over three spacial dimensions and $P(i)$ is the momentum of the i^{th} particle. T_{ab} is a symmetric matrix and can therefore be diagonalized. These eigenvalues are normalized to

$$Q_1 + Q_2 + Q_3 = 1 \ ,$$

where their ordering is

$$0 \leq Q_1 \leq Q_2 \leq Q_3 \ .$$

Two combinations of these eigenvalues are used in this analysis.

It is possible that a non-normalized matrix similar to T_{ab} would be useful. We did not investigate such a matrix here but a future analysis might consider the use of such a matrix [60].

Aplanarity

Aplanarity ($\mathcal{A}$) is defined in terms of these eigenvectors and eigenvalues of T_{ab}. The eigenvectors $\vec{n}_2$ and $\vec{n}_3$ define the event plane and $\mathcal{A}$ is closely related to the momentum out of this plane and the eigenvalue associated with the eigenvector out of that plane $\vec{n}_1$,

$$\mathcal{A} \equiv \frac{3}{2} Q_1 \ .$$

Usually this quantity will be normalized to lie between values of 0 and $\frac{1}{2}$, however this normalization will be changed before being input into our network as described in Section 4.5. If $\mathcal{A}$ is small the event is very coplanar (pancake shaped) or collinear (cigar shaped). This is equivalent to a ratio of the total momentum out of the plane of greatest momentum to the total momentum in that plane,[4]

$$\mathcal{A} = \frac{3}{2} \frac{\sum P^2(out)}{\sum P^2(in\ plane)} \ .$$

The aplanarity uses the five highest E_T jets passing low-threshold cuts and the 3-momenta of the leptonic W.

Sphericity

The sphericity (S) is defined in terms of the first and second eigenvalues of T_{ab},

$$\mathcal{S} \equiv \frac{3}{2}(Q_1 + Q_2) \ .$$

Sphericity lies in a range of 0 to 1, though we scale it before being input into the *ANN* according to the procedure described in Section 4.5. This is equivalent to a ratio of the total transverse momentum (in plane) to longitudinal momentum (out of plane), however in this case the plane is the plane of minimum momentum,[4]

$$\mathcal{S} = \frac{3}{2} \frac{\sum_i P_i^2(in\ plane)}{\sum_i P_i^2} \ .$$

As $\mathcal{S} \rightarrow 1$ the event becomes more spherical and events with $\mathcal{S} << 1$ are cigar shaped.

Summed Transverse Energy

The total energy in a $t\bar{t}$ production event with $m_t = 175\ GeV/c^2$ tends to be higher than background events. Even when all our cuts are applied the $t\bar{t}$ events will tend to be significantly higher above the cut thresholds for total transverse energy (H_T) therefore this quantity,

$$H_T = P_T(lepton) + \not{E}_T + \sum_i^N jetsP_T(i)\ ,$$

proves to provide good discrimination between signal and background.

Separation in η-ϕ Space

Every pair of jets in a given event has some separation in η-ϕ space (ΔR_{jj}). We find the minimum separation between jets in η-ϕ space (ΔR_{jj}^{min}), considering all the jets passing low-threshold cuts. This quantity, ΔR_{jj}^{min}, is used to define a variable,

$$\Delta R_{jj}^{min} P_T^{min}/P_T(l) = \frac{\Delta R_{jj}^{min} P_T^{min}}{P_T(lepton)}\ ,$$

where P_T^{min} is the P_T of the lower P_T jet in the pair used in ΔR_{jj}^{min}.

It is not possible for two jets to be less than $\Delta R = 0.4$ apart, two such jets would be considered a single jet. If two jets are less than $\Delta R = 0.7$ they will be merged approximately half the time.[3]

Minimum Dijet Mass

The minimum dijet mass utilizes the leading four passing jets. For each of the jet pairs the dijet mass is,

$$\begin{aligned} dijet\ mass &= \sqrt{(\tilde{p}_i + \tilde{p}_j)^2} \\ &= \sqrt{(E_i + E_j)^2 - (\vec{p}_i + \vec{p}_j)^2}\ . \end{aligned}$$

The $MinMass(jj)$ then is just the minimum dijet mass for the four jet system. Given the harder nature of the jets in signal events the $MinMass(jj)$ will tend to be higher for these events.

Longitudinal Momentum : Transverse Momentum

The ratio of the longitudinal momentum to the transverse momentum $(\sum P_z/\sum P_T)$ is constructed using the leptonic W and the four leading jets. The leptonic W is constructed by vector sum of the lepton and the reconstructed neutrino. Then the final sum is a scalar sum of the W and the jets. The neutrino reconstruction is described in the next section.

4.4 Reconstruction of the Leptonic W

Although the transverse and longitudinal components of the lepton and the jets are measured in the *CDF* detector, reconstruction of the missing energy is only possible in the transverse direction ($\not{E}_T$). The problem with measuring the missing longitudinal energy ($\not{E}_z$) is that the initial momentum in $'z'$ is not well known due to the lack of information about parton momentum fractions. If the $\not{E}_z$ is to be used it must be reconstructed with some additional knowledge.

Fortunately we do have knowledge about the missing energy in a $t\bar{t}$ event. We assume that most of the missing energy ($\not{E}$) in a $t\bar{t}$ event is due to the neutrino from the leptonically decaying W boson decay. This effect may be mimicked by large energy losses in detector cracks. In the *lepton+jets* channel only one neutrino (ν) is present from the leptonically decaying W, see Figure 1.16. Since only one neutrino is present and we know the mass of the originating particle (W) the third component of the neutrino momentum can be reconstructed by,

$$(\tilde{P}_l + \tilde{P}_\nu)^2 = M_W^2, \tag{4.1}$$

where $\tilde{P}_l$ is the 4-momenta of the lepton, $\tilde{P}_\nu$ is the 4-momenta of the neutrino, and M_W is the known mass of the W-boson. This mass is given by the Particle Data Book to be $\approx 80.4\ GeV/c^2$ [32].

$$\tilde{P}_l^2 + \tilde{P}_\nu^2 + 2\tilde{P}_l\tilde{P}_\nu = M_W^2$$

$$m_l^2 + m_\nu^2 + 2\tilde{P}_l\tilde{P}_\nu = M_W^2$$

The mass of the neutrino is zero or approximately zero (m_ν). And at the W mass scale the lepton mass is $m_l^2 \approx 0$ ($M_W >> m_l$). Therefore,

$$2\tilde{P}_l\tilde{P}_\nu = 2(E_lE_\nu - \vec{P}_l \cdot \vec{P}_\nu) = M_W^2\ . \tag{4.2}$$

The neutrino energy (E_ν) contains the 'z' component of the neutrino ($\not{P}_z$). Therefore it is expanded,

$$E_\nu^2 = \vec{P}_\nu^{\,2} + m_\nu^2 \Rightarrow E_\nu = \sqrt{\not{P}_T^{\,2} + \not{P}_z^{\,2} + m_\nu^2} \tag{4.3}$$

Equation (4.2) may be included in Equation (4.3) to yield

$$E_l\left(\not{P}_T^{\,2} + \not{P}_z^{\,2}\right)^{1/2} = \frac{M_W^2}{2} + \vec{P}_T \cdot \vec{\not{P}}_T + P_z\not{P}_z\ .$$

Squaring and expanding this equation, and then collecting terms results in a quadratic equation,

$$\left(E_l^2 - P_z^2\right)\not{P}_z^{\,2} - 2\left(\frac{M_W^2}{2}P_z + \vec{P}_T \cdot \vec{\not{P}}_T P_z\right)\not{P}_z - \left(\frac{M_W^4}{4} + \vec{P}_T \cdot \vec{\not{P}}_T M_W^2 - E_l^2\not{P}_T^{\,2} + (\vec{P}_T \cdot \vec{\not{P}}_T)^2\right) = 0\,,$$

therefore

$$\not{P}_z = \frac{-P_z\left(2\vec{P}_T \cdot \vec{\not{P}}_T + M_W^2\right) \pm \sqrt{4E_l^2\left(\frac{M_W^4}{4} + \vec{P}_T \cdot \vec{\not{P}}_T M_W^2 - E_l^2\not{P}_T^{\,2} + \not{P}_T^{\,2}P_z^2\right)}}{2\left(P_z^2 - E_l^2\right)}\ .$$

This gives two possible solutions for $\not{P}_z$ and can be complex. If the values are not real we set the value of the radical portion of the equation to zero. Once real solutions are produced the smaller solution is chosen for the use in our kinematic variables.[57]

By using $t\bar{t}$ Monte Carlo and examining generator level variables we can make an association between the generator level neutrino and our reconstructed solutions. In this way the proper solution of the two found can be identified. After comparing the solution selected with generator level information we can say that the efficiency on this procedure is $\approx 64\%$.

The 'z' component of the missing energy ($\not{E}_z$) appears in the following kinematics: $\sum P_z / \sum P_T$, $\mathcal{A}$, and $\mathcal{S}$.

4.5 Summary

The kinematics discussed in this chapter are summarized in Table 4.1. Distributions for each kinematic variable being considered are generated for both signal and background with selection criteria described in Section 6.2. These variables are shown, scaled, in Figures 4.5, 4.5, 4.5, 4.5.

The input variables must be initially scaled to be consistent across the entire input pattern. Our inputs have been scaled in a standard way, each input is shifted to its mean and scaled by the standard deviation of the distribution [57, 83].

How particular variables were considered and prioritized for inclusion in the input to our neural net is discussed in the next chapter.

Variable	Definition
$P_T(lepton)$	The transverse momentum of the lepton.
$\not{E}_T$	The missing transverse energy.
H_T	The scalar sum of the transverse energy in the event ($P_T(lepton) + \not{E}_T + \sum P_T(jets)$).
$\mathcal{A}$	The aplanarity using the 5 highest E_T jets passing low-threshold cuts and the 3-momenta of the leptonic W. The plane of greatest $\sum P^2$ is determined and then $\mathcal{A} = \frac{2}{3} \sum P^2(out) / \sum P^2(in\ plane)$.
$\mathcal{S}$	The sphericity is a measurement of how spherical, coplanar, or collinear is a given event.
$MinMass(jj)$	The minimum di-jet mass.
$\sum P_z / \sum P_T$	The sum of the linear momentum, $\|P_z\|$, for the leptonic W and the 4 leading jets in the event divided but the sum of the transverse momentum, P_T, for the same objects. The $\|P_z\|$ of the neutrino must be reconstructed by using the W mass. The minimum $\|P_z\|$ neutrino solution is used.
η_{max}	The η of the highest η jet in the leading four jets.
$< \eta^2 >$	The average η square of the leading four jets.
ΔR_{jj}^{min}	The minimum separation in η-ϕ space.
$\Delta R_{jj}^{min} P_T^{min} / P_T(l)$	The $\Delta R_{jj}(min)$ times the P_T of the lower P_T jet of the pair that make the $\Delta R_{jj}(min)$ divided by the P_T of the lepton.
$P_T(1)/H_T$	The P_T of the leading jet divided by the H_T.
$Njets$	The number of jets in the event passing low threshold cuts.
$P_T(i)$	The P_T of the given jet ($i = 1, 4$).
$P_T(i) + P_T(j)$	The sum of the two jets specified ($i = 1, 3$ and $j = 2, 4$).

Table 4.1: The Kolmogorov-Smirnov Statistic *(KS)* for each kinematic variable considered for inclusion in this analysis may be found in Table 5.1. Although all of these variables were examined only those in Figures 4.1, 4.2, 4.3, 4.4 were included in the final *ANN*.

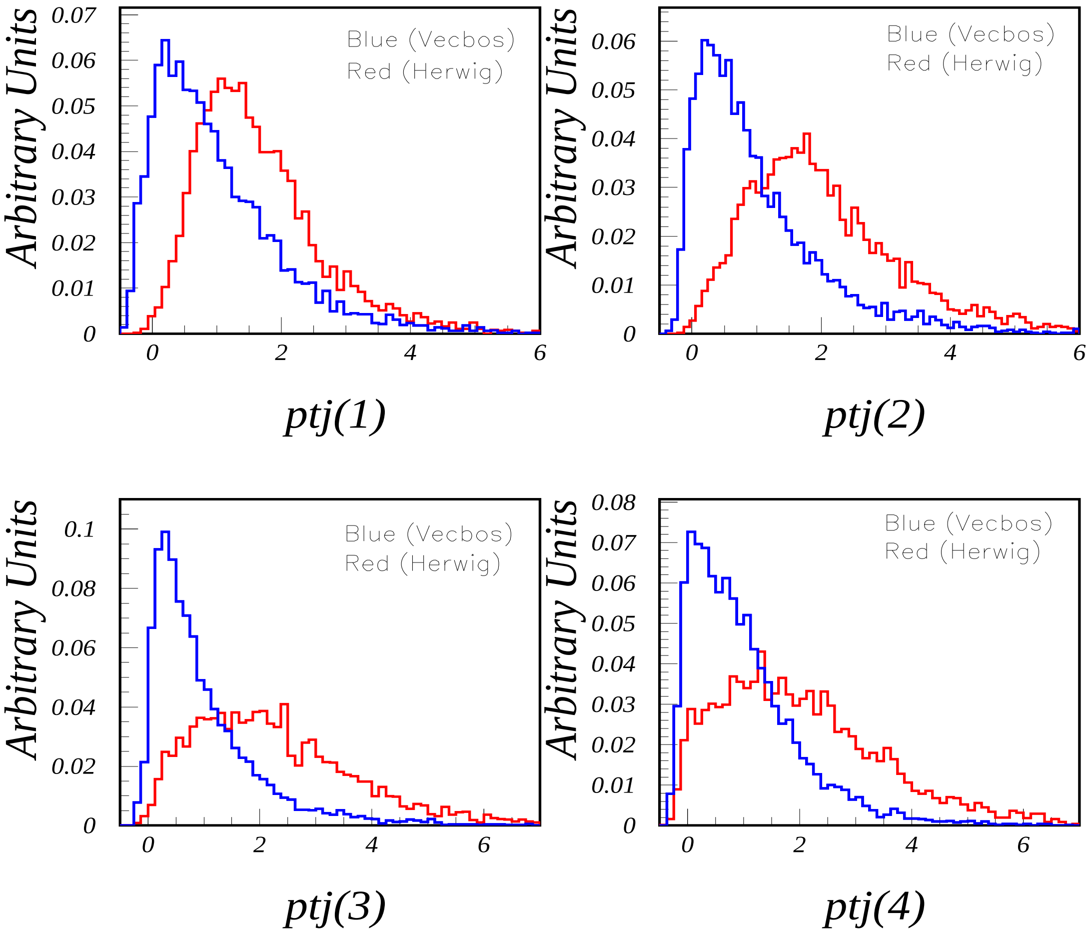


Figure 4.1: Input distributions used in the training of our 16 node ANN.

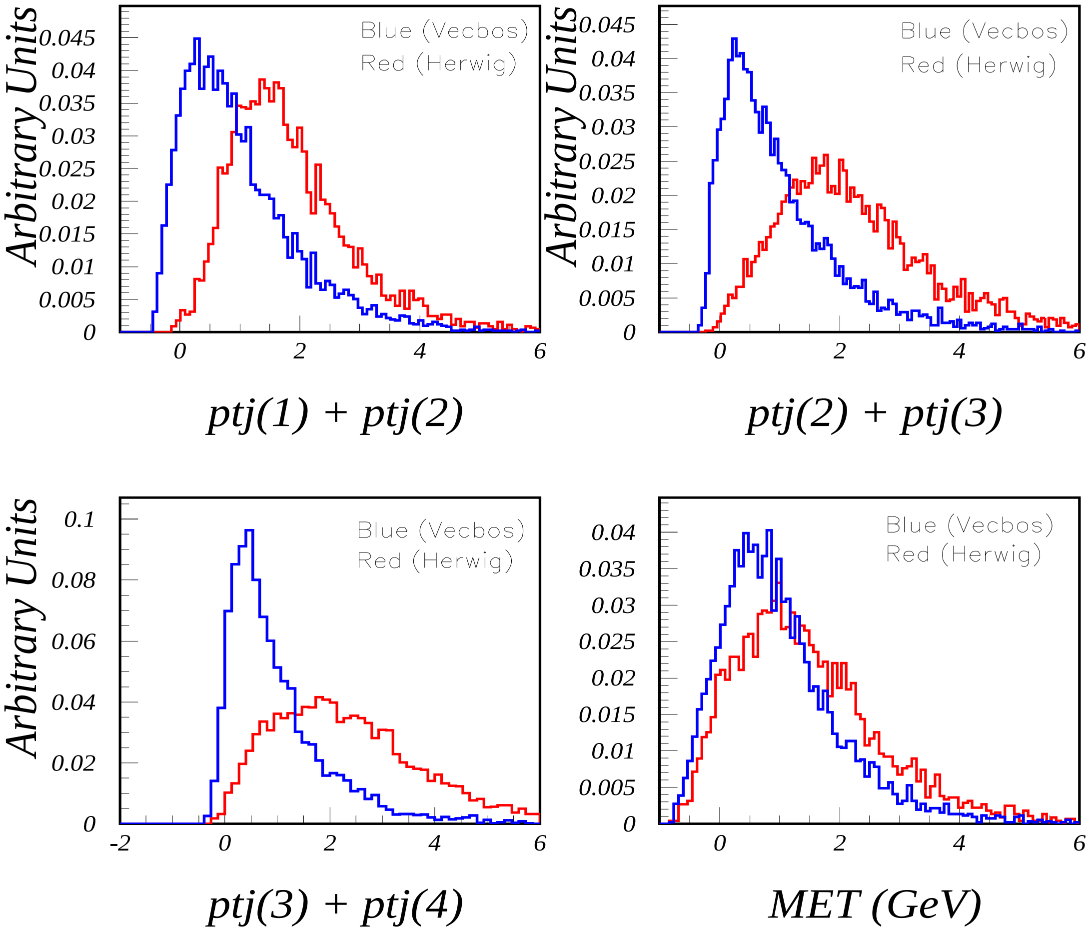


Figure 4.2: Input distributions used in the training of our 16 node ANN.

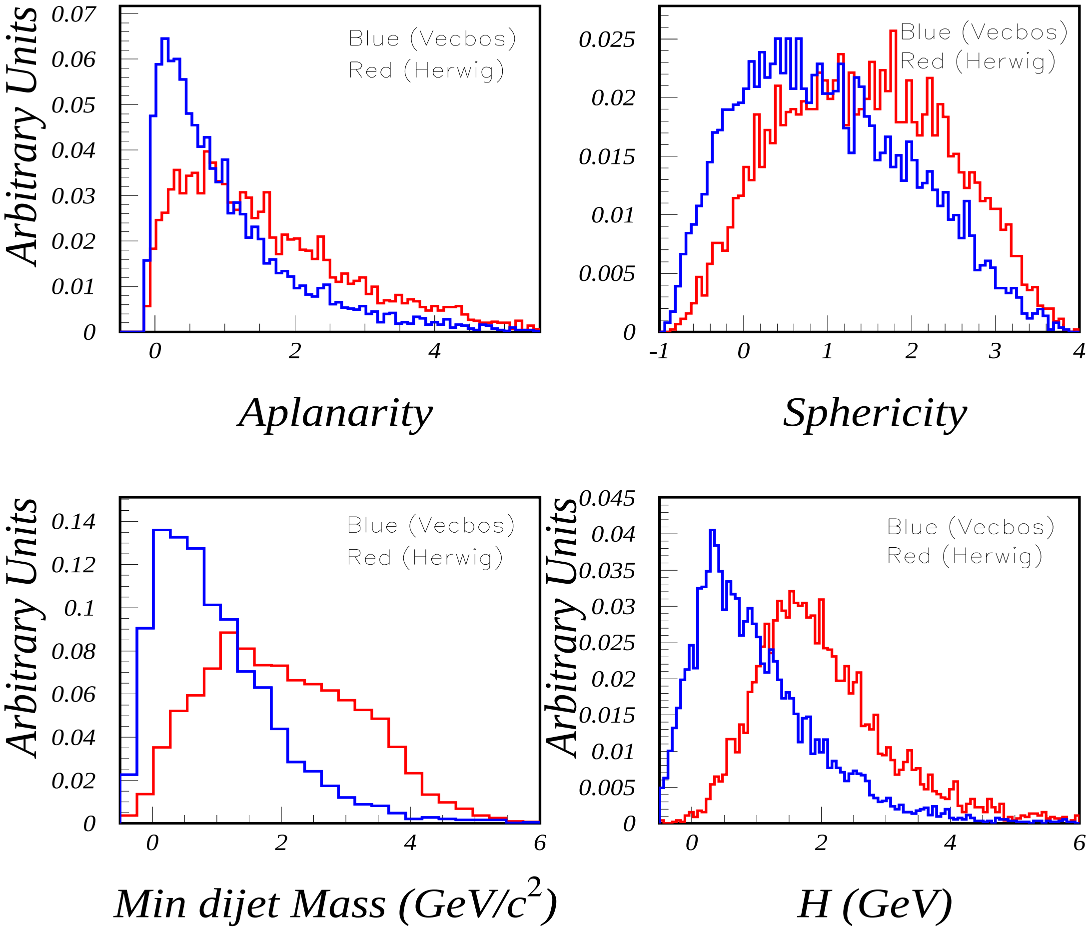


Figure 4.3: Input distributions used in the training of our 16 node ANN.

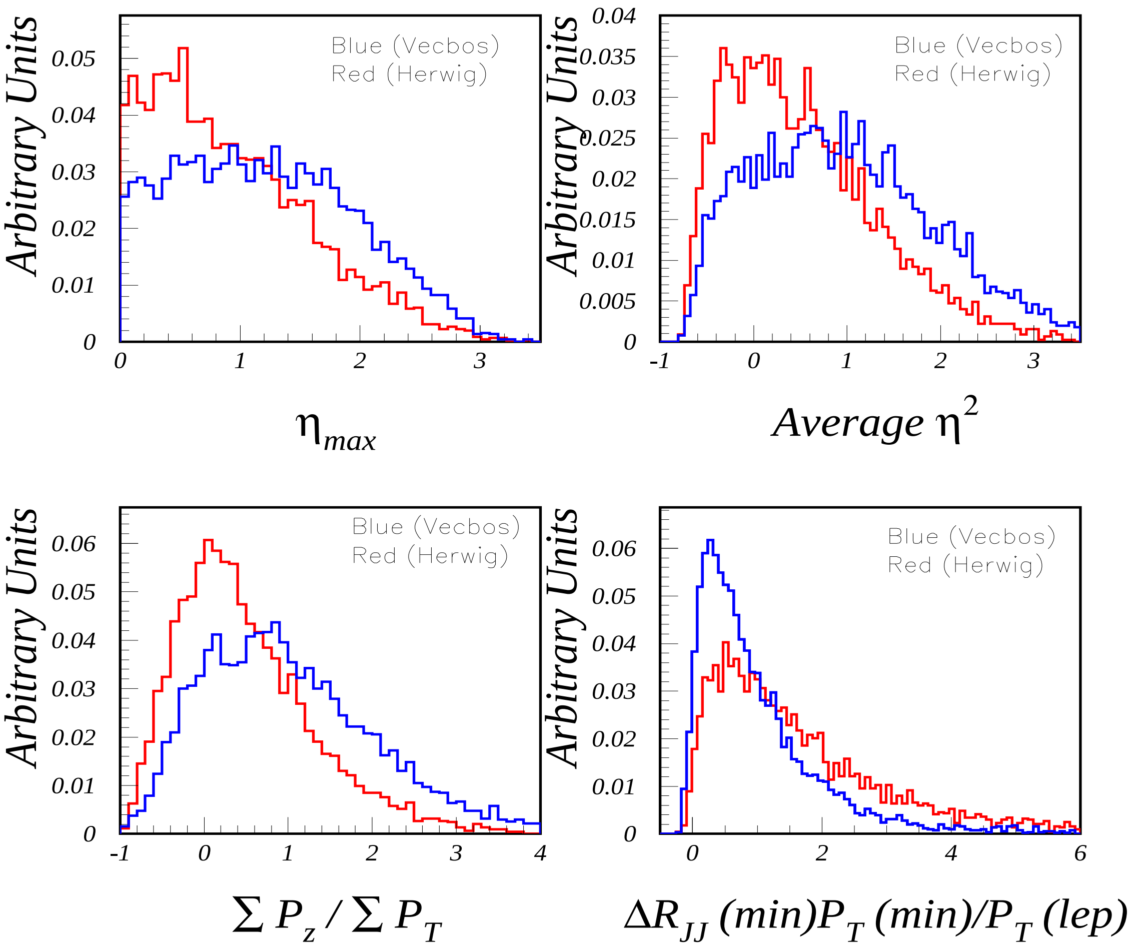


Figure 4.4: Input distributions used in the training of our 16 node ANN.

Chapter 5

Neural Network Optimization

Every *ANN* has it's own architecture. This architecture will be the key to the efficiency of the input to output mapping created. We have studied appropriate kinematics and neural architectures for the study of $t\bar{t}$ production in the $lepton+jets$ channel. The separating power of many different inputs were considered. In addition, the large effective size of the network employed warranted significant investigation of the general network architecture and network parameters to be employed.

5.1 Introduction

The use of *Artificial Neural Networks (ANN)* for the study of heavy quark identification has been extensively studied [59, 56]. Some of these analyses employed traditional "intelligent" variables (*Ex:* Aplanarity) and others utilized large numbers of "simple" variables (*Ex:* Jet P_T). The top quark itself has been searched for using intelligent variables as input pattern to an *ANN* by the DØ Collaboration [82]. Many of the intelligent kinematic variables which may be of value in $t\bar{t}$ searches have been carefully examined for their ability to separate $t\bar{t}$ from its primary background of W+$jets$ [57].

A *feed-forward Multi-Layered Perceptron* (MLP) network is used in this analysis. This is certainly the most commonly employed type of neural network for particle physics applications. It is particularly effective in the case ($t\bar{t}$) were a single signal type is being separated from a single dominant background. A training sample is put together using HERWIG generated $t\bar{t}$ signal and VECBOS generated, with HERPRT fragmentation, $W + jets$ background.

An independent sample had to be generated for the testing and optimization of the network. This sample was generated using the same generators and parameters as the training sample. The quality of the input variables were examined for their overall effect on the network performance, and the error on the signal returned in fitting the *ANN* outputs. This was done to eliminate variables which were not contributing to the overall discriminating power of the *ANN*. The internal layers were also tested to determine the optimal number of internal layers and nodes for the discrimination of $t\bar{t}$ from $W{+}jets$. This is essential in our case due to the large number of overall nodes in the network. Minimizing the number of nodes used in the network decreases the probability that convergence may fail, however over simplification is sure to result in a degradation of the network performance. Additional adjustable network parameters are examined for their effect on performance, which should always be done regardless of the relative complexity of the network being used.

5.2 *ANN* Training

This analysis uses JETNET[83] which is a neural network package commonly used in particle physics (see Appendix D). We use a feed-forward MLP network. The *activation function* used is the standard, Figure 5.2

$$g(x) = \frac{1}{1 + e^{-2x}}. \tag{5.1}$$

The nodes in the first layer (input layer) of a feed-forward *MLP* form an input vector $(x_1, x_2, ..., x_n)$ or pattern. In our case each of these nodes, x_n's, represent one kinematic value, and the input vector represents all the information used for a given

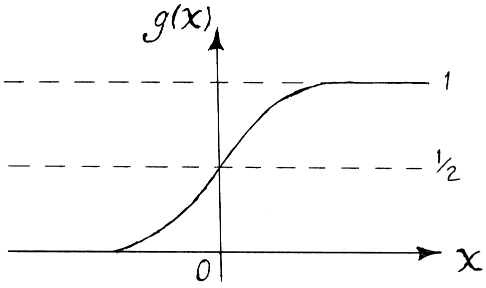


Figure 5.1: This is the activation function used in our neural network.

event. Each of the input nodes connects to each of the nodes in the *hidden* layer, $(h_1, h_2, ..., h_n)$, each of these connections has an associated *weight* (ω_{ji}). A node in the hidden layer performs a weighted sum of these connecting weights, passing this sum through a *transfer function*,

$$h_j = g(\frac{1}{T}\sum_i \omega_{ji}x_i + \theta_j) \ , \tag{5.2}$$

where θ_j is a bias weighting associated with each node, and the T (temperature) is a simple scaling effecting the steepness of $g(x)$. The output layer then performs a simple weighted sum of the inputs sent to it from the hidden layer,

$$f_k = \sum_j \omega_{kj}h_j. \tag{5.3}$$

The output layer corresponds to a vector $(f_1, f_2, ..., f_n)$ just as the input layer does. In our case this output will be a single number (node) with a value between 0 and 1.

Every network has some number of free parameters, corresponding to the number of network weights. The number of weights correspond to the number of connections between nodes plus a *bias* (θ) weight for each node in the input and output layers,

$$N_{weights} = N_1 + N_n + \sum_{i=2}^{n} N_i N_{i-1}. \tag{5.4}$$

As an example there will be 166 weights in a 5-10-10-1 network. The network architecture that will be used in this analysis 16-22-22-1 has 875 weights. The specific

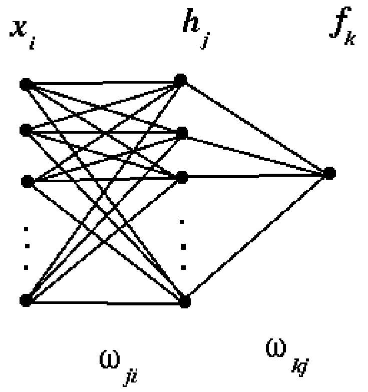


Figure 5.2: The x_i, h_j and f_k (dots) represent the input, hidden, and output nodes in the network. As shown here, a typical network has only one output node. In general there can be multiple hidden nodes. The links represent the weights ω_{ji} whereby the i^{th} node in the preceding layer influences the j^{th} node of a given layer. In a feed-forward network, nodes receive inputs only from the immediately preceding layer.

object of training the *ANN* is to set the network weights to values which will provide the proper mapping, **G**, between the input and output vectors,

$$\vec{f} = \mathbf{G}\vec{x}. \tag{5.5}$$

This mapping is "learned" by a procedure were the weights and thresholds are then adjusted by minimizing an error function E,

$$E(\omega, \theta) = \frac{1}{2N_p} \sum_{p=1}^{N_p} (f^p(\omega, \theta) - t^p)^2. \tag{5.6}$$

where t^p are the known (or target) values of the p^{th} training event (or pattern) and N_p is the total number of training events.

For a more thorough discussion of *feed-forward MLP* networks refer to Appendix D or to the Lund documentation of JETNET[83].

5.3 Kinematic Variables

The kinematic variables used in the network are fully described in Chapter 4. Many of the variables examined where selected for possible inclusion due to the studies of Andy Beretvas and Morris Binkley [57] and/or due to DØ studies [82].

Distributions for each kinematic variable being considered are generated for both signal and background with selection criteria described in Section 5.4.3. The input variables must be initially scaled to be consistent across the entire input pattern. Our inputs have been scaled in a standard way, each input is shifted to its mean and scaled by the standard deviation [57, 83].

The kinematic variables which were considered for inclusion in the network input pattern were prioritized by the Kolmogorov-Smirnov (KS) Statistic between the signal and background distributions.

The KS statistic is a value which tests whether two samples are consistent with being drawn from the same distribution. A number consistent with zero indicates that the two samples may be drawn from the same sample, the larger the KS statistic the less likely that hypothesis becomes. The usefulness of the KS statistic arises from the fact that a resulting non-zero value has a certain significance which is quantifiable.

Table 4.1 provides a listing and definitions for the kinematics considered for inclusion in the neural network. Not all these variables were selected for inclusion. The procedure by which some of these variables were chosen and others left out

is described in Section 5.5.1. Those kinematics that have been retained for inclusion in this analysis all appear in Figures 4.1, 4.2, 4.3, 4.4 .

5.4 Event Generation & Selection

The signal events used in the training and testing of the network were generated using HERWIG at $m_t = 175\ GeV$ with QFL detector simulation. The background events used were generated using VECBOS w/ HERPRT fragmentation and QFL. For a complete description of event generation and selection refer to Chapter 6.

The use of pseudo-experiments are common in HEP. Standard event distributions are constructed, usually by using event generators (Monte Carlo), see Chapter 6 These standard event distributions are randomly sampled in order to create artificial (pseudo) experimental data sets. These pseudo-experiments may then be used to test methods to be applied to the real data. These tests are valuable because the expected outcome is known in advance, that is the type of events are known in advance.

5.4.1 Training Sample

Approximately 9000 events are used to train the network. Every training event has a target value of 1 for a signal event or 0 for a background event.

5.4.2 Testing Sample

Pseudo-experiments used for testing the network are created from the samples of $t\bar{t}$ and W+$jets$ events already described but independent of the training sample. Each pseudo-experiment contains approximately 50 signal events and 80 background events. The number of signal and background events in each pseudo-experiment are independently Poisson fluctuated; therefore the total number of events in each pseudo-experiment is unconstrained.

5.4.3 Event Selection

A complete discussion of the event selection may be found in Chapter 6. The majority of the selection cuts are those used in the standard top (lepton+jets) analysis [95]. These are as follows:

- A single high P_T, greater than 20 GeV/c , electron or muon which is isolated.
- Missing transverse energy, $\not{E}_T$, of at least 20 GeV , before jet correction.
- At least 3 “high-threshold” jets with uncorrected $E_T > 15$ GeV and $|\eta| < 2.0$.
- Events consistent with Z bosons or γ or π^o conversions are removed.

- Events consistent with dileptons are removed.

In addition to these "standard" cuts there is one additional cut which events are required to pass for this analysis.

- At least 1 additional jet which must pass "low-threshold" with uncorrected $E_T > 8\ GeV$ and $|\eta| < 2.4$.

The most important difference in selection between this and the standard *lepton+jets* top analysis is that this analysis does not make use of b-tagging.

5.5 Network Optimization

In this section we discuss the variations possible in a network used to distinguish $t\bar{t}$ signal from $W{+}jets$ background. The first element of interest is the input vector itself, therefore we test to determine which if any of our possible input variables are not of use in the network training. Once an optimal input pattern is determined, proper architecture for the best possible mapping from that input to the desired output is established. Network parameters are then examined to attempt to improve performance. And finally, with an established *ANN*, the stability of the network is examined to determine how many training cycles (*epochs*) will be needed for stable results.

Clearly it is not possible to strictly follow this ordering. Some number of epochs must be used initially in the testing of the architecture just as an architecture must be selected in the determination of the usefulness of a given variable. Therefore intermediate selections must be made and re-examined several times during the course of the overall optimization procedure.

The determination of important optimization issues depend on the network being employed. The complexity of this network, the large number of weights, warrant careful examination of issues which may be less critical in a simpler network.

5.5.1 Input Pattern Selection

In what order various variables will be examined for inclusion in the input vector of the network is roughly determined by the value of the *KS* statistic between signal and background distributions for that variable. In this way variables were added one by one until the final set of variables was decided on. It might be argued that testing in a more individual fashion instead of this cumulative one would be a better way to judge the individual merits of a given kinematic variable. However, this reasoning does not take into account that correlations with other variables in the network may render any given variable useless to the overall discrimination power of the network.

Variable	*KS* Statistic
$P_T(lepton)$	0.068 ± 0.034
$\not{E}_T$	0.177 ± 0.026
H_T	0.464 ± 0.041
$\mathcal{A}$	0.233 ± 0.029
$\mathcal{S}$	0.175 ± 0.030
$MinMass(jj)$	0.377 ± 0.033
$\sum P_z/\sum P_T$	0.235 ± 0.029
η_{max}	0.204 ± 0.030
$< \eta^2 >$	0.216 ± 0.027
$\Delta R_{jj}^{min} P_T^{min}/P_T(l)$	0.233 ± 0.026
$Njets$	0.158 ± 0.026
$P_T(1)/H_T$	0.112 ± 0.038
$P_T(1)$	0.344 ± 0.044
$P_T(2)$	0.440 ± 0.035
$P_T(3)$	0.451 ± 0.034
$P_T(4)$	0.347 ± 0.028
$P_T(1) + P_T(2)$	0.405 ± 0.040
$P_T(2) + P_T(3)$	0.478 ± 0.039
$P_T(3) + P_T(4)$	0.477 ± 0.033
$\not{E}_z$	0.016 ± 0.078

Table 5.1: Kolmogorov-Smirnov Statistic *(KS)* for each kinematic variable considered for inclusion in this analysis.

When a given variable shall be included in the input pattern is determined by the value of the *KS* statistic for the $t\bar{t}$ vs. *W+jets* distributions for that variable. A higher *KS* statistic indicates greater difference between the two distributions, See Table 5.1.

Starting out with three variables with relatively low correlation $(H_T, \mathcal{A}, MinMass(jj))$ we examined the performance of the network. As more variables were added to the input pattern this performance was re-examined to determine the kinematics which were having an impact on the overall performance of the *ANN*, and those which were not. Although some possible variables were dropped from consideration because they did not improve the networks ability to discriminate signal from background, this does not imply that those variables are not of general interest in the study of $t\bar{t}$. A good discriminator may be of little or no value to the *ANN* simply because all the information contained in that variable has been extracted from other correlated variables which were included in the network earlier. Figure 5.3 gives some idea of how the performance was enhanced as the sixteen inputs were gradually incorporated into the network.

Further testing was done to see the effect of extracting individual variables from the input vector. This was done because some of the variables incorporated earlier might have become redundant with the building of the final set. Only one of seventeen variables $(P_T(1)/H_T)$ was eliminated in this fashion. Each time one of the remaining sixteen variables was removed the performance of the network was, to some degree, degraded.

5.5.2 Network Architecture

An input vector of sixteen nodes is large and will probably require a complex network in order to map the correct output vector. As an *ANN* gets more complex the risk of non-convergence and the probability of getting stuck in a local minimum increases. Therefore it is reasonable to seek to minimize the size and complexity of the network. However the oversimplification can degrade the performance and can also make a proper mapping impossible to achieve.

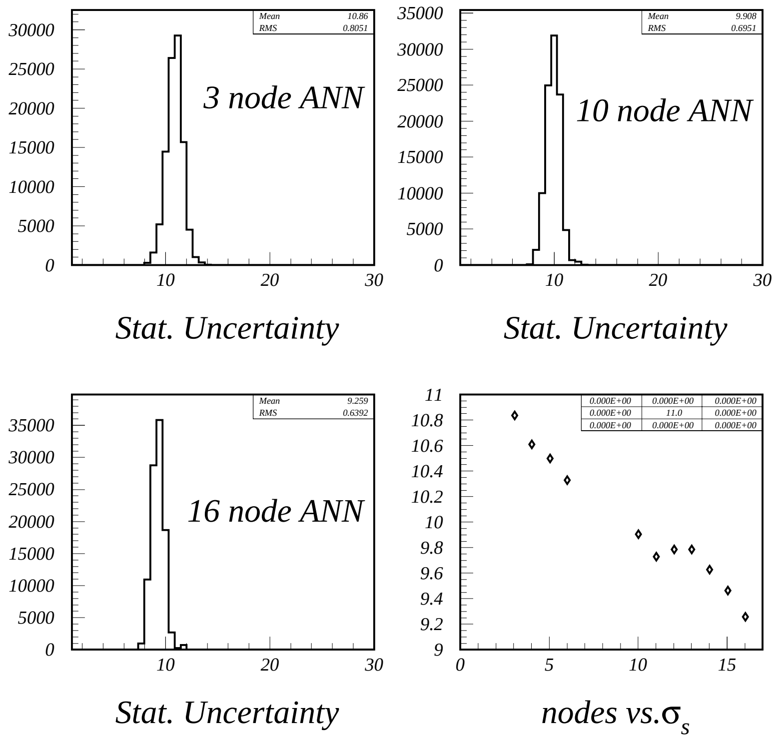


Figure 5.3: During optimization as nodes are added the uncertainty on the number of signal events returned by the fitter diminishes. The bottom-right plot shows how the uncertainty changes as nodes were added.

The significance,

$$Significance \equiv \frac{1}{\sqrt{\left(\frac{\sigma_S}{S}\right)^2 + \left(\frac{S - S_{exp}}{S_{exp}}\right)^2}},$$

of various configurations was examined, where S_{exp} is the actual number of signal events in a given pseudo-experiment and S is the number of signal events returned from the fitter, and σ_S is the uncertainty on that fit. Although it was not the only factor examined during the testing of the network architectures, significance was documented in all the architectures and parameters discussed.

Number of Nodes & Layers

Networks with both one and two hidden layers were tested. Although any mapping should be achievable with one hidden layer this configuration does not always provide the most efficient solution to a given mapping [56]. Some single hidden layer performances and some double hidden layer performances are shown in Figure 5.4. The 16-22-22-1 configuration was finally selected for our network, where this

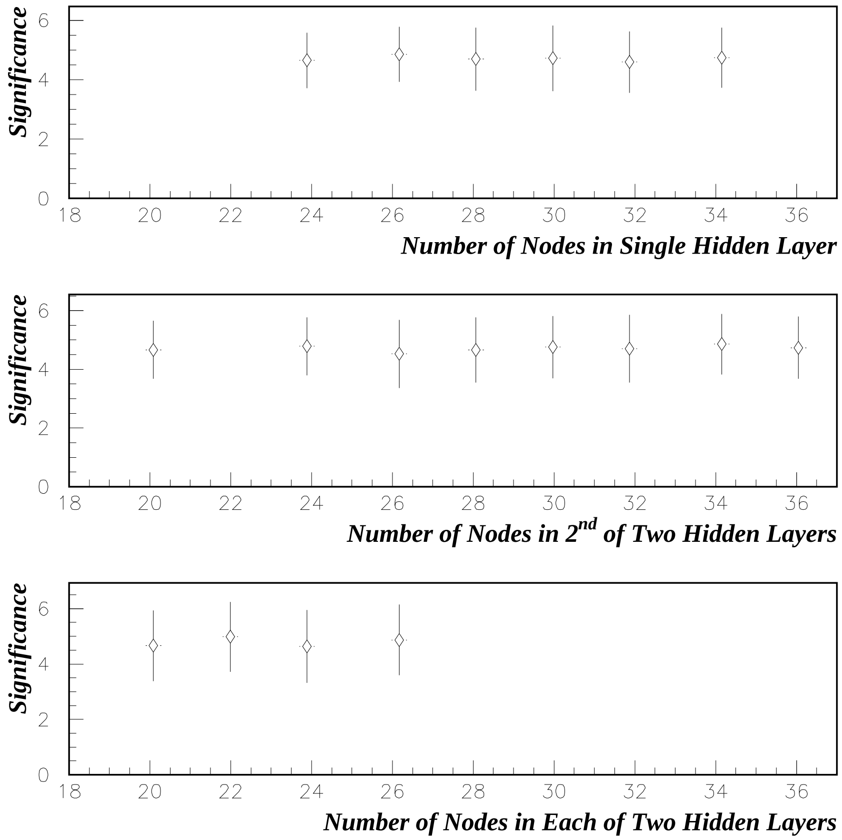


Figure 5.4: Significance of various architectures. All the networks have sixteen inputs and one output but the number of hidden layers and nodes is varied. The middle histogram includes all networks with 16-16-x-1 architectures where the second hidden layer is varying. The bottom histogram has a 16-x-x-1 architecture where both hidden layers have the same number of nodes. There are large bin-to-bin correlations due to the fact that the same training samples are used in each case.

corresponds to the nodes in the input-hidden-hidden-output layers. This gives a network with 875 weights to adjust, see Equation 5.4.

Although it is possible to achieve the same performance level with a single hidden layer that single layer must be so large that the resulting network is more complex (has a greater number of weights) than our two layer architecture.

5.5.3 Network Parameters

There are a number of network parameters which can have considerable impact on the performance of an *ANN*. These are studied to determine both whether the best possible performance has been achieved and to prevent specific problems, such as the error function finding only a local minima instead of the true function minimum.

Learning Rate (η)

The learning rate (η) is the factor that governs the incremental changes in the updating of the network weights. This we varied from 0.001 to 0.2 the results of which are shown in Figure 5.5. Once η exceeded 0.01 it remained stable and a value of 0.1 was selected for the network learning rate.

Momentum Parameter (α)

The momentum parameter (α) typically varies from 0 to 1. By adjusting α stability can be enhanced and convergence facilitated. Throughout the range examined the network performed acceptably except at $\alpha = 1.0$. At $\alpha = 1.0$ the network converged but the significance was reduced by half, see Figure 5.5. An $\alpha = 0.5$ was selected for our *ANN* parameter.

Inverse Temperature (β)

The inverse temperature ($\beta = 1/T$) is the determining factor in the steepness of the transfer function $g(x)$ shown in Equation 5.1. We examined β in a range from 0.5 to

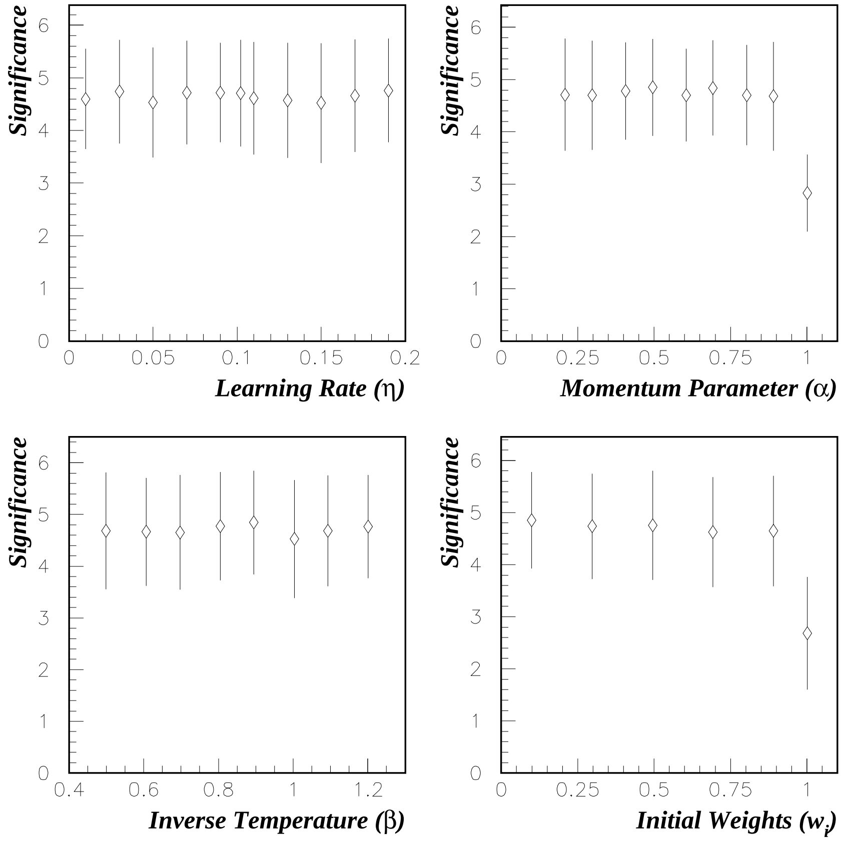


Figure 5.5: Significance of various network input parameters: learning rate, momentum parameter, inverse temperature, and initial weights. There are large bin-to-bin correlations due to the fact that the same training samples are used in each case.

1.2 and again found it to provide consistent results across this range, see Figure 5.5. We selected a value of 0.9 for the inverse temperature.

Initial Weights (w_i)

All the weights in the network must be set to an initial value before training begins. Frequently an *ANN* is insensitive to changes in these initial weights, however adjustment of these values can be particularly useful in determining whether a network is properly minimizing the error function, Equation 5.6. If instead of a global minimum

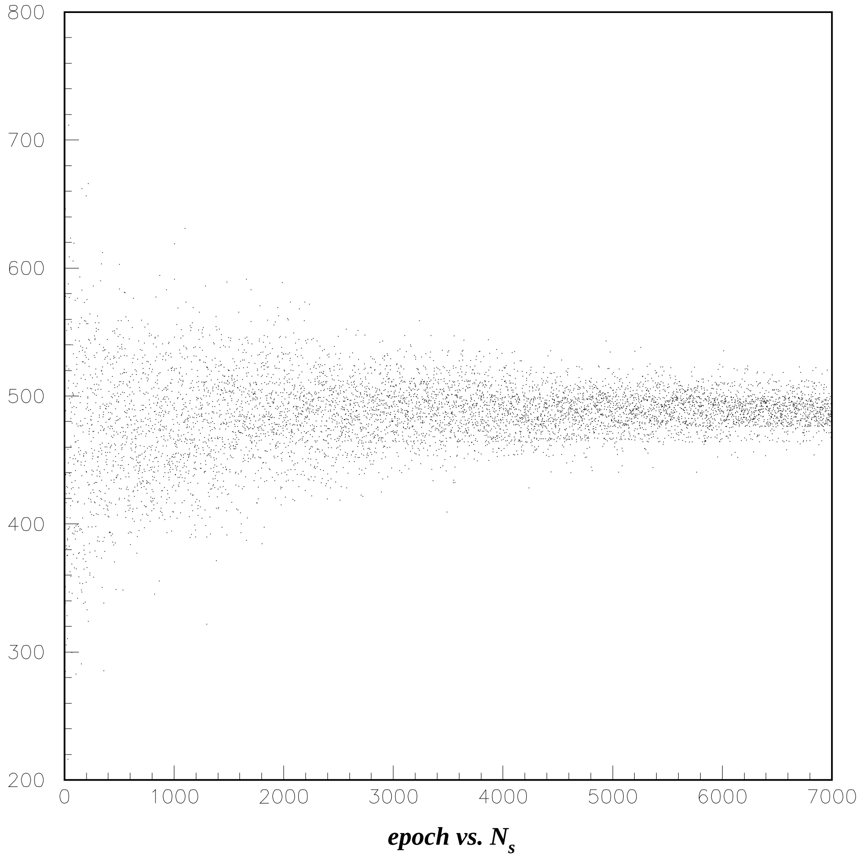


Figure 5.6: Stability requires several thousand learning cycles (epochs). Each dot represents the number of signal events (N_s) returned by the fitting of a pseudo-experiment output to signal & background output distributions after the number of epochs shown on the x-axis. This pseudo-experiment contains 468 signal events.

for the function the network is being trapped in some local minima then varying the initial weights should have a significant effect on the *ANN's* performance. Typically, weights are varied either from 0 to 1 or from -1 to 1 depending on the error function being used. We tested our network from 0 to 1 and except for $w_i = 1.0$ it showed no sign of having converged to an improper minimum. We selected 0.1 for the values of our initial network weights.

5.5.4 Number of Training Cycles

The number of training cycles necessary for stable performance increases with the complexity of the network. We varied our training cycles, “epochs”, from 10 to 10,000 cycles. Network performance is stable after 4000 cycles. To eliminate the effect of instabilities in training 6000 to 8000 cycles will be used in the balance of the analysis. This stability can be seen in Figure 5.6.

5.6 Testing Selected *ANN*

All the optimization tests yielded individual parameters and architectures which were deemed best or acceptable, depending on whether or not a given parameter was a factor in performance. Once all these factors were set together the resulting *ANN* was re-examined for performance. The result of this test was equal or increased significance over any individual network tested during optimization. And it performed better in terms of stability than most of the other networks tested.

5.7 Summary

Of the nineteen variables examined for possible inclusion in the network sixteen were selected for inclusion based on the relative significance of the results after their inclusion. Of the eighteen network architectures shown in Figure 5.4 16-22-22-1 was selected as a reasonable compromise between performance and complexity. Many network parameters were varied to determine their impact on network performance, including learning rate (η), momentum (α), inverse temperature (β), and initial weights (w_i), from these tests the following values were selected. The testing of these parameters indicated that the network was relatively insensitive to these parameters. These result of course can not be easily generalized to other networks.

A thorough examination of the network to be employed is critical if the result is to be trusted, especially for a large (and therefore complex) network where comparison to a correlation matrix method is impractical. In the case of small network where the

Parameter	Selected Value
epochs	4000 to 5000
η	0.1
α	0.5
β	0.9
w_i	0.1

calculation of the all of the correlations is reasonable that should be done, in which case a less thorough examination of some optimization issues should be acceptable. However certain parts of the optimization must always be examined, notably the network parameters. Other issues like the network architecture may be less critical.

Despite the difficulties inherent in the optimization of a large neural network, it is precisely these large pattern recognition situations in which the neural network approach is indispensable. It is complex input pattern situations in which no other method of analysis is possible.

Output distributions for our signal and background training samples are shown in Figure 5.7.

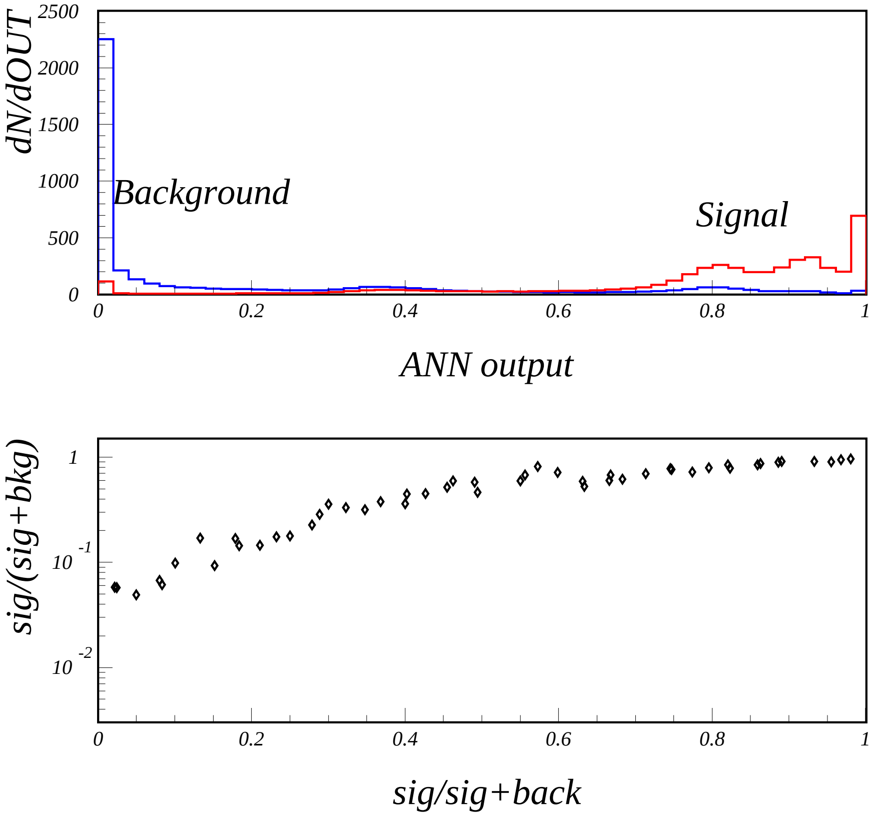


Figure 5.7: Outputs from 16 node ANN.

Chapter 6

Event Simulation & Data Selection

6.1 Monte Carlos

The use of simulated events has become an indispensable tool in high energy physics analysis. The ways in which these *Monte Carlo* events are used are many fold, but the basic purpose is always to see how your detection process registers actual events. This is true whether you are interested in detector acceptance, systematic errors or any other type of event comparison.

The most central application of Monte Carlo (MC) in this analysis comes with the testing and training of our neural network. Clearly training of the *ANN* must be done with MC events given that the trained network is a prerequisite for discrimination of signal from background events. In our case the *ANN* implemented is trained with a signal sample of $t\bar{t}$ generated using HERWIG [66], and a single type background sample of *W+jets* generated using VECBOS [67]. Other background samples are also generated for later background subtraction. The reason for this approach is two fold. Firstly, our *Non-W* background samples have statistics which were too limited for *ANN* training. Secondly, since the $W{+}jets$ is a much more dominate background than the other backgrounds it is not clear that inclusion of those small backgrounds would have been appropriate mis-identification: would then have been difficult to quantify.

In this analysis, as with many others, the use of MC events was indispensable for the determination of acceptances and systematics. However training and optimization of the neural network was the most central function of these MC samples as was described in Chapter 5.

Every event simulation used in this analysis has certain basic elements in common. In all the Monte Carlos discussed here the events are generated first as a QCD hard scattering using the production cross sections for a given process alone with parton *structure functions* which model the internal parton dynamics according to theoretical models of parton structure. *Initial* and *Final State Radiation* (*ISR & FSR*) is incorporated, and *hadronization* of bare quarks takes place. It may then be necessary to fragment jets depending on the particular MC being employed. This is the basic procedure followed by each of the event generators used in this analysis.

Once events are generated one more essential step remains before those events can be *stripped* and utilized in our analysis, this final step is detector simulation. The detector simulation used in our analysis is QFL, it takes into account the eccentricities inherent in the detector. Cracks in the detector, efficiencies, smearing effects, are all simulated using QFL to make events look like real *CDF* events.

After the events are generated and put through QFL they may be treated exactly like detector data and be made subject to all the same selections. These selections are discussed in some detail in Section 6.2.

6.1.1 The Right Monte Carlo

Every event simulator has its own advantages and disadvantages. The various strengths and weaknesses of a given MC may make one more suited to the simulation of a certain process than another, because of this several different types of MCs have been employed to simulate different types of events though all of them have been subject to the same detector simulation (QFL).

6.1.2 HERWIG

The signal for this analysis is modeled using HERWIG. This is the generator most commonly used to model $t\bar{t}$ events at *CDF* and at DØ. This $t\bar{t}$ sample is used to train and test our *ANN*. In addition to this, HERWIG is used to model one of our secondary backgrounds (QCD). This background is the largest of the non-W backgrounds and is described in Chapter 7.

The top quark mass is taken to be $m_t = 175\ GeV/c^2$. HERWIG includes initial and final state radiation, the amount of this emission is determined by the Q scale of the hard scattering. It is the uncertainty in this emission that is the greatest HERWIG contribution to our systematic error. Due to this uncertainty PYTHIA samples without initial and final state radiation will be generated to quantify this uncertainty, Chapter 8.

The elementary hard sub-process can be calculated perturbatively to leading order. This sub-process is $(q\bar{q} \rightarrow g \rightarrow q'\bar{q}')$ in general, and $(q\bar{q} \rightarrow g \rightarrow t\bar{t})$ in our case. The momentum transfer scale Q, sets the limits on the ensuing initial and final state parton showers.

Once these partons are calculated they must be converted to hadrons. The momentum transfers in these processes are in general low and the strong coupling is large, as with the b-quarks common in $t\bar{t}$ events. In this situation perturbative calculations are not applicable and a phenomenological hadronization model must be used for fragmentation of the partons. Clearly this is only the case after the top quarks have decayed since they decay weakly but very quickly via a real W-boson in a time much shorter than the hadronization timescale.

6.1.3 VECBOS

Vecbos is used to model the $W + jets$ background which is the primary background for this analysis. This is the generator most commonly used to model W+$jets$ events at *CDF* and at DØ. A large sample was needed for both training and testing of the neural network. An additional sample was needed to measure the uncertainty associated with the Q scale, this systematic uncertainty is discussed in Chapter 8.

In order to generate VECBOS events from an existing grid weighted events must be generated first. A sampling from these weighted distributions is then performed thereby producing unweighted events. And finally jets must be fragmented before the events may be sent on to QFL.

Grids

Events are sampled from a Vegas integration grid.[68] Both large and small grids were tried in the generating of VECBOS events.

Although it is normal to utilize a small grid in *CDF* VECBOS event generation it is worth noting that the grid statistics seem to play a role in features which appear in distributions of the generated samples, see for example the upper right figure in Figure 6.2. The large grid requires considerably more time to generate than the smaller grid. However this is a one time computational cost, the event generation time itself is unaffected. This means that generating events from already existing grids requires the same time regardless of the grid size.

The grids will be sampled at a rate of 1,000 points for each iteration. In order to determine the proper number of iterations to employ during the creation of weighted events the cross sections and χ^2 are examined. The smaller grid has much

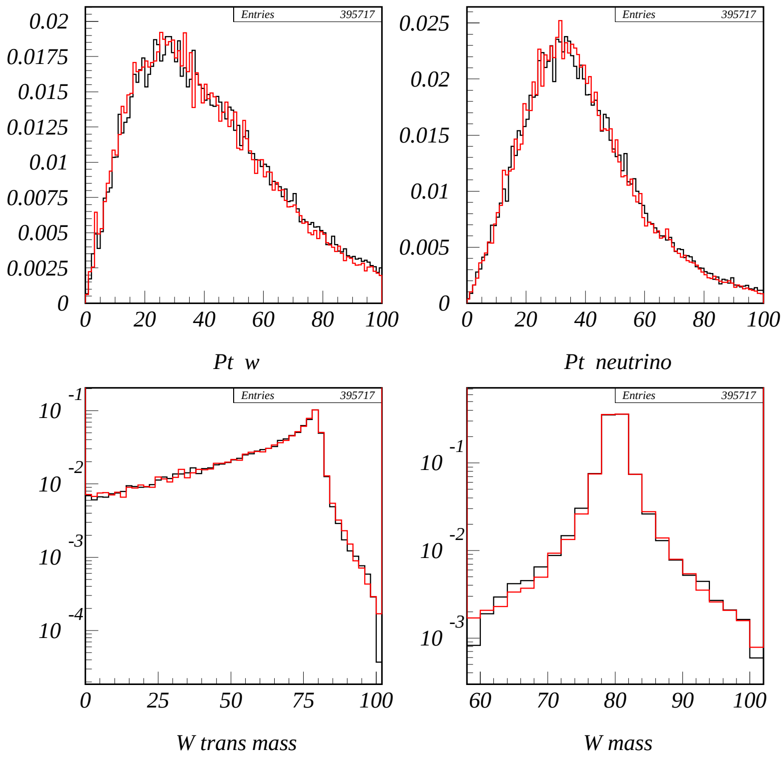


Figure 6.1: Events generated from the large grid (red) are compared with events generated from the small grid (black).

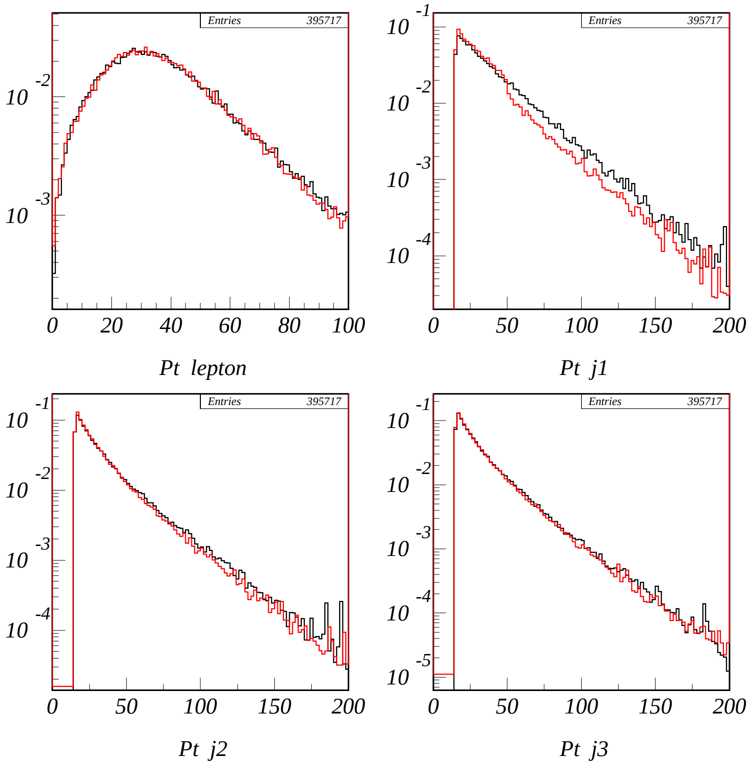


Figure 6.2: Events generated from the large grid (red) are compared with events generated from the small grid (black).

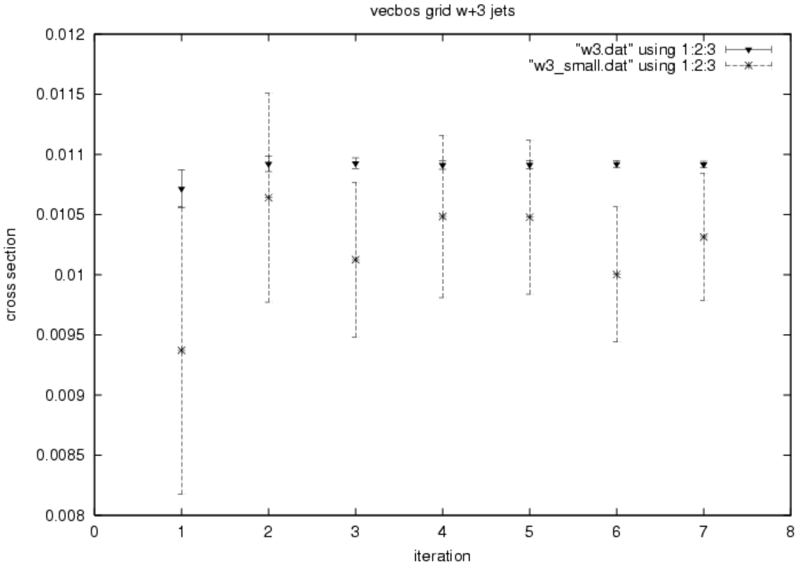


Figure 6.3: Each iteration represents 1,000 additional samplings from the VECBOS grid during the creation of the weighted events. Clearly the error on cross section is much higher for the smaller grid.

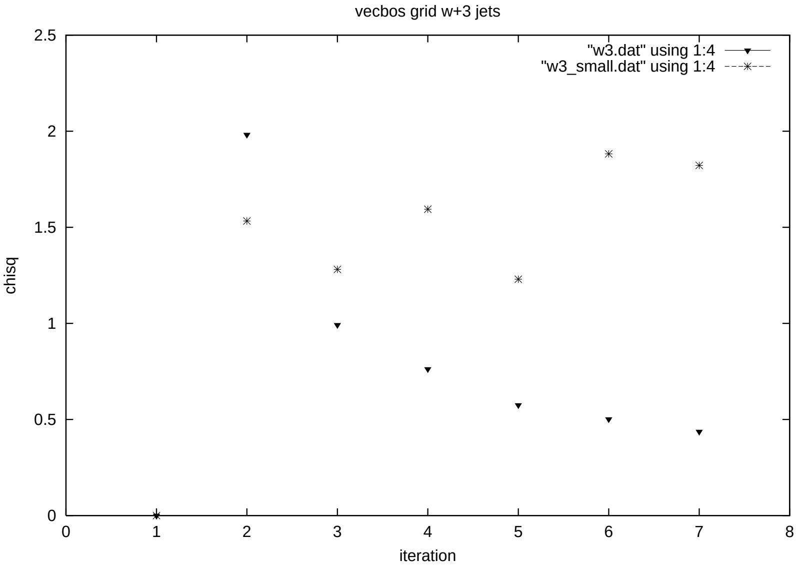


Figure 6.4: Each iteration represents 1,000 additional samplings from the VECBOS grid during the creation of the weighted events. The stability of χ^2/dof is closely related to the error on the cross section and therefore on the grid size.

greater error on the cross section. It is not surprising then that the chi square is correspondingly less stable, Figures 6.3, 6.4 .

Weighting

Each event is generated with an associated weight. This weight is related to the probability of occurrence of the event. The sum of the weights of all the events generated is the total cross section. Weighted events are first histogrammed by their respective weights, and then

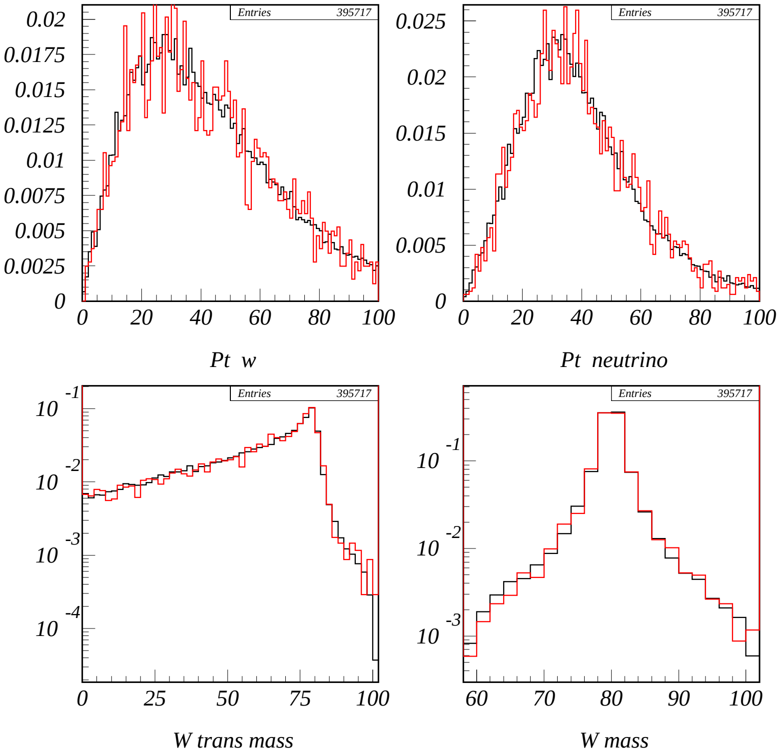


Figure 6.5: The weighted events are shown in (red), the unweighted events are shown in (black).

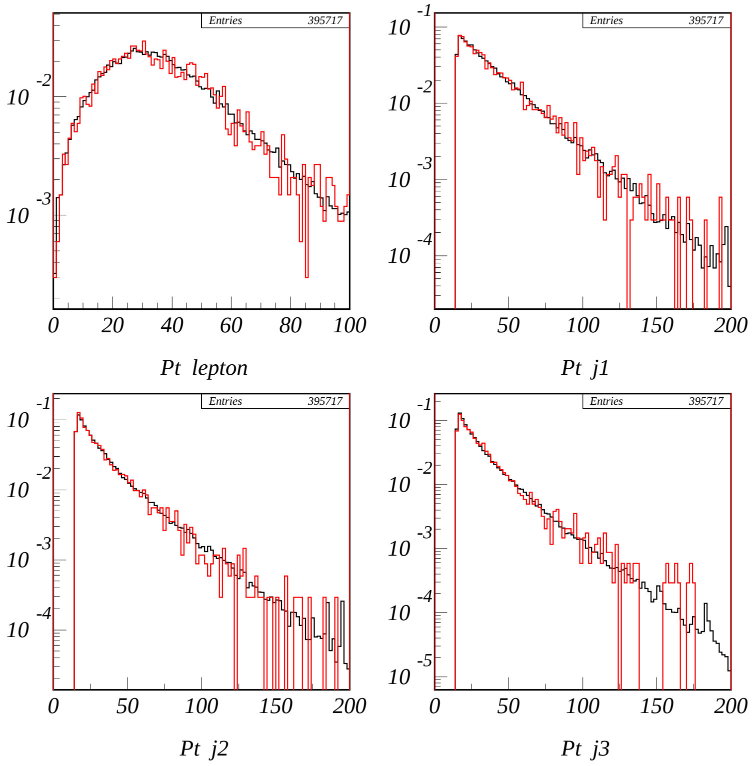


Figure 6.6: The weighted events are shown in (red), the unweighted events are shown in (black).

in order to unweight events the weighted event distributions are sampled. If the ratio of the events weight (w_i) to a maximum weight (w_{max}) is greater than a random number (X_{rand}) between 0 and 1,

$$\frac{w_i}{w_{max}} > X_{rand} \ ,$$

then the event is kept. Efficiency may be improved by setting w_{max} to a value of less than the actual maximum weight as long as not too many of the generated events are discarded [79]. In our case 1.9% of the total cross section of events have weights which are greater than the selected maximum weight, Figure 6.7.

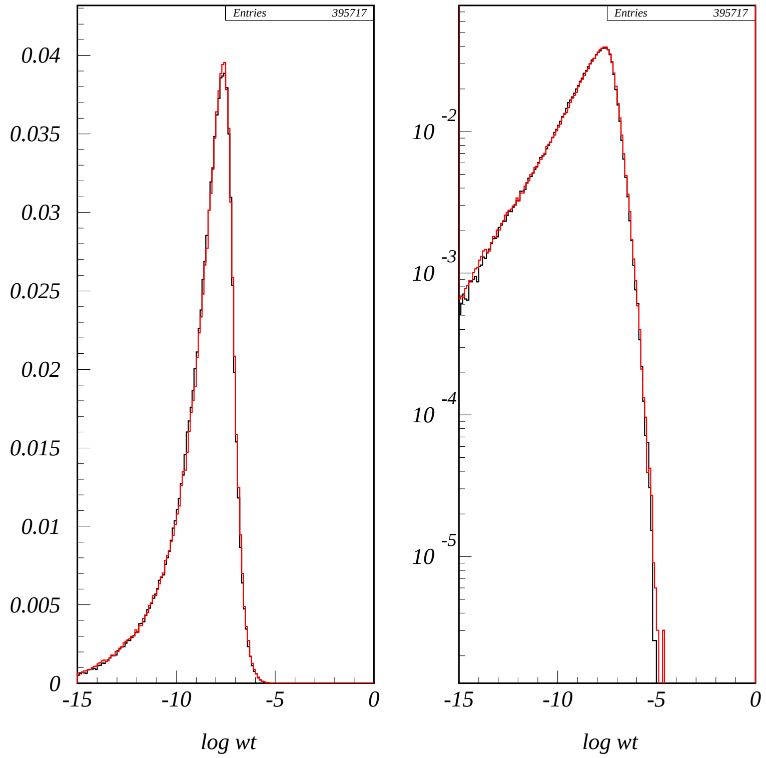


Figure 6.7: The weights seem to drop off in very nearly the same way for both grids the maximum weight is approximately the same. Again the large grid is shown in (red), and the small one in (black).

Fragmentation

During the aforementioned generation *W+3jet* events are created. Although it is possible to generate *W+4jet* events the computation time is far too great for that kind of generation to be practical. In order to generate *W+* $\geq$*4jets* it was necessary to generate *W+3jet* events and fragment. This fragmentation of jets is done by putting the events through HERPRT[66].

6.1.4 ISAJET

The backgrounds not already mentioned are simulated using ISAJET. These include diboson production and Drell-Yan event samples. Although there are other possible backgrounds to our process, specifically $Z^o \rightarrow \tau\tau$ and $W \rightarrow \tau\nu$, the estimates on these backgrounds are lower (less than 4/100)[89] and were not included in our background removal. It should be remembered that only a fraction of the background events present in our sample will appear *signal like* and the rest will be discounted as *W+jets like*, see Chapter 7 for a discussion of background removal.

Due to our limited statistics [90] it was decided to combine several different structure functions to create background samples. Both the input and output distributions for different structure functions were compared, the *KS* statistic between the output distributions for structure functions: *GRVL0*, *CTEQ2L*, and *MRSD0* were compared as well as being compared with data samples, Table 6.1. The distributions used to make all these comparisons are shown in figures 6.8.

Distributions	***KS* statistic**
CTEQ2L vs. *MRSD0*	0.039 ± 0.004
CTEQ2L vs. *GRVL0*	0.046 ± 0.004
GRVL0 vs. *MRSD0*	0.022 ± 0.004
CTEQ2L vs. Data	0.240 ± 0.006

Table 6.1: Kolmogorov-Smirnov Statistic *(KS)* comparing diboson background samples generated using different structure functions.

6.1.5 PYTHIA

Of the two MC generators used to simulate $t\bar{t}$ events HERWIG is used to model $t\bar{t}$ for purposes of training and testing the *ANN* and thereby separate signal events from background events. The primary uncertainty associated with our use of HERWIG is in its modeling of the initial and final state radiation, this is where the use of PYTHIA becomes necessary. Initial and final state radiation cannot be turned off in

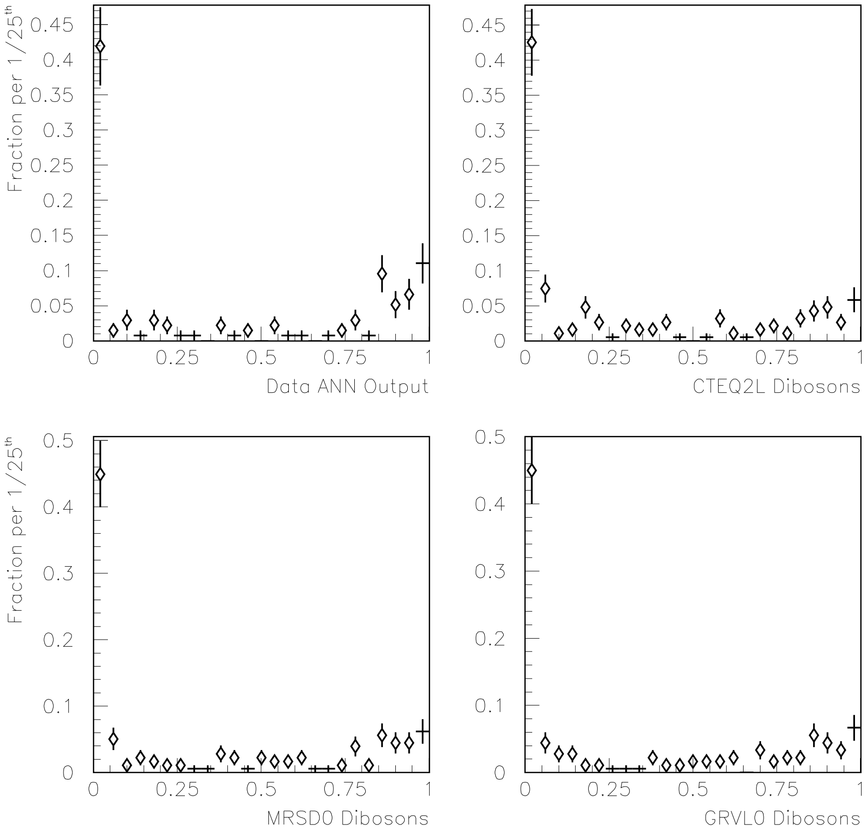


Figure 6.8: The *ANN* output distributions for data is shown in the upper left. Each of the other three distributions are diboson samples utilizing a different structure function.

HERWIG and therefore the systematic must be determined using some other Monte Carlo.

PYTHIA like HERWIG calculates leading-order *QCD* matrix elements for hard-scattering processes. However unlike HERWIG, PYTHIA's use of JETSET to model *ISR & FSR* can be turned off [70].

6.2 Data Selection

The channel and signature of the data sample used in this analysis have already been discussed in Chapter 1. The $lepton + jets$ channel for $t\bar{t}$ production is characterized by a high E_T isolated lepton, significant $\not\!\!E_T$, and several high E_T jets. A preliminary data reduction is made by focusing on an inclusive lepton set. This will be further reduced by requiring high missing transverse energy in the W-sample. And of the W- samples only those with the appropriate number of jets will be considered.

6.2.1 Inclusive Electron Sample

An electron candidate must get through three trigger levels to be included in the inclusive electron sample. The level 1 trigger requires only a CEM cluster of $E_T > 8\ GeV$. The level 2 trigger requires a CEM cluster of $E_T > 16\ GeV$ and a matching CFT track with $P_T > 12\ GeV/c$. At level 3 various cuts are applied to improve the signal to background efficiency for real electrons.

Events passing level 3 triggers are then put through slightly tighter versions of the level 3 cuts to eliminate non-electrons, photon conversions, and electrons within jets. The 16 GeV CEM cluster requirement is raised to 20 GeV. If the level 2 cluster in the electromagnetic calorimeter is a true match to the CFT track the ratio of the energy in the CEM(E) to the momentum in the CFT(P) should not be to great ($E/P < 1.8$). A particle shower that is truly from an electron will deposit most of its energy in the electromagnetic calorimetry, therefore we require the ratio of hadronic calorimetry(E_{had}) to the electromagnetic calorimetry(E_{em}) should be small ($E_{had}/E_{em} < 0.05$). Electron tracks are required to match with hits in the CES. Distances between the CES hits and tracks (Δz & $\Delta x(r-\phi)$) must pass certain cuts. Additionally the CES shower profile (χ^2_{strip}) must be matched to testbeam profiles. A sum over adjacent towers in the CEM(L_{shr}) is used to indicate how well a CEM energy profile matches a testbeam profile,

$$L_{shr} = 0.14 \sum \frac{E_i^{obs} - E_i^{pred}}{\sqrt{(0.14\sqrt{E})^2 + \sigma^2_{pred}}} \ .$$

Where E_i^{obs} and E_i^{pred} are the observed and predicted energy in tower i, σ_{pred}^2 is the uncertainty on the predicted energy, and $0.14\sqrt{E}$ is the uncertainty on the measured energy. In order to demonstrate that the electron comes from the primary vertex the electrons z position (z_{elec}) is compared to the event vertex as determined in the VTX (z_{vtx}). In addition, the event vertex must be within 60 cm of the center of the detector. Finally, any energy clusters which are too close to detector boundaries are discarded (fiducial cuts). These cuts are summarized in Table 6.2 and shown in Figure 6.9.

$E_T > 20\ GeV$
$E/P < 1.8$
L_{shr}
Track Strip/Wire Matching:
$\Delta x < 1.5$ cm
$\Delta z < 3.0$ cm
$\chi^2_{strip} < 10$
$\lvert z_{elec} - z_{vtx}\rvert < 5$ cm
$\lvert z_{vtx}\rvert < 60$ cm
Fiducial Requirements

Table 6.2: Inclusive electron sample requirements.

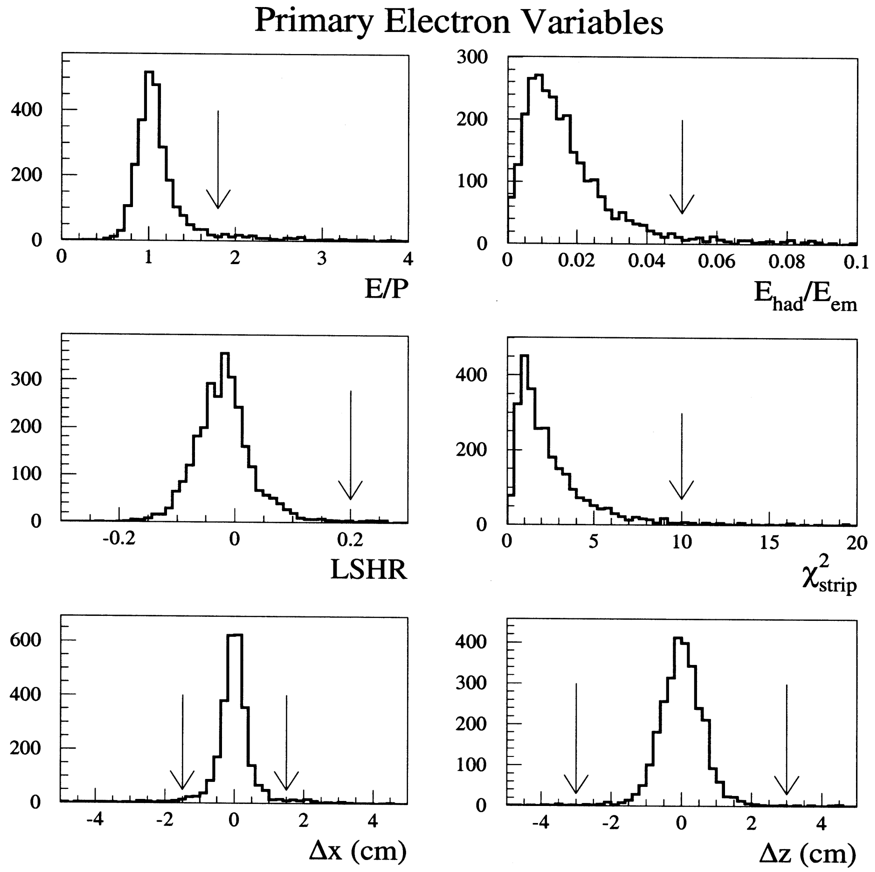


Figure 6.9: Selection cuts used to define the inclusive muon sample. This plot is of the secondary electrons in a $Z \to e^+e^-$ candidate sample, additional selection for explicit Z removal will be made. Arrows indicate cut values see Table 6.2. This figure taken from reference [42].

6.2.2 Inclusive Muon Sample

A muon candidate is identified via a different set of parameters than that used for electrons. Muons are much more penetrating (essentially min. ionizing) due to their large mass ($m_\nu >> m_e$). At *CDF* muon candidates are classified according to which of the detector subsystem(s) in which they appear. A muon may be a CMX muon, CMU-only muon, CMP-only muon, or a CMUP muon (both CMU & CMP hits).

A level 2 muon must have a CFT track with $P_T > 12\ GeV/c$ and lie within 5^o of a level 1 muon stub in one of the muon subsystems. Due to very high event rates several of the level 2 muon triggers are prescaled, meaning 1 of every n triggers is kept. The value of n scales with luminosity. Combining the muon trigger with a level 2 calorimetry cluster trigger keeps the efficiency for $t\bar{t}$ events high despite the prescaling. Details of level 2 trigger requirements for various types of muons are shown in Table 6.3.

Several of the inclusive level3 muon requirements are similar to those for electrons, this can be seen in Table 6.4. However, as muons are minimum ionizing and very little of the muons energy is deposited in the calorimetry. The total energy deposited in the electromagnetic and hadronic calorimetry is less than 2 GeV and 6 GeV respectively. In addition to this the distance between tracks in the CFT and stubs in the muon components (Δx) must be less than a certain amount. It should be noted that the cut on Δx varies by detector component. The distance traversed through various components is different, for example more material is traversed by a CMP muon than a CMU muon because they are more deflected by multiple Coulomb scattering. These cuts are shown in Figure 6.10 and summarized in Table 6.4.

6.2.3 The W sample

The W sample is drawn from the inclusive lepton samples. The only requirements of the W sample is an isolated lepton and a significant amount of corrected $\not{E}_T$, Appendix A. It should be noted that implicit in this selection is a 2/9 factor for the

Triggers	Prescaled
CMX Muons	
1. $\not{E}_T > 35\ GeV$ and two jets w/ $E_T > 3\ GeV$	No
2. CFT track w/ $P_T > 12\ GeV/c$ matching a CMX stub	Yes
3. CFT track w/ $P_T > 12\ GeV/c$ matching a CMX stub and one jet w/ $E_T > 15\ GeV$	Yes
CMU-only Muons	
1. $\not{E}_T > 35\ GeV$ and two jets w/ $E_T > 3\ GeV$	No
2. CFT track w/ $P_T > 12\ GeV/c$ matching a CMU stub	Yes
3. CFT track w/ $P_T > 12\ GeV/c$ matching a CMU stub and one jet w/ $E_T > 15\ GeV$	No
CMP-only Muons	
1. $\not{E}_T > 35\ GeV$ and two jets w/ $E_T > 3\ GeV$	No
2. CFT track w/ $P_T > 12\ GeV/c$ matching a CMP stub	Yes
3. CFT track w/ $P_T > 12\ GeV/c$ matching a CMP stub and one jet w/ $E_T > 15\ GeV$	No
CMUP Muons	
1. $\not{E}_T > 35\ GeV$ and two jets w/ $E_T > 3\ GeV$	No
2. CFT track w/ $P_T > 12\ GeV/c$ matching a CMU and CMPstub	Yes
3. CFT track w/ $P_T > 12\ GeV/c$ matching a CMU and CMP stub and one jet w/ $E_T > 15\ GeV$	Yes

Table 6.3: Muon level2 trigger requirements.

$W \rightarrow l\nu$ ($l = e, \mu$) branching ratio ($Br(W \rightarrow l\nu) = 2/9$), 2/3 will go into quarks. Individual sets with different numbers of jets in addition to the W-boson are made. A jet is required to have $E_T > 15\ GeV$ to be included in these samples, since we require 3 jets this hard the only relevant data samples will be $W + 3jets$ and $W+ \geq 4jets$.

The only additional lepton requirement not already discussed in the inclusive lepton sample sections above is that the lepton should be isolated. In order to quantify the isolation of a lepton we use a single parameter which is defined as the energy difference between the lepton and the energy in a cone surrounding the lepton divided by the lepton energy,

$$I_{so}^{elec} = \frac{E_T^{cone} - E_T^{elec}}{E_T^{elec}} ,$$

where E_T^{cone} is the calorimetry energy contained within a cone of radius $\Delta R = 0.4$

$P_T > 20\ GeV/c$
$E_{em} < 2\ GeV$
$E_{had} < 6\ GeV$
Track Stub Matching:
$\|\Delta x\|_{CMU} < 2.0$ cm
or
$\|\Delta x\|_{CMP} < 5.0$ cm
or
$\|\Delta x\|_{CMX} < 5.0$ cm
Impact Parameter < 33 mm
$\|z_{muon} - z_{vtx}\| < 5$ cm
$\|z_{vtx}\| < 60$ cm
Fiducial Requirements

Table 6.4: Inclusive muon sample requirements.

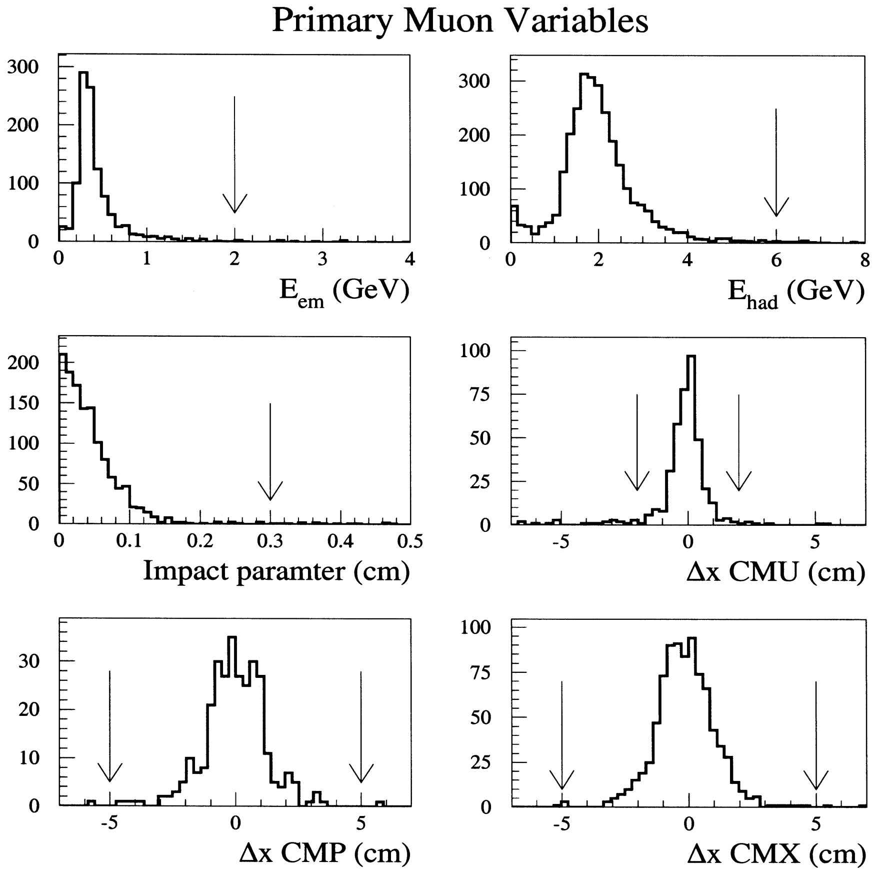


Figure 6.10: Selection cuts used to define the inclusive muon sample. This plot is of the secondary muons in a $Z \to \mu^+\mu^-$ candidate sample, additional selection for explicit Z removal will be made. Arrows indicate cut values see Table 6.4. This figure taken from reference [42].

around the lepton, and E_T^{elec} is the energy of the electron itself. An isolation of $I_{so} < 0.1$ is required for primary lepton.

Leptonically decaying W-bosons emit a neutrino which naturally escapes the detector and thereby creates a significant imbalance in the event energy. This "missing" energy, or more specifically missing transverse energy ($\not{E}_T$) can be reconstructed from known quantities in the interaction. The negative of a simple vector sum of the transverse components of energy deposited in the calorimetry yields the raw $\not{E}_T$.

Events which are consistent with Z-boson decays are removed from the W samples. In order to identify these events a looser set of cuts are applied to a secondary lepton in the event. The event is thrown out if the combined mass of the two leptons is between 75 and 105 GeV. Additionally there is a removal of all runs which were known to be "bad" runs for various reasons related to the detector performance.

6.2.4 Additional Stripping

In addition to those cuts already specific to our W subset of the inclusive lepton samples we apply three further cuts. Firstly, only the $W + 3jets$ and $W+ \geq 4jets$ samples are used, these require "high" threshold jets ($E_T > 15$ GeV and $|\eta| < 2.0$). Secondly, all multiple leptonic events are discarded preventing contamination from dileptonic $t\bar{t} \rightarrow q\bar{q}$ events (events where both W's decay to leptons). And finally, we require an additional fourth jet in our event passing looser quality cuts ($E_T > 8$ GeV and $|\eta| < 2.4$). It is worth noting again that unlike previous *CDF* $t\bar{t}$ measurements this analysis does not require the tagging of one of the b-quarks. The lack of b-tagging will have a positive effect on the overall acceptance and means that our data sample will be partially independent of the previous measurements.

6.3 Efficiencies

There are may efficiencies which need to be calculated in order to determine the total acceptance of our data selections. Fortunately many of these efficiencies are the same as efficiencies used in previous analyses [94]

$\epsilon_{\mathbf{geom}}$ This includes the efficiencies for the fiducial acceptance of the *CDF* detector. The deficiencies, such as cracks must be taken into account. Also the efficiency for being within the tracker($\epsilon_{|zvtx|}$) is included.

ϵ_{kin} The kinematic efficiencies for the standard $t\bar{t}$ analysis must be determined including: high $\not{E}_T$, high P_T lepton, and 3 high threshold jets.

ϵ_{lepID} In addition to the high P_T lepton requirement the lepton must also be isolated as described in Section 6.2.3.

ϵ_{veto} Certain types of events need to be specifically rejected. The removal of events consistent with Z-bosons have an associated efficiency, as does the rejection of dileptonic events (were both W's in the event decay to leptons).

ϵ_{trig} The triggers described in previous sections have certain acceptances associated with them.

ϵ_{tag} This efficiency is commonly seen and has to do with the efficiency for the tagging of b-quarks in the events. This efficiency is excluded in our calculation because we do not tag for bottom quarks.

All the above efficiencies except for ϵ_{trig} & ϵ_{tag} and our branching ratio are summarized into a single acceptance $A_{t\bar{t}}$. The ϵ_{trig} will be included separately. One additional efficiency must be included in our calculation and that is for our requirement of a soft fourth jet ($\epsilon_{4^{th}jet}$). This value is calculated from the $t\bar{t}$ training sample. Of 5,007 events passing other cuts 3,928 pass the fourth jet requirement. The efficiency for this cut is 0.785 ± 0.005. This compares with the b-tagging efficiencies in the previous $t\bar{t}$ lepton+jets analysis of $\epsilon_{SVX} = 0.405 \pm 0.033$ and $\epsilon_{SLT} = 0.184 \pm 0.018$.[94]

These efficiencies are summarized in Table 6.5. The total efficiency for $t\bar{t}$ events in this analysis is then,

$$\epsilon_{t\bar{t}} = A_{t\bar{t}} \cdot \epsilon_{trig} \cdot \epsilon_{4^{th} jet} \ .$$

Efficiencies

$A_{t\bar{t}}$	0.1040 ± 0.0104
ϵ_{trig}	0.9200 ± 0.05
$\epsilon_{4^{th} jet}$	0.7845 ± 0.005
Total Efficiency: $\epsilon_{t\bar{t}}$	0.0751 ± 0.008

Table 6.5: Summary of analysis efficiencies. The acceptance ($A_{t\bar{t}}$) includes the branching ratio for $t\bar{t} \to l{+}jets$.

Chapter 7

Signal Extraction & Background Removal

7.1 Fitting

Using our *ANN* we have created a data distribution which provides the maximum possible separation between signal and background. There is more than one way to extract from this distribution a characteristic number of signal events. One reasonable approach would be to set a simple cut. By examining the signal and background distributions one could find the optimal place to set such a cut, Figure 7.3. Once that cut location is set it could be applied to the data sample and thereby determine the number of signal events. Although this approach would be reasonable given the good separation between signal and background it still would not take into account shape information contained within the distributions. There is a standard method which exploits this shape information to improve discrimination between signal and background. This is by the use of a binned likelihood fit. Using distributions for signal ($t\bar{t}$) and background ($W{+}jets$), a binned maximum likelihood fit of the data to the two MC distributions is performed. Our data sample is small with several

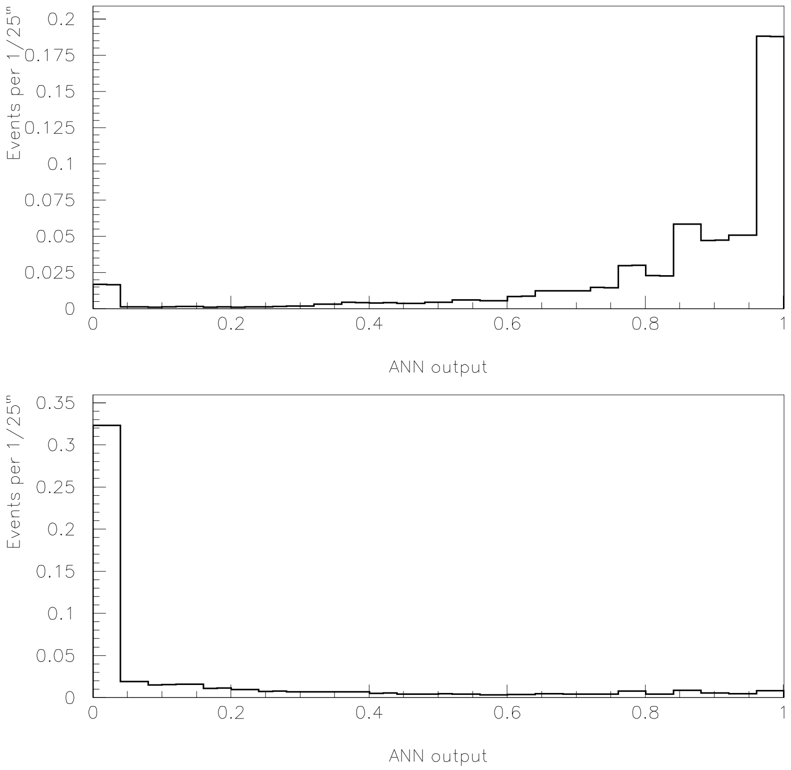


Figure 7.1: The top shows the output from the *ANN* for a sample of HERWIG $t\bar{t}$ events. The bottom shows the same for VECBOS W+*jet* events. The features in the signal distribution are statistically significant and indicative of correlations in various input kinematics. In each case 1000 different *ANN*s (training epochs) are run, the normalized average values for those runs is shown.

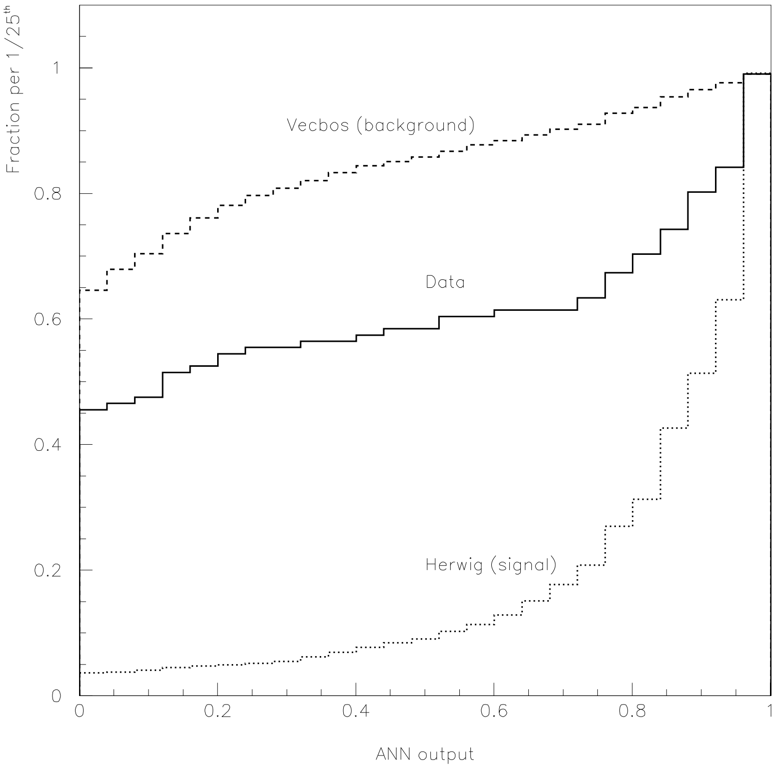


Figure 7.2: Cumulative distributions for our data output compared to the signal and background outputs we fit against.

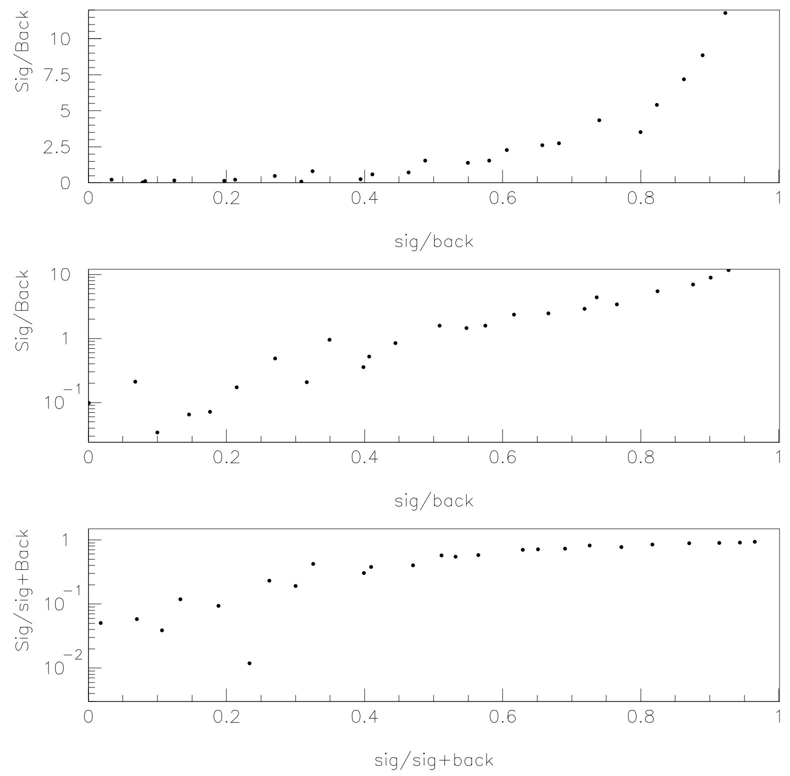


Figure 7.3: The signal content versus background content of bins is compared. Each bin in the distributions in Figure 7.1 are compared. Note that the bottom two histograms are on a logarithmic scale.

empty bins. In this case Poison statistics are normally assumed and the likelihood function used is,[73]

$$\mathcal{L}(\alpha, \beta, W_i, T_i) = \prod_{i=1}^{N} \frac{(\alpha T_i + \beta W_i)^{k_i}}{k_i!} e^{-\alpha T_i - \beta W_i} \ ,$$

where α and β are fit parameters, N is the number of bins in the distributions. The k_i is the number of data events in the i^{th} bin. T_i and W_i are the number of *signal* and *background* events respectively in the i^{th} bin of each respective Monte Carlo. Instead of trying to maximize this function the negative log of this function,

$$L(\alpha, \beta, W_i, T_i) = \sum_{i=1}^{N} \{-k_i ln(\alpha T_i + \beta W_i) + (\alpha T_i + \beta W_i)\} \ ,$$

is minimized. The factorial term is constant and can be dropped without affecting the minimization. This function will then be minimized, which amounts to the same thing as maximizing the previous function.

$$-\frac{\delta L}{\delta N_i} = 0 \ .$$

The *CERN* library package *MINUIT* is used to perform this minimization [74]. The fitting procedures used here have been tested using Monte Carlo pseudo-experiments and the test results may be found in Appendix C. The number of signal like events returned by the fitting procedure is

$$49.91 \pm 8.17 \ .$$

This includes the statistical uncertainty only.

This N_s would be the number of $t\bar{t}$ events in our data sample if we did not already know that the data sample has non-W background contamination. The extraction of this non-W background contamination from N_s will be the subject of the remainder of this chapter.

7.2 Non-W Background Removal

The removal of our primary background (W+$jets$) happens organically in the course of the fitting of the data to the Monte Carlo distributions. The background distribution used in the fitting above includes only the W+jet events, other backgrounds, though small by comparison, are not negligible, Table 7.3. There are two basic procedures used for the non-W background removals. Both methods for the subtraction of the non-W backgrounds are relatively simple. The second method serves as a cross check on the former.

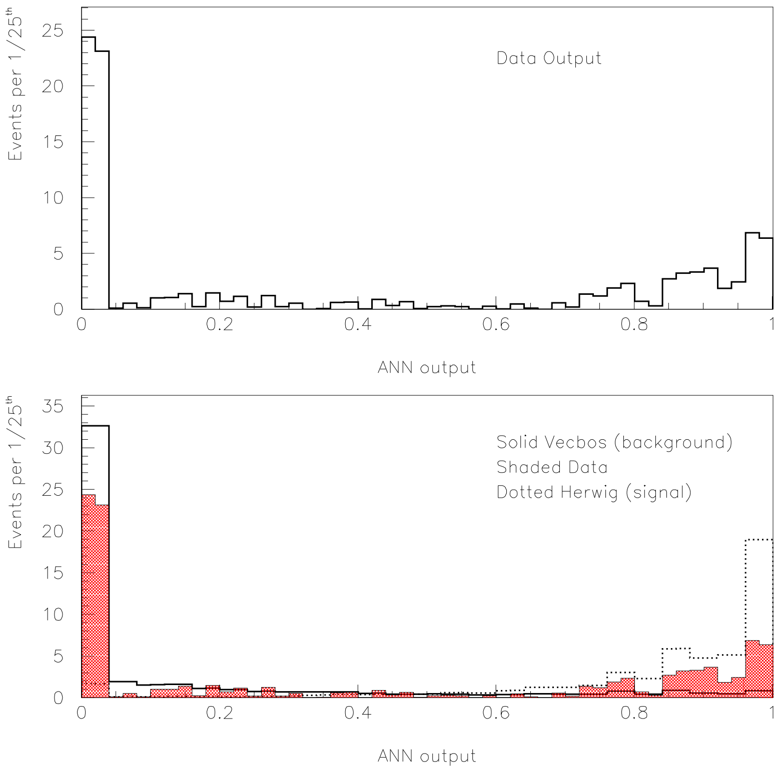


Figure 7.4: The top shows the output from the *ANN* for the passing data sample and the number of "background like" events. The bottom shows the number of "signal like" events and the statistical error on that number. In each case 1000 different *ANN*s (training epochs) are averaged.

Any given background will contaminate in some measure the number of signal events determined above. Although we can estimate the number of non-W background events in our data sample it is not trivial to then determine how many of those events are designated by the fitting procedure to be signal events. Previous kinematic $t\bar{t}$ analyses have neglected this because the backgrounds were more $W+jets$ like than $t\bar{t}$ like, but this can only be said to be the case in analyses where the variables' inter-relationships are well understood. This need not necessarily be the case, and is certainly not true in this analysis.

The procedure devised here for removing these backgrounds is fairly straight forward. Samples of each background are generated as was described in Chapter 6. The sample will then be treated just as our data sample was in the previous section: stripped, put through our *ANN*, and then fit against the $W+jets$ and $t\bar{t}$ distributions. A number of background events consistent with signal like events is determined, this number is then scaled down to make it consistent with the expected number of events in the stripped data sample. This same procedure is followed with each of the non-W backgrounds.

7.2.1 Dibosons

A sample of diboson events(WW, WZ, ZZ) were generated using ISAJET[69]. The standard model cross sections for these processes are,[4]

$$\sigma(p\bar{p} \rightarrow WW) \approx 9.5\ pb$$

$$\sigma(p\bar{p} \rightarrow WZ) \approx 2.5\ pb$$

$$\sigma(p\bar{p} \rightarrow ZZ) \approx 1.0\ pb\ .$$

The luminosity of our data set is 86.3 pb^{-1}. The efficiencies for these dibosons can be determined by putting the full sets through our stripping cuts. From this information the number of diboson events in our data sample may be determined. Of the generated events 546 events pass selections, and are sent through the *ANN* and the fitter. Of these 131, or 24%, are tagged as signal like, Figure 7.6. Therefore the estimate for this background will be scaled by 0.24.

Determining the estimate for this background must be done carefully. There are several factors which must be taken into account, the calculation inputs are summarized in Table 7.1. The Next to Leading Order (NLO) correction factors can be found in reference [75]. The justification for combining various structure functions may be found in Chapter 6. Given the diboson background estimate of 9.61 events in the stripped data set, the correction factors, and luminosity, it is possible to estimate the number of diboson events which contaminate the estimated number of signal events (N_S) found in Section 7.1.

Parton Distribution Function	**Process**	**$\int \mathcal{L}$dt of MC (pb^{-1})**	**Passing**	**NLO Corr. Factor**	**Corr. Passing**	**Est.for 86.3 pb^{-1}**
MRSD0	WW	2154	122	1.5	183	7.33
	WZ	5169	34	2.95	100.3	1.67
	ZZ	6590	23	1.45	33.4	0.44
	$Total$	–	179	–	–	9.44 ± 1.51
CTEQ2L	WW	2343	138	1.5	207	7.62
	WZ	5819	30	2.95	88.5	1.31
	ZZ	6843	19	1.45	27.6	0.35
	$Total$	–	187	–	–	9.28
GRVL0	WW	2373	132	1.5 ± 0.3	198	7.20
	WZ	5696	31	2.95 ± 0.59	91.5	1.39
	ZZ	7149	17	1.45 ± 0.29	24.7	0.30
	$Total$	–	180	–	–	8.89
Combined	–	–	546	–	–	9.20 ± 0.91

Table 7.1: Combined diboson sample, consisting of three processes and three structure functions. The Next to Leading Order(NLO) correction factors can be found in Ref.[75]. The source of the uncertainty on the final value is the uncertainty on the ISAJET correction factors. The statistical uncertainty on the samples is taken into account later.

The 546 diboson events are then put through the same *ANN* used for the data, yielding the output distribution shown in Figure 7.5. The output distribution is fit against the same samples that the data was fit against, Figure 7.1. By following this procedure once again a number of signal like events is determined. The ratio of signal like to background like events may then be multiplied by the number of expected dibosons, Figure 7.6.

$$N_{cont} = \frac{N_S^{cont}}{N_{total}^{cont}} N_{exp}^{cont} \ . \tag{7.1}$$

The N_S^{cont} is the number of signal like events returned by the fitter. The N_{total}^{cont} is the total number of diboson events fit to, 546 for dibosons. Of the 546 total diboson events 136.1 ± 14.8 are signal like. And the N_{exp}^{cont} is the number of expected diboson events in the data sample (9.20 ± 0.91). Therefore the number of dibosons contaminating the previously determined N_s is

$$N_{cont}^{diboson} = 2.29 \pm 0.34 \ .$$

7.2.2 Drell-Yan

A Drell -Yan sample has been generated with ISAJET as discussed in Chapter 6. Using the correction factors and luminosity information in Table 7.2 the estimated number of Drell-Yan events in our data may be calculated.

Parton Distribution Function	**Process**	**$\int \mathcal{L}dt$ of MC (pb^{-1})**	**Passing**	**NLO Corr. Factor**	**Corr. Passing**	**Est.for 86.3 pb^{-1}**
MRSD0	$DY\ \gamma$	431.8	7	2.19	15.33	3.07
	$DY\ Z^o$	579.7	42	1.79	75.18	11.20
	Total	–	49	–	–	14.25
CTEQ2L	$DY\ \gamma$	483.7	10	2.16	21.6	3.85
	$DY\ Z^o$	620.3	57	1.94	110.6	15.39
	Total	–	67	–	–	19.24
GRVL0	$DY\ \gamma$	501.7	13	2.07	26.91	4.63
	$DY\ Z^o$	*N/A*	–	–	–	–
	Total	–	13	–	–	4.63
Combined	–	–	129	–	–	17.14 ± 3.19

Table 7.2: Combined Drell-Yan sample, consisting of three processes and three structure functions. The Next to Leading Order(NLO) correction factors can be found in Ref.[75]. The source of the uncertainty on the final value is the uncertainty on the ISAJET correction factors. The statistical uncertainty on the samples is taken into account later.

An output distribution of the Drell-Yan sample is generated by putting the events through the same *ANN* as the data and diboson samples. This is once again fit to get a number of "signal like" result for Drell-Yan, Figure 7.7.

This then provides the scaling for the number of Drell-Yan in the sample as in Eq.7.1. Of the 129 diboson events 20±6.3 are signal like. The structure functions add an additional 2% uncertainty to the number of signal like events. The total number of signal like Drell-Yan events is then estimated to be,

$$N_{cont}^{DY} = 2.66 \pm 0.97\ .$$

7.2.3 QCD Contamination

The procedure for our HERWIG generated QCD events was the same as that for the proceeding two cases. A sample with a total luminosity of 2107 pb^{-1} was generated. Of these events 253 passed our cuts. A HERWIG estimated number of 10.8 events should be in our data set. The estimate from HERWIG is then corrected for the known ratio of cross sections for HERWIG $QCD(2 \to 2)$ versus the same value for the Next-to-Leading Order (NLO) calculation giving a correction factor of 1.43 ± 0.43. [76] This value is then compared to an estimate based on data alone done previously with a subsample (67.3 pb^{-1}). [92] The weighted mean of these two values yields a total QCD estimate of 14.07 ± 3.05.

The QCD output distribution is then fit again against the same $W + jets$ and $t\bar{t}$ distributions, Figure 7.1. Of all the backgrounds QCD yield the highest rate of signal contamination, this can be understood only in the context of comparing the entire input pattern to the inputs used in training the neural network. The fit returns 225.8 ± 15.8 signal like events, Figure 7.8. Using Eq.7.1 the contamination of N_S by QCD events is estimated to be,

$$N^{QCD}_{cont} = 12.56 \pm 3.05 \ .$$

7.3 Summary: 1^{st} Method of Background Removal

The summary of non-W background contamination is shown in Table 7.3. The errors shown here include statistical errors due to limited sample sizes for each Monte Carlo set. The total number of $t\bar{t}$ events is then this background subtracted quantity,

$$N_{t\bar{t}} = 32.5 \ . \tag{7.2}$$

This $N_{t\bar{t}}$ is the number of signal events in our data without systematic errors. The final errors on the measurement of the backgrounds are shown in Table 7.3. This non-W background total is subtracted from the N_S from Section 7.1 to get the $N_{t\bar{t}}$ above.

Background	**Events Expected**	**Events in N_s**
Dibosons (WW, WZ, ZZ)	9.20 ± 0.97	2.29 ± 0.34
Drell-Yan	17.14 ± 3.19	2.66 ± 0.97
QCD/Heavy Quark	14.07 ± 3.27	12.56 ± 3.05
Total	40.41 ± 4.67	17.51 ± 3.22

Table 7.3: Summary of backgrounds in the data sample and the contamination of the N_s for subtraction. The error will be shown in Chapter 8 as MC statistics and corrections.

7.4 2^{nd} Method of Background Removal

In the first method of background removal the data was first fit. When N_S was determined the amount of contamination was found and this contamination was the subtracted. The second method essentially reverses the process.

As in the first case output distributions for the data and each of the three non-W backgrounds are generated. The background distributions are scaled down to the expectation value for their occurrence in our data sample, these values are shown in Table 7.3. After scaling, the background distributions are subtracted from the data distribution. The new output distribution (with backgrounds subtracted) is fit against the same MC distributions used previously. The resulting number of signal like events extracted during the fit is already background subtracted. Therefore the N_S from this fit is equivalent to the $N_{t\bar{t}}$.

$$N_{t\bar{t}} = 33.11 \pm 6.68 \tag{7.3}$$

This shows very good agreement with the previous method, Equation 7.2. The data (136 passing events) is shown in Figure 7.9, and this is compared to the cumulitive contributions, as determined by the fit, of signal and backgrounds.

7.5 Three Component Fit

As described above we subtracted the secondary backgrounds from our data sample before (or after) doing an unconstrained two component fit. The fit is then only to $t\bar{t}$ signal and our primary background (W+$jets$). It is possible to do a multiple component fit to other secondary backgrounds. Since QCD is the primary contaminating background to our signal we attempt a three component fit of the data, minus dibosons & Drell-Yan, to the usual $t\bar{t}$ & W+$jets$ distributions but also including a third QCD distribution. This QCD component is Gaussian constrained to our background estimate for QCD. The constraint on QCD is necessary due to its similarity to the $t\bar{t}$ output distribution, Figure 7.10. The result of this three component method is,

$$32.91 \pm 7.76 \; (stat.) \; ,$$

events are found to be $t\bar{t}$. The increased statistical error with this method is to be expected due to the similarity between the signal and QCD output distributions.

7.6 Cross Section Calculation

Once the number of signal events is determined, extracting the cross section becomes a simple matter. The efficiencies ($\epsilon_{t\bar{t}}$) for our data selection are shown in Chapter 6. The total integrated luminosity ($\int \mathcal{L}dt$) in Run IB was discussed in Chapter 2 [93]. These can be combined to determine the total $t\bar{t}$ cross section,

$$\sigma_{t\bar{t}} = \frac{N_{t\bar{t}}}{\epsilon_{t\bar{t}} \cdot \int \mathcal{L}dt} \tag{7.4}$$

The resulting cross section with statistical errors only is,

$$\sigma_{t\bar{t}} = 5.1 \pm 1.0 \; (stat.only) \; pb \; .$$

The systematic uncertainties are the subject of the next chapter.

The figure above compares the contributions (as found by the 3 component fit) to the overall selected data sample. Figure 7.11

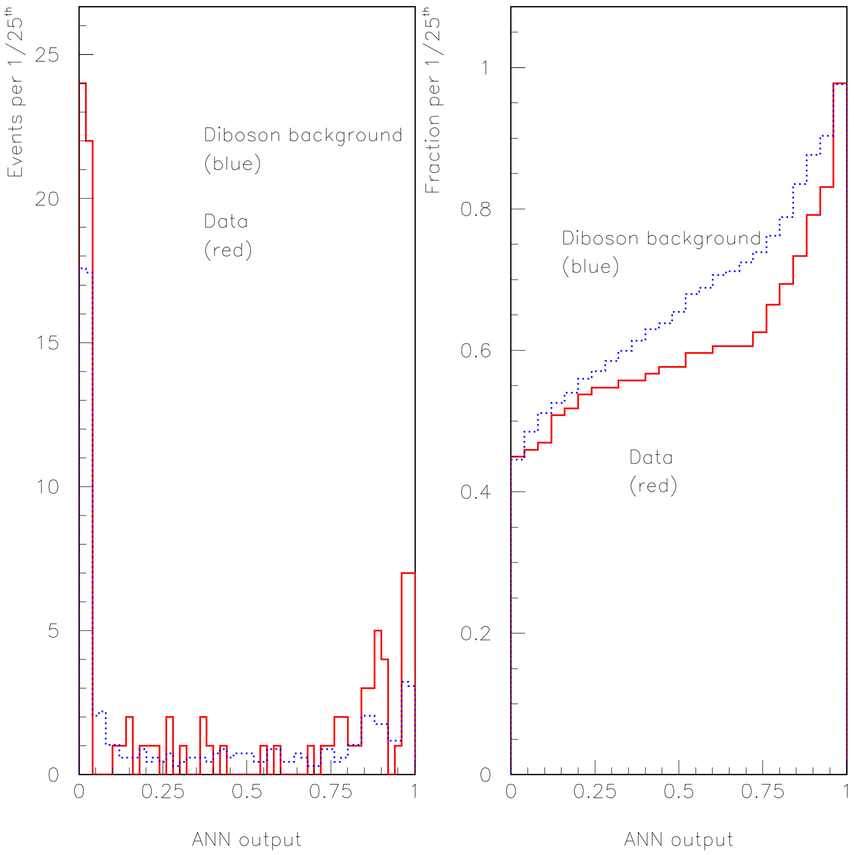


Figure 7.5: The outputs from the diboson sample and data sample are compared on the left. The same distributions are compared in a cumulative fashion on the right.

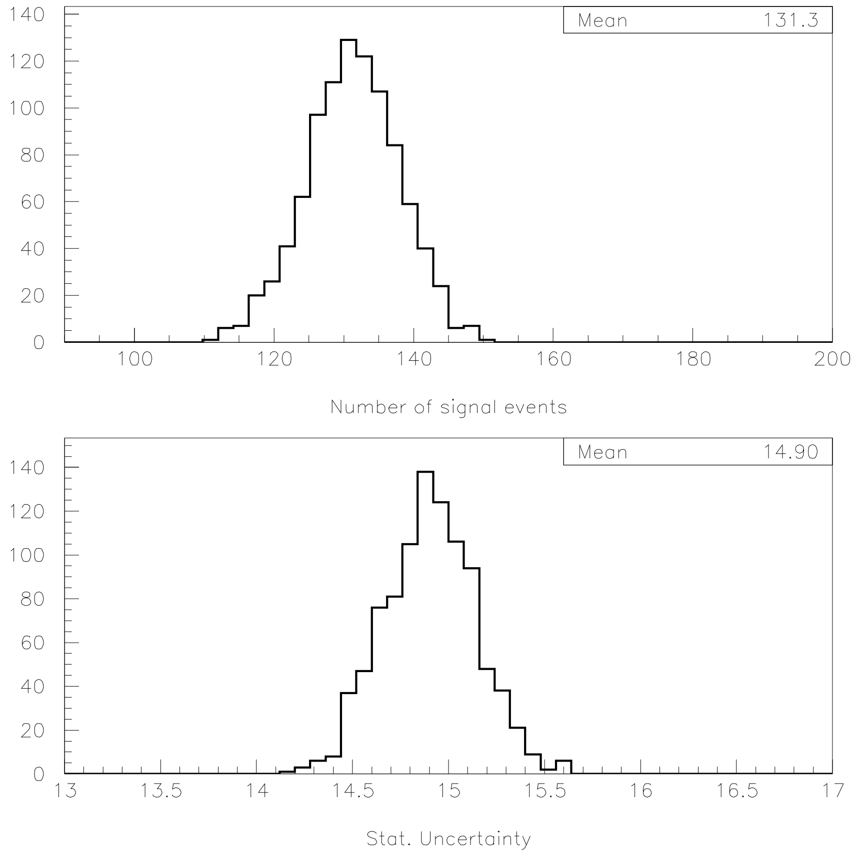


Figure 7.6: The top plot shows the number of “signal like” events in the diboson sample according to the the fitter. The lower plot shows the statistical uncertainty on the number of events. In each case 1000 different *ANN*s (training epochs) are shown.

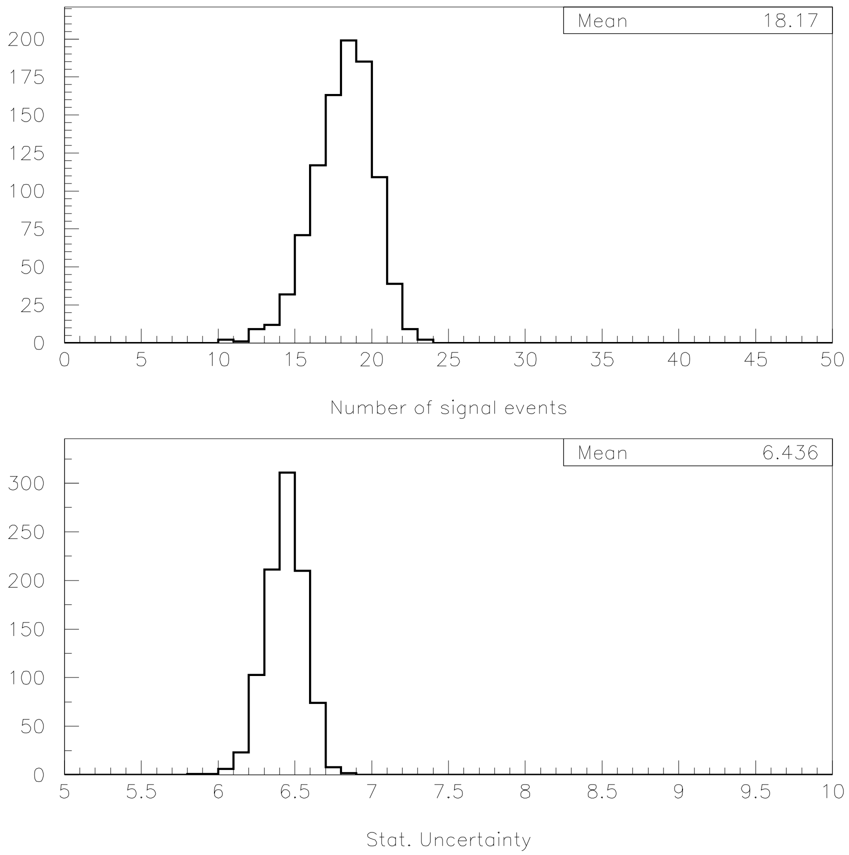


Figure 7.7: The top plot shows the number of "signal like" events in the Drell-Yan sample according to the the fitter. The lower plot shows the statistical uncertainty on the number of events. In each case 1000 different *ANN*s (training epochs) are shown.

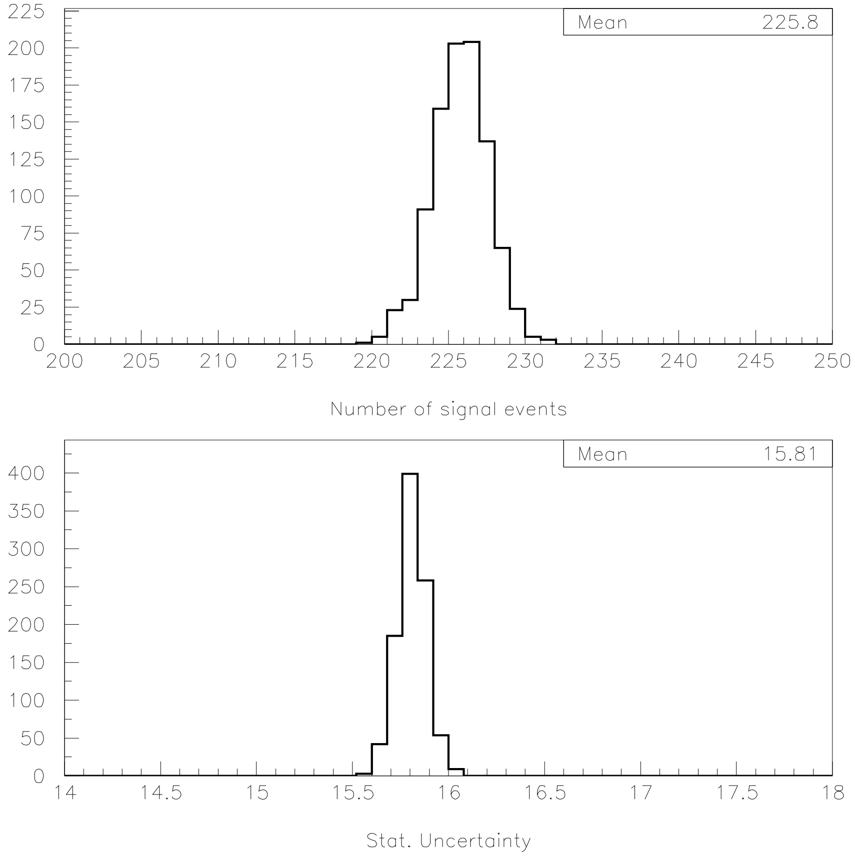


Figure 7.8: The top plot shows the number of “signal like” events in the QCD sample according to the the fitter. The lower plot shows the statistical uncertainty on the number of events. In each case 1000 different *ANN*s (training epochs) are shown.

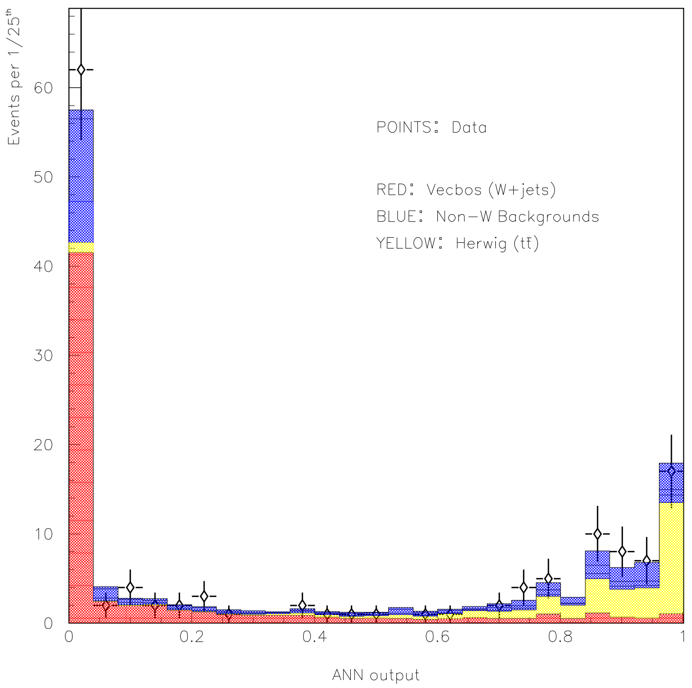


Figure 7.9: The data output compared to determined signal and background component contributions.

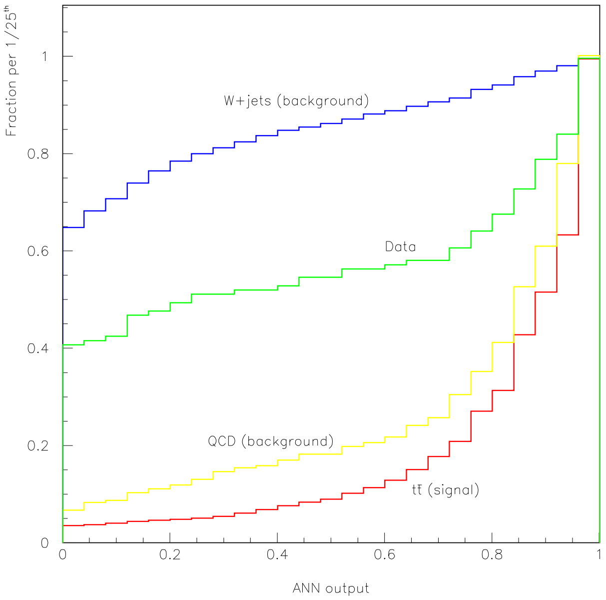


Figure 7.10: Cumulative distributions for our data output, HERWIG $t\bar{t}$, VECBOS W+$jets$, and HERWIG QCD. are compared.

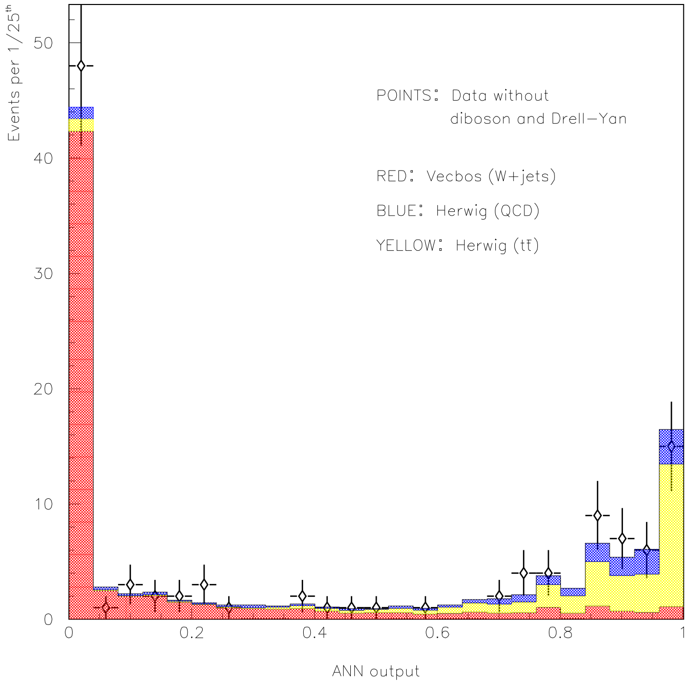


Figure 7.11: The data output, with diboson and Drell-Yan subtracted, compared to determined signal and background component contributions.

Chapter 8

Systematics

Systematic errors have already been adjusted for in the selection and reconstruction of the data samples. As an example of this the jet energies are known to be systematically mis-measured in the detector. The jets appearing in our events are “corrected” to account for this systematic error. Although systematic *errors* are usually dealt with in this fashion not all systematic effects can be eliminated in this way because some effects are not completely understood (not certain). Therefore it is important to quantify these systematic uncertainties in order to give relevance to the measurement.

No analysis can calculate every associated systematic uncertainty. However this is not necessary, if the primary systematics are taken into account then remaining uncertainty may be considered negligible. This analysis has six dominate uncertainties: (1) detector jet energy scale, (2) limited MC statistics & corrections, (3) initial and final state radiation, (4) proper Q^2 scale, (5) integrated luminosity, and (6) efficiencies(trigger, lepton id, etc...).

8.1 Method

It is not possible to outline a completely general method to deal with the systematics. However most of the systematics will be dealt with in similar ways. Usually the calculation of the systematic will involve comparing the results when a systematic parameter is shifted by $\pm 1\sigma$ with the results when the parameter is unshifted. Because other cross section measurements have been made with similar data sets some of these uncertainties have been measured before and we will certainly take advantage of that fact.

8.1.1 Efficiencies

The systematic on our efficiencies including: triggers, lepton identification, kinematics, etc... have been calculated previously as has been discussed in Chapter 6. These uncertainties have already been shown in Table 6.5. These calculations are are based on PYTHIA & HERWIG studies [94].

8.1.2 Integrated Luminosity

A substantial amount of work has been done to calculate the uncertainty on the integrated luminosity. The uncertainty on the luminosity has two sources, the Beam-Beam counters cross section (σ_{BBC}) and the normalization (N_{BBC}),

$$L = \frac{N_{BBC}}{\sigma_{BBC}} \ ,$$

this may be compared to Equation 2.1. The total uncertainty on Run IB luminosity is 4.1% [93].

$\|\eta\|$ **interval**	**Relative Corr. Uncertainty**
0.0 - 0.1	2%
0.1 - 1.0	0.2%
1.0 - 1.4	4%
1.4 - 2.2	0.2%
2.2 - 2.6	4%

Table 8.1: The percentage of the relative jet energy correction uncertainty for varying $|\eta|$.

8.1.3 Jet Energy Scale

The jet energy scale uncertainty comes from the systematic corrections applied to the jets in the samples of both data and Monte Carlo. These corrections are described in detail in Appendix A. This uncertainty has several components related to the different aspects of jet energy which must be corrected. The various jet correction uncertainties are applied to the data in two different parts. The relative correction (the η dependent correction) is applied to the data and then fit against the same distributions which the data was fit against. The result of this fitting is then compared to data the difference is taken as 1σ. The $+1\sigma$ and -1σ uncertainties will be applied separately. The fit

results of these two cases ($\pm\sigma$) are shown in Figure 8.1. The relative corrections applied are summarized in Table 8.1.

Other jet correction uncertainties are dominated by the absolute energy scale and soft gluon radiation. These two corrections are applied together to the data. The values of theses corrections come from a parameterization of the Behrends curves[96]. Once again the correction uncertainties are applied in both the $+1\sigma$ and -1σ directions. As with the relative uncertainty the adjusted data is fit against the usual signal & background distributions. The fit results are shown in Figure 8.2.

8.1.4 Monte Carlo Statistics

There are two sources of Monte Carlo statistic uncertainties. The first comes from the finite statistics of the non-W background distributions, these are shown in Table 7.3. The other source is from the training samples themselves. In order to estimate this systematic 1000 separate ANNs were trained and our data sample was then put through these networks. The outputs of these networks are then fit, the data against the MCs. The number of signal events (N_S) for these fits are plotted in Figure 8.3. The width of this distribution we estimate to be the systematic uncertainty on the network training.

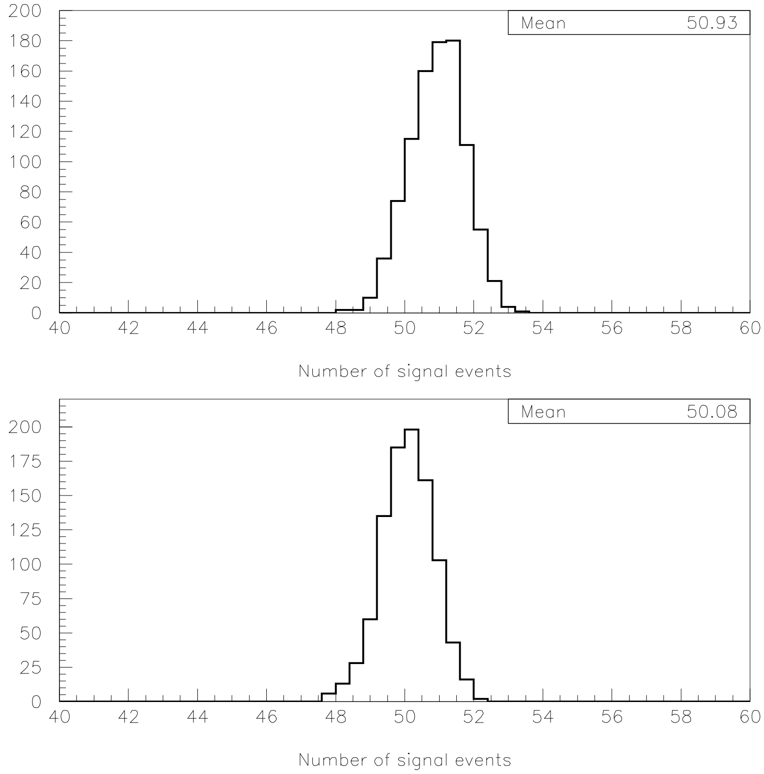


Figure 8.1: The top shows the fit of the data to the usual signal and background distributions when the relative jet correction has been shifted by $+1\sigma$. The bottom plot shows the same quantity when the shift is -1σ. In each case 1000 different *ANN*s (training epochs) are shown.

8.1.5 Initial & Final State Radiation (ISR & FSR)

There is substantial systematic uncertainty associated with the modeling of the *ISR & FSR* in our HERWIG $t\bar{t}$ events. The $t\bar{t}$ events used in this analysis are generated using HERWIG. In HERWIG *ISR & FSR* can not be turned off. Therefore we use PYTHIA to estimate this systematic uncertainty. Two independent PYTHIA $t\bar{t}$ samples are generated, one with *ISR & FSR* the other without. These samples are then used to create pseudo-experiments. Each pseudo-experiment used is a combination of $t\bar{t}$ and VECBOS generated $W + jets$ events. We then fit the data to the usual MC samples. In this way we create two distributions of N_S and take the difference

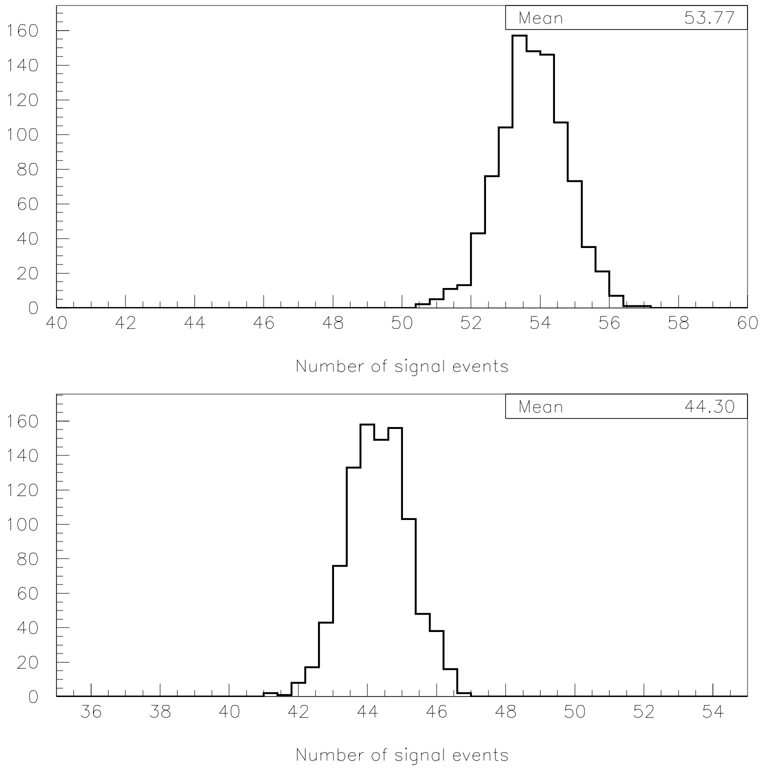


Figure 8.2: The top shows the fit of the data to the usual signal and background distributions when the absolute and soft gluon radiation jet correction has been shifted by $+1\sigma$. The bottom plot shows the same quantity when the shift is -1σ. In each case 1000 different *ANN*s (training epochs) are shown.

between these distributions to be half our range of error this is assumed to be a flat distribution and therefore divided by $\sqrt{12}$, this is then taken as 1σ uncertainty on the *ISR & FSR*, Figure 8.4.

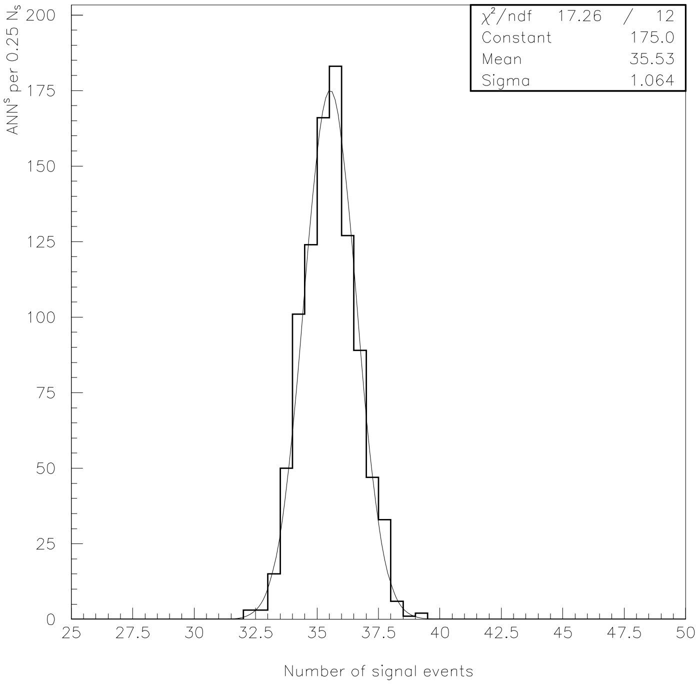


Figure 8.3: This plot shows 1,000 pseudo-experiments. Each pseudo-experiment represents a different training of the *ANN*. Each pseudo-experiment is trained using the same general network parameters and architecture.

8.1.6 Momentum Transfer Scale

There is a systematic uncertainty associated with the appropriate Q^2 scale used in the generation of VECBOS background events. Q^2 is the measure of the momentum transfer in the collision of the original partons in the event. A Q^2 of P_T^2 is used in the samples used for *ANN* training. In order to determine the magnitude of the uncertainty from the choice of Q^2 we set it to M_W^2 and compare the results in that case with the result for the P_T^2 case.

Independent Vecbos samples are generated where Q^2 is set to P_T^2 and M_T^2 respectively. From these sets pseudo-experiments are generated and combined with signal events. These pseudo-experiments are put through our ANN and the output distributions are then fit against the same signal & background distributions used to fit the data, Figure 7.1. The fits produce results shown in Figure 8.5. The difference in the amount of signal produced by each of the two sets yield the systematic uncertainty on the Q^2 scale.

8.2 Summary of Uncertainties

The dominate systematic uncertainties of this analysis are summarized in Table 8.2. Of these systematics there is no single systematic which seems to dominate the overall uncertainty. The systematic uncertainties are summed in quadrature to get the total systematic uncertainty. The resulting total systematic uncertainty on the number of $t\bar{t}$ events is,

$$N_{t\bar{t}} = 33.1 \pm 6.7\ (stat)\ ^{+6.3}_{-6.8}\ (syst)\ . \tag{8.1}$$

This naturally does not include the efficiency and luminosity uncertainties. These uncertainties will be included in the cross section uncertainty in Chapter 9.

Systematic	Value of Uncertainty	
Jet Corr.:		
relative	$+2.0\%$	$+0.3\%$
absolute	$+7.7\%$	-11.3%
Total	$+8.0\%$	-11.3%
ISR	$\pm 6.2\%$	
MC stat. & Corr.		
non-W backgnds	$\pm 9.7\%$	
training	$\pm 2.9\%$	
Total	$\pm 10.1\%$	
Vecbos Q^2 scale	$\pm 12.3\%$	
Total syst. ($N_{t\bar{t}}$)	$+18.9\%$	-20.5%
Efficiencies ($\epsilon_{t\bar{t}}$)	$\pm 10.7\%$	
Luminosity ($\mathcal{L}$)	$\pm 4.1\%$	
Total syst. ($\sigma_{t\bar{t}}$)	$+22.1\%$	-23.5%

Table 8.2: Summary of the systematic uncertainties on the measurement of the $t\bar{t}$ cross section $\sigma_{t\bar{t}}$.

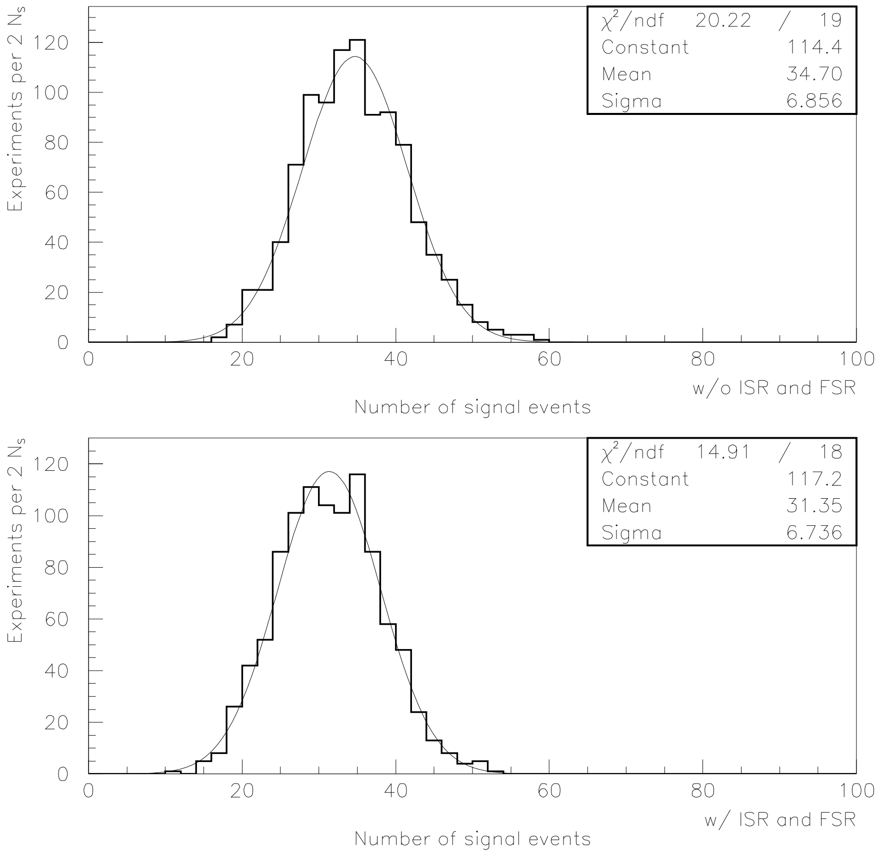


Figure 8.4: Each plot shows 1,000 pseudo-experiments. Each pseudo-experiment is a combination of PYTHIA ($t\bar{t}$) and VECBOS ($W + jets$). The $t\bar{t}$ events used in the top plot have *ISR & FSR* turned off. The $t\bar{t}$ events used in the bottom plot have *ISR & FSR* turned on.

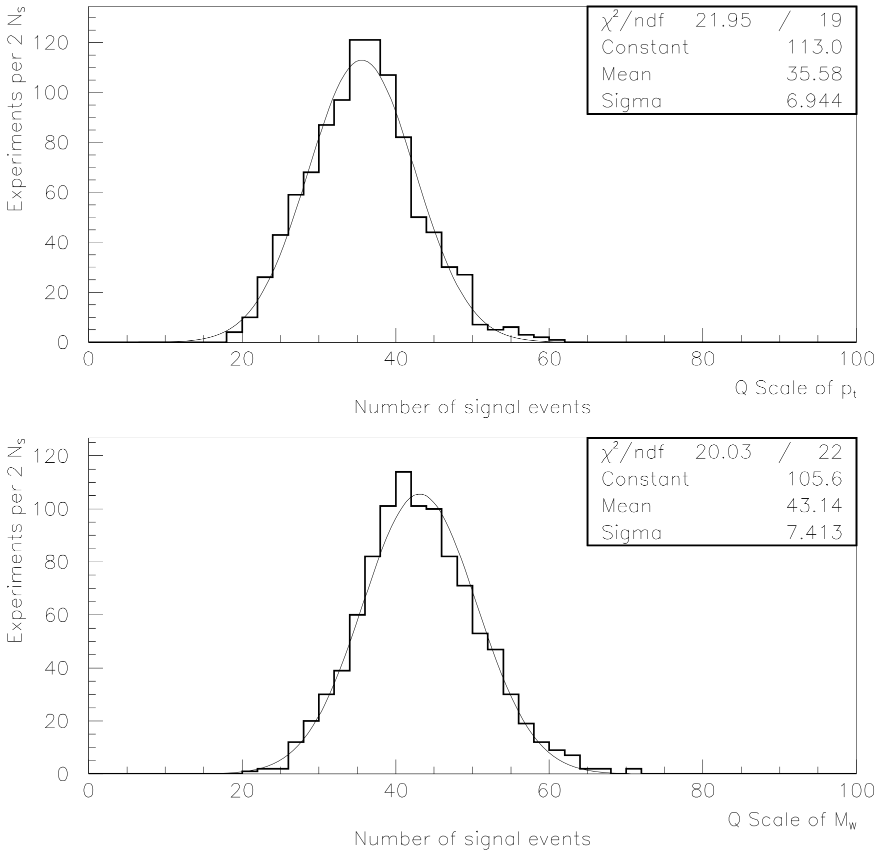


Figure 8.5: Each plot shows 1,000 pseudo-experiments. Each pseudo-experiment is a combination of HERWIG ($t\bar{t}$) and VECBOS ($W + jets$). The $W + jet$ events used in the top plot where generated with $Q^2 =< P_T^2 >$. The $W + jet$ events used in the bottom plot where generated with $Q^2 = M_W^2$.

Chapter 9

Conclusion

We have performed a new measurement of the $t\bar{t}$ cross section for $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$. This measurement was performed in the *lepton+jets* channel at *CDF* on the Run IB data set ($90\ pb^{-1}$). Unlike previous measurements we employed an artificial neural network to distinguish our signal, utilizing a large number of kinematic variables in the process. Our lack of b-tagging has provided a significant gain in acceptance over the earlier *lepton+jets* analyses performed at *CDF*.

In the $90\ pb^{-1}$ Run IB sample with 136 events passing our selection criteria we have calculated

$$33.1^{+9.1}_{-9.5}$$

events to be $t\bar{t}$. When this value is put into Equation 7.4 the total cross section for $p\bar{p} \to t\bar{t}$ is found to be

$$\sigma_{t\bar{t}} = 5.1^{+1.5}_{-1.6}\ pb. \tag{9.1}$$

This is in very good agreement with published $t\bar{t}$ cross section measurements. In addition, it matches the standard model predictions, Figure 9.1, 9.2.

The technique developed in this analysis will be useful in Run II for improving the current values of $\sigma_{t\bar{t}}$. In the coming run an estimated $2\ fb^{-1}$ of data will be taken at $\sqrt{s} = 2.0\ TeV$. In addition to this 20+ fold improvement in statistics there are detector systematics which are very likely to improve. The Monte Carlo systematics are more uncertain, but they too may be improved in future calculations. Run 2 measurements are likely to yield uncertainties which are compatible with the current theoretical uncertainties even without taking into account neural network measurements. *ANN* measurements in this and other channels are likely to have a significant effect on this measurement, and may result in uncertainties which are significantly less than the theoretical values.

Clearly, these techniques are extendible to other measurements which were not easily accessible at Run IB statistics. In fact for certain processes (Ex: single top production) *ANN* techniques will probably be the most important techniques to be employed.

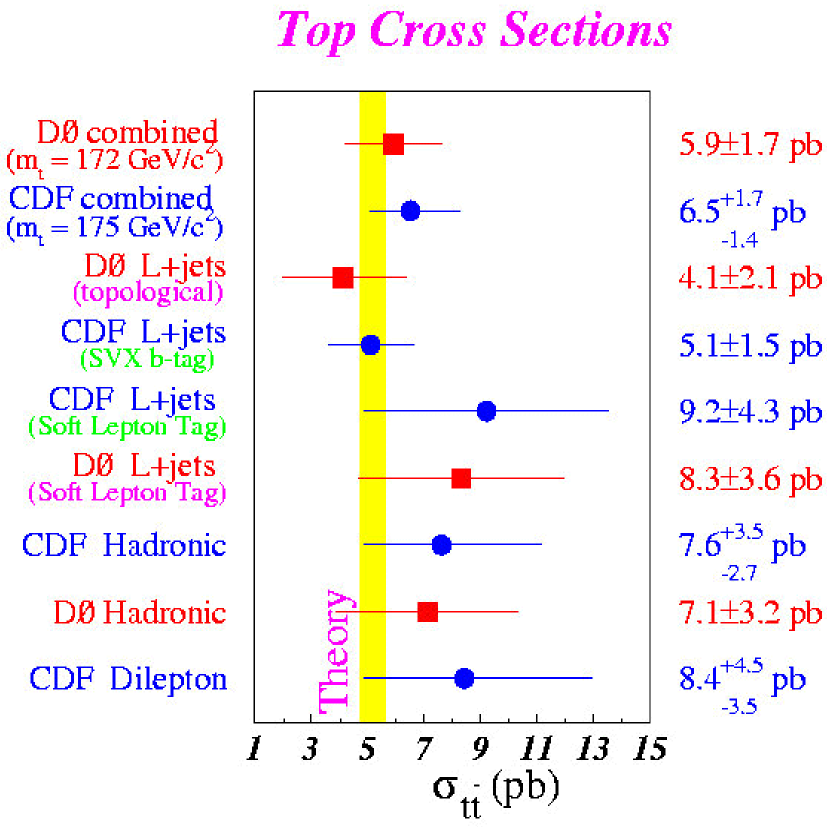


Figure 9.1: The measured $t\bar{t}$ cross sections in all channels at all experiments [97]. Our measurement is not shown, it is most similar in technique to the D topological measurement in lepton+jets.

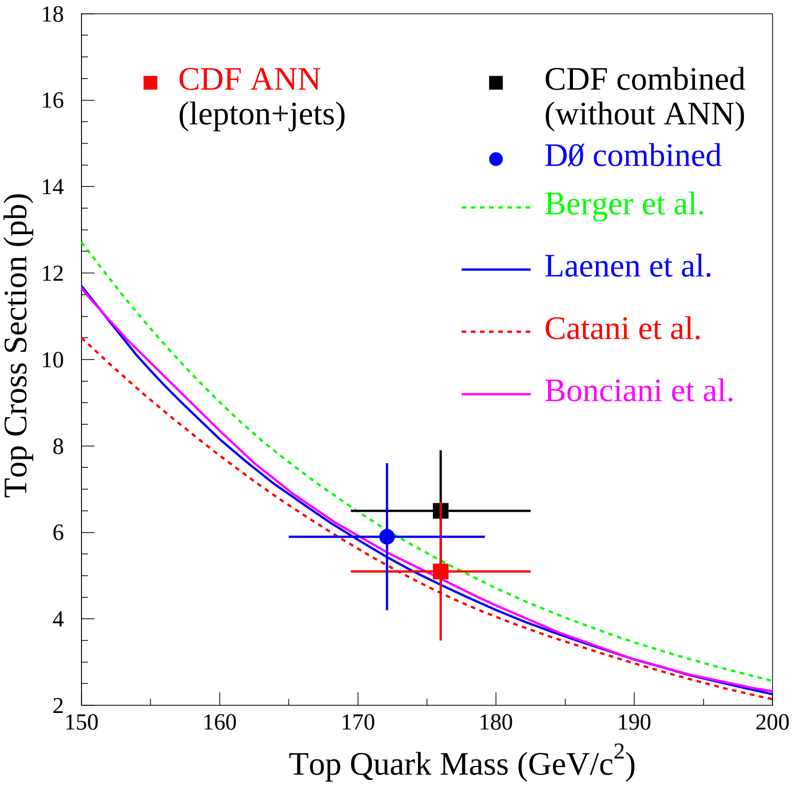


Figure 9.2: The measured $t\bar{t}$ cross sections in this (ANN) analysis is shown in red. The combined measurements for all channels at CDF & DØ are shown in black & blue respectively. Once again calculated cross section and masses are shown.

Epilogue

There has already been some discussion of what may lie ahead in Run 2, but it is probably worth relating some ideas on how this work might proceed and differ for Run 1.

Now that this analysis is over there are several issues which, with hind-sight, may be worth revisiting, these fall basically into two categories: network optimization and input patterns. It is not clear whether more work on these issues will improve the given results and re-examining them is only worthwhile in the context of the necessity of repeating the analysis on Run 2 data.

The effects of re-optimization are hard to guess, but the first step in this process would be the generation of larger and more varied Monte Carlo samples. Larger MC samples would allow the investigation of many alternative network architectures. The application of separate networks to distinguish various secondary backgrounds is a possible option, it should be remembered that this may adversely impact the systematic uncertainties. Generation with different MCs might be of value (Ex.- PYTHIA $t\bar{t}$). One possible change in the optimization which is particularly interesting is to focus on certain systematics and optimize to reduce these. Unfortunately, there is no single dominate systematic to focus on in an effort to drive the overall uncertainty down by optimization.

The issue of input pattern may hold more promise than optimization in the attempt at improving this measurement. The most obvious place to start as far as inputs are concerned is a reduction in QCD background. The *ANN* as it now stands not only removes the $W{+}jets$, but also removes over eighty percent of the diboson and Drell-Yan backgrounds. Therefore, the trick in any effort to reduce QCD is to do so without harming the current performance of the network, this may prove impossible.

The first step toward reducing the QCD background will be to identify kinematic variables which discriminate QCD from $t\bar{t}$. Another systematic which could possibly be reduced is jet corrections. Decreasing the ANN s dependence on jet energies may have a positive effect on the jet systematic uncertainty. The jet systematic may well be driven down in Run 2 through the efforts of other groups. The reduction of the uncertainty on Q^2 scale may be possible by examining other event generators for $W{+}jet$ events, this however must be done in consultation with experts in this area.

Finally, it is worth noting that our current systematics are comparable to our statistical errors. Working in Run 2 to keep the systematics competitive with the statistics will be difficult, however the fact that several of our systematics are competitive for dominance gives more opportunities to be successful at this effort.

Appendices

Appendix A. Jet Energy Corrections

Appendix B. b-tagging Comparison

Appendix C. Kinematic Analysis - without ANN

Appendix D. JETNET

Appendix A

Jet Energy Corrections

Before jets are used to calculate the kinematic variables used by the *ANN* they are put through certain jet corrections. The corrections take into account a wide variety of detector effects. These effects include things like absolute energy loss in the calorimetry, underlying event effects, cracks in the detector, and energy loss outside the jet "cone".

Relative jet energy scale correction:(f_{rel}) The relative jet energy scale correction corrects for the variation in detector response. The detector response varies with the η value of the jet.

Underlying event correction:(UEM) This correction takes into account the extra energy in events due to multiple interactions. For single interactions UE is applied for the primary vertex.

Absolute jet energy correction:(f_{abs}) The absolute jet energy correction takes into account the overall limitations on detector response in the calorimetry and fragmentation effects. Fragmentation effects account for the fact that the originating particles are fragmenting in the calorimeter and are then observed as jets.

Outside of the jet cone correction:(OE) This correction corrects for the energy lost outside of the acceptance cone of the jet ($\Delta R = 0.4$). This is due to cone size and to soft gluon radiation (splash-out). This splash-out portion of the correction dominates.

These corrections are applied to the jets in a prescribed way and are dependent on cone size used. This analysis uses a cone size of 0.4.

$$P_T^{corr}(R) = [P_T^{raw}(R) \times f_{rel} - UEM(R)] \times (R) f_{abs}(R) - UE(R) + OE(R)$$

The values P_T^{corr} and P_T^{raw} are the corrected and raw P_T's of the jet.

The missing transverse energy ($\not{E}_T$) is corrected to take into account the these jet corrections, as well as correction related to μ mismeasurements.

Appendix B

b-tagging Comparison

As a cross check to our *ANN* method of signal identification we examine the b-tagged events in our sample. It is expected that these events will, if our network is doing its job, appear more "top like" than the general event sample of 136 passing events. Therefore outputs from the events with *SVX* and *SLT* tags are compared to the output distribution for the whole sample, Figure B.1. The *SVX* and *SLT* tagging procedures are well documented elsewhere [1].

A simple cut or an actual fit both return the same result, from the tagged events in our sample 23 of 32 (72%) are "signal like". This compares with 50 of 136 (37%) "signal like" events in the sample as a whole. This can be compared to the conventional analysis, which utilizes b-tagging. In that analysis 37 b-tagged $W\pm \geq 3$-jet events containing 27 *SVX* tags with 6.7 ± 2.1 expected and 23 *SLT* tags with 15.4 ± 2.0 expected. This comparison should be considered rough, the selection for that analysis did not require a soft 4^{th} jet.

The input distributions for tagged events are shown in Figure B.2. The nine "background like" events are shaded. The mean values of the nine events are shown compared to signal and background MC mean values in Table B.1. This tends to indicate that the *ANN* is behaving as expected.

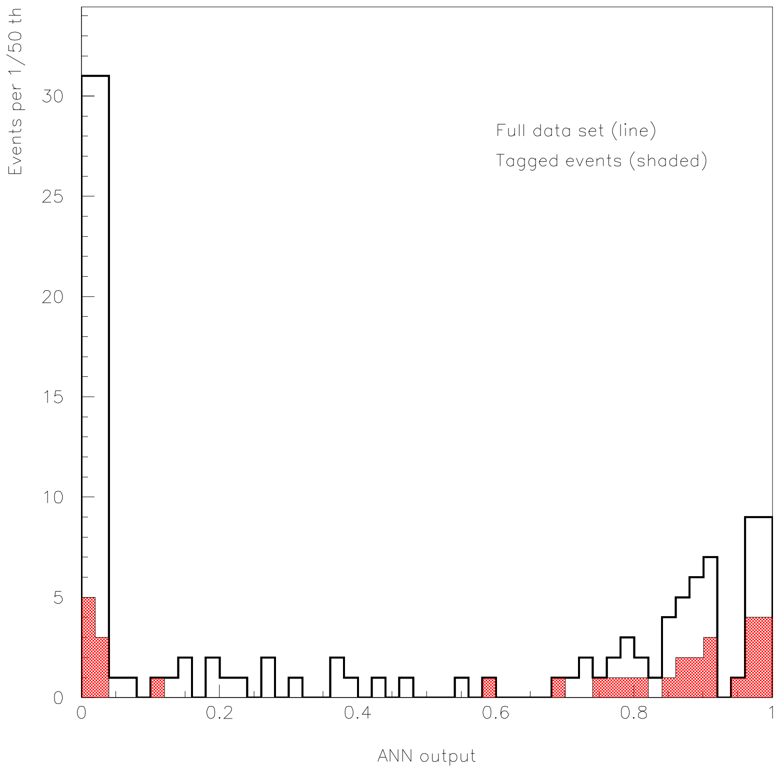


Figure B.1: The output distribution from the *ANN* for the data is shown (solid line). This is the full 136 passing events. The shaded distribution are the events with *SVX* or *SLT* tags.

Kinematic		**Means**	
	Data	Signal MC	Background MC
$P_T(jet1)$	63 GeV/c	96 GeV/c	75 GeV/c
H_T	270 GeV	344 GeV	292 GeV
$\mathcal{A}$	0.067	0.073	0.060
$<\eta^2>$	0.84	0.81	1.07

Table B.1: The mean values for signal and background MC's are contrasted with the means for the 9 "background like" *SVX* or *SLT* tagged data events.

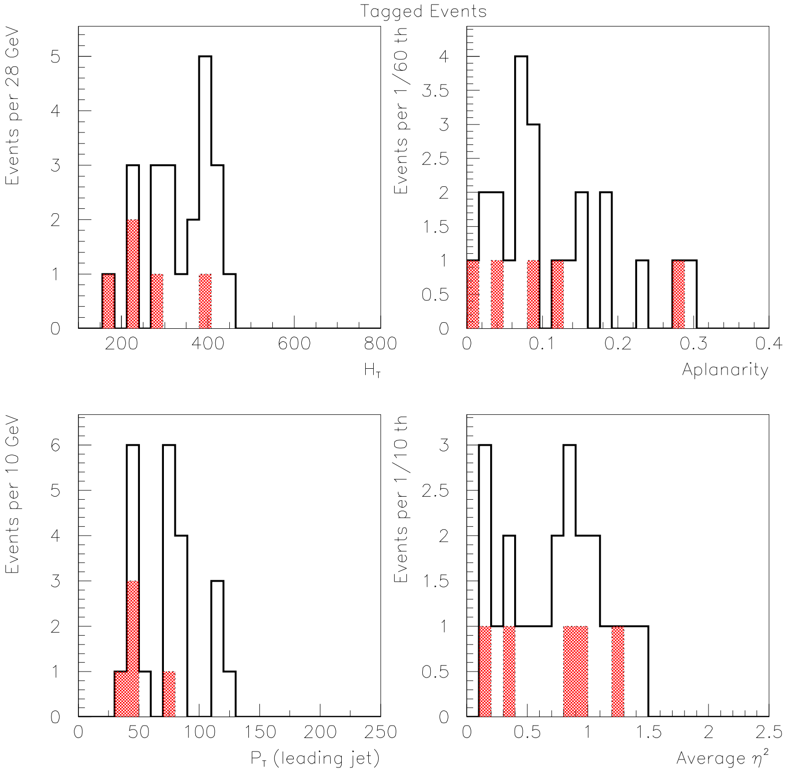


Figure B.2: A selection of input distributions are shown for the tagged events (solid line). This is the 32 *SVX* or *SLT* tagged events. The shaded distribution are the nine events with *SVX* or *SLT* tags which the *ANN* determined to be "background like".

Appendix C

Kinematic Analysis - without *ANN*

The work presented here was done as an extension of a previous *CDF* analysis [89]. The analysis was done without the use of a neural network and the results are given for individual kinematic variables. These results may be compared with the result in Chapter 7 which shows a statistical error only value for the *ANN* analysis which is the subject of this dissertation.

The event selection for the analysis of this appendix is the same as the *ANN* analysis as presented in Chapter 6. The definitions of the kinematic variables used are shown in Chapter 4.

The discriminating power of each variable is examined by generating a large pure $t\bar{t}$ "data" sample. This sample contains the number of signal events expected in 2 fb^{-1} of actual data. This Monte Carlo generated pseudo experiment is used to produce templates to fit against the signal and background MCs, Figure D.3, D.3. The error on the $N_{t\bar{t}}$, Δ, then serves as a measure of the discriminating power of each variable Table C.1. This rather crude measure of discriminating power was the first used in the *ANN* analysis and was latter replaced by the more conventional *KS* statistic. Cross section extraction from $N_{t\bar{t}}$ has been described in Chapter 9.

The $N_{t\bar{t}}$ are derived by means of a likelihood fit as described in Chapter 7. The various variables are consistent with each other. The $N_{t\bar{t}}$ for the $\mathcal{H}$ variable (48 ± 12)

Variable	Δ	$N_{t\bar{t}}$	$\sigma_{t\bar{t}}$ (pb)
$\mathcal{H}$	3.7	48 ± 12	7.3
$P_T(1) + P_T(2)$	12	42$\pm$12	6.4
$P_T(2) + P_T(3)$	3.4	44$\pm$11	6.8
$P_T(1)$	22	36$\pm$14	5.5
$P_T(2)$	3.5	40$\pm$12	6.1
$P_T(3)$	7.7	40$\pm$11	6.1
$\not{E}_T$	22	63$\pm$32	9.6
$\mathcal{A}$	7.2	63$\pm$22	9.6
$\mathcal{S}$	7.6	38$\pm$27	5.8

Table C.1: The Δ, discriminating power, is the error on the number of tops in a fit of a large pure $t\bar{t}$ "data" sample. The other two columns relate to the results for the actual data sample with only statistical errors shown.

is consistent with the value for $\mathcal{A}$ (63 ± 22). These yield $\sigma_{t\bar{t}} = 7.3 \pm 1.8$ *pb* for $\mathcal{H}$ and $\sigma_{t\bar{t}} = 9.6 \pm 3.3$ *pb* for $\mathcal{A}$, with statistical errors only as shown in Table C.2. Both of these numbers are consistent with the measurement from the standard lepton + jets $\sigma_{t\bar{t}}$ analysis at CDF which gives a value of $6.7^{+2.0}_{-1.7}$ *pb*.[95] The data sample as a whole may be compared with the SVX and SLT tagged events (b-tagged) in figs. D.3, D.3.

	$\mathcal{H}$	$\mathcal{A}$
N_{events}	135	
$N_{t\bar{t}}$	48 $\pm$ 12	63 $\pm$ 22
Luminosity	86.3	
Efficiency	7.6	
$\sigma_{t\bar{t}}(M_t = 175\ GeV)(pb)$	7.3 $\pm$ 1.8	9.6 $\pm$ 3.3

Table C.2: The $\mathcal{H}$ is the sum of the transverse momenta and $\mathcal{A}$ is aplanarity as described in section 3. The errors shown are purely statistical.

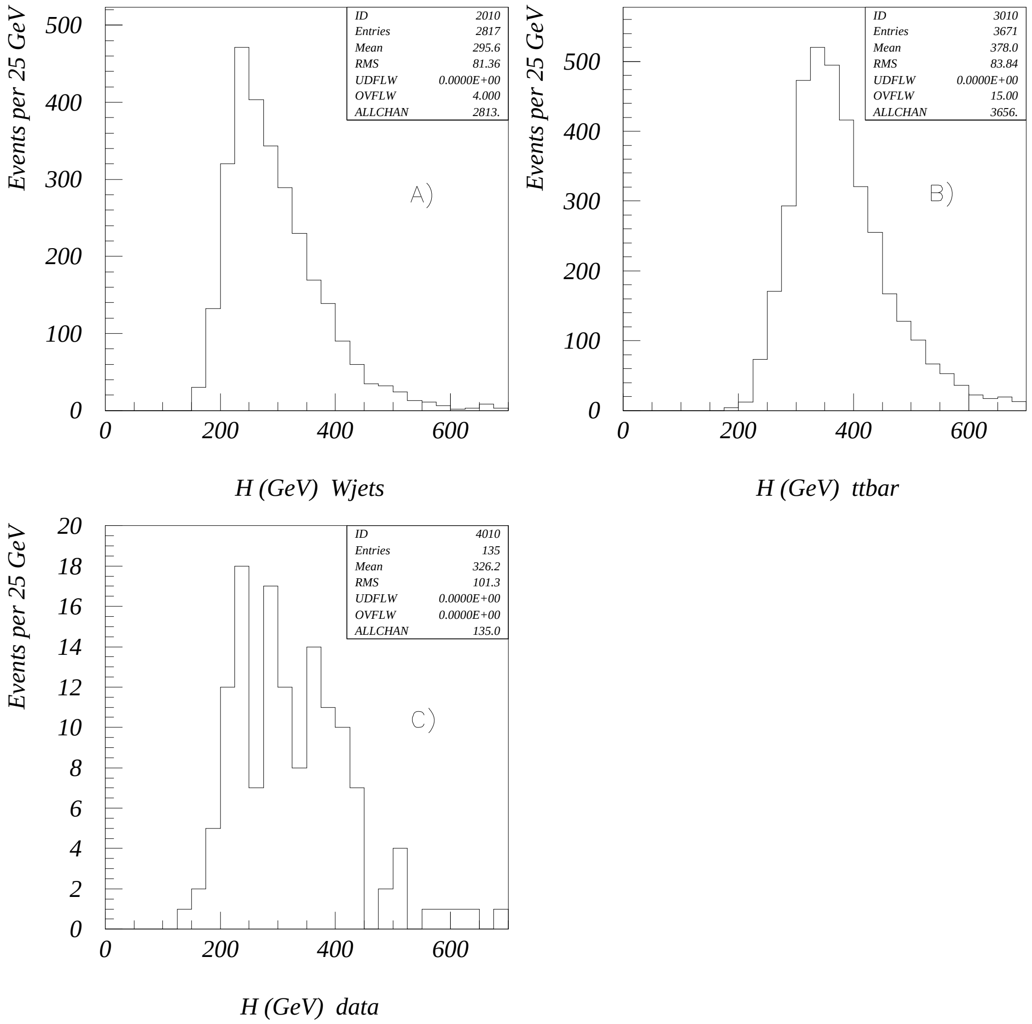


Figure C.1: The $\mathcal{H}$ distributions A) and B) are respectively the Vecbos $W+\geq 4jets$ background template and the Herwig $t\bar{t}$ signal template. The $\mathcal{H}$ distribution for the $t\bar{t}$ enriched data sample is shown in C).

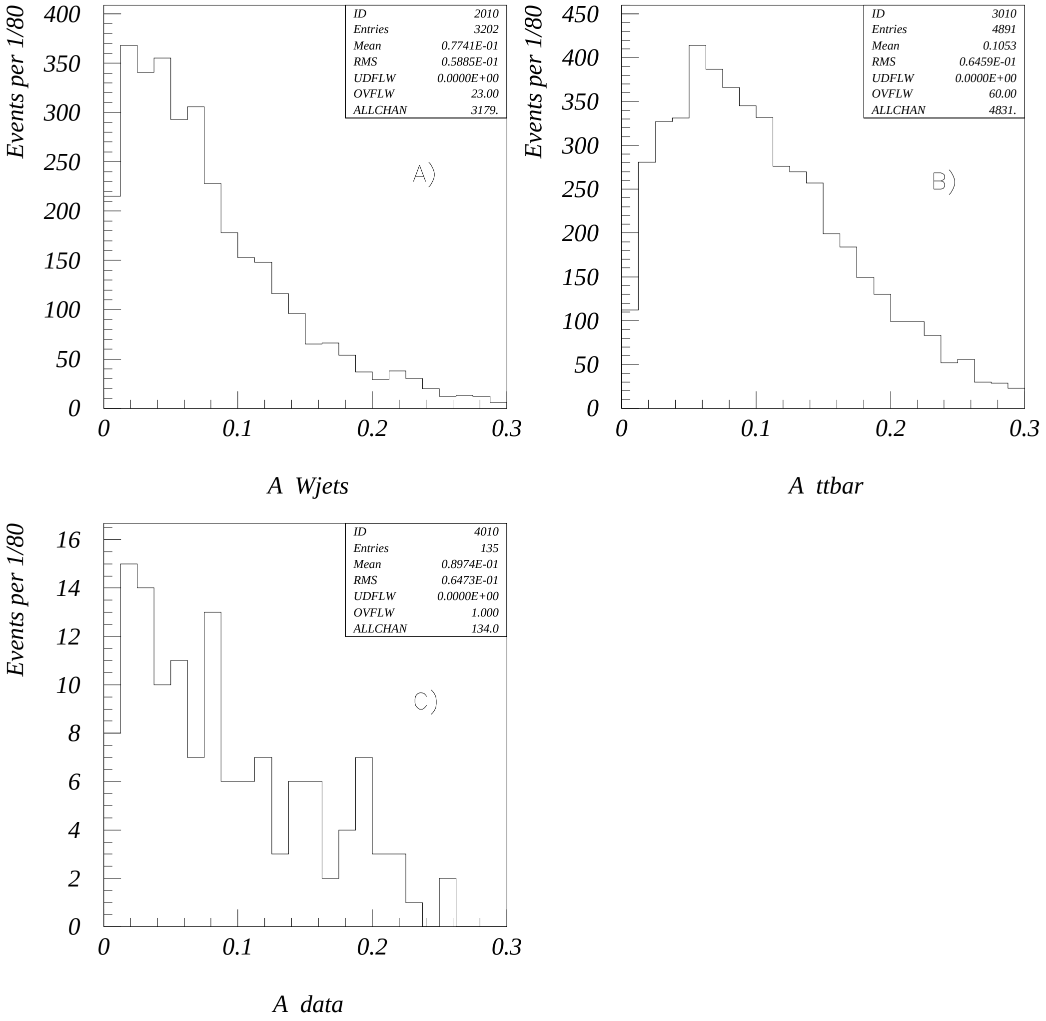


Figure C.2: The $\mathcal{A}$ distributions A) and B) are respectively the Vecbos $W+\geq 4jets$ background template and the Herwig $t\bar{t}$ signal template. The $\mathcal{A}$ distribution for the $t\bar{t}$ enriched data sample is shown in C).

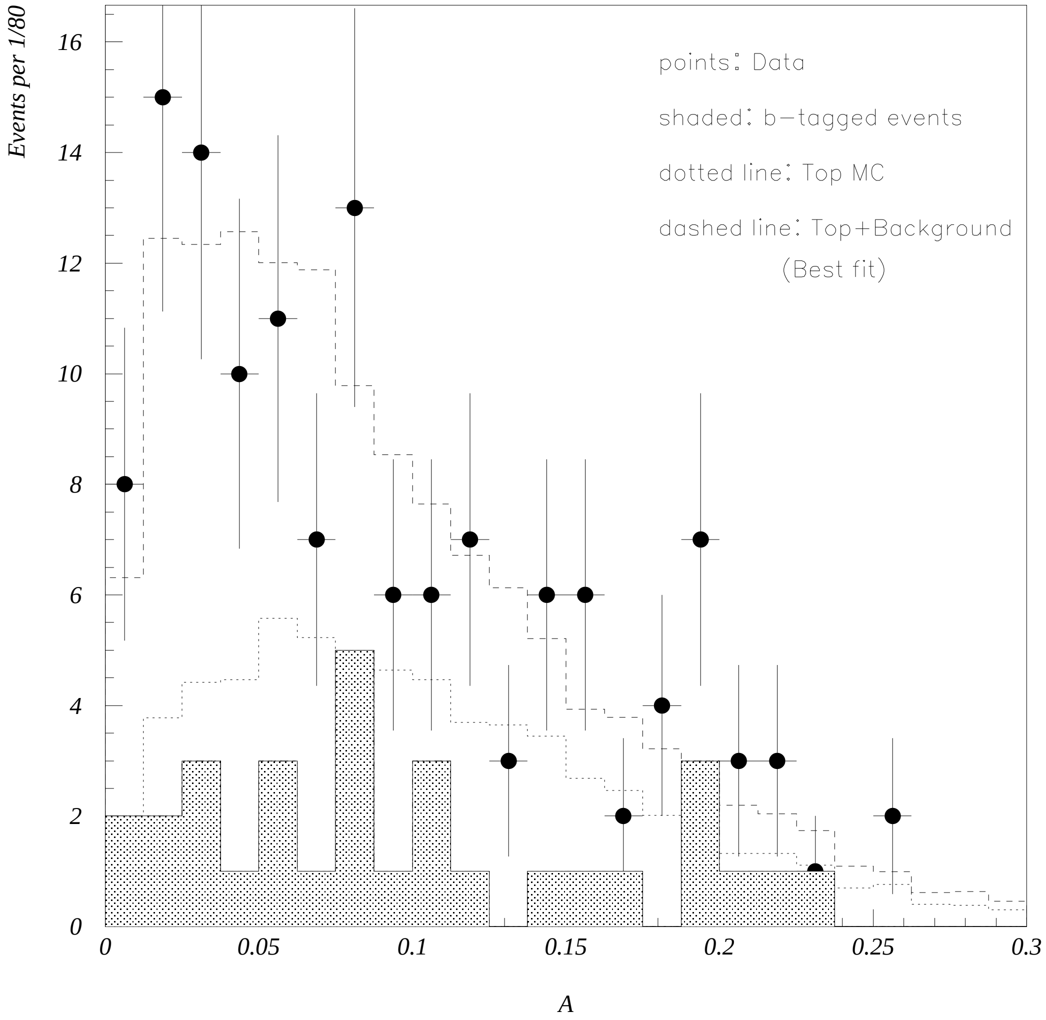


Figure C.3: The data points are our $t\bar{t}$ enriched data sample. The dashed line is a linear combination at the best fit ratio of Vecbos $W+ \geq 4jets$ and Herwig $t\bar{t}$ Monte Carlos normalized to data sample size. The dotted line is the $t\bar{t}$ component alone. The $\mathcal{A}$ distribution of the SVX and SLT tagged events in the sample are shaded.

A more complete discussion of the analysis in this appendix may be found in the CDF note issued at the time [71] and presented at the Centennial meeting of the *American Physical Society* 1999.

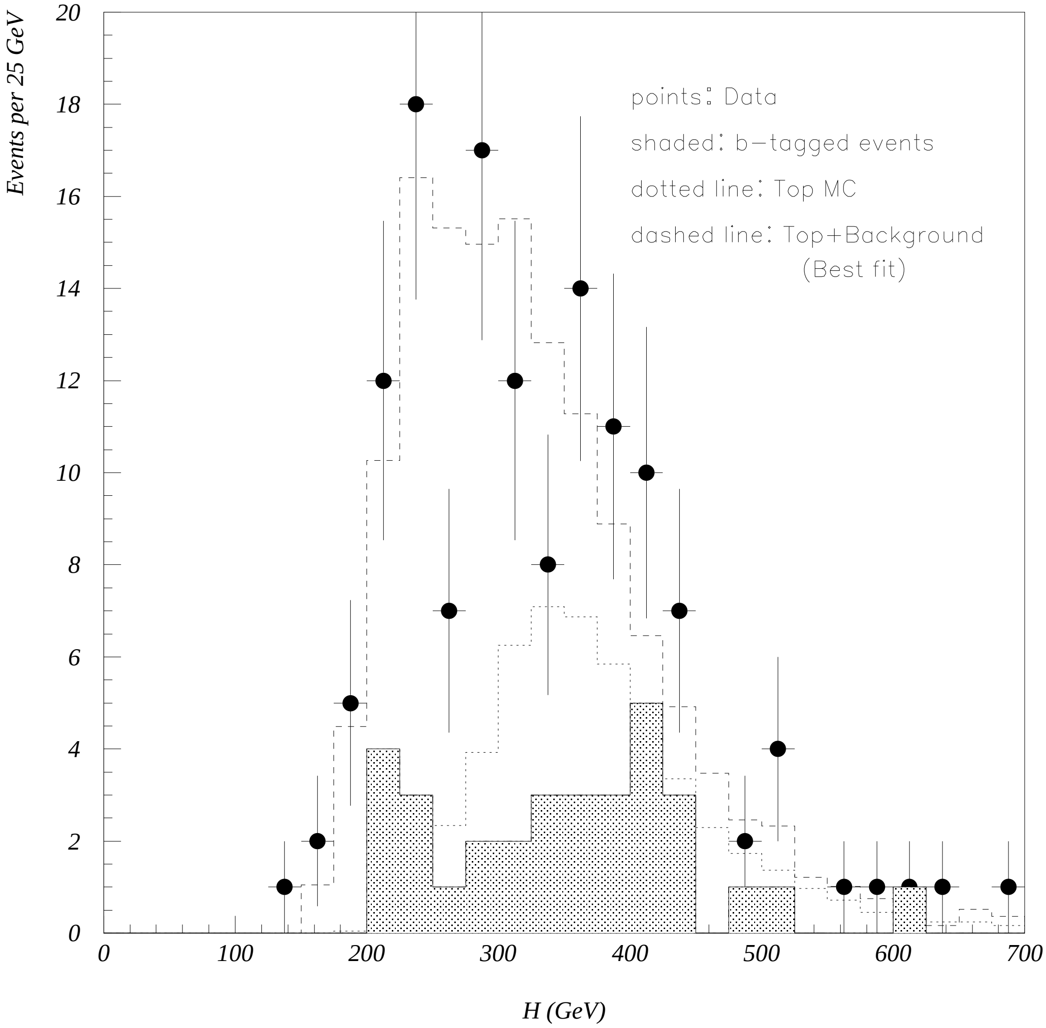


Figure C.4: The data points are our $t\bar{t}$ enriched data sample. The dashed line is a linear combination at the best fit ratio of Vecbos $W+ \geq 4jets$ and Herwig $t\bar{t}$ Monte Carlos normalized to data sample size. The dotted line is the $t\bar{t}$ component alone. The $\mathcal{H}$ distribution of the SVX and SLT tagged events in the sample are shaded.

Appendix D

JETNET

This appendix draws on our CDFnote [64]. We study the *ANN* performance using toy Gaussian data, utilizing JETNET version 3.4. [83] An event (or pattern) consists of some number of variables that characterize the event as either signal or background. The toy data consists of events with five random numbers selected from one of two Gaussian distributions, one of which we take to be signal and the other background, Figure D.2.

D.1 Learning with Back-propagation in *ANN*

The error back-propagation learning rule involves changing the weights in a network in an iterative fashion to minimize the error function. This method is known as the gradient descent method, because the changes induced trace the functions direction of greatest descent.

The weights ω and the thresholds θ are changed by gradient descent, where the change in weight is given by,

$$\Delta\omega(t+1) = -\eta\frac{\partial E}{\partial \omega} + \alpha\Delta\omega(t),$$

and $\omega(t)$ refers to the previous updating. [85]

The learning weight parameter, η, determines how the gradient effects the change. The α term is known as the momentum term and is used to prevent oscillation around a minimum and thereby interfere with convergence. Unfortunately there is no general criteria on which to base the selection of these two quantities– they are highly problem specific.

The frequency with which the network weights are changed is set by the user. One epoch is a single training over all available events. The weights may be changed as often as after each event or as rarely as the end of each epoch.

D.1.1 Saturation

Training will continue for as many epochs as the user has specified, ideally it will be terminated after the error function E as reached its minimum. A "saturation" parameter may be examined to determine when the proper termination point for the learning process has been reached.

The saturation measure, s, provides a guide as to whether or not a set of nodes have "made up their minds" for a given set of data. Saturation for layer l (s^l) is defined as a sum over all nodes in layer l as,[84]

$$s^l = \sum_j (1 - 2h_j^l)^2 \ ,$$

where h_j^l are the nodes in layer l.

The saturation parameter allows us to know whether or not the network is learning. When the network is initialized the weights are set randomly. After training on all or part of the data the new weights determined during this first training period are feed back into the network as initial weights, and the same or additional data is trained on with these new weights used as the initialization. Training continues in this iterative way for as long as is desired. All the data will be examined many times; each iteration of which is referred to as an epoch. At some point the saturation parameter will converge on its maximum value and stabilize, this is the point at which the network is no longer learning.

D.2 Testing Procedure

This section pertains to the results obtained using toy data in the type of *ANN* described above.

D.2.1 Data Generation

As has been previously stated our events consist of toy data generated from one of two Gaussians representing signal and background. The two Gaussians are of different widths with separated peaks. As would be expected changing the separation between the two Gaussians increases the networks ability to distinguish signal from background, Figure D.9. An event is made up of five, uncorrelated samplings from one of the two Gaussians. It was randomly determined whether any particular event would be selected from the signal or background Gaussian, Figures D.2, D.3.

D.2.2 Event Axis

In our simple example the symmetry naturally leads to two simple variables that distinguish signal and background. The event axes are defined as the axis through the peaks of both Gaussian distributions, d_0; and the perpendicular distance to this axis, d_1, Figure D.1. The selection of which Gaussian is signal is irrelevant to the outcome. In our case the signal Gaussian is one sigma and the background is two sigma. The region where the signal and background Gaussians overlap is the region where the network begins to misidentify events, Figures D.4, D.5. These newly defined axes provide the parameters on which cuts to distinguish signal from background would be applied in a traditional analysis.

D.2.3 Training Process

During the first training epoch 5000 "test" events were run through the network. The initial weights of the network are selected randomly. It was these random weights that were first used as the network weights, the test events were put through the network

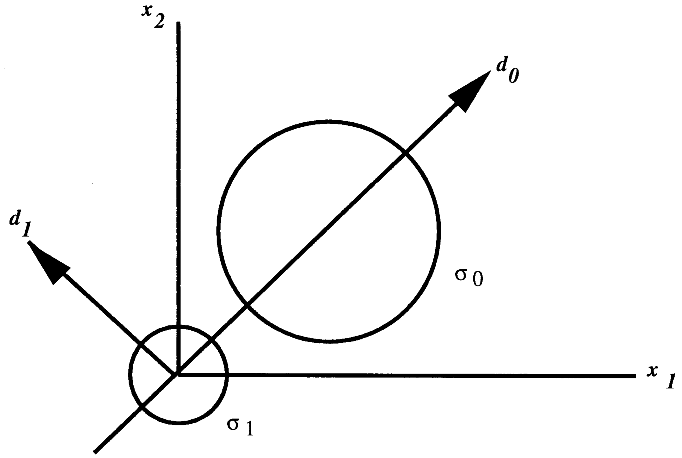


Figure D.1: Circles representing signal (σ_1) and background (σ_0) distributions. Only two of the five x_i event variables are shown. Due to the high degree of symmetry, the cylindrical variables d_0 and d_1 contain all the information available to separate the two distributions.

after its first training (1 epoch) and the network output f for signal and background events are histogrammed, Figures D.6, D.7. This may then be contrasted on the same figure with the same background/signal events put through the completely trained network (100 epochs). Clearly the training procedure is working to create a network that can distinguish signal from background in the simply case of our Gaussians. By the end of the training, the output value is sharply peaked at $f = 0$ and $f = 1$ for background and signal events respectively. This shows that the majority of events are well separated into signal and background.

D.2.4 Saturation

The saturation parameter's behavior over epochs can be gauged by looking at the set of each layers nodes, and comparing it to a parameter we call the performance, defined as the fraction of correctly identified (as either signal or background) events. Saturation was examined as a function of epochs, and saturation is compared with performance, Figure D.8. The saturation converges to its peak quickly which may be due to the simple nature of this case. The saturation has been shown to be consistent with the minimization of the error function E, Figure D.10.

D.2.5 Modified Network Architecture

Much of the network architecture is set by the user before training begins. Here we examine two elements of the architecture which the user must specify before training.

Additional Hidden Layer

Most HEP applications of ANNs can be performed with only one hidden layer in the network. This is likely to be the case in our application of the network also; however JETNET allows the introduction of more hidden layers. We have tested the effect of additional hidden layers in the case of our two Gaussians. However in this case the addition of the extra hidden layers has had no noticeable effect.

Additional Nodes

It is harder to guess what the effect might be of adding additional nodes to our hidden layer. We have examined the performance of a network with varying numbers of nodes in the hidden layer. The network does seems to perform better when the number of hidden nodes exceeds some minimum number of nodes but this effect soon levels off. Therefore adding too many nodes becomes counterproductive, the computational time goes up with no corresponding improvement in performance, Figure D.11.

D.3 Summary

As would be expected from a traditional analysis, it is the region were the signal and background Gaussians overlap were the network has difficulty distinguishing signal from background. In our simple test case the saturation parameter converges to its peak level quickly. This is not inconsistent with the performance or the behavior of the error function E. The output of the network appears insensitive, in this case, to the number of hidden layers. However adding additional nodes does seem to have an effect up to a number slightly greater than the number of nodes in the input layer. We see that least in this simple test case, the JETNET code does a good job of training the network and produces an output feature f with the correct behavior.

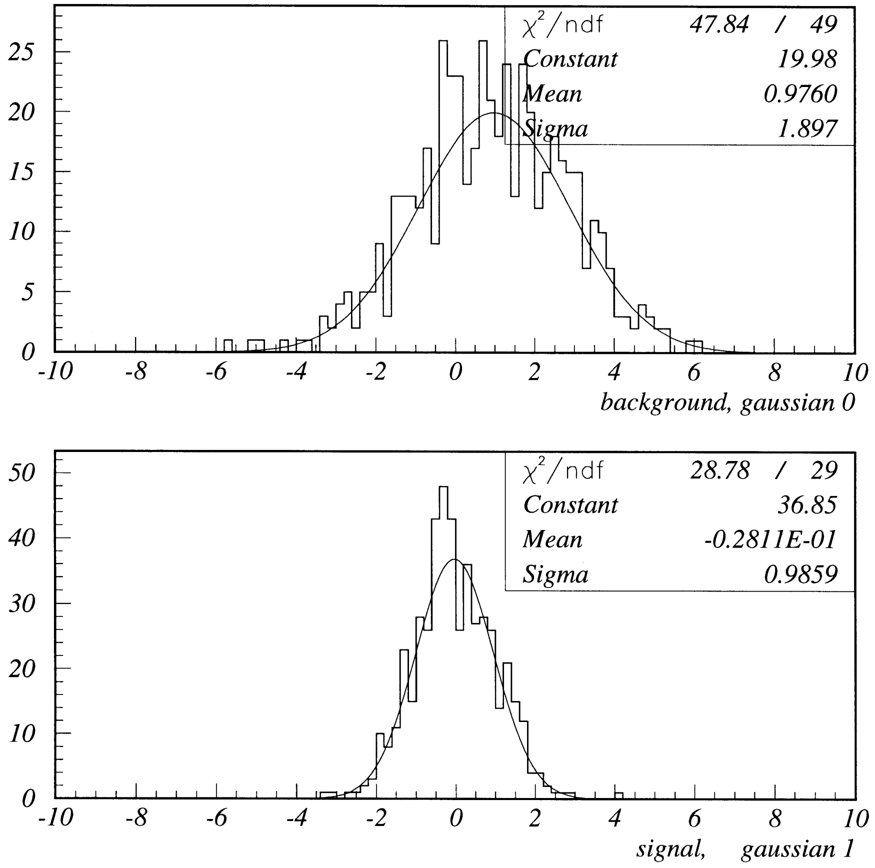


Figure D.2: Profiles of the two Gaussian distributions.

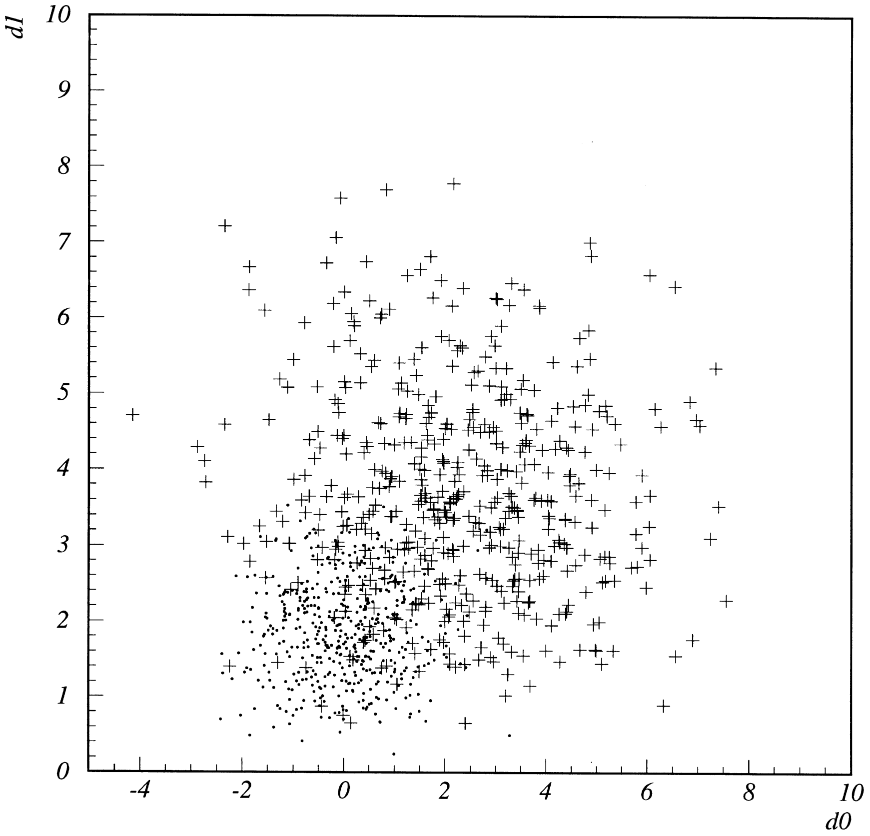


Figure D.3: Dots and crosses represent the data drawn from the signal and background Gaussians respectively.

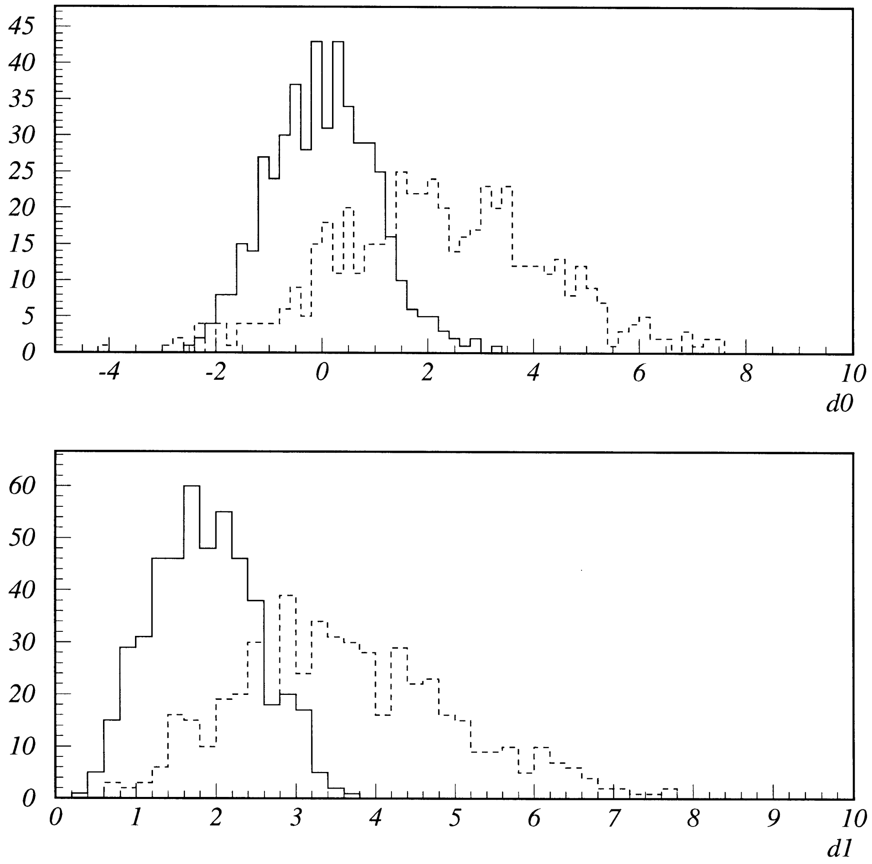


Figure D.4: The d_0 parameter is the axis through the peaks of both of the two Gaussian distributions shown in Figure D.2. The d_1 parameter is the perpendicular axis to d_0.

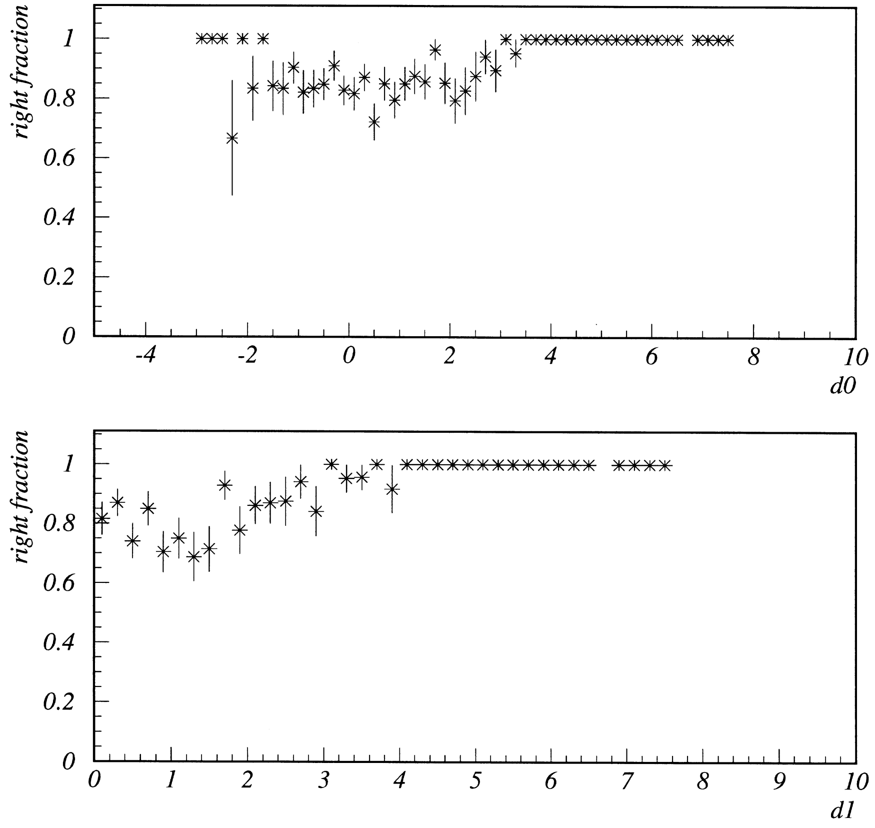


Figure D.5: The region where the network has difficulty distinguishing signal and background is the same region were the Gaussians overlap, Figure D.4.

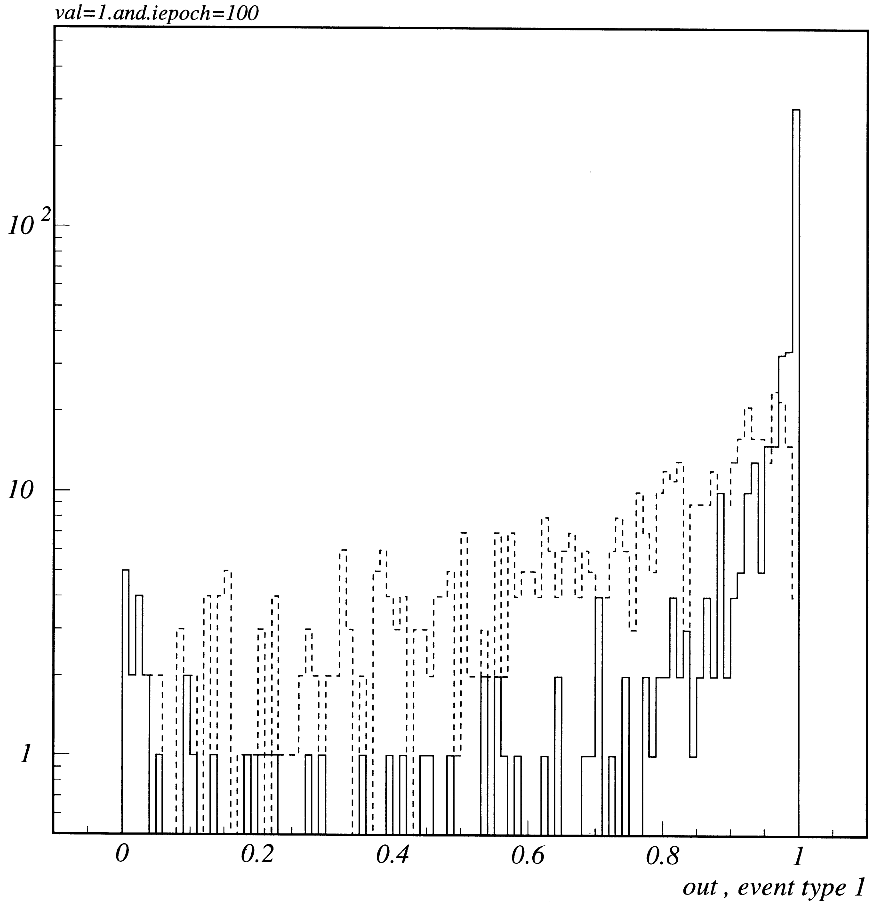


Figure D.6: The dotted line represents output for "signal" events after only one epoch (training iteration). The solid line represents output for "signal" events after one-hundred epochs (training iterations).

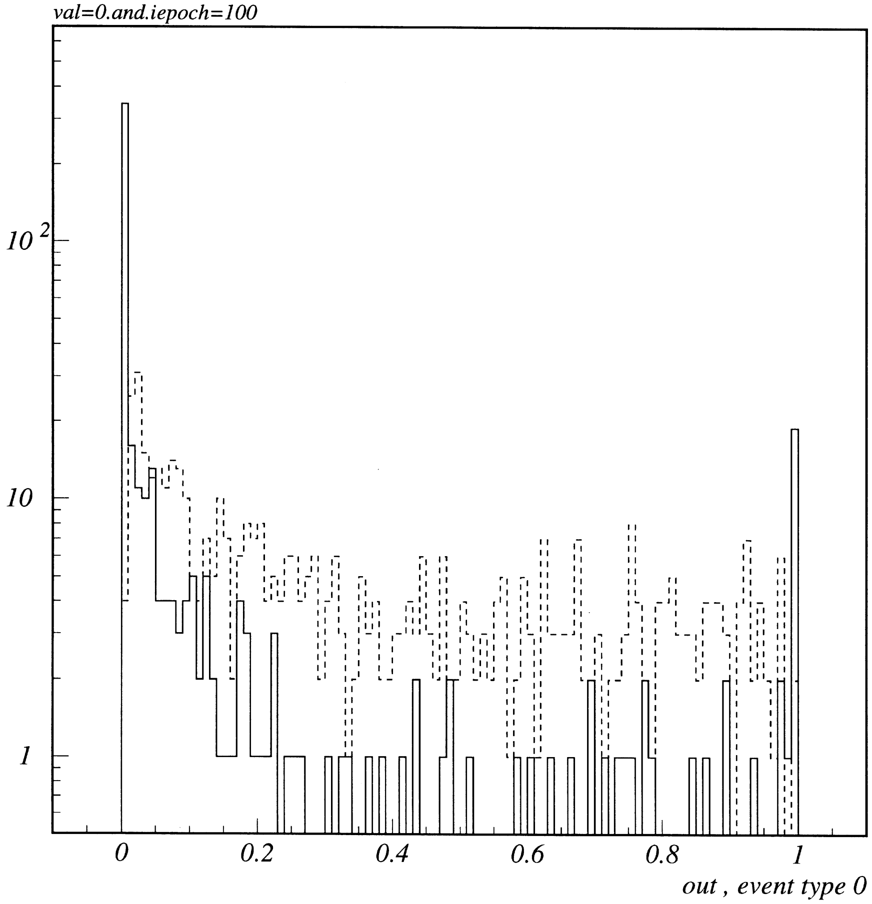


Figure D.7: The dotted line represents output for "background" events after only one epoch (training iteration). The solid line represents output for "background" events after one-hundred epochs (training iterations).

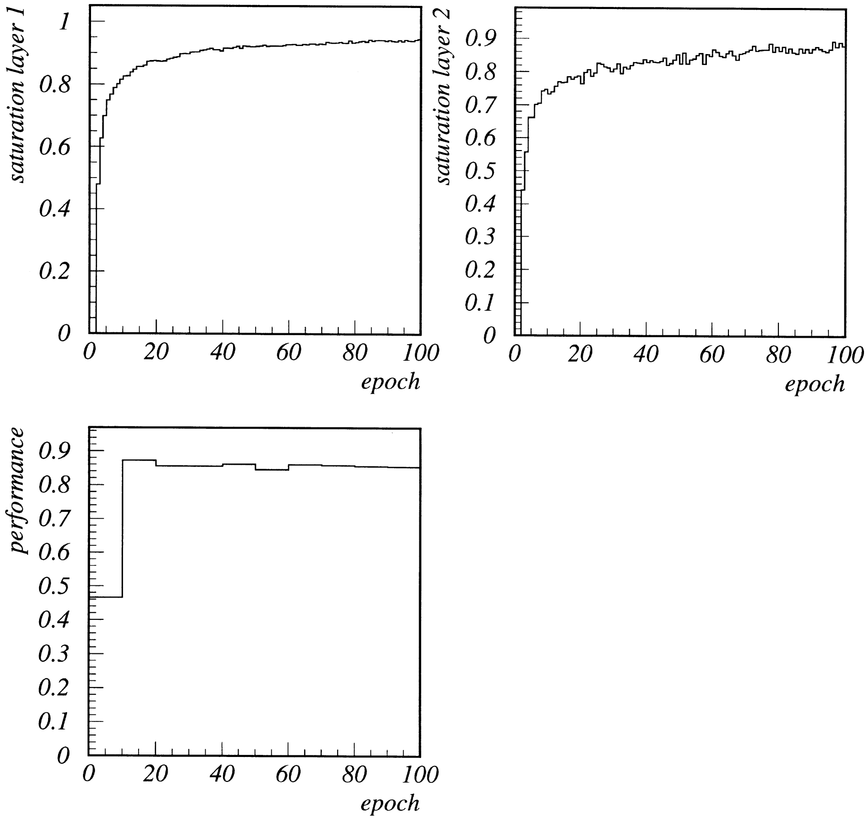


Figure D.8: Saturation as a function of epoch (training iteration) for first and second layers. And performance as a function of epoch.

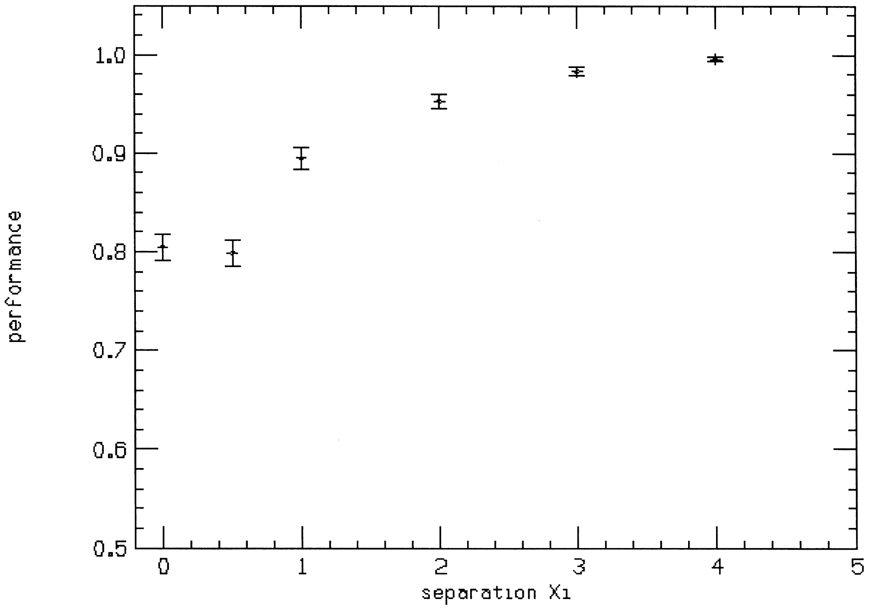


Figure D.9: The separation is between the peaks of the signal and background Gaussians. The performance is the fraction of correctly identified (signal or background) events.

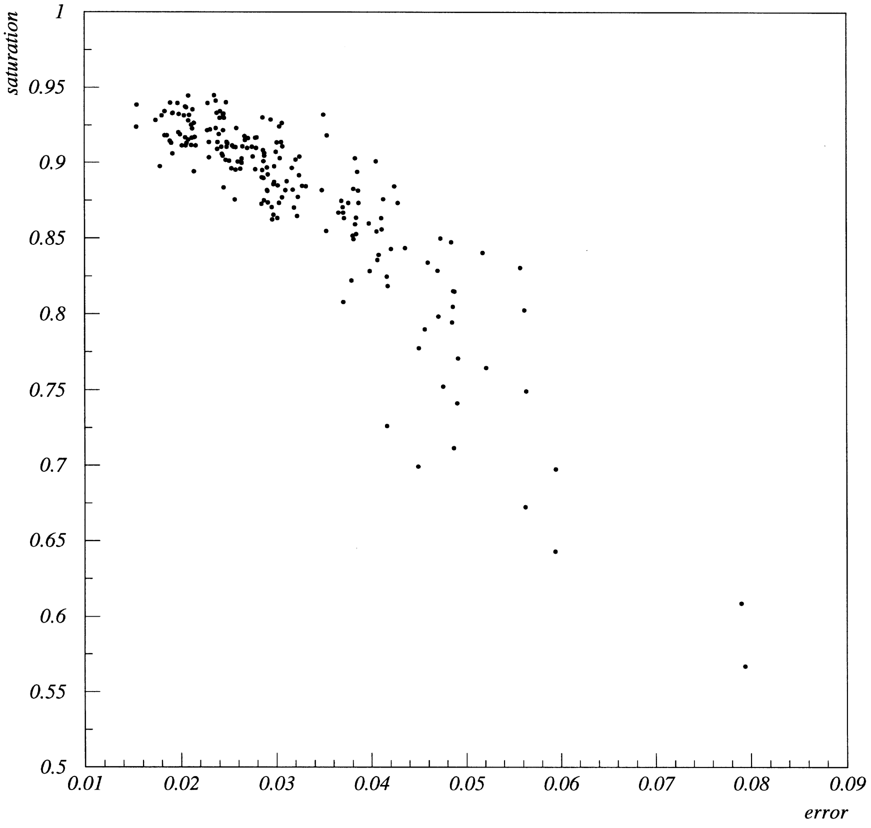


Figure D.10: The saturation parameter is shown plotted against the value of the error function at the end of each epoch.

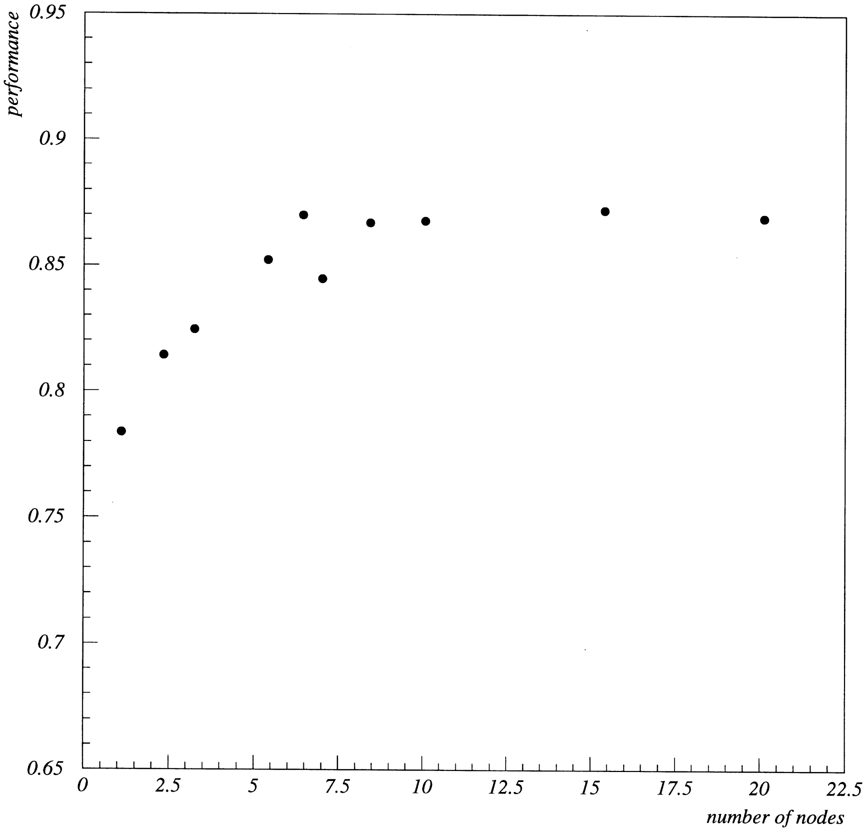


Figure D.11: This is the number of nodes in the hidden layer. The performance is the fraction of correctly identified (signal or background) events.

References

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

[2] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **74**, 3548 (1995).

[3] A. Brandl,, Ph. D. Thesis, University of New Mexico, unpublished (estimated 2012).

[4] V. Barger and R. Phillips, *Collider Physics*, Updated Ed. (Addison-Wesley Publishing Company, New York, NY, 1997).

[5] E. Simmons, Talk: *The Significance of the Heavy Top Quark*, given at Beyond the Standard Model V, Balholm, Norway (April 1997).-LANL:hep-ph/9707452.
E. Simmons, Talk: *Top Theories*, given at 8th International Symposium on Heavy Flavor Physics, Southampton, England (July 1999).-LANL:hep-ph/9908488.
E. Simmons, Talk: *Top Physics*, given at TASI 2000, Boulder, Colorado (June 2000).-LANL:hep-ph/0011244 v3.

[6] F. Abe *et al.*, P. Bhat, H. Prosper, S. Snyder, FERMILAB-PUB-98/236 (1998).

[7] PLUTO Collaboration, Ch. Berger *et al.*, Phys. Rev. Lett. **B76**, 243 (1978); C.W. Darden *et al.*, Phys. Rev. Lett. **B76**, 246 (1978); J.K. Bienlein *et al.*, Phys. Rev. Lett. **B78**, 360 (1978).

[8] D. Schaile and P.M. Zerrwas, Phys. Rev. **D45**, 3262 (1992).

[9] G. Kane, *Modern Elementary Particle Physics*, Updated Ed. (Addison-Wesley Publishing Company, New York, NY, 1993).

[10] S.W. Herb *et al.*, Phys. Rev. Lett. **39**, 252 (1977).

[11] F. J. Gilman, K. Kleinknecht, and B. Renk, *Review of Particle Properties* (1988).

[12] Private communication with Michael Gold. (2002).

[13] AMY Collaboration, H. Sagawa *et al.*, Phys. Rev. Lett. **60**, 93 (1988); AMY Collaboration, S.Igarashi *et al.*, Phys. Rev. Lett. **60**, 2359 (1988); TOPAZ Collaboration, I. Adachi*et al.*, Phys. Rev. Lett. **60**, 97 (1988); VENUS Collaboration, H. Yoshida *et al.*, Phys. Rev. Lett. **B198**, 570 (1987); VENUS Collaboration, K. Abe *et al.*, Phys. Rev. Lett. **B234**, 382 (1990).

[14] UA1 Collaboration, G. Arnison *et al.*, Phys. Lett. **B147**, 493 (1984).

[15] UA1 Collaboration, C. Albajar *et al.*, Z. Phys. **C37**, 505 (1988).

[16] ALEPHI Collaboration, D. Decamp *et al.*, Phys. Lett. **B236**, 511 (1990); DELPHI Collaboration, P. Abreu *et al.*, Phys. Lett. **B242**, 536 (1990); OPAL Collaboration, M.Z.. Akrawy *et al.*, Phys. Lett. **B236**, 364 (1990).

[17] UA1 Collaboration, C. Albajar *et al.*, Z. Phys. **C48**, 1 (1990); UA2 Collaboration, T. Åkesson *et al.*, Z. Phys. **C46**, 179 (1990).

[18] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **68**, 447 (1992); *ibid*, Phys. Rev. **D45**, 3921 (1992).

[19] DØCollaboration, F. Abachi *et al.*, Phys. Rev. **D52**, 4877 (1995).

[20] DØCollaboration, F. Abachi *et al.*, Phys. Rev. Lett. **72**, 2138 (1994).

[21] CDF Collaboration, F. Abe *et al.*, Phys. Rev. **D50**, 2966 (1994).

[22] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **73**, 225 (1994).

[23] S. Dawson, R.K. Ellis, and P. Nason, Nucl. Phys. B **303**, 607 (1988).

[24] R.K. Ellis, Phys. Lett. B **259**, 492 (1991).

[25] E. Laenen, J. Smith, and W.L. van Nerven, Phys. Lett. B **321**, 254 (1994).

[26] E. Laenen, J. Smith, and W.L. van Nerven, Phys. Lett. B **321**, 254 (1994).

[27] S. Dawson, R.K. Ellis, and P. Nason, Nucl. Phys. B **303**, 607 (1988).

[28] M. Jezabek and J. H. Kühn, Nucl. Phys. B **314**, 1 (1989).

[29] DØCollaboration, P.C. Bhat, in *Proceedings of the 8th Meeting of the Division of Particles and Fields of the American Physical Society, Albuquerque, New Mexico, USA* (World Scientific, River Edge, NJ, 1994), p. 705.

[30] M. Smith and S. Willenbrock, Phys. Rev. D **54**,6696 (1996).

[31] T. Stelzer, Z. Sullivan, and S. Willenbrock, Phys. Rev. D **56**,5919 (1997).

[32] The Particle Data Group, *The European Physical Journal c: Particles and Feilds* "The Particle Data Book", Vol.15, Number 1-4, Springer (2000)

[33] H. Singh, Ph. D. Thesis, University of California, Riverside, unpublished (1999).

[34] P. Azzi, G. Busetto, A. Castro, S. Dusini, I. Lazzizzera, A. Sidoti, J. Wyss, CDF/ANAL/TOP/CDFR/4722.

[35] J. Da Costa, *Standard Model Higgs Combined Result from CDF*, Inter. Jour. of Mod. Phys. A, (submitted).

[36] J. Cassada, CDF/DOC/EXOTIC/PUBLIC/4979.

[37] C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*, (The Benjamin/Cummings Publishing Company, Reading, Massachusetts, 1983).

[38] E. Guillian, Ph. D. Thesis, University of Michigan, unpublished (1999).

[39] J. Valls, *CDF Run II Discovery Reach for Neutral MSSM Higgs Bosons via* $p\bar{p} \to bb\phi \to b\bar{b}b\bar{b}$, CDF/ANAL/EXOTIC/4964 (1999).

[40] C. Pagliarone, *SUSY Searches at Tevatron Collider*, CDF/ANAL/EXOTIC /4925 (1999).

[41] T.J. LeCompte, *Supersymmetry Searches at the Tevatron*, unpublished.

[42] K. Tollefson, Ph. D. Thesis, University of Rochester, unpublished (1997).

[43] *CDF*, *The Collider Detector at Fermilab: A Compilation of reprinted articles from NIM*, North-Holland, Amsterdam (1988).

[44] *CDF*, *The CDF II Detector: Technical Design Report*, FERMILAB-Pub-96/390-E (1996).

[45] B. Müller, J. Reinhardt, and M. T. Strickland, *Neural Networks: An Introduction* 2nd *Ed.*, Springer-Verlag, New York (1995).

[46] D. O. Hebb, *The Organization of Behavior: A Neurophysiological Theory*, Wiley, New York (1949).

[47] C. E. Kahn Jr., *Artificial Intelligence in Radiology: Decision Support Systems*, Radiographics **14**, 849 (1994).

[48] G. Hripcsak, *Using Connectionist Modules for Decision Support*, Meth. of Info. in Medicine **29**, 167 (1990).

[49] J. A. Gentry, P. Newbold and D. T. Whitford, *Predicting Backruptcy: If Cash Flow's Not the Bottom Line, What is ?*, Finacial Analysts Journal, pp.47-56 (1985).

[50] F. Wong and P. Y. Tan, *Neural Networks and Genetic Algorithms for Economic Forecasting*, *AI* in Econ. and Bus. Admin., (1992).

[51] G. Swales and Y. Yoon, *Proceedings of the IEEE* 24th *Annual International Conference of Systems Sciences*, pp.156-162 (1991).

[52] H. J. Kim and H. S. Yang, *A Neural Network Capable of Learning and Inference for Visual Pattern Recognition*, Pattern Recognition Lett. **27**, 1291 (1994).

[53] L. Lönnblad, C. Peterson, and T. Rögnvaldsson, Phys. Rev. Lett. **65**, 1321 (1990).

[54] J. P. Tuttle, C. Hays, A. V. Kotwal, CDF/DOC/ELECTRON/CDFR/5791, (2001).

[55] L. Lönnblad, C. Peterson, and T. Rögnvaldsson, Phys. Letters B **278**, 181 (1992).

[56] K. H. Becks *et. al.*, Nucl. Instr. and Meth. **A329**, 501 (1993).

[57] A. Beretvas, M. Binkley, CDF/PUB/TOP/CDFR/3985, (1998).

[58] DØ Collaboration, hep-ex/9907041, (1999).

[59] P. Abreu *et al.*, Phys. Lett. **B295**, 383 (1992).

[60] P. Azzi *et. al.*, internal *CDF* note, CDF/ANAL/TOP/CDFR/4722 (1999).

[61] C. Peterson, T. Rögnvaldsson, and L. Lönnblad, CERN-TH.7135/94 (1993).

[62] F. Abe et al.,Phys. Rev. Lett. **80**,2773 (1998).

[63] C. Peterson, T. Rögnvaldsson, and L. Lönnblad, Computer Physics Communication **70** (1992).

[64] M. Gold, E. Moore and T. Palmer, internal *CDF* note, CDF/ANAL/TOP /CDFR/5163 (1999).

[65] C. A. L. Bailer-Jones *et al.*, astro-ph/0102224, (2001).

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

[67] F. A. Berends *et al.*, Nucl. Phys. **B357**, 32 (1991).

[68] W. H. Press *et al.*, *Numerical Recipess in C* 2^{nd} Ed., Cambridge Press, New York (1988).

[69] F. Paige *et al.*, in *Supercollider Physics*, editted by D. Soper (World Scientific, Singapore, 1986), p. 41.

[70] T. Sjostrand and M. Bengtsson, Comput. Phys. Commun. **39**, 347 (1986); T.Sjostrand and M. Bengtsson, Comput. Phys. Commun. **43**, 367 (1987).

[71] E. Moore and M. Gold, internal *CDF* note, CDF/ANAL/TOP/CDFR/4906 (1999).

[72] E. Moore, M. Gold and S. Worm, internal *CDF* note, CDF/ANAL/TOP /CDFR/5680 (2001).

[73] E. Moore and M. Gold, internal *CDF* note, CDF/ANAL/TOP/CDFR/5372 Ver.2 (2001).

[74] F. James, *MINUIT: Function Minimization and Error Analysis*, CERN Program Library Long Writeup D506 (1994).

[75] J.P. Done, M. Chertok, and T. Kamon, *ISAJET Monte Carlo Validation*, CDF internal note 4903 (1999).

[76] *CDF* Collaboration, F. Abe *et al.*,Phys.Rev.Lett. 80 3461 (1998).

[77] T. Stelzer, Z. Sullivan, and S. Willenbrock, *Single Top Quark Production via W-gluon Fusion at Next-to-Leading Order*, hep-ph/9705398 (1997).

[78] F. Abe *et al.*, The CDF Collaboration, Phys. Rev. Lett. **76**, 3070 (1996)

[79] M. Mangano, *The Unweighting of Matrix Element Monte Carlos*, CDF internal note 1665 (1991).

[80] D. Gerdes, *Improved Unweighting of VECBOS Monte Carlo Events*, CDF internal note 3180 (1995).

[81] P. Abreu et al., “Classification of the Hadronic Decays of the Z^0 into b and c Quark pairs using a Neural Network”, *Physics Letters B* **295**, 383 (1992).

[82] B. Abbott et al., “Neural Networks for Analysis of Top Quark Production”, hep-ex/9907041.

[83] Leif Lonnblad et al., JETNET-3.4, University of Lund (1995).

[84] C. Peterson and T. Rognvaldsson, “JETNET 3.0 - A Versatile Artificial Neural Network Package”, CERN-TH 7135/94.

[85] L. Lonnblad C. Peterson and T. Rognvaldsson, “Pattern Recognition in High Energy Physics with Artificial Neural Networks - JETNET 2.0”, *Comput. Phys. Commun.* **70**, 167(1992).

[86] K.H. Becks et al., “B-quark tagging using neural networks and multivariate statistical methods”, *Nucl. Instrum. Methods*, **A329**, 501 (1993).

[87] B. Müller et al., “Neural Networks 2nd Edition”, *Springer-Verlag*, New York (1995).

[88] A. Beretvas,M. Binkley, CDF/PUB/TOP/CDFR/3985, version 5.0, June 1998.

[89] F. Abe et al.,Phys. Rev. Lett. **75**, 3997 (1995).

[90] Isajet samples generated curtessy of the SUSY working group at *CDF* acquired through privat communication with Stephan Lammel.

[91] L. Galtieri, J. Lys., CDF/ANAL/TOP/CDFR/3501, version 2.5, March 1996.

[92] L. Demortier,*et. al.*, CDF/PHYS/TOP/CDFR/3089, version 1.4, Apr.1995.

[93] D. Cronin-Hennessy,A. Beretvas, CDF/DOC/CDF/4721, version 1.0, Nov (1998).

[94] T. Liss, R. Roser, CDF/ANAL/TOP/CDFR/3481, version 4.0, June (1997).

[95] F. Abe et al.,Phys. Rev. Lett. **80**,2773 (1998).

[96] S. Behrends, internal *CDF* note, CDF/ANAL/JET/CDFR/2151 (1993).

[97] "Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8\ TeV$ ", published is available from hep-ex/0101036.

2017 ProQuest Distribution Agreement

This Agreement is between the author (Author) and ProQuest LLC, through its ProQuest Dissertation Distribution business (formerly ProQuest/UMI). Under this Agreement, Author grants ProQuest certain rights to preserve, archive and distribute the dissertation or thesis (the Work), abstract, and index terms provided by Author to ProQuest.

Section I. License for Inclusion of the Work in ProQuest Publishing Program

Grant of Rights. Author hereby grants to ProQuest the **non-exclusive**, worldwide right to reproduce, distribute, display and transmit the Work (in whole or in part) in such tangible and electronic formats as may be in existence now or developed in the future. Author further grants to ProQuest the right to include the abstract, bibliography and other metadata in the ProQuest Dissertations & Theses database (PQDT) and in ProQuest Dissertation Abstracts International and any successor or related index and/or finding products or services.

ProQuest Publishing Program - Election and Elements. The rights granted above shall be exercised according to the publishing option selected by Author in Section III, Author Options, and subject to the following additional Publishing Program requirements:

- Distribution of the Work. Except as restricted by Author in the publishing option selected, the rights granted by Author automatically include (1) the right to allow sale and distribution of the Work, in whole or in part, by ProQuest and its sales agents and distributors and (2) the right to make the Abstract, bibliographic data and any meta data associated with the Work available to search engines and harvesters.
- Restrictions. ProQuest will use commercially reasonable efforts to restrict the distribution of the Work as provided under the publishing option selected by Author or *as later elected by Author* through direct contact with ProQuest. Such election is subject to Author's Degree Granting Institution Directives. With respect to restrictions requested after submission of the Work, Author acknowledges that ProQuest cannot recall or amend previously distributed versions of the Work.
- Removal of Work from the Program. ProQuest may elect not to distribute the Work if it believes that all necessary rights of third parties have not been secured. Refer to the website http://www.proquest.com/products_umi/dissertations/ for information about copyright and your dissertation or thesis. If Author's degree is rescinded, and/or the degree-granting institution so directs, ProQuest will expunge the Work from its publishing program in accordance with its then current publishing policies.
- Degree Granting Institution Directives. Author is solely responsible for any conflict between policies and directions of Author's degree-granting institution, Author's choice of publishing model, and/or any restriction Author places on the use of the Work. For the avoidance of doubt, ProQuest is not responsible for access to the Work that is provided by Author's degree-granting institution through its library or institutional repository. Author must work directly with Author's degree granting institution to ensure application of any restrictions to access relating to the Work by Author's degree granting institution.

Delivery of the Work. Author shall provide to ProQuest the Work and all necessary supporting documents during the online submission process, according to the Instructions accompanying this agreement.

Rights Verification. Author represents and warrants that Author is the copyright holder of the Work and has obtained all necessary rights to permit ProQuest to reproduce and distribute third party materials contained in any part of the Work, including all necessary licenses for any non-public, third party software necessary to access, display, and run or print the Work. Author is solely responsible and will indemnify ProQuest for any third party claims related to the Work as submitted for publication.

Open Access Publishing Plus

☑ I want the broadest possible dissemination of my work, and I want to provide free global access to the electronic copy of my work via the internet.

☑ I understand that I will not be eligible to receive royalties.

I want major search engines (e.g. Google, Yahoo) to discover my work. Learn more: http://www.proquest.com/en-US/products/dissertations/google.shtml

☑ Yes
☐ No

Acknowledgment: I have read, understand and agree to this ProQuest Publishing Agreement, including all rights and restrictions included within the publishing option chosen by me as indicated above.

REQUIRED Author's signature________ Date 10 JULY 2017

(Print Name) ERIC MOORE

Institution conferring degree The University of New Mexico

This page must accompany your manuscript and the rest of the submission materials

Please check type of manuscript:

☐ M (Master's Thesis)

☑ D (Dissertation)

Dissertation/Master's Thesis Submission Form

Please print clearly in block letters

Personal Information

Last Name: MOORE

Middle Name or Initial: THOMAS

First Name: ERIC

Country (ies) of Citizenship: USA

Degree & Dissertation Information

Title of Dissertation/ Thesis: A Neural Network Measurement of the $t\bar{t}$ Pair Production Cross section in the Lepton+Jets Channel

Institution conferring degree: The University of New Mexico

Degree awarded (abbreviate; e.g., Ph.D.): Ph.D.

College, School, or Division: College of Arts & Sciences

Year degree awarded: 2002

Department or Program: Department of Physics & Astronomy

Year manuscript completed: 2002

Advisor/Committee Chair: Michael Gold

Committee Member: John Matthews

Committee Member:

Committee Member: Gerald Stephenson

Committee Member:

Committee Member:

Committee Member:

Language of manuscript: English

Primary Subject Category: Enter the 4-digit code and category name from the Subject Category Guide that most closely describes the area of your research.

Code 0798 Category Particle Physics

You may suggest two additional subject categories that may aid in the discovery of your work in our digital database.

Code_____ Category_____ Code_____ Category_____

Provide up to 6 keywords or short phrases for citation indices, library cataloging, and database searching.

top quark	neural network	Run I CDF
top pair production	lepton + jets	top anti-top

Current Contact Information

Current Email Address: mooreet@gmail.com

Street Address:

Please provide your postal address if you are interested in receiving royalties on sales of your thesis.

City:

Province:

Daytime Phone: 703.853.7994

Country:

Postal Code:

Evening Phone:

Permanent Contact Information

Permanent Email Address:

Street Address (line 1):

City:

Province:

Future Phone:

Country:

Postal Code:

Alternate Future Phone: