

MEASUREMENT OF THE TOP PAIR PRODUCTION
CROSS SECTION AT CDF USING NEURAL NETWORKS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Radu Marginean, B.S., M.S.

* * * * *

The Ohio State University

2004

Dissertation Committee:

Prof. Richard E. Hughes, Adviser

Prof. Michael A. Lisa

Prof. Samir D. Mathur

Prof. Brian L. Winer

Approved by

Adviser
Graduate Program in Physics

ABSTRACT

In the Tevatron accelerator at Fermilab protons and antiprotons are collided at a 1.96 TeV center of mass energy. CDF and DØ are the two experiments currently operating at the Tevatron. At these energies top quark is mostly produced via strong interactions as a top anti-top pair ($t\bar{t}$). The top quark has an extremely short lifetime and according to the Standard Model it decays with $\sim 100\%$ probability into a b quark and a W boson. In the “lepton+jets” channel, the signal from top pair production is detected for those events where one of the two W bosons decays hadronically in two quarks which we see as jets in the detector, and the other W decays into a electrically charged lepton and a neutrino. A relatively unambiguous identification in the detector is possible when we require that the charged lepton must be an electron or muon of either charge. The neutrino does not interact in the detector and its presence is inferred from an imbalance in the transverse energy of the event. We present a measurement of the top pair production cross section in $p\bar{p}$ collisions at 1.96 TeV, from a data sample collected at CDF between March 2002 and September 2003 with an integrated luminosity of 193.5 pb^{-1} . In order to bring the signal to background ratio at manageable levels, measurements in this channel traditionally use precision tracking information to identify at least one secondary vertex produced in the decay of a long lived b hadron. A different approach is taken here. Because of the large mass of the top quark, $t\bar{t}$ events tend to be more spherical and more energetic

than most of the background processes which otherwise mimic the $t\bar{t}$ signature in the “lepton+jets” channel. A number of energy based and event shape variables can be used to statistically discriminate between signal and background events. Monte Carlo simulation is used to model the kinematics of $t\bar{t}$ and most of the background processes. A neural network technique is employed to combine multiple variables in order to enhance signal versus background separation. Such a measurement takes advantage of a larger data sample than the b-tagging based analyses and achieves a comparable level of precision. A binned likelihood fit to the neural network output distribution for a 519 events data sample yields a $17.6 \pm 3.0(stat)\%$ fraction of $t\bar{t}$ events. The inclusive top pair production cross section is measured to be $\sigma_{t\bar{t}} = 6.6 \pm 1.1(stat) \pm 1.5(syst)$ pb.

ACKNOWLEDGMENTS

First, I would like to thank the members of the CDF Collaboration, the people of Fermilab Staff and those who devoted time, work and financial resources into building and running this experiment. I appreciate the opportunity I was given to work at CDF.

I would like to say thank you to my advisers, Richard Hughes and Brian Winer. I must say, my life as a graduate student was a more exciting experience than I ever would have hoped for.

I thank the former members of the OSU group, Dr. Evelyn Thomson and Dr. Phil Koehn for their support. The former OSU students Catalin Ciobanu, Carlos Sanchez, Chris Neu, Jong-young Chung and Corey Stambaugh have helped me make a good start and I appreciate the help they have given me in many occasions after their graduation. I wish them all the best.

My parents Lucia and Ioan, my sister Mara, have always been there for me. *Va multumesc.*

Finally I want to thank my wife Camelia who is the most important to me.

VITA

June 1997 B.S. in Physics, University of Bucharest, Bucharest, Romania.

June 1999 M.S. in Physics, University of Bucharest, Bucharest, Romania.

November 1998– June 2000 Research Assistant, Institute for Space Sciences, Bucharest, Romania.

July 2000–June 2001 University Fellow, Brown University, Providence RI.

July 2001–June 2002 Physics Department Fellow, The Ohio State University, Columbus, OH.

July 2002–December 2004 Graduate Research Assistant, The Ohio State University, Columbus, OH.

FIELDS OF STUDY

Major Field: Physics

TABLE OF CONTENTS

	Page
Abstract	ii
Acknowledgments	iv
Vita	v
List of Tables	viii
List of Figures	x
Chapters:	
1. Introduction	1
1.1 The Standard Model of Particle Physics	1
1.2 Production and Properties of the Top Quark	7
2. The CDF Experiment at the Tevatron	13
2.1 The Fermilab Accelerators Complex	13
2.2 The CDF Detector	16
2.2.1 The Silicon Detectors	17
2.2.2 The Central Drift Chamber	19
2.2.3 The Calorimeter System	21
2.2.4 The Muon System	22
2.2.5 The Event Trigger	23
3. Event Reconstruction and Simulation	27
3.1 Introduction	27
3.1.1 Reconstruction of Jets	27

3.1.2	Electron and Muon Identification	32
3.1.3	Missing Transverse Energy Measurement	34
3.2	The Event Selection	37
3.2.1	The Data Sample	40
3.3	Monte Carlo Event Simulation	41
4.	Neural Networks	44
4.1	Introduction	44
4.2	Artificial Neural Networks	46
4.2.1	A Theoretical Perspective on Classification Problems	48
4.2.2	Finding the Optimum Solution	50
4.2.3	The Choice for the Network Architecture	54
4.3	Some Aspects of Artificial Neural Networks	55
5.	Cross Section Measurement	59
5.1	Selecting $t\bar{t}$ Events in the “Lepton+Jets” Sample	59
5.2	Simulation of the W +jets Background	60
5.3	The Model for the Multijet Background	63
5.4	The Signal Acceptance and Efficiency	66
5.5	Principle of Measurement	67
5.5.1	The Binned Likelihood Fit	68
5.5.2	Running Pseudo-Experiments	69
5.6	Kinematic Variables	73
5.7	ANN Training and Testing	77
5.7.1	Choosing an Optimum Network Configuration	79
5.8	Comparisons Between Data and Simulation	83
5.9	Systematic uncertainties	85
5.10	The Fitting Technique	95
5.11	Cross Section Calculation	97
5.11.1	Results for $M_{top}=175 \text{ GeV}/c^2$	97
5.11.2	Results as a Function of the Assumed Top Mass.	98
5.11.3	Consistency Checks	98
6.	Summary and Conclusions	104
	Bibliography	106

LIST OF TABLES

Table	Page
1.1 The elementary particles in the Standard Model of particle physics. . .	2
1.2 The $SU(2)$ multiplets for the three generations of fermions.	3
3.1 Summary of the event selection criteria.	35
3.2 The number of selected events as a function of jet multiplicity for 193.5 pb ⁻¹ of CEM and CMUP and 175 pb ⁻¹ of CMX data samples.	41
5.1 Contributions from other electroweak processes relative to the dominant $W+3p$ process in the $W+\geq 3$ jets mode. The cross section numbers are leading order estimates and thus are expected to have large uncertainties.	60
5.2 The fraction of multi-jet background in the W +jets data sample as a function of jet multiplicity. The uncertainties shown are statistical only.	65
5.3 Signal acceptance in the $W+\geq 3$ jets sample assuming $m_{top} = 175$ GeV/ c^2 . The number of generated events after the parton p_T cuts at the generator level was 384,875.	66
5.4 Signal efficiency in the $W+\geq 3$ jets sample assuming $m_{top} = 175$ GeV/ c^2 . The overall signal efficiency weighted with respect to the CEM/CMUP/CMX luminosity is 0.0711 ± 0.0056	67
5.5 The definition for all the kinematic variables considered in this analysis. The particular order for listing the variables was obtained from starting with the H_T and adding variables one by one in such order that performance in training a ANN on PHYTIA $t\bar{t}$ and ALPGEN $W+3p$ was maximized during each step.	75

5.6	systematic effects on $t\bar{t}$ contribution extracted from fits using the H_T shape.	93
5.7	Table for systematic uncertainties in the $W+\geq 3$ jets mode. The overall uncertainty is obtained by adding in quadrature the individual effects.	94
5.8	Table for systematic uncertainties in the $W+\geq 4$ jets mode. The overall uncertainty is obtained by adding in quadrature the individual effects.	94
5.9	The $t\bar{t}$ production cross section measured in the $W+\geq 3$ jets data sample for fits to HERWIG $t\bar{t}$ MC generated at different top masses. The uncertainties shown here are statistical only.	101

LIST OF FIGURES

Figure	Page
1.1 Leading order Feynman diagrams for $t\bar{t}$ production in $p\bar{p}$ collisions. . .	8
2.1 A schematic of the accelerator complex at Fermilab.	15
2.2 Elevation view of one half of the CDF detector.	17
2.3 The CDF tracking system.	18
2.4 Partial view of the COT showing the alternating axial and stereo layers.	20
2.5 The Run2 trigger block diagram.	25
3.1 The angle between the \cancel{E}_T and the leading jet in the transverse plane versus the \cancel{E}_T for (a) our model of the multi-jet background from non-isolated lepton data sample, (b) $t\bar{t}$ Monte Carlo, and (c) W +jets Monte Carlo.	38
4.1 A graphical representation of a biological neuron.	45
4.2 Schematic view of a three layer feed forward network.	47
5.1 The \cancel{E}_T versus isolation distribution for events with a lepton and 3 or more jets. The structure apparent near 20 GeV comes from removing events where missing transverse energy lies close to direction of the first jet.	64
5.2 Comparison plots of kinematic variables using ALPGEN W +3p and PYTHIA $t\bar{t}$ for a variety of kinematic variables in the reconstructed $W+\geq 3$ jets mode. The two histograms are normalized to equal area.	70

5.3	Comparison plots of kinematic variables using ALPGEN $W+3p$ and PYTHIA $t\bar{t}$ for a variety of kinematic variables in the reconstructed $W+\geq 3$ jets mode. The two histograms are normalized to equal area.	71
5.4	Comparison plots for some shape variables using ALPGEN $W+3p$ and PYTHIA $t\bar{t}$ for a variety of kinematic variables in the reconstructed $W+\geq 3$ jets mode. The two histograms are normalized to equal area.	72
5.5	Comparison plots for some shape variables using ALPGEN $W+3p$ and PYTHIA $t\bar{t}$ for a variety of kinematic variables in the reconstructed $W+\geq 3$ jets mode. The two histograms are normalized to equal area.	73
5.6	Expected statistical sensitivity from fitting variable distribution in the $W+\geq 3$ jets data sample. Points mark the median of the distribution of fit the uncertainty relative to the fitted $t\bar{t}$ fraction in pseudo-experiments. The error bars mark the 16-84 percentile interval in the relative uncertainty distribution. One should note that many of these variables are correlated.	76
5.7	Expected statistical sensitivity from fitting variable distribution in the $W+\geq 4$ jets data sample. Points mark the median of the distribution of fit the uncertainty relative to the fitted $t\bar{t}$ fraction in pseudo-experiments. The error bars mark the 16-84 percentile interval in the relative uncertainty distribution. One should note that many of these variables are correlated.	77
5.8	Typical variation of the training and testing ANN error with the number of training epochs.	80
5.9	The output for ANN's trained in the $W+\geq 3(4)$ jets mode using PYTHIA $t\bar{t}$ and ALPGEN $W+3(4)p$ MC.	80
5.10	Average correlations coefficients for some 26 different kinematic variables in $W+3p$ (left) and $t\bar{t}$ (right) simulated events. The ordering of variables in the two plots is the same.	81
5.11	Predicted statistical and systematic uncertainties relative to a measured $t\bar{t}$ fraction for fitting the ANN output in the 3 ore more jets mode. The average number events in pseudo-experiments corresponds to a 195^{-1}pb data sample. Statistical error bars, not shown here, should be slightly smaller in size than the triangles in the figure. . .	84

5.12	The leading jet E_T , \cancel{E}_T and the ANN input variables distributions in $W+3$ jets (exclusive) events compared to the prediction from ALPGEN+HERWIG MC, multi-jet background and PYTHIA $t\bar{t}$ MC. These are fit results, the MC model histogram is normalized to the number of events in the data sample. The KS test statistic is shown as a measure for the compatibility between the two shapes.	86
5.13	Distribution of the correlation coefficients in data compared to a mix of 10% PYTHIA $t\bar{t}$ and 90% ALPGEN $W+3p$ events. Plots are in the $W+3$ jets exclusive mode. Histograms are normalized to equal area. For a quantitative evaluation of the compatibility between the two shapes, each plot has attached the KS test result.	87
5.14	Distribution of the correlation coefficients in data compared to mix of 10% PYTHIA $t\bar{t}$ and 90% ALPGEN $W+3p$ events. Plots are in the $W+3$ jets exclusive mode. Histograms are normalized to equal area. For a quantitative evaluation of the compatibility between the two shapes, each plot has attached the KS test result.	88
5.15	Distribution of the correlation coefficients in data compared to mix of 10% PYTHIA $t\bar{t}$ and 90% ALPGEN $W+3p$ events. Plots are in the $W+3$ jets exclusive mode. two histograms are normalized to equal area. For a quantitative evaluation of the compatibility between the two shapes, each plot has attached the KS test result.	89
5.16	Distribution of the correlation coefficients in data compared to mix of 10% PYTHIA $t\bar{t}$ and 90% ALPGEN $W+3p$ events. Plots are in the $W+3$ jets exclusive mode. Histograms are normalized to equal area. For a quantitative evaluation of the compatibility between the two shapes, each plot has attached the KS test result.	90
5.17	Test of the fitting technique under different signal hypotheses for the $t\bar{t}$ cross section. The error bars mark the 16-84 percentile interval in the distribution of measured cross section in pseudo-experiments. . .	96
5.18	Fit pull distribution from pseudo-experiments for fits using the ANN's trained in the $W+\geq 3$ and $W+\geq 4$ jets mode.	96

5.19	Fit to the data using the $t\bar{t}$ W +jets and QCD background ANN templates in the $W+\geq 3$ (top) and $W+\geq 4$ (bottom) jets samples assuming a top mass of $175 \text{ GeV}/c^2$. The relative uncertainty returned by the fitter is compared to relative uncertainty distribution in pseudo-experiments (right).	99
5.20	Measured $t\bar{t}$ fraction in the $W+\geq 3$ jets data sample (top) and $W+\geq 4$ jets data sample (bottom) for 20 different single variables assuming a top mass of $175 \text{ GeV}/c^2$. The uncertainty shown here is statistical only, the vertical line shows the expected fraction for a theory motivated $6.7 \text{ pb } t\bar{t}$ cross section.	100
5.21	The $t\bar{t}$ production cross section in the $W+\geq 3$ jets data sample from fits to HERWIG $t\bar{t}$ MC samples generated at different top masses. The uncertainties shown are statistical, except for the 175 GeV point where the systematic uncertainties are shown in black.	101
5.22	The SECVTX b-tagged data compared to the tagged PYTHIA $t\bar{t}$ MC in the $W+\geq 3$ jets sample. This is not a fit, instead the two histograms are normalized to equal area.	102
5.23	Comparison between data and MC simulation ANN output in $W+\geq 3$ jets exclusive mode. No b-tagging requirement was made here. This is not a fit, instead the two histograms are normalized to equal area. . .	103

CHAPTER 1

INTRODUCTION

1.1 The Standard Model of Particle Physics

The current understanding of the elementary particles and their interactions forms the basis for the Standard Model of particles and fields (SM). At the atomic level ($< 10^{-10}$ m), matter is composed of point-like electrons orbiting atomic nuclei which are tightly packed structures of proton and neutrons. Reducing the scale under the characteristic size of a proton (10^{-15} m), proton and neutrons are revealed to be composed of spin 1/2 fermions called quarks. Six quark flavors have been discovered so far: u , d , c , s , b and t . The electron is one of the three known charged leptons: e , μ and τ . The lepton content of the SM is completed by three light neutral leptons ν_e , ν_μ , ν_τ . Like the quarks, all leptons are spin 1/2 fermions. Within the current experimental resolution of 10^{-19} m, the quarks and leptons are known to be structureless elementary particles.

The matter particle in the Standard Model along with the force carriers are shown in Table 1.1. Four types of forces are known in nature. Listed here in the order of their strength starting with the weakest, these forces are: gravitation, weak, electromagnetic and strong. Except for the case of gravitational forces which will not be discussed here, the carriers of the interactions are spin 1 gauge bosons.

particle generation			force carriers
1'st	2'nd	3'rd	
u	c	t	γ
d	s	b	g
ν_e	ν_μ	ν_τ	Z^0
e	μ	τ	W^\pm

Table 1.1: The elementary particles in the Standard Model of particle physics.

As far as the strong force is concerned, the state of a quark can be represented by a vector in a 3 dimensional space where elements of the $SU(3)$ symmetry group act as linear operators. A quark can carry one of the three color charges, conventionally named: red (R), green (G) and blue (B). The three color states form a fundamental triplet representation for the $SU(3)$ color gauge group. Strong interactions are mediated by 8 massless gluons which carry two color charges and form an octet representation of the color $SU(3)$. Being a nonabelian gauge theory, QCD exhibits the phenomenon of asymptotic freedom [4, 5]: the color forces are weak at small distances (high momentum transfers) and strong at large distances (low momentum transfers). It also exhibits the property of color confinement: the only finite-energy asymptotic states of the theory are color singlets. Thus the physical bound states of quarks such as mesons or baryons must exist in a singlet representation of $SU(3)$: RGB , $R\bar{R}$ etc. Since quarks carry color charge they are not believed to exist as free on-shell particles, which is consistent with all the observations made so far [6].

Both the leptons and quarks come in six flavors and can be grouped in three generations as shown in Table 1.1. Leptons do not carry color charges and thus have no strong interactions. They interact with other leptons and quarks via electromagnetic

and weak forces. The electromagnetic and weak forces have a unified description in the $SU(2)_L \times U(1)_Y$ gauge theory [1, 2, 3], with the force carrier particles being a massless photon and the massive W^\pm and Z^0 bosons. In SM different helicity states behave differently with respect to the weak interactions, as shown for particle states in Table 1.2. The corresponding antiparticle states feature the opposed sign for additive quantum numbers and reversed helicity. The negative helicity (left-handed) particle states form $SU(2)$ doublets, the same being true for the positive helicity (right-handed) antiparticle states. The positive helicity particle states and negative helicity antiparticle states are $SU(2)$ singlets and thus do not couple to the $SU(2)$ charged gauge fields W^\pm . They interact with other particles exchanging Z bosons and photons. The neutrinos carry no electric charge. Only left-handed neutrinos (ν_L) and right handed antineutrinos ($\bar{\nu}_R$) have been directly observed and are accommodated by the SM [7].

$$\begin{array}{ccc}
 \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L & \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L & \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L & & \begin{pmatrix} u \\ d \end{pmatrix}_L & \begin{pmatrix} c \\ s \end{pmatrix}_L & \begin{pmatrix} t \\ b \end{pmatrix}_L \\
 e_R^- & \mu_R^- & \tau_R^- & & u_R, d_R & c_R, s_R & t_R, b_R
 \end{array}$$

Table 1.2: The $SU(2)$ multiplets for the three generations of fermions.

A compact¹ representation of the interaction terms in the SM Lagrangian [8] for the first generation of quarks and leptons is given bellow:

¹This is not the complete Lagrangian, also the gauge invariance is not manifest here.

$$\mathcal{L} = \sum_{f=\nu_e, u, d} eQ_f (\bar{f}\gamma^\mu f) A^\mu + \frac{g_2}{\sqrt{2}} [(\bar{u}_L\gamma^\mu d_L + \bar{\nu}_{eL}\gamma^\mu e_L) W_\mu^+ + h.c.] + \frac{g_2}{\cos\theta_w} \sum_{f=\nu_e, u, d} [\bar{f}_L\gamma^\mu f_L (T_f^3 - Q_f \sin^2\theta_w) + \bar{f}_R\gamma^\mu f_R (-Q_f \sin^2\theta_w)] Z_\mu,$$

where Q_f is the fractional electric charge and T_f^3 is the weak isospin for each fermion species, e is the unit electric charge², g_2 is the weak coupling constant and θ_w is the Weinberg angle, all free parameters in the Standard Model. As a consequence of the gauge structure of the SM Lagrangian, the electroweak properties should be identical for all fermion generations, a fact that is verified experimentally [9].

It was found that charged current exchanges (W^\pm) do not only couple different particle (antiparticle) states within the same generation but also states from different generations. This is possible because quark mass eigenstates and weak eigenstates are not identical. A quark momentum eigenstate can be written as a mix of three different weak eigenstates. Mixing is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix using an expression like:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \quad (1.1)$$

where the (d', s', b') and (d, s, b) are the weak and the mass quark eigenstates respectively. The only constraint that the standard model makes on the CKM matrix is the unitarity condition, made in order to preserve the anticommutation relations

²The unit electric charge $e = 1.6 \times 10^{-19}$ C.

between the quark field operators. Values for any particular element of this matrix must be obtained from experiment. The diagonal elements of the matrix are close to unity, which means that the weak current is dominant for quarks within the same generation. The weak currents between different generations are suppressed due to much smaller off-diagonal elements. No flavor changing neutral currents reactions corresponding to tree level enhanced transitions have been detected experimentally. This observation is accommodated by the Glashow-Iliopoulos-Maiani (GIM) mechanism [10] which requires that quarks in each generation come in pairs. Soon after the b quark was discovered, the existence of the t quark was automatically inferred from the limits on flavor changing neutral b decays [11]. Precision measurements of the elements in the CKM matrix is today the subject of a very active area of research. The presence of some complex elements in this matrix is believed to be responsible for the CP nonconservation observed in c and b meson decays [12].

The principle of gauge invariance is a central feature of the standard model. It accounts for the local conservation of charges that we observe in nature, provides the framework for building a renormalizable theory of weak and strong interactions, makes definite and testable predictions on the phenomenology of electroweak interactions. In order to keep the gauge invariance of the theory, the gauge bosons are required to be massless. On the other hand it is a known experimental fact that the W^\pm and Z bosons are massive. One method proposed to resolve this contradiction is the Higgs mechanism. It requires adding to the SM at least one additional field. The Higgs field must be a spin zero complex $SU(2)$ doublet, with weak hypercharge $Y=1$ and a singlet in the color space:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} \quad (1.2)$$

and contribute to the SM Lagrangian with a potential of the form:

$$V = \mu^2(\phi\phi^\dagger) + |\lambda|(\phi\phi^\dagger)^2. \quad (1.3)$$

If the value for μ^2 is negative the potential has a minimum for a nonzero value of the Higgs field. Vacuum is by definition the lowest possible energy state and thus has a nonzero expectation value for the Higgs field. As a result, while the Lagrangian of the theory is still $SU(2)_L$ invariant, the vacuum is no longer invariant and the $SU(2)_L$ symmetry becomes spontaneously broken. Gauge bosons interact with the Higgs field and acquire an effective mass. Out of the 4 degrees of freedom of the Higgs field, 3 will generate mass for the W^\pm, Z^0 bosons while the fourth is expected to produce a massive scalar neutral particle, the Higgs boson. The lepton and quark masses can be generated using a similar mechanism introducing Yukawa couplings between the Higgs and fermion fields. The mass of the Higgs itself is not predicted from the theory since it depends on the unknown Higgs self-coupling constant λ . No Higgs particle has been discovered yet. The best lower mass limit is 114 GeV established at the 95% confidence level by the LEP2 experiment [15].

The most comprehensive tests of the SM have been performed in collider experiments for center of mass energies up to ~ 200 GeV in e^+e^- collisions, ~ 2 TeV in $p\bar{p}$ collisions and ~ 300 GeV in $e^\pm p$ collisions. While so far the agreement with SM predictions has been good, it is expected that the SM is an effective approximation to whatever new physics may exist at higher energy scales. Possible hints for new physics

beyond SM come from the neutrino oscillation³ experiments [16] and astrophysical observations [17].

1.2 Production and Properties of the Top Quark

The top quark is the heaviest of the known six quarks. It is expected to have electric charge $2/3$, weak isospin $T_3 = +1/2$ and to transform as a triplet under the color $SU(3)$. While the mass of the top quark ($M_{top}=178.0\pm 4.3 \text{ GeV}/c^2$) is known with better precision than for any other quark, none of its quantum numbers, including spin, have been directly measured so far. The first measurements are expected to be performed in the Run2 experiments at the Fermilab $p\bar{p}$ collider.

The dominant mechanism for the top quark production in $p\bar{p}$ collisions is QCD pair production processes Figure 1.1. At $\sqrt{s} = 1.96 \text{ TeV}$ center of mass energy $\sim 85\%$ of top pairs are produced in $q\bar{q} \rightarrow t\bar{t}$ processes while $\sim 15\%$ will be produced in a gg fusion process. In addition to the hard scattering processes, pair production can also occur through Z^0 or photon exchange but with a much smaller probability. Contribution from the qg and $\bar{q}g$ channels are negligible [18] at Tevatron.

The top pair production cross section can be calculated using perturbative QCD. The leading order (LO) expression can be written as a sum over parton-parton cross sections weighted with factors corresponding to parton distribution probability functions (PDF's) for proton (p) and antiproton (\bar{p}) respectively:

$$\sigma(p\bar{p} \rightarrow t\bar{t}) = \sum_{partons} \int dx_i F_i^p(x_i, \mu^2) \int dx_j F_j^{\bar{p}}(x_j, \mu^2) \bar{\sigma}_{ij}(\bar{s}, \mu^2, M_{top}) \quad (1.4)$$

³In SM ν_R does not exist so the neutrino cannot acquire mass via the standard Higgs mechanism.

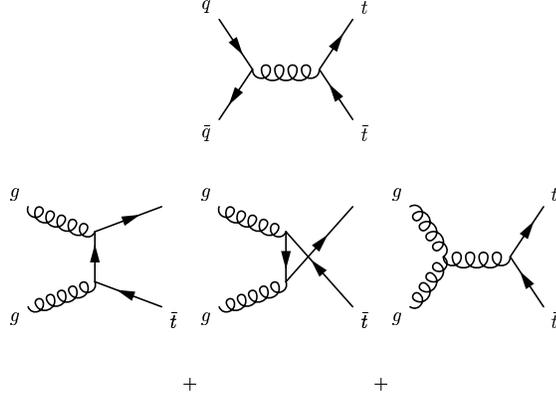


Figure 1.1: Leading order Feynman diagrams for $t\bar{t}$ production in $p\bar{p}$ collisions.

Here $F_i^p, F_j^{\bar{p}}$ are the PDF's for quarks and gluons in p and \bar{p} , x_i is the fraction of the proton (antiproton) momentum carried by the i -th parton and $\bar{s} \sim sx_i x_j$ is the center of mass energy for the parton-parton hard scattering. The cross section is dependent on the center of mass energy s , the mass of the top quark M_{top} , the strong coupling α_S and the momentum scale μ^2 at which both α_S and PDF's are evaluated. In QCD calculations higher order contributions are often significant. As a simple example, the n -th order contributions to a physical cross section for a process like $e^+e^- \rightarrow hadrons$ is of the form:

$$\alpha_S(\mu_r)^n (\ln(\mu_r^2/s))^{n-1} + \dots \quad (1.5)$$

where $\alpha_S(\mu)$ is the QCD running coupling calculated at the renormalization scale μ_r . The impact of higher order contributions can be reduced by using a renormalization scale μ_r relevant for this process. Of course the full QCD cross section is independent of the choice of the renormalization scale but since the calculation is

performed up to a finite order in perturbation expansion, the calculated cross section is dependent on μ_r^2 .

In $p\bar{p}$ collisions, the PDF's functions account for soft processes inside the proton where the coupling constant becomes strong and which can not be calculated using the perturbation theory. They are derived from fits to experimental data at some particular energy scale Q^2 and perturbative equations are used to evolve those PDF's up to a factorization scale μ_f relevant to the $t\bar{t}$ production [22]. For top cross section calculation usually both the factorization and renormalization scales are fixed at $\mu_r = \mu_f = \mu = M_{top}$. The dependence of PDF's and α_s on the choice of scale is logarithmic and uncertainties are estimated by varying the μ by a factor of 2 around the top mass. This procedure yields a estimate of the uncertainty due to missing higher order terms.

In ref. [18] the full NLO calculation is found to be 25% higher than the LO cross section. The α_s^3 corrections to the $q\bar{q}$ channel are small whereas for the gg channel corrections are 70% of the LO term. The cross section is a strong function of top quark mass. It has been argued that much of the NLO contribution comes from initial state gluon radiation near the $t\bar{t}$ threshold. This situation suggests that the soft gluon contribution corresponding to higher order corrections might be important. Resummation techniques can be used to improve the treatment of soft gluons beyond NLO, resulting in corrections ranging from 10% [18, 19] to 1% in [20]. Electroweak corrections are at the level of 1% for a large range of Higgs masses [20].

Significant uncertainties in the predicted cross section arise due to limited knowledge of PDF's and the assumed value of Λ_{QCD} which affects the μ^2 dependence of both α_s and PDF's.

The most recent estimations for top quark cross section at Tevatron were presented in ref. [23] and [26]. Cacciari et al [23], perform a NLO calculation and take into account resummation effects at NLL level. They evaluate the factorization and renormalization scale uncertainty to be $\sim 5\%$. They use PDF's extracted using NLO calculations and propagated using NLO evolution equations. Using the most recent sets of PDF's (MRST[24] and CTEQ[25]) which come with a set of $\pm\sigma$ uncertainties for each parameter in the fit, they evaluate the overall impact of the PDF's uncertainties on the total $t\bar{t}$ cross section. The largest uncertainty comes from the gg channel which, depending on the particular choice of PDF's, accounts for 10 – 20% of the total cross section. Uncertainties from the strong coupling constant were evaluated by varying $\alpha_S(M_Z)$ between 0.117 and 0.121. At $s = 1.96$ TeV $\sigma_{t\bar{t}}$ is estimated to be ~ 6.7 pb with an overall uncertainty of 15%.

Kidonakis and Vogt [26] perform a NNLO calculation (actually include only dominant NNLO terms) and find a similar central result with a factorization and renormalization scale dependence of only 3%. In this calculation the PDF's uncertainties are not evaluated (complete NNLO PDF's are not yet available).

Within the SM, the top quark decays with $\sim 100\%$ branching ratio via $t \rightarrow W^+b$ with a width $\Gamma \sim 1.4$ GeV. The top lifetime is $\sim 4 \times 10^{-25}$ s while the hadronization process should occur on $\sim 10^{-23}$ s time scale. As a result the top quark decays before hadronizing. The $t\bar{t}$ final state contains a b and a \bar{b} quark which materialize as jets in the detector. The W bosons decay on the same time scale as the parent top with almost the same probability in any of the following channels: (e, ν_e) , (μ, ν_μ) , (τ, ν_τ) , (\bar{u}, d') , (\bar{c}, s') . Since quarks come in 3 colors the probability for producing a particular species of leptons in a W boson decay is $\sim 11\%$ while the probability to observe 2

jets (regardless of the original quark flavor), is $\sim 66\%$. Thus the signal from top pair production can be detected in 3 different channels:

- *Dilepton channel:* both W 's bosons decay leptonically with e^+e^- , $\mu^+\mu^-$ or $e^\pm\mu^\mp$ in the final state. Backgrounds tend to be small in this channel but the branching ratio is also small: $4/81$. Some of the τ lepton events where τ decays leptonically will contribute to signal in this channel. The $t\bar{t}$ signature is 2 high p_T leptons, large missing transverse energy due to the neutrinos⁴ and possibly two jets from the b quarks.
- *Lepton+jets channel:* one W boson decays leptonically while the other decays hadronically. The branching ratio for e and μ channels is $8/27$. A small contribution from $W \rightarrow \tau\nu$ decays is also expected. One high p_T lepton, considerable missing energy, four jets, at least two them heavy flavor jets, are produced in the final state. This channel has a large background from inclusive W boson production associated with jets and a smaller backgrounds from QCD jet production, Z +jets and dibosons WW, WZ, ZZ + jets and single top production.
- *All jets channel:* both W 's decay hadronically, the final state has 6 possibly distinct jets. It has the largest branching ratio ($\sim 44\%$) but the background from multijet QCD processes is 2-3 orders of magnitude larger than the signal.

In the following chapters of this Thesis will report on a measurement of the cross section for the top pair production in $p\bar{p}$ collisions at 1.96 TeV performed in the “lepton+jets” channel by the CDF experiment at the Tevatron.

⁴Neutrinos do not interact with the detector and their presence is often inferred from an imbalance in the transverse energy.

In the next chapter will briefly discuss the Fermilab Tevatron and the CDF detector. Event reconstruction and simulation of events in the CDF detector are discussed in Chapter 3. The Chapter 4 contains a brief introduction to the field of Artificial Neural Networks and their applications to high energy experiments. The $t\bar{t}$ the cross section measurement and results are presented in Chapter 5.

CHAPTER 2

THE CDF EXPERIMENT AT THE TEVATRON

2.1 The Fermilab Accelerators Complex

The Fermilab Laboratory is located 40 miles west of Chicago in Batavia, Illinois. The Tevatron is the largest of the Fermilab accelerators with a circumference of about 4 miles. Here, during Run1 (1992-1996), the top quark was jointly discovered by the CDF and the DØ experiments in $p\bar{p}$ collisions at 1.8 TeV center of mass energy [27]. After the success of Run1, the Tevatron was upgraded at 1.96 TeV center of mass energy and higher $p\bar{p}$ luminosity while both the CDF and the DØ experiments improved their detectors in order to pursue a detailed study of the top and bottom quarks, perform various QCD and electroweak measurements and search for new physics near the TeV scale.

In Figure 2.1 a schematic for the Run2 accelerator complex is presented. The protons are extracted from hydrogen gas placed under an electrically charged dome. Negatively charged hydrogen ions are produced and accelerated under the dome-ground potential difference of 750 KeV. A 500-foot linear accelerator is used to further accelerate the ions up to 400 MeV. Then, the hydrogen ions pass through a carbon foil where the electrons are stripped off leaving only the protons. Protons are transferred to a 75 m radius synchrotron called the Booster, where their energy is increased

to 8 GeV. A synchrotron is a circular array of radio-frequency (RF) cavities. As charged particles move through the synchrotron, they are accelerated by the electric component of the RF oscillating electromagnetic field in the cavities. The accelerated particles travel in bunches⁵ which are tens of centimeters long and less than 10^{-2} mm² thin. Particles build up energy progressively, traveling thousands of times through the accelerator ring each second. As the velocity of the particles changes, so does the frequency for each bunch to pass through any given region of the accelerator. The oscillation of the RF fields is synchronously adjusted with the increasing speed of the particles, in order to produce always a positive boost upon the particles in the beam.

The next phase of the acceleration takes place in the Main Injector, a synchrotron having seven times the circumference of the Booster which can accelerate protons up to 150 GeV.

Some protons are extracted at an energy of 120 GeV from the Main Injector and used to bombard a nickel target and produce antiprotons which are transferred to the accumulator and then to the debuncher⁶. Here antiprotons are stochastically cooled and initially stored. When enough antiprotons are accumulated they are transferred into the Main Injector and accelerated up to an energy of 150 GeV.

The Tevatron can accept both protons and antiprotons from the Main Injector and accelerate them from 150 GeV to 980 GeV. Tevatron magnets are made of superconducting niobium/titanium alloy and are kept to a temperature of 4K. In the collider mode, after the protons and antiprotons are fully accelerated the beams can be stored for hours and the Tevatron is functioning as a storage ring. During a normal

⁵A typical bunch contains up to $10^{11} - 10^{12}$ protons or antiprotons.

⁶The accumulator and debuncher complex is sometimes referred to as the Antiproton Source.

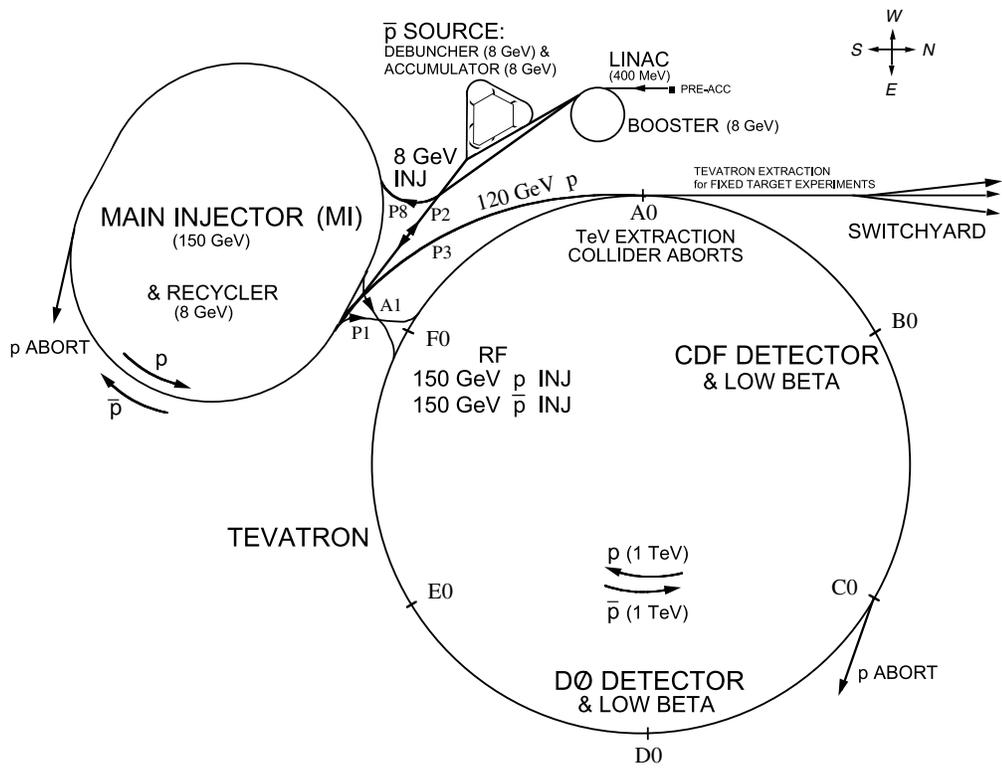


Figure 2.1: A schematic of the accelerator complex at Fermilab.

operation mode, 36 bunches of protons and 36 bunches of antiprotons are circulating the Tevatron ring. Beam crossings take place every 396 ns at two different locations in the ring named B0 and D0 where the CDF and respectively the DØ experiments are built. The record instantaneous luminosity achieved so far is $10.3 \times 10^{31} \text{ cm}^{-2}\text{s}^{-1}$ (July 2004). This number will hopefully increase over the next few years with the addition of an antiproton recycler, improved antiproton production, cooling and collection efficiency.

2.2 The CDF Detector

The Collider Detector at Fermilab (CDF) is a general purpose experiment combining tracking, calorimetry and muon detection. It was built with the intention of exploring a broad range of physics topics in $p\bar{p}$ collisions at 1.96 TeV. Pending the start of the LHC scheduled toward the end of this decade, this is the world highest energy available in the laboratory.

A schematic view of the detector can be found in Figure 2.2. The main tracking device is a drift chamber having cylindrical symmetry around the beam axis. It is contained in a 4.8 m long superconducting solenoid, 1.5 m in radius that generates a 1.4 T magnetic field parallel to the beam axis. Next, outside of the solenoid, sit the calorimeters where most of the particles are absorbed and their energy is measured. Muons interact weakly with matter and usually escape the detector. They are identified in the muon drift chambers, constructed at the periphery of the detector. Inside the solenoid, closest to the beam pipe, is placed the silicon system yielding precise position information on charged particle tracks.

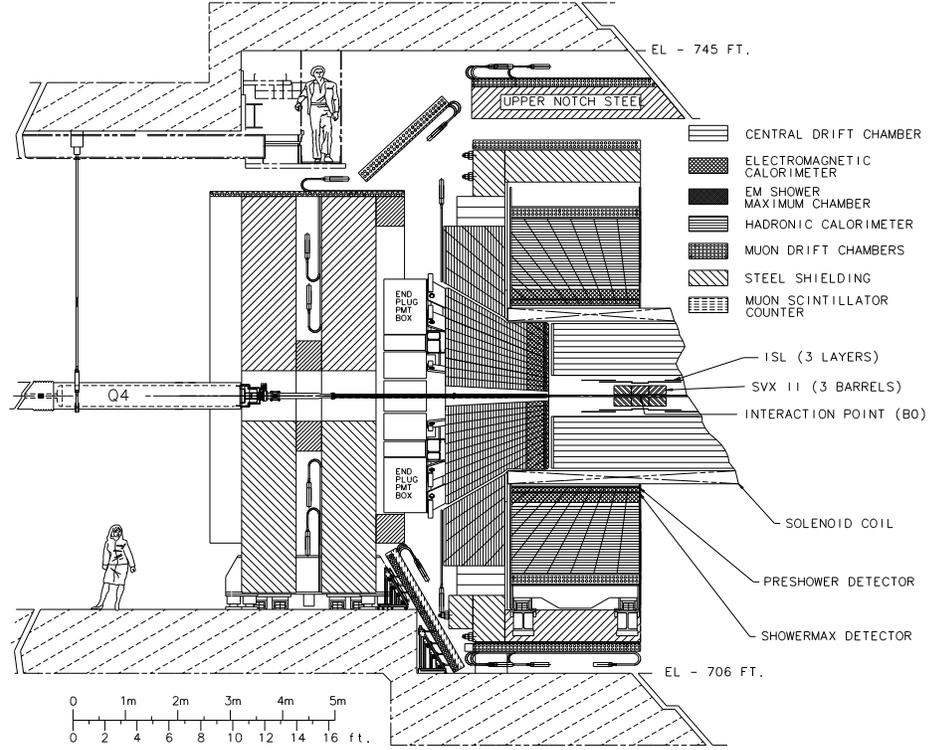


Figure 2.2: Elevation view of one half of the CDF detector.

A more detailed picture the tracker system system is shown in Figure 2.3 where the detector coverage as a function of pseudorapidity is indicated. The pseudorapidity η is defined as a function of the polar angle θ made with the proton beam direction.

$$\eta = -\ln(\tan(\theta/2)).$$

2.2.1 The Silicon Detectors

In Run2 the silicon system (SVX2) has a larger geometric acceptance than the silicon device used during in Run1. With a total length of 96 cm it covers 2.5σ of the $p\bar{p}$ luminous region. It is composed of five layers of double sided silicon microstrip

CDF Tracking Volume

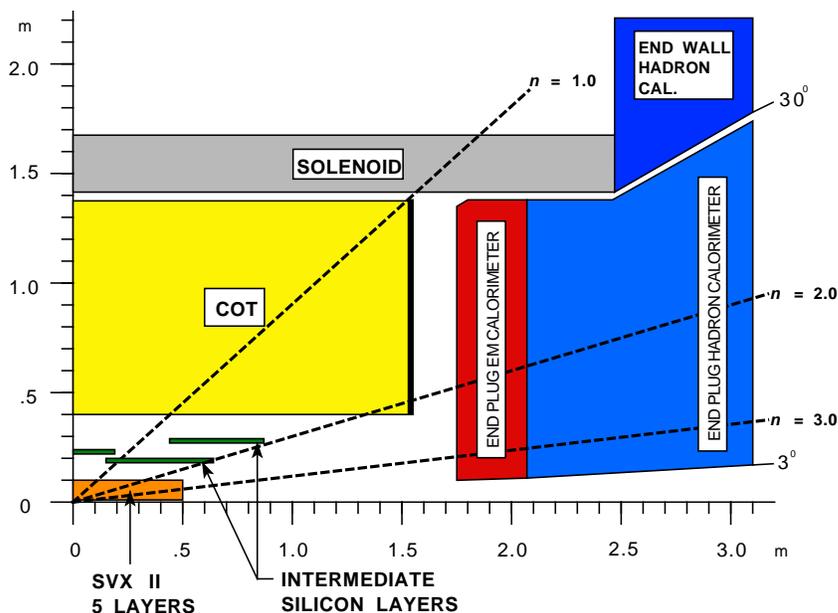


Figure 2.3: The CDF tracking system.

detectors with radii between 2.4 and 10.7 cm. The strips are oriented⁷ for $r - \phi$ measurements on one side and for $r - z$ or stereo measurements on the other. In order to reduce residual interactions, silicon crystals are supported by low mass substrates called ladders. Ladders are mounted on precision machined beryllium bulkheads that also support the water colling system for the read-out electronics. The radiation hardened read out chips are mounted on electrical hybrids on the surface of the silicon detectors. A chip has 128 channels, each channel having charge sensitive amplifiers, pipelined buffers and ADC circuits. The system has a total of 400505 channels and is read out by a highly parallel data acquisition system in $10\mu\text{s}$. In order to increase the hit efficiency an additional single sided silicon microstrip detectors

⁷ z, r, ϕ are here the cylindrical coordinates where z axis lies along beam direction

layer called Layer00, has been mounted on the beryllium beam pipe at a radius of 1.7 cm, generating $r - \phi$ hit information.

A supplementary tracking device is the Intermediate Silicon Layer (ISL). It consist of a single layer of double sided silicon chips placed at 22 cm radius in the central region and two layers of double sided silicon chips placed at 20 and 28 cm in the plug region ($1 \leq |\eta| \leq 2$), which is otherwise poorly covered by tracking in the drift chamber. The ISL system has a total 268800 channels.

The integrated silicon system allows impact parameter resolution as precise as 15 μm for high p_t tracks which is essential for identification of long lived particles. The performance of the device versus time will depend on the integrated radiation dose. The innermost system (Layer00) will receive some 0.5 MRad/fb. The device is expected to have stable performance up at least to 1.5 MRad. More information on the silicon system can be found in [28].

2.2.2 The Central Drift Chamber

The Central Outer Tracker (COT) is a cylindrical 3.1 m long drift chamber, with a 40 cm radius on the interior and 138 cm on the exterior, as shown in Figure 2.4. It is filled with 50:50 Ar-Et gas which has a drift velocity of $\sim 50\mu\text{m}/\text{ns}$. As a charged particle passes through, it ionizes the gas. Anode wires in the chamber collect the ionization electrons. By measuring the drift time of the electrons, 140 μm single hit resolution can be achieved. Wires are grouped on 8 alternating axial⁸ and stereo⁹ superlayers. Each superlayer has 12 wires, drift times are of the order of 200 ns. The total number of channels is 30240 and the material in the device accounts

⁸oriented along the z axis

⁹oriented at $\pm 2^\circ$ away from the z axis

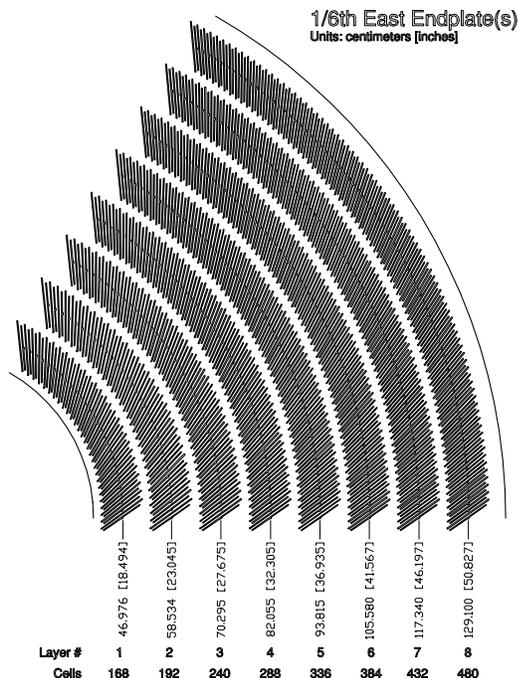


Figure 2.4: Partial view of the COT showing the alternating axial and stereo layers.

for 1.3% radiation length at normal incidence. By matching the hits within each layer, track segments are reconstructed. Track segments in different layers are then used to reconstruct the helix of a charged particle trajectory. From the curvature of a charged particle track moving in magnetic field, the transverse momentum p_t is measured with a resolution of $\delta p_T/p_T^2 \sim 0.0015 \text{ (GeV/c)}^{-1}$. The longitudinal momentum p_z can be deduced from the p_t value and the helix geometry. Charged particle tracks are measured down to 300 MeV/c transverse momenta. When silicon information is included the momentum resolution is: $\delta p_T/p_T^2 \leq 0.001 \text{ (GeV/c)}^{-1}$. Reduced efficiency may arise not so much from nonexistent signal but from attaching wrong hits to a track. Tracking reconstruction efficiency depends on the number of minbias events in the detector and as a consequence on luminosity. For luminosity as

high as $L = 5 \times 10^{31} \text{ cm}^{-2}\text{s}^{-1}$, for isolated tracks with $p_T > 10 \text{ GeV}$ the COT tracking efficiency is measured to be 98% up to $|\eta| < 1$ [29]. The overall tracking efficiency up to $|\eta| < 1$ is $\sim 95\%$. A more detailed description of the COT can be found in [30].

2.2.3 The Calorimeter System

Calorimeters are used to measure particle energies. They are composed of a dense material where the incident particle loses its energy by successive scatterings producing a shower of secondary particles. In the course of showering, most of the incident particle energy is eventually converted into “heat”, which explains the name calorimeter. A scintillator material is used to collect the energy. Signal from the scintillators is sent to photomultipliers, amplified and measured allowing reconstructing the energy of the incident particle.

The calorimeters are grouped in 2 regions: central ($|\eta| < 1.1$) and end-plug ($1.1 < |\eta| < 3.6$). Cerenkov counter luminosity monitors are placed in the forward region ($3.6 < |\eta| < 4.2$). The electromagnetic calorimeter system (EC) is placed next to the tracking system, the hadronic calorimeters (HC) are positioned further to the exterior. Lead absorber is used for the electromagnetic calorimeter (total of 18 radiation lengths), iron for the HC (total of 6.8 absorption lengths). The geometry of the tower is projective pointing to the nominal interaction point with cell size $\Delta\eta \times \Delta\phi \approx 0.1 \times 0.15$. Located six radiation lengths inside the central electromagnetic calorimeter, proportional wire chambers provide a shower position measurement in z and $r\phi$ view, used to distinguish electromagnetic from hadronic showers. Proportional chambers near the solenoid detect early shower development in the detector coil. By

matching jets with the tracks reconstructed in the COT, one can infer if the shower was produced by a charged or neutral particle.

In the case of an incident photon or electron, the incident particle will produce a shower of secondary electron-positron pairs and photons which is often contained and thus measured within the electromagnetic calorimeter. The energy resolution for the central electromagnetic calorimeter (CEM) is $14.0\%/\sqrt{E}$. Incident hadrons produce also nuclear fragments as secondary particles. While the showering may start inside the EC, the energy will be absorbed fully only in the later layers of the hadronic calorimeter. The energy resolution for a single jets in the HC is $50.0\%/\sqrt{E}$. An objective at CDF for Run2 is to calibrate the absolute jet energy scale to a 2.5% precision. A detailed description of CEM and HC can be found in [33] and [34] respectively. A review of the latest calorimetry upgrade for Run2-b can be found in [35].

2.2.4 The Muon System

Muons are identified in drift chambers surrounding the calorimeters, for rapidities up to $|\eta| < 1.0$ with a lower momentum threshold of $p_t > 1.4$ GeV/c. Muon system is composed of the Central Muon Detector (CMU) with a rapidity coverage $|\eta| < 0.6$, which consist of muon drift chambers, placed parallel to the beam outside the HC but protected from residual hadronic interactions by a 5.5 absorption lengths steel shielding. Further to the exterior, the Central Muon Upgrade (CMP) is also placed parallel to the beam, protected by an additional 60 cm thick layer of steel, having a $|\eta| < 0.6$ rapidity coverage. The chambers form a “box” around the central detector and consist of four layers of single wire drift tubes. The tubes run in proportional

mode with a drift time of $1.4 \mu\text{s}$. The signal from these chambers is read out by a single TDC per wire and trigger information is formed from coincidences with signal from nearby wires in the CMU chambers. Also for triggering purposes a layer of scintillator counters (CSP) is installed outside the surface of the drift chambers wall.

The Central Muon Extension consists of drift tubes (CMX) and scintillator counters (CSX) located at each end of the central detector between $42^\circ - 55^\circ$ polar angles, extending the rapidity coverage between $0.65 < \eta < 1.0$. No additional steel shielding was added, at these large angles, since there is already enough material in the structure of the detector.

The Intermediate Muon System (IMU) extends rapidity coverage up to 1.5. It consists of a combination of drift chambers and scintillator counters placed behind toroid magnets. There is no trigger associated with the IMU yet and we will not use information from IMU for this analysis. More information on the CDF muon system can be found in [36].

2.2.5 The Event Trigger

The collision rate inside the CDF detector is about 1.7 MHz. However the rate available for recording useful information on tape for analysis is much smaller: 50-80 Hz. The CDF trigger system is designed to select online events of physics interest and discard the much more abundant minimum bias events where p and \bar{p} hardly interact with each other. One should keep in mind that the $t\bar{t}$ cross section is 9 orders of magnitude smaller than the minbias¹⁰ cross section. Having an efficient and accurate trigger system is vital to any high luminosity collider experiment. The

¹⁰The minbias data is collected with almost no trigger bias, and typically consist of events where the proton and antiproton hardly interact and a few low p_T particles are produced.

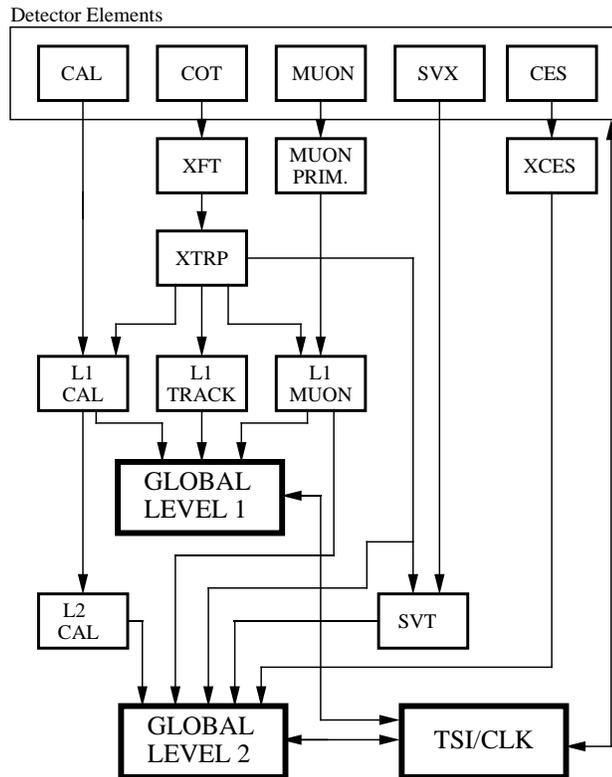
trigger system is designed to look for events where a hard interaction takes place, resulting in high P_T particles, jets, energetic electrons or muons and events with a large imbalance in transverse energy signaling the presence of a weakly interacting neutral particle like a neutrino.

Event triggering is performed at three successive levels as shown in Figure 2.5. The Level-1 trigger has a 40KHz accept rate. Events that pass Level-1 are further analyzed by the Level-2 trigger which has a 300 Hz accept rate. Finally the Level-3 trigger will allow events to be written to offline storage at a rate of 50-80 Hz. For luminosities smaller than $4 \times 10^{31} \text{ cm}^{-2}\text{s}^{-1}$, the trigger dead-time was less than 10%.

The Level-1 trigger input comes from the calorimeters, the COT and the muon detectors. An important upgrade made for Run2 is that now tracks are reconstructed within $2.7 \mu\text{s}$ after each beam crossing collision by a dedicated processor called the **eXtremely Fast Tracker** (XFT). A detailed description of the XFT can be found in [31]. Matching tracks to a electromagnetic cluster in the calorimeter or to track segment in the muon chambers allows electron and muon identification to be done on line in time for the Level-1 decision to be made. Tracks are also used in triggers such as $B \rightarrow \pi^+\pi^-$. The decision to keep an event is based on numbers and energies of electrons, muons, jets, missing transverse energy and tracks. The decision time for the Level-1 trigger is $5.5 \mu\text{s}$.

The Level-2 trigger has a $20 \mu\text{s}$ decision time. It is made of custom electronics using all the information from the Level-1 but with increased precision. Data from the shower max detector allows improved identification of photons and electrons. Jet reconstruction is done online. An important component of the Level-2 trigger is the Silicon Vertex Trigger (SVT), which triggers on tracks having large impact

RUN II TRIGGER SYSTEM



PfW 9/23/96

Figure 2.5: The Run2 trigger block diagram.

parameter, often a indication for the presence of heavy flavor quarks in the event. The SVT system is described in detail in [32].

CDF front end electronics is fully pipelined allowing it to store information for multiple beam crossings. A Level-2 accept initiates a full detector read out. At the Level-3, a computer farm performs a complete event reconstruction using the

offline reconstruction software to validate the Level-1 and the Level-2 decision. More information on CDF trigger system can be found in [37].

CHAPTER 3

EVENT RECONSTRUCTION AND SIMULATION

3.1 Introduction

Using the event trigger to select events of physics interest out of the 3 million beam crossings that take place every second inside the CDF detector is the first step toward physics analysis. After the detector information for the triggered events was read on-line and delivered to permanent storage, each event is later reconstructed offline with increased accuracy. In trying to further select collisions of interest for a given kind of physics, a few classes of objects are reconstructed within each event: electrons, muons, jets and the missing transverse energy. These objects are the flags we use for offline event selection before more specialized physics analysis is performed.

3.1.1 Reconstruction of Jets

The initial phase of $p\bar{p}$ collisions at these energies is dominated by large Q^2 hard interactions, which can be described by the perturbative QCD. A number of high transverse momentum quarks and gluons may be produced, on which the effect of the p, \bar{p} remnants is minimal. Quarks and gluons do not exist as free particles and they fragment into a spray of hadrons collimated along the direction of the initial parton momentum. These particles are detected as a cluster of energy in the calorimeters,

forming a jet. Jet structure has been identified in scatterings with transverse energy as low as 6-8 GeV [38]. At Tevatron in Run2 jets are expected to form with transverse energy as high as 600 GeV.

Jet reconstruction algorithms aim to find clusters of energy in the calorimeters and associate these to the energetic partons produced during the initial phase of the hard scattering process, before their fragmentation into many lower energy particles. For this analysis the JETCLU [39] reconstruction algorithm has been used. A brief description of the JETCLU follows here. First, a list of the towers in the calorimeter and their energy is formed. Towers containing an energy larger than 1 GeV are called seed towers. Starting from the highest seed tower, additional seed towers are added within a 7×7 tower window, forming preclusters. The preclusters in an event are ordered according to their E_T . For each precluster, a cone of radius $\Delta\Phi \times \Delta\eta = 0.4$ is reconstructed around the E_T weighted centroid. Within this cone, additional towers above 0.1 GeV are added to the precluster, and in this way a cluster is formed. The process is repeated until the tower list is stable. At this point towers are not uniquely assigned to clusters. Two clusters are merged if common towers carry more than 75% of the energy of the smallest cluster. Otherwise the clusters are split, the common towers being assigned to the closest tower. In the end towers are uniquely assigned to clusters. The process is iterated until the tower list for all clusters is stable. Finally each cluster is defined as a jet.

This algorithm suffers from a few drawbacks. It is not infrared safe, meaning that soft radiation between two jets may cause a jet merging when above seed threshold. It is not collinear safe: jets are not reconstructed if the seed energy is below the threshold, when jet energy is split among several detector towers. Also sometimes

the number of reconstructed jets depends on the initial E_T seed ordering. Improved jet algorithms have been recently developed in order to correct for some of these drawbacks by looking for stable cones around every seed tower (Midpoint), by not using seeds at all (Seedless) and by getting away from the cone clustering paradigm (K_T). A brief description of all these algorithms can be found in [40].

In the CDF detector $p\bar{p}$ collisions take place inside a luminous region having a Gaussian distribution with $\sigma_z = 30$ cm along the beam axis and $\sigma_{xy} \sim 30$ μm in the transverse plane. The spread of the collision points in the plane transverse to the beam direction has some dependence on z coordinate, which is slightly different along the x and the y directions. The jet E_T is reconstructed with respect the z position of the event. The event z position is found using COT+SVX reconstructed tracks which were fitted using a χ^2 minimization algorithm PrimeVtx [41] which allows vertex reconstruction with a resolution as high as 15 μm in the $x - y$ plane and 200 μm in the z direction, depending on the number of tracks and the topology of the event. More than one $p\bar{p}$ collision can take place during a beam crossing and multiple primary vertices may be reconstructed. Since we select $t\bar{t}$ events in the “lepton+jets” channel requiring one isolated high p_T lepton, the z position for the hard scattering event is chosen to be the nearest primary vertex within 5 cm of the z position of the point of closest approach between the lepton track and the beam line. In order to ensure good event reconstruction the primary vertex position is required to be within 60 cm from the center of the detector, which covers $\sim 95\%$ of the $p\bar{p}$ luminous region.

Jet energy measurement is affected by both detector and physics effects. Detector effects include nonuniformities in the detector response as a function of rapidity, cracks in the calorimeter and differences between central calorimeter response and

plug calorimeters. Physics effects arise from the underlying soft interaction of the p, \bar{p} remnants, multiple $p\bar{p}$ interactions and the energy that flows out of the $\Delta R = 0.4$ cone radius. Corrections are used to reproduce as accurately as possible, *on average*, the energy of the initial partons from the energy deposition measured by the calorimeters. Jet energy correction are performed on 6 successive correction levels:

- The first correction level accounts for nonuniformities in the detector and are designed to make detector response a uniform function of rapidity. These are determined mostly from dijet balancing and are applied to both data and Monte Carlo simulation (MC). Dijet balancing is performed in two jet events by balancing energy in two jets, one being measured in a precisely calibrated portion of the detector (central), the other being measured in some other rapidity region that needs to be calibrated. The CEM absolute energy scale was set using the $Z \rightarrow e^+e^-$ mass peak, the CHA scale was taken from Run1.
- The second correction level accounts for time dependent modifications in the calorimeter response due to aging phototubes and is applied only to the data.
- The third correction level takes care of scale differences between detector simulation and the real detector response. Such differences appear close to the detector calorimeter boundaries due to energy leaking out of the detector.
- The fourth correction level subtracts the average contribution due to multiple interactions as estimated from minbias trigger data. The number of reconstructed secondary vertices is used as a measure for the number of additional soft $p\bar{p}$ scatterings.

- The fifth level of corrections are the absolute energy scale corrections. They account for nonlinear calorimeter response due to many lower energy particles in a jet, energy leakage and nuclear absorption. The calorimeter simulation is based on a parametrization of the calorimeter response to single particles using data as a function of rapidity, azimuthal angle and particle energy. The central calorimeter response to single particles (e, π) was determined from test beam data at high energy and minbias events at low energy. The detector simulation is tuned to reproduce single particle response. In order to model the response of the detector to jets, fragmentation models are tuned to reproduce charged particle multiplicity and p_T distribution from data.
- The sixth level of correction accounts for the energy outside the jet clustering radius and energy from soft interactions of the p and \bar{p} remnants (the so called underlying event). These are applied to both data and MC when the energy of the original parton needs to be known.

The 5-th and 6-th correction levels are applied in those cases where the full energy of the initial partons is needed, for instance in a top quark mass reconstruction analysis. Since our strategy for this analysis is to compare data to MC simulation in order to estimate a $t\bar{t}$ contribution to the data sample, jets used in this analysis will be only corrected up to the 4th level which include: relative corrections (data and MC), time dependent corrections (data), scale corrections (data and MC) and multiple interaction (data and MC).

The total uncertainty on the jet energy for this analysis is between 5% and 12%, for jet E_T in the 15-100 GeV range [42], with jets in the central calorimeter being slightly better measured than jets in the plug region. Out of this, the uncertainty

on the absolute energy scale is $\sim 4.5\%$. In the initial phase of Run2, the jet energy corrections at CDF have been driven by the previous Run1 experience using the JETCLU cone algorithm. This does not take into account tracking information. Since some 2/3 of particles in a jet are charged and since the track momentum resolution is much better than the calorimeter resolution, including tracking information might improve the jet energy resolution in the future.

3.1.2 Electron and Muon Identification

In a W boson decay e , μ or τ leptons may be produced with the same probability. The electron and the muon leave distinctive signals in the detector. Tau leptons have a relatively short lifetime (0.29 ps) and decay soon after being produced. A fraction of the τ 's can decay into a e or μ and two neutrinos and thus contribute to the signal we observe in the “lepton+jets” channel. In many cases though, the τ 's decay hadronically¹¹ producing jets. Tau jets can in principle be separated from hadronic jets but this cannot be trivially done. Here, we will limit ourselves to select events where one high p_T , isolated e or μ is detected accompanied with jets and missing transverse energy. The selection criteria described below are designed to select the leptonic W decays and reject softer e or μ produced in b, c meson decays, jets misidentified as leptons, conversion electrons and cosmic ray muons. Other physics analysis may use different definitions and cuts for an object identified as electron or muon.

Electrons are identified by matching a COT track to a cluster of energy in the EM calorimeter. Only electrons detected in the central calorimeter are considered for this analysis. A lower 20 GeV cut is made on the lepton energy in order to eliminate

¹¹The τ lepton decays hadronically $\sim 65\%$ of the time.

softer electrons from background processes. In order to make sure the electron shower is fully contained within the calorimeter boundaries and well measured, fiducial requirements are made on the electron shower position as identified by the proportional chambers located inside the CEM calorimeter at the approximate shower maximum. The momentum of the track associated with the lepton can not be smaller than 1/2 of the electron energy. This is not a tight requirement and allows for photons being radiated off the electron and detected in the EM calorimeter at the same time with the original electron. To reject stray charged hadrons like π^\pm or hadronic jets having only one charged track, the energy deposited in the hadronic calorimeter can not exceed a certain fraction of the energy deposited in the electromagnetic calorimeter. Also, the lateral shower profile in the calorimeter cluster is compared against that observed for test beams electrons. Jets reconstructed in the electromagnetic calorimeter which are identified as electrons are removed from the hadronic jets list. A series of quality requirements are made on the COT electron track, see Table 3.1 for more details. Sometimes a photon close to the detector components (wires, pipes, metal walls) may produce a pair of conversion electrons. Electrons from photon conversion are identified by looking for tracks with opposite charge lying very close in the direction of the electron track. Events with identified conversion electrons are removed from the event list.

Muons are identified by matching a COT track to track segments in the muon chambers. Quality requirements are made on the COT muon track. The muon transverse momentum is measured from the reconstructed COT track which must be consistent with being a minimum ionizing particle leaving only a small amount of energy in the electromagnetic and hadronic calorimeters. The momentum of the track

must be larger than 20 GeV, and is subsequently corrected for the small deviation in beam position from the center of the detector. In order to separate muons produced in a $p\bar{p}$ event from cosmic ray muons, a cut is made on the distance of closest approach between the track and the beam line. Further cosmic muon identification is achieved by using the timing capabilities of the COT to distinguish particles traveling into from those traveling out of the detector. Events with muons consistent with being cosmic ray muons (traveling from the exterior into the detector) are removed from the event list.

Leptons produced in a W boson decay tend to be more isolated than leptons produced in background processes, mostly b and c quarks decaying leptonically and misidentified jets. For both electron and muons, we require that the extra calorimeter energy in a cone of radius $R = 0.4$ around the lepton direction to be less than 10% of the lepton energy (the lepton isolation $I < 0.1$).

3.1.3 Missing Transverse Energy Measurement

Neutrinos do not interact in the detector. Since the initial proton and antiprotons have negligible transverse motion with respect to the beam¹², the presence of a neutrino in an event is assumed when large imbalance in the transverse energy is detected. The transverse energy is defined as $\cancel{E}_T = -|\sum_i E^i \hat{n}_i|$ where E^i is the energy in the calorimeter cell i within $|\eta| < 3.2$ and \hat{n}_i is the direction unit vector of the tower in the plane transverse to the beam direction. Muons are not typically included when doing the missing energy budget since many identified muons can be produced by cosmic rays. Some muon tracks may be fake muons produced by punch-through into the muon chambers from the hadronic jets in the calorimeter and since they

¹²Also the partons inside p, \bar{p} have negligible transverse momentum.

Variable	Value
Global event quantities	
Z and cosmic veto	applied
Missing E_T	> 20 GeV
$\Delta\phi(\text{MET,Lead Jet})$	> 0.5 and < 2.5 if $\cancel{E}_T < 30$ GeV
# tight leptons	1
Jets	
E_T	≥ 15 GeV (corrected to level 4)
$ \eta $	< 2
Tight electrons	
Fiducial and CEM	= 1
E_T	≥ 20 GeV
p_T	≥ 10 GeV/ c
E/p (if $p_T \leq 50$ GeV/ c)	< 2.0
E_{had}/E_{EM}	$\leq 0.055 + 0.00045 * E$
L_{shr}	< 0.2
$ \Delta z $	< 3 cm
$ \Delta x $	> -1.5 and < 3.0 cm
χ_{Strip}^2	< 10
$ \Delta z_0 $	< 60 cm
$ Z_{vtx} - z_0 $	< 5 cm
# axial segments	≥ 3
# stereo segments	≥ 3
Isolation	< 0.1
Not a conversion	$\Delta\Lambda > 0.04$ and $\Delta r > 0.2$
Tight muons	
Region	CMUP or CMX
p_T	≥ 20 GeV/ c
E_{EM}	$< \max(2, 2 + 0.0115 * (p-100))$
E_{HAD}	$< \max(6, 6 + 0.0280 * (p-100))$
$ \Delta x _{CMU}$	< 3.0 cm
$ \Delta x _{CMP}$	< 5.0 cm
$ \Delta x _{CMX}$	< 6.0 cm
$ \Delta z_0 $	< 60 cm
$ Z_{vtx} - z_0 $	< 5 cm
d_0	$\geq 0.02(0.2)$ cm with (without) SVX
# axial segments	≥ 3
# stereo segments	≥ 3
Isolation	< 0.1

Table 3.1: Summary of the event selection criteria.

have no corresponding COT track, their momentum can not be precisely measured. Sometimes real muons resulting from in-flight meson decays may be matched to a wrong COT track and be mismeasured. However when an isolated energetic muon originating from the primary vertex is identified, its momentum is subtracted from the missing energy and to avoid double counting, the \cancel{E}_T calculation is corrected for the energy in calorimeter cells that contain the muon track.

There are situations where the missing transverse energy is not the result of a neutrino from a W boson decay: since the calorimeter energy resolution is limited¹³ it may happen that the energy of a jet is mismeasured, resulting in fake missing transverse energy; also neutrinos can be produced in the decay chain of b particles producing real missing energy. Sometimes a jet containing a single charged track and several neutral particles like π^0 's, can fake a lepton. When \cancel{E}_T is accompanied by a real or misidentified lepton, the event has the characteristics of a $W \rightarrow l\nu$ event. Such events where \cancel{E}_T is not from a W boson decay are generically called multijet fake events (or QCD fakes).

In Figure 3.1 the distributions in \cancel{E}_T versus the azimuthal angle between missing energy transverse direction and the direction of the leading jet, $\Delta\phi$, are shown for a nonisolated data sample ($I > 0.2$), the Monte Carlo simulated W +jets background and the $t\bar{t}$ signal. The kinematics of the nonisolated data sample are expected to be similar to the kinematics of multijet fake events. One can see in Figure 3.1 that most of the events with $I > 0.2$ are clustered at \cancel{E}_T having small angles with respect to the leading jet. In order to reduce the occurrence of such events in our data sample ($I < 0.1$), will try to apply a cut at low \cancel{E}_T by removing events having the direction

¹³The energy resolution is $14.0\%/\sqrt{E}$ (GeV) for CEM and $50.0\%/\sqrt{E}$ (GeV) for HC.

of the \cancel{E}_T close in azimuth to the direction of the leading jet in the event. Our estimates¹⁴ for the number of QCD fake events in our data sample show that such a cut is indeed efficient, removing $\sim 40\%$ of the QCD fakes in our data sample with the price of only a 5% loss on signal efficiency.

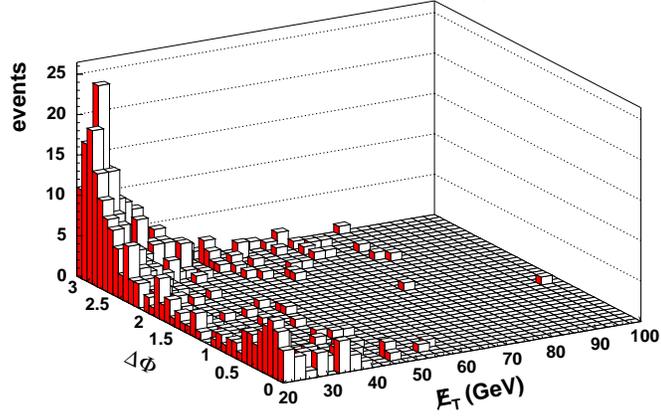
3.2 The Event Selection

We present in this section the event selection requirements for this analysis, see also Table 3.1 for a compact summary and precise numerical values. We correct jet energy up to Level 4. We ask for at least three jets with corrected $E_T \geq 15.0$ GeV and rapidity $\eta < 2.0$. In order to make sure the jets and the leptons are reconstructed from the same interaction point, the event z position is defined by the closest reconstructed vertex with at least two good quality COT tracks within a 5 cm window from the z position given by the identified tight lepton track. In order to ensure good event reconstruction the primary vertex position is required to be within 60 cm from the center of the detector.

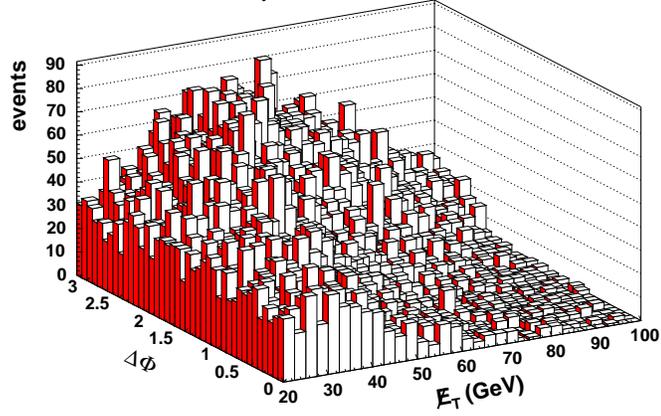
The missing transverse energy after corrections must be larger than 20 GeV, with the above mentioned cut on \cancel{E}_T direction if $20 \leq \cancel{E}_T \leq 30$ (GeV). The various cuts for electron and muon selection have been studied and optimized by the Top Group at CDF. In this analysis we use CEM electrons, CMX and CMUP muons. Including plug electrons could produce a 30% increase in the acceptance for $t\bar{t}$ events. But since there is no established method to identify conversion electrons from real plug electrons, we do not yet include plug electrons here. We require a single energetic lepton ($E_T \geq 20$ GeV), either a electron or a muon, with isolation $I < 0.1$. The

¹⁴See Section 5.3.

(a) Data Sample (nonisolated leptons), $N_j \geq 3$



(b) PYTHIA $t\bar{t}$ MC, $N_j \geq 3$



(c) ALPGEN+HERWIG W+jets MC, $N_j \geq 3$

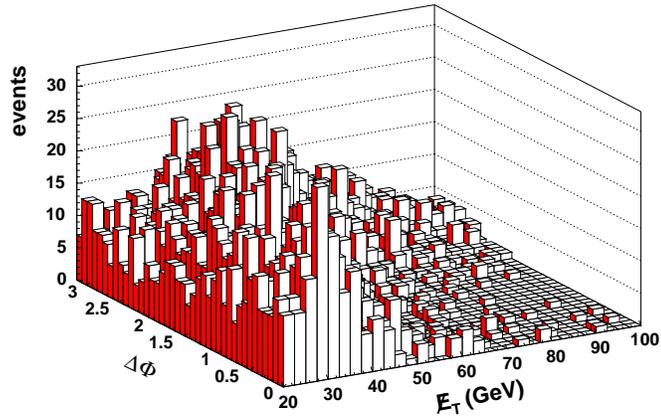


Figure 3.1: The angle between the \cancel{E}_T and the leading jet in the transverse plane versus the \cancel{E}_T for (a) our model of the multi-jet background from non-isolated lepton data sample, (b) $t\bar{t}$ Monte Carlo, and (c) W +jets Monte Carlo.

COT track associated with the lepton must have at least 3 axial and 3 stereo good¹⁵ segments.

In order to distinguish electrons from hadronic jets, a cut is made on the ratio of energy deposited in the CEM to the energy deposited in the hadronic calorimeter: E_{had}/E_{EM} . The momentum of the track associated with the electron can not be smaller than 1/2 of the energy measured in the CEM. Cuts are made on the distance between the extrapolated track position and the CES shower position as measured in the $r - \phi$ and z views: Δx and respectively Δz . The χ^2_{Strip} is a measure for the compatibility of the CES shower profile with that from test beam electrons.

Conversion electrons are identified by looking for a nearby oppositely charged track within $\Delta r < 0.2$ cm distance in the $r-\phi$ plane and having up to $\Delta\Lambda < 0.04$ difference between the cotangent of polar angles. Sometimes the conversion partner has an additional conversion track itself. This may happen when an electron radiates a virtual photon which is converted into a pair of real electrons. We do not remove events where multiple conversion partners are found in order to reduce the over-efficiency of the conversion algorithm.

To reject cosmic ray muons a cut on the muon track impact parameter d_0 , is made. Muon tracks in the COT must extrapolate to a track segment in the muon chambers within a maximum distance Δx , and leave an energy deposition in the electromagnetic (hadronic) calorimeter E_{EM} , (E_{HAD}) consistent with a minimum ionizing particle.

In order to keep the cross section measurement orthogonal with the $t\bar{t}$ cross section measurement using dilepton decays, events with two isolated leptons are rejected. When an isolated electron is detected in the plug in addition to a central isolated

¹⁵A good track segment must have at least 7 hits out of a total of 12 possible hits.

electron or muon, the event is also excluded. A measurement of the $t\bar{t}$ cross section in the dilepton channel has been published elsewhere [43].

We veto events consistent with a Z decay when in addition to the primary lepton a second object of opposite charge exists forming an invariant mass within the 76-106 GeV window. The second object can be an isolated track or a jet with EM energy fraction $\geq 95\%$ if the primary lepton was an electron.

3.2.1 The Data Sample

The data sample for this analysis was collected by the CDF experiment during Run2, between March 2002 and September 2003. The datasets were based on the 4.8.4 offline release with calorimeter-dependent reconstruction remade in 4.11.1. In order to exclude those runs where parts of the detector did not function properly, quality requirements are made for all runs. Such requirements cover most of the detector components, however the silicon is not required to be active in this data sample. The integrated luminosity of the sample was measured [75] to be $193.5 \pm 11.4 \text{ pb}^{-1}$ for CEM and CMUP triggers and $175.3 \pm 10.3 \text{ pb}^{-1}$ for CMX triggers, where the uncertainty of 5.9% comes from 4.4% uncertainty on the acceptance of the z cut and 4.0% from the calculation of the total $p\bar{p}$ cross section ($60.7 \pm 2.4 \text{ mb}$).

Table 3.2 shows the number of observed events in the data as a function of reconstructed jet multiplicity bin. All of the selection criteria described above have been applied. The data is broken out in terms of electron+jets and muon+jets.

Njets	CEM	CMUP	CMX	Total
0	99454	49672	26531	175657
1	9407	4617	2365	16389
2	1442	705	349	2496
3	254	101	46	401
≥ 4	78	33	7	118

Table 3.2: The number of selected events as a function of jet multiplicity for 193.5 pb^{-1} of CEM and CMUP and 175 pb^{-1} of CMX data samples.

3.3 Monte Carlo Event Simulation

Monte Carlo (MC) methods use a set of computer generated pseudo-random numbers to sample the probability distribution functions for physical processes and stochastically simulate interactions that take place inside the detector.

Event simulation is critical for the physics analysis at CDF. Because the detector does not cover the full 4π solid angle and because of inevitable inefficiencies, the information obtained from any given event is always incomplete. In addition to $t\bar{t}$ events, background events that partially resemble the $t\bar{t}$ final state are more likely to be produced. In the “lepton+jets” channel the signal to background ratio (S/B) is $\sim 1/5$. Thus it is very hard, if not impossible, to categorize any given event as signal or background. However certain event features statistically show different behavior for signal and background events. MC simulation is the instrument we use to estimate the behavior of signal and background events for a statistical significant sample which can be reliably translated to probability distribution functions. By analyzing the data set we obtain after event selection, the fraction of signal events in the sample can be estimated using a statistical procedure like a maximum likelihood fit.

Event simulation is performed at three successive levels:

- The generation level is where physical processes like $p\bar{p} \rightarrow t\bar{t}$ are simulated. Particles are produced and allowed to decay according to their known properties, usually those compiled by the Particle Data Group [6] or in some cases those given by the tree level SM predictions, though the user is allowed to vary most of the parameters.
- At the second level, MC simulation is used to model the interaction between particles and the detector “producing” hits in the detector¹⁶, energy deposition in calorimeters etc. The detector simulation is performed with the CDFSim package [44] developed by the CDF collaboration. It is based on a parametrization of the detector response to particle kinematics which is derived from data.
- At the third level, events are reconstructed using the standard offline reconstruction software as would be done for real data.

The $t\bar{t}$ events are generated using the PYTHIA [45] and HERWIG [46] MC simulation programs. Most of the background processes are simulated with ALPGEN [47] MC simulator. The events are generated starting with the highest momentum scale interaction described by tree level diagrams with few virtual energetic final state particles. Shower algorithms are used to evolve the initial partons down in the energy scale by emitting other partons according to perturbative calculated probability distributions. Infrared and collinear singularities are avoided by requiring the final state partons to have a transverse momentum exceeding some fixed cut off value and to be separated by more than some minimum ΔR value. When the evolution energy scale

¹⁶The COT chamber simulation is not directly a hit level simulation.

becomes small, the running coupling grows, the phase space fills with mostly soft partons and the perturbation theory breaks down. To deal with this, the perturbative emission is cut off at a fixed infrared cutoff when partons are confined into hadrons according to phenomenological models, in a process referred to as hadronization. Both PYTHIA and HERWIG have their own hadronization models.

The top quark in particular lives a short time ($\sim 4 \times 10^{-25}$ s) and it decays before the cut-off of the parton shower: its decay is just one of the steps in the shower evolution. The evolution for other heavy quarks is modeled using fragmentation models. In PYTHIA the b and c quarks are fragmented according to the Peterson parametrization [48]. HERWIG uses its own fragmentation model. After hadronization, heavy flavor hadrons are decayed using a dedicated MC calculation developed by the CLEO experiment [49].

For $p\bar{p}$ collisions, in addition to the hard scattering portion of the event, the soft interaction of the remnants of the proton and antiproton is modeled using data from minimum bias events.

The ALPGEN MC event generator is designed for simulation of multijet processes accompanying W^\pm and Z boson production, having exact leading order matrix element calculation for up to 6 hard separated partons in the final state. The subsequent evolution of the final state, including the approximate evolution of the parton shower and hadronization, is handled using HERWIG MC simulation.

CHAPTER 4

NEURAL NETWORKS

4.1 Introduction

The original motivation of the research done on neural networks were the early attempts to model the processes that take place inside the brain. The biological neuron: Figure 4.1, is a single cell that receives electric signal through dendrites from ~ 1000 neighboring neurons, performs a integration of the received input on a limited spatial and temporal domain and produces an output electric potential that propagates through the axon and, if above a certain threshold, initiates through synaptic connections similar activity in other neurons. More information on the anatomy and the function of a biological neuron can be found in [50]. McCulloch and Pitts are credited to have initiated the investigations on neural networks by proposing a simple model of a neuron as a binary threshold unit [51]. In this model, a neuron “ i ” computes a weighted sum of the inputs it receives from other neurons and produces the output 1 or 0 depending whether the sum is bellow or above a certain threshold μ_i :

$$n_i(t + 1) = \Theta(\sum_j \omega_{ij} n_j(t) - \mu_i). \quad (4.1)$$

The state of the neurons in this model is synchronously updated, $n_i(t + 1)$ is the output of the neuron i at time $t + 1$, n_j is the input received at time t from the j -th

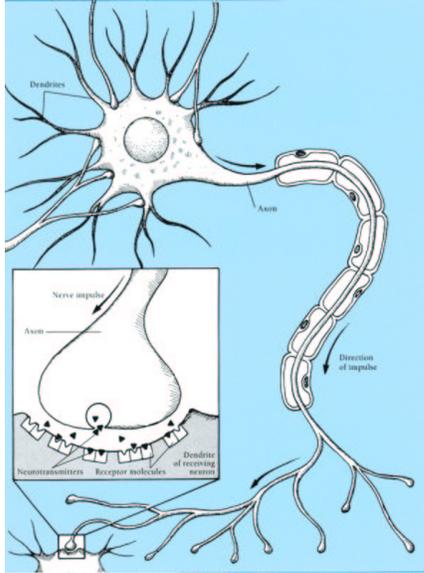


Figure 4.1: A graphical representation of a biological neuron.

neuron, weighted with the connection strength ω_{ij} . The output of a neuron is modeled by the neuron activation function, chosen in this case to be a unit step function: $\Theta(x) = 1$ if $x \geq 0$ and $\Theta(x) = 0$ if $x < 0$. Provided that it is complicated enough, if adequate values for thresholds and connection weights are chosen, a synchronous network of such neurons is capable of performing any computation accessible to a digital computer [52]. The biological realism of this model can be enhanced by implementing continuous nonlinear activation function functions, asynchronous neuron updating and stochastic neuron response to a given input.

From a slightly different perspective, it is often argued that human pattern recognition is mainly learned and it is believed that by using the large degree of parallelism and redundancy among the neurons in the brain, inter-neuron connection strengths are used to store memory.

While such ideas are still the subject of active investigation [53], today there is also considerable interest in the study of artificial neural networks (ANN). These systems were inspired by the original McCulloch and Pitts paradigm but their objective shifted toward creating structures capable of performing calculations useful for practical applications. By varying the network parameters (weights, thresholds) these systems are able to modify their behavior. Algorithms have been identified which allow modifying parameters in order to produce within some approximation almost any desired result. In other words, with proper “training” these networks can “learn” to execute various tasks. ANN inspired methods are especially useful in optimization [52] and pattern recognition [54] problems.

4.2 Artificial Neural Networks

Often in high energy physics one has to separate the signal for a process of physics interest out of a sample dominated by background events. A class of ANN useful for solving such problems are the feed-forward¹⁷ networks. Figure 4.2 shows the schematic for a simple feed-forward network: the units (neurons), are organized in layers, the output of each unit is transferred forward to the units in the next layer. The i -th unit produces an output of this form:

$$y_i = g \left(\sum_{j=1}^k \omega_{ij} y_j + \omega_i \right), \quad (4.2)$$

where y_j are inputs from units in previous layers, ω_{ij} the connection weights and ω_i is a unit bias weight. The neuron activation function $g(x)$, is chosen to be a continuous

¹⁷ANN are also used in track reconstruction algorithms, where the recurrent network paradigm [52] may be preferred.

nonlinear function, often the sigmoid function:

$$g(x) = \frac{1}{1 + e^{-x}}. \quad (4.3)$$

The neurons in the first layer (input units), are initialized with variables that discriminate between signal and background events while the network output values can be used to better distinguish among various classes of input events.

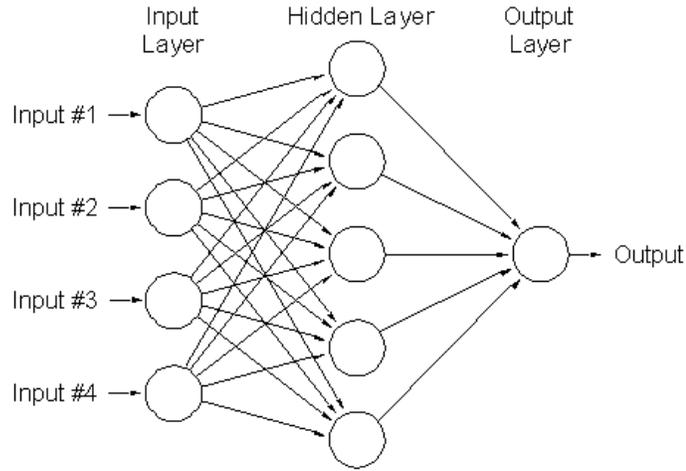


Figure 4.2: Schematic view of a three layer feed forward network.

Before training, the network parameters are assigned random values. These are adjusted by training, which is typically an iterative process. The network is presented with random signal and background events and the network parameters are adjusted in order to minimize an error function like:

$$E = \frac{1}{N_t} \sum_{i=1}^{N_t} (NN_{out}(i) - target(i))^2. \quad (4.4)$$

In case of a single output unit, target values can be chosen 1 for signal events and 0 for background events. The summation \sum_i is performed over all events N_t in the

training sample. If the network is properly trained, the network output provides better separation between signal and background events than any of the input variables alone.

In Ref. [53] it is argued that in terms of computing capabilities, the computational power of such networks are rather modest. Their utility and relevance for solving practical problems comes from the large computational effort necessary to find an appropriate set of parameters that accurately describes the training data. The resulting network can be looked at as a compact representation of the training data set. When presented with a new event, by making a few simple calculations the network will be able to predict if the event is more likely to be a signal or a background event.

As a bonus, in contrast to many parametric methods, no model assumptions need to be made regarding the relationship and correlations among various input variables. From a slightly different perspective, it is sometimes argued [52] that neural networks can be successfully applied to solving problems where the functional relation between various examples in a data sample is unknown but believed to exist.

4.2.1 A Theoretical Perspective on Classification Problems

When discussing classification problems, the Bayes rule is often invoked to judge the quality of a classification procedure [54], [55]. A cost (risk) function can be defined by assigning 0 for an event correctly classified and 1 for a misclassified event. For any problem there is a lower theoretical limit on the value of the cost function that can be achieved for a random event sample. Out of a collection of possible outcomes $\{C_k\}$, a so-called minimum risk classifier, ascribes an unknown event that is characterized

by a vector \mathbf{x} in a feature space, to the class C_k having the highest Bayes posteriori probability:

$$P(C_k|\mathbf{x}) = \frac{P(\mathbf{x}|C_k)\pi(C_k)}{P(\mathbf{x})} \quad (4.5)$$

where :

- $P(\mathbf{x}|C_k)$ - is the class conditional probability distribution function, or the so called likelihood (the probability for the event \mathbf{x} if that event was chosen from the class C_k).
- $P(\mathbf{x})$ - is the class unconditional probability distribution function (the probability for the event \mathbf{x} regardless of the class of the event).
- $\pi(C_k)$ is the *a priori* belief in the classification of the data event \mathbf{x} . For a set containing an equal number of signal and background events, $\pi(C_k)$ is constant and may be taken to be 1.

For the purpose of minimizing such a cost function, the Bayes rule (Equation 4.5) gives an optimum solution to any classification problem. Applying it in practice is not always easy because determining the class conditional likelihood function becomes computationally challenging as the dimension of the feature space where \mathbf{x} is defined, increases. It is often said that ANN's sidestep such difficulty by automatically accounting for variable correlations during the training process. If the error function, learning algorithms and network architecture are adequately chosen it can be shown [58] that ANN output provides an estimation for the posteriori probability density $P(C_k|\mathbf{x})$.

4.2.2 Finding the Optimum Solution

A legitimate question to ask after the training was completed is how close are we to the theoretical Bayes bound. This problem is connected to that of finding the global minimum of an error like Equation 4.4, defined as a function of the network parameters ω over a given training set. In order that the error function adequately reflects the features of the problem at hand, the size of the training sample should be large enough. In this subsection we will try discuss the relevance of training algorithm, requirements for the size of the training sample and the number network of parameters.

One could easily image that the landscape of the training function gets complicated as the number of network parameters increases, presenting lots of local minima. Most training algorithms do a reasonable job in finding a minimum but there is no guarantee that the true global minimum was found or not. Of course, finding a local minimum instead of the global minimum, means that the separation provided by the network is not optimum, but does not automatically invalidate a measurement obtained when applying the network to the new data.

The most popular minimization algorithms are variations of the gradient descent method where the network parameters ω are modified in small steps like:

$$\Delta\omega_{t+1} = -\eta \frac{\partial E}{\partial \omega} + \alpha \Delta\omega_t \quad (4.6)$$

where η is the learning rate and is usually chosen to be smaller than 2. The momentum parameter ($0 \leq \alpha \leq 1$) is added in order to improve learning stability at the end of training. This is also called the back-propagation method because the partial derivatives $\frac{\partial E}{\partial \omega}$ are calculated starting from the output unit and going to the input

units of the network, the opposite direction of the normal flow of the calculation in the network. These are iterative learning algorithms and follow a path in the weight space until a reasonable minimum for the error function is found. It is quite likely that such minimum will not be the global minimum. Using different starting points (the random values given to the initial weights) is a method to search for a better solution. So called global search methods have been developed such as genetic algorithms [56] or simulated annealing [57] which are not so easily “fooled” by local minima.

There is actually a consensus in the literature that the back propagation methods fail to find an absolute minimum in most of the cases. The convergence of this method is relatively slow even on a fast PC. There is another side of the coin here: it is reasonable to expect that for any problem, there is a asymptotic form of the error function landscape that can be represented using an infinite sized data set. On the other hand, for a finite sized training sample, the landscape of the error function might be slightly displaced. A minimum in the landscape of the training sample may or may not be the same with a minimum on the “ideal” landscape which, after all, is the correct representation of the problem. Getting too close to a minimum on the training sample is not necessarily the best thing to do and to some degree incomplete convergence will certainly improve generalization.

Using the standard back propagation method in the JETNET [59] package on a 8000 events training sample ($t\bar{t}$ and W +jets), we found that while the training error keeps on decreasing even after a few thousand training epochs¹⁸, the errors measured on the training sample are consistently smaller than the errors measured on an statistically independent testing sample after 200-300 epochs.

¹⁸An epoch corresponds to making weight adjustments by reading once all events in a given training set.

Another useful way to look at ANN can be as a mapping function between the input variable space to the output variable space. The above discussion is sometimes formulated in terms of the overfitting problem: using a complicated network having many free parameters might result in an mapping function that is a very good representation of the training set but exhibits poor generalization (i.e. behaves poorly when applied to new data). Overfitting can happen when the number of weights is too large and training was performed for too many iterations. A procedure that inhibits overfitting to some degree is to use regularization methods [52]. One such method supported by JETNET adds a penalty term to the error function (4.4):

$$E = E_0 + \lambda \sum_{i,j} \frac{\omega_{ij}^2}{1 + \omega_{ij}^2} \quad (4.7)$$

where λ is a Lagrange multiplier. Performing gradient descent on this new error function:

$$\Delta\omega_{ij} = -\eta\partial E/\partial\omega_{ij} \quad (4.8)$$

is equivalent to subtracting an extra term $\frac{2\lambda\eta}{(1+\omega_{ij}^2)^2}\omega_{ij}$ out of the weight ω_{ij} at each minimization step. Using this procedure, small weights are driven to zero, removing unnecessary connections. An alternative procedure uses early stopping algorithms [54] where a very large number of weights may be used.

Another claim often made in the literature is that back propagation is not an optimal training algorithm since it is not sensitive to the metric of the error function landscape and thus does not find the best path toward minimum [55]. Certainly normalizing the input variables to some common scale often helps in this direction [59]. More advanced algorithms attempt to use the Hessian of the error function in order to figure out the best path toward the minimum. Since calculating the full Hessian is

computationally demanding, various approximation methods can be used [52]. This is now a well known result but to this day back propagation is the most often method employed in training neural networks. The JETNET authors found it often to be at least as good as more sophisticated training methods [59]. The claim made in [60] is that regardless of the number of parameters, feed-forward ANN, using back-propagation, are biased toward smoother solutions which helps generalization. This is an optimistic evaluation which agrees with most of the experience we found in NN literature but there is no rigorous proof for this statement.

A circumspect person might as well interpret a ANN as a very large look-up table, providing some connection rule between the input variable space and the output space according to the content of the training sample. Assuming that we trained the network to reproduce correctly the training sample, what guarantee do we have that when presented with a new event it will find the input-output relation we actually had in mind? There may be, after all, more then one input-output rule consistent with a given finite sized training set. The connection between the number of network weights, the number of events in the training sample and the network generalization ability has been extensively discussed in the literature. A treatment of the problem from a statistical point of view can be found in [61, 62], also a short summary can be found in [52] page 153-157. For problems having discrete outputs, a theoretical expression is found for the minimum number of events in the training sample needed to put an upper bound on the generalization error, although calculating this number for real problems is difficult, and that upper bound may much overestimate the size of the samples that are found to give good results in practice [60]. A similar treatment

for a network having continuous outputs, using the methods borrowed from statistical mechanics can be found in [63]. The theoretical results mentioned above are difficult to apply for real problems. However a certain pattern emerges: for any given problem, there seem always to be a critical size in the statistics of the training sample beyond which improvements in network training and generalization performance slowly saturate.

4.2.3 The Choice for the Network Architecture

The layers placed between the input and the output unit layer are called the hidden layers and the choice for the number of the hidden layers as well as the choice for the number of units in each hidden layers is somewhat arbitrary. It has been shown that a network with no hidden layer (perceptron) can be used to separate signal vs. background only when the problem is linearly separable, e.g. a plane can be found in the feature space that separate two regions having two possible answers. Yet there are quite simple problems which are not linearly separable and thus can not be solved with a simple perceptron: the boolean AND function is linearly separable, while the exclusive OR (XOR) function is not (for details see [52], pag 94).

The universal approximation theorem for neural networks states that a continuous function mapping the input variable space to an interval on the real axis can be approximated arbitrarily closely by a multi-layer perceptron with just one hidden layer, provided that a large enough number of hidden nodes is used. This result holds for a nonlinear activation functions (e.g. sigmoidal). Nonlinearity of the activation function is important because with linear activation functions a multilayer feed forward network can be shown to be equivalent to a single layer network.

We will consider here only networks having one hidden layer. Provided that there is a certain degree of separation in the input space between the regions mostly populated by signal and the region mostly populated by background, in other words provided that the S/B separation can be achieved, a feed forward network with one hidden layer can be trained to separate the signal from background.

A network with two hidden layers can be used to map arbitrarily closely any function having finite discontinuities. In certain instances two hidden layers might be useful: using two hidden layers can produce in some instances a reduction of the total number of hidden units needed to fit a function [59]. On the other hand using two hidden layers increases the number of local minima. As already mentioned, finding a local minimum instead of the global minimum is not necessarily a problem.

In general, if a problem is linearly separable the application of ANN methods can not be expected to result in an improvement in performance. A simple way to test this is to check the performance of a network having one hidden layer. If the performance of a network with $N_{hidd} = 1$ is close to the network having $N_{hidd} > 1$, the problem is probably linearly separable and traditional linear regression methods might be a better choice. But since many classification problems are not linearly separable, it is expected for such problems that ANN methods outperform the more traditional classification methods like linear discriminants etc.

4.3 Some Aspects of Artificial Neural Networks

The problem of overfitting, finding the minimum training size, using a training algorithm versus another, have all been extensively discussed in literature and could make ANN's seem unreliable tools or as providing approximate solutions at best.

We try here to argue that this is not the case and that ANN can be used for making precise and accurate measurements. The problem of overfitting is acute when example data is scarce. In many applications, especially high energy physics problems, simulating new data is relatively easy and such difficulties can be avoided in a straightforward manner. All one has to do, after training and testing is completed, is to use a statistically independent sample for producing the templates for fitting the data. This way, the particular choices made during the training process: training algorithm, the point where training is stopped, the number of network weights, the number of hidden layers, whether regularization methods¹⁹ were used or not, become irrelevant. An extreme but still true statement is that one might as well think of pulling the network parameters out of a hat provided that good S/B separation is obtained when the signal and background MC are processed with the trained network.

One often asks how large these MC-simulated samples need to be, making the assumption that the more variables are being used in the ANN input, the larger the training samples need to be in order to have a fine enough sampling of the multi-dimensional input variable space. Actually it is not the dimension of the input space important here, but it is the dimension of the output space. For a single dimensional output, several thousand events should be enough to obtain reliable templates, regardless of the number of input variables²⁰ used by the network [64],[65]. From this point of view, using a “complicated network” (many input variables, more hidden nodes) versus a “simpler one” should not make any difference with respect to our confidence in the final result.

¹⁹Regularization methods and/or pruning are in general recommended.

²⁰As long as the number of input variables is finite.

On the other hand, if the distribution of the input variables in the feature space is different in data versus MC simulation, that is, when the correlations between variables in the training set are different from the correlations in data, erroneous results will be obtained. This time the problem of “simple” vs. “complicated” can be put in these terms: which network is more resistant to slight variations in the variable correlations? A naive answer would be that a simpler network is more likely to be less affected by such problems. We can answer this question by looking at the output function of a 1-hidden layer feed forward ANN using the neuron activation function in Equation 4.3 and having k input variables:

$$NN_{out}(y_1, \dots, y_k) = g \left(\omega^{(2)} + \sum_{i=1}^k \omega_{2i}^{(2)} g \left(\sum_{j=1}^k \omega_{ij}^{(1)} y_j + w_i^{(1)} \right) \right)$$

where $\omega_i^{(1)}, \omega^{(2)}$ are the threshold parameters for neurons in the hidden layer and the output layer respectively. The $\omega_{ij}^{(1)}, \omega_{2i}^{(2)}$ are the connection weights from the input to the hidden layer and from the hidden to the output layer respectively. In order that the above mapping is less sensible to slight variable shifts, it must provide a smooth mapping:

$$\frac{\partial NN_{out}}{\partial y_i} \simeq 0, \quad \frac{\partial^2 NN_{out}}{\partial y_i \partial y_j} \simeq 0.$$

The first derivative can be written as:

$$(NN_{out} - NN_{out}^2) \left(\sum_{i=1}^k \omega_{2i}^{(2)} \omega_{ij}^{(1)} \left(g_i \left(\sum_{j=1}^k \omega_{ij}^{(1)} y_j + \omega_i^{(1)} \right) - g_i^2 \left(\sum_{j=1}^k \omega_{ij}^{(1)} y_j + \omega_i^{(1)} \right) \right) \right)$$

From the expression above one can see that the derivative is small in those regions of the variable space where the $NN_{out} \rightarrow 1$ or $NN_{out} \rightarrow 0$. One could also try to monitor the behavior of the sum shown above in parenthesis for all events in the training sample. The benefit of using regularization methods is obvious here since such

methods tend to drive weights to small values. Since the sign of the weights is more or less random, summing over a large number of hidden nodes, that is choosing a more “complicated” network, may provide sometimes a smoother input-output mapping.

The representation of the input variables is important and preprocessing the input data can help. One should not expect that after throwing all available information into a network, a typical training procedure will find the best separation that there is possible. Preprocessing the data sample may be important in order to find better separation results. Training a network to distinguishing whether a certain number is divisible by 3 or not, might be difficult if the input numbers are decimals. However, if the input number are given in the base 3, a network can learn after a few examples to recognize the last digit.

In those cases where training data is scarce, one often cited “rule of thumb” on how many independent parameters a network is allowed to have in order to avoid overtraining, is that the number of training events be at least 10 times the number of independent parameters of the network. This is a “rule of thumb”, it often holds when early stopping training algorithms are being used, but there is no real justification for it.

CHAPTER 5

CROSS SECTION MEASUREMENT

5.1 Selecting $t\bar{t}$ Events in the “Lepton+Jets” Sample

This chapter describes a measurement of the cross section for top pair production in $p\bar{p}$ collisions at 1.96 TeV performed in the “lepton+jets” channel. The data sample has a integrated luminosity of 193.5 pb^{-1} and was collected between March 2002 and September 2003 at the Fermilab Tevatron by the CDF experiment. A description of the CDF detector was presented in Section 2.2. The event reconstruction technique and a detailed description of the event selection requirements were presented in Section 3.1 and Section 3.2 respectively.

In the “lepton+jets” channel we select $t\bar{t} \rightarrow W^+b W^- \bar{b}$ events where one W boson decays hadronically producing two jets while the other W decays leptonically. The branching ratio for $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ channels is $8/27$. Additional contributions to this channel are expected from the $W \rightarrow \tau\nu$ decay mode when τ decays leptonically and from a small fraction of the dilepton²¹ events. One high p_T isolated lepton, considerable missing transverse energy and four jets (at least two heavy flavor), are expected in the final state. Sometimes a jet may be reconstructed under the 15 GeV threshold or may lie outside the calorimeter geometric coverage. So we select events

²¹In dilepton events both W 's from the $t\bar{t}$ pair, decay leptonically.

having at least three central, energetic jets. Extra jets may sometimes be produced due to initial or final state QCD radiation effects (ISR, FSR). A total of 519 events pass our selection requirements where for a theory motivated $\sigma_{t\bar{t}} = 6.7$ pb we expect a signal to background ratio $S/B \sim 1/5$.

The $t\bar{t}$ signal in this channel has a large background from inclusive W boson production associated with jets and smaller backgrounds from QCD multijet jet production, Z + jets, WW, WZ, ZZ + jets and single top. Table 5.1 list the expected contributions from these electroweak backgrounds relative to the dominant $W+3p$ background as calculated with the ALPGEN event generator in the $W+\geq 3$ jets mode.

Process	Generator σ (pb)	$W \rightarrow e\nu$ (%)	$W \rightarrow \mu\nu$ (%)
$W \rightarrow (e, \mu)\nu+3$ parton	179.8	87.3	84.8
$W \rightarrow \tau\nu+3$ parton	89.9	4.6	4.6
$Z \rightarrow \ell^+\ell^- + 2$ parton	46.6	1.5	4.2
$Z \rightarrow \tau^+\tau^- + 2$ parton	23.3	1.3	1.3
$WW+1$ parton	4.38	3.8	3.7
$WZ+1$ parton	2.37	0.4	0.4
single top	3.0	1.0	1.0

Table 5.1: Contributions from other electroweak processes relative to the dominant $W+3p$ process in the $W+\geq 3$ jets mode. The cross section numbers are leading order estimates and thus are expected to have large uncertainties.

5.2 Simulation of the W +jets Background

The W/Z +jets processes are important for many collider physics studies and the modeling of such processes via Monte Carlo (MC) simulation has received considerable attention. A short pedagogical discussion of the MC simulation methods for

hadron collider physics can be found in [73]. A brief description of the MC simulation programs used in this analysis can be found in Section 3.3.

From a theoretical point of view, W/Z +jets production is an important test for the QCD calculations and for the knowledge of PDF's. The use of perturbative QCD may be justified here since the production of heavy bosons require considerable parton-parton scattering energy: $\alpha_S(Q = M_Z) \sim 0.118$.

Tree level matrix element based simulation programs are currently used to model both the $t\bar{t}$ signal and the background processes. However, since we concerned here with QCD processes, where α_S is relatively large, LO cross sections are not precise enough to predict the the absolute normalization for each of these process in our data sample. We will trust the numbers shown in Table 5.1 only as a reliable estimation for the relative contribution from each individual background.

For simulating heavy boson production, the state of the art today is given by the ALPGEN [47] and the Madevent [66] generators which calculate tree level matrix elements for production of W/Z plus as many as 6(5) additional jets in case of ALPGEN (Madevent). These programs generate event information at the parton level: bare quarks and gluons. Shower MC programs like those from PYTHIA or HERWIG are used to evolve partons down the energy scale and fragment the outgoing quarks and gluons into jets. The shower algorithms are found to model well the energy and multiplicity of the particles in a jet and can be thought of as simulating the effects of the missing higher order terms in the matrix element (ME) calculation. The fragmentation model is based on leading log terms in perturbation theory which account for most of the infrared and collinear singular terms in the perturbation expansion but

do a not so good job in accounting for nonsingular terms: e.g. emission of a hard gluons at large angles. The parton shower algorithms are combined with ME calculated at the tree level which provide the exact LO results in the hard scattering region of the event phase space. The tree level matrix element should include most if not all the hard partons expected in the event: $W + 2, 3, 4$ hard parton processes should be relevant for data in the $W+3$ jets mode. This procedure is expected to yield good predictions, however combining different parton level processes is not exactly trivial if one wants to avoid double counting events in certain regions of the phase space because the parton shower can generate the same diagrams with those already considered in tree level ME. Also, in order to avoid divergences at tree level, a minimum separation between jets must be required, a step which may produce nonphysical dependences in the simulation results. Very recently there has been considerable theoretical progress on combining shower MC with tree level ME calculations [67, 68].

For our analysis here however, we will use $W+n$ parton process to simulate data in the $W+\geq n$ jets mode, relying on parton shower MC to model the emission of extra jets in the event. Studies performed within the CDF collaboration have shown this is an acceptable choice for up to one extra jet in the event relative to the number of hard, well separated partons. The ALPGEN generator was also used for simulation of diboson+jets processes (Table 5.1). For simulating the electroweak production of single top we use PYTHIA.

During Run1 most of the comparisons between W/Z +jets data and theoretical prediction were made for LO calculations. Recently a calculation for NLO correction effects in the $W/Z+2$ jets events has been performed [69]. Complete NLO calculations

for larger jet multiplicities are not available yet, see [70] for some latest advances on this problem.

Combining next to leading order (NLO) exact ME calculations with parton shower algorithms in MC event simulation is likely to improve the quality of the predictions since NLO effects will model the radiation of energetic partons at large angles better than the parton shower algorithms and produce more reliable and more stable²² predictions for relative normalization of $W+n$ and $W+(n+1)$ jet rates. The progress in this field has been limited so far, see [71] for a basic discussion of the problem. Recently a NLO simulation package which can be interfaced with HERWIG has become available for a number of QCD processes [72].

5.3 The Model for the Multijet Background

For the QCD multijet fake events we do not have an adequate simulated sample. A plausible representation for the kinematic properties of this data sample can be obtained by applying our standard event selection criteria to data but instead of requiring one isolated lepton we require one lepton having isolation $I > 0.2$. In order to estimate the systematic uncertainty associated with this modeling, an alternative sample is built from identified conversion electrons where cuts like lepton identification, \cancel{E}_T , number of jets, etc. are kept unmodified.

The fraction of multijet fakes in our data sample can be obtained by assuming that the lepton isolation and the missing transverse energy \cancel{E}_T are uncorrelated. Using the usual selection requirements but dropping off the cuts on lepton isolation and \cancel{E}_T ,

²²Stable with respect to the choice for factorization and renormalization scale.

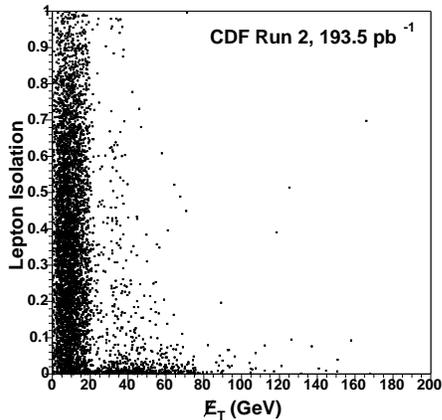


Figure 5.1: The \cancel{E}_T versus isolation distribution for events with a lepton and 3 or more jets. The structure apparent near 20 GeV comes from removing events where missing transverse energy lies close to direction of the first jet.

a 2-d plot of lepton isolation versus \cancel{E}_T is constructed as shown in Figure 5.1. The plot can be divided in four distinct regions:

- *A*: lepton isolation >0.2 and $\cancel{E}_T < 10$ GeV
- *B*: lepton isolation <0.1 and $\cancel{E}_T < 10$ GeV
- *C*: lepton isolation >0.2 and $\cancel{E}_T > 20$ GeV
- *D*: lepton isolation <0.1 and $\cancel{E}_T > 20$ GeV

Assuming there is no correlation between Iso and \cancel{E}_T , the number of multijet fake events in the signal region *D* can be deduced from the event population found in the control regions which are dominated by generic QCD events:

$$n_D^{multijet} = n_C \times (n_B/n_A) \quad (5.1)$$

In performing this calculation we correct the number of events in the control regions for contributions from $t\bar{t}$ and W +jets processes using the relative fractions determined from simulation and from estimating event numbers in the D region using a $\sigma_{t\bar{t}} = 6.7$ pb. An effective method to reduce the amount of multijet fakes in our data sample has proved to be vetoing events where the azimuthal direction of \cancel{E}_T lies close to the direction of the leading jet in the event. The calculated amount of background in the W +jets data sample is presented in Table 5.2 where the uncertainties shown are statistical only.

Jet multiplicity	$W \rightarrow e\nu$	$W \rightarrow \mu\nu$	Total
1 jet	$3.8 \pm 0.2\%$	$2.9 \pm 0.2\%$	$3.4 \pm 0.3\%$
2 jets	$6.1 \pm 0.5\%$	$2.0 \pm 0.2\%$	$4.3 \pm 0.5\%$
≥ 3 jets	$7.7 \pm 1.3\%$	$3.1 \pm 0.9\%$	$6.3 \pm 1.6\%$

Table 5.2: The fraction of multi-jet background in the W +jets data sample as a function of jet multiplicity. The uncertainties shown are statistical only.

One distinct component of the multijet fake background in the electron sample is made from unidentified conversion electrons. The efficiency of the conversion removal algorithm was measured from data to be $72.6 \pm 0.07\%$ [74]. Using the standard event selection requirements, we find 111 events with identified conversion electrons which implies a residual background from unidentified conversions of $\sim 9\%$ in the electron channel.

5.4 The Signal Acceptance and Efficiency

The acceptance for signal in the lepton+jets channel can be measured using Monte Carlo simulated $t\bar{t}$ events. For $t\bar{t}$ simulation there are two generators which are currently being used: PITHIA and HERWIG. The default HERWIG generator is known to lack QED radiation for the leptons in the final state. This is why PYTHIA has been chosen to be the default generator for this analysis.

The acceptance is calculated using $t\bar{t}$ generated events to whom we apply the standard event selection as we would do for real data except for applying conversion electron and cosmic ray removal algorithms. At the generator level a lower 8 GeV cut is made on the parton p_T . We use the standard LO set of PDFs available in PYTHIA, with a factorization and renormalization scale chosen at $\mu = M_{top}$. The acceptance is calculated separately for the CEM, CMUP and CMX triggers because the luminosity of the data is different in CEM and CMUP samples with respect to CMX data sample.

Quantity	CEM	CMU/CMP	CMX
events	17800	10908	4025
$\epsilon_{t\bar{t}}^{\text{MC}}$	0.0462 ± 0.0004	0.0283 ± 0.0003	0.0104 ± 0.002

Table 5.3: Signal acceptance in the $W + \geq 3$ jets sample assuming $m_{top} = 175 \text{ GeV}/c^2$. The number of generated events after the parton p_T cuts at the generator level was 384,875.

For a number of 384,875 generated $t\bar{t}$ events, after event selection we found the contributions from various decay modes of the $t\bar{t}$ system to the acceptance in the lepton+jets channel as listed in Table 5.3.

In order to obtain the signal efficiency in data further corrections must be applied [76]. These numbers must be corrected with the luminous region efficiency, the data versus MC tracking efficiency scale factor, the lepton trigger efficiency and the efficiency for lepton identification in data relative to the simulation as determined from $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ data and MC samples. The contributions from various decay modes of the $t\bar{t}$ system to the signal efficiency in the lepton+jets channel are listed in Table 5.4.

Quantity	CEM	CMU/CMP	CMX
$\varepsilon_{t\bar{t}}^{\text{MC}}$	0.0462 ± 0.0004	0.0283 ± 0.0003	0.0104 ± 0.002
Trigger Eff.	0.962 ± 0.006	0.887 ± 0.007	0.954 ± 0.004
Lepton Id.	0.965 ± 0.014	0.887 ± 0.014	1.001 ± 0.017
Data - MC	1.009 ± 0.002	1.009 ± 0.002	1.009 ± 0.002
Track. Scale			
Lum. Reg.	0.948 ± 0.003	0.948 ± 0.003	0.948 ± 0.003
$\varepsilon_{t\bar{t}}$	0.0412 ± 0.0033	0.0213 ± 0.0017	0.0095 ± 0.0008

Table 5.4: Signal efficiency in the $W + \geq 3$ jets sample assuming $m_{top} = 175 \text{ GeV}/c^2$. The overall signal efficiency weighted with respect to the CEM/CMUP/CMX luminosity is 0.0711 ± 0.0056 .

5.5 Principle of Measurement

The traditional method for measuring $t\bar{t}$ cross section in the “lepton+jets” channel at CDF relies on the identification of heavy flavor jets which brings the S/B ratio

close to 3. Because of its large mass, top pair production results in central, spherical, energetic events. As an alternative to the b-tagging method we will separate top events using kinematic and topological variables. Since the b-tagging efficiency is $\sim 55\%$ and since we do not make silicon quality requirements for the data runs, this method takes advantage of a larger data set and provides a precise cross check of the b-tag method result.

We use a neural network (ANN) to combine multiple kinematic and topological variables in order to better discriminate $t\bar{t}$ from background events. We are considering feed forward neural networks with some number N_i of input variables, one hidden layer having N_h nodes and one output. The shape of the ANN output distribution, obtained after processing our standard MC samples with a trained network, will be used as discriminating variable. A binned maximum likelihood fit to data is performed and a $t\bar{t}$ fraction is extracted. The cross section is calculated using the following expression:

$$\sigma_{t\bar{t}} = \frac{n_{t\bar{t}}}{\epsilon_{t\bar{t}} \mathcal{L}} \quad (5.2)$$

where $\epsilon_{t\bar{t}}$ is the signal efficiency in the “lepton+jets” channel and \mathcal{L} is the measured luminosity. The luminosity [75] and the $t\bar{t}$ signal efficiency [76] number was extracted by the Top Group at CDF.

5.5.1 The Binned Likelihood Fit

Using any particular discriminating variable we can perform a binned likelihood fit to the data sample and find a most likely number of events from $t\bar{t}$ production, $n_{t\bar{t}}$ from minimizing a binned likelihood function like this:

$$L(\mu_{t\bar{t}}, \mu_w, \mu_q) = \prod_{i=1}^{N_{bins}} \frac{e^{-\mu_i} \mu_i^{d_i}}{d_i!}, \quad (5.3)$$

where $\mu_{t\bar{t}}$, μ_w , μ_q are the parameters of the fit, representing Poisson means for the number of $t\bar{t}$, W-like, and multi-jet events in our data sample. The expected number of events in the i -th bin is $\mu_i = (n_{t\bar{t}}P_{t\bar{t},i} + n_wP_{w,i} + n_qP_{q,i})$, where $P_{t\bar{t},i}$, $P_{w,i}$, $P_{q,i}$ are the probability of observing an event in bin i from $t\bar{t}$, W+jets²³ and multi-jet processes, d_i is the number of observed data events in the i -th bin. Since the fraction of QCD events was evaluated with an independent method, we constrain n_q in the fit to the measured fraction of 6.3%. For minimizing of the likelihood the MINUIT [80] package is used.

5.5.2 Running Pseudo-Experiments

Pseudo-experiments have been performed to test the sensitivity of the method and estimate the systematic uncertainties for a measurement using some particular discriminating variable: V . For each pseudo-experiment, the signal and background V distributions are randomly sampled and a “data” histogram is filled with events. The numbers for signal and background events are drawn from Poisson distributions with average contributions corresponding a theory motivated 6.7 pb $t\bar{t}$ cross section. Subsequently a fit is performed as would be done for real data and a most likely fraction of signal events is determined. By performing a few thousand pseudo-experiments one can verify if the fit is indeed unbiased and yields Gaussian errors. This way we also estimate the expected measurement uncertainty and decide *a priori* which is the best variable to perform fits with and extract a cross section.

²³The W+jets background is defined as the sum from W+3(4)p and the other electroweak processes in the proportions listed in Table 5.1.

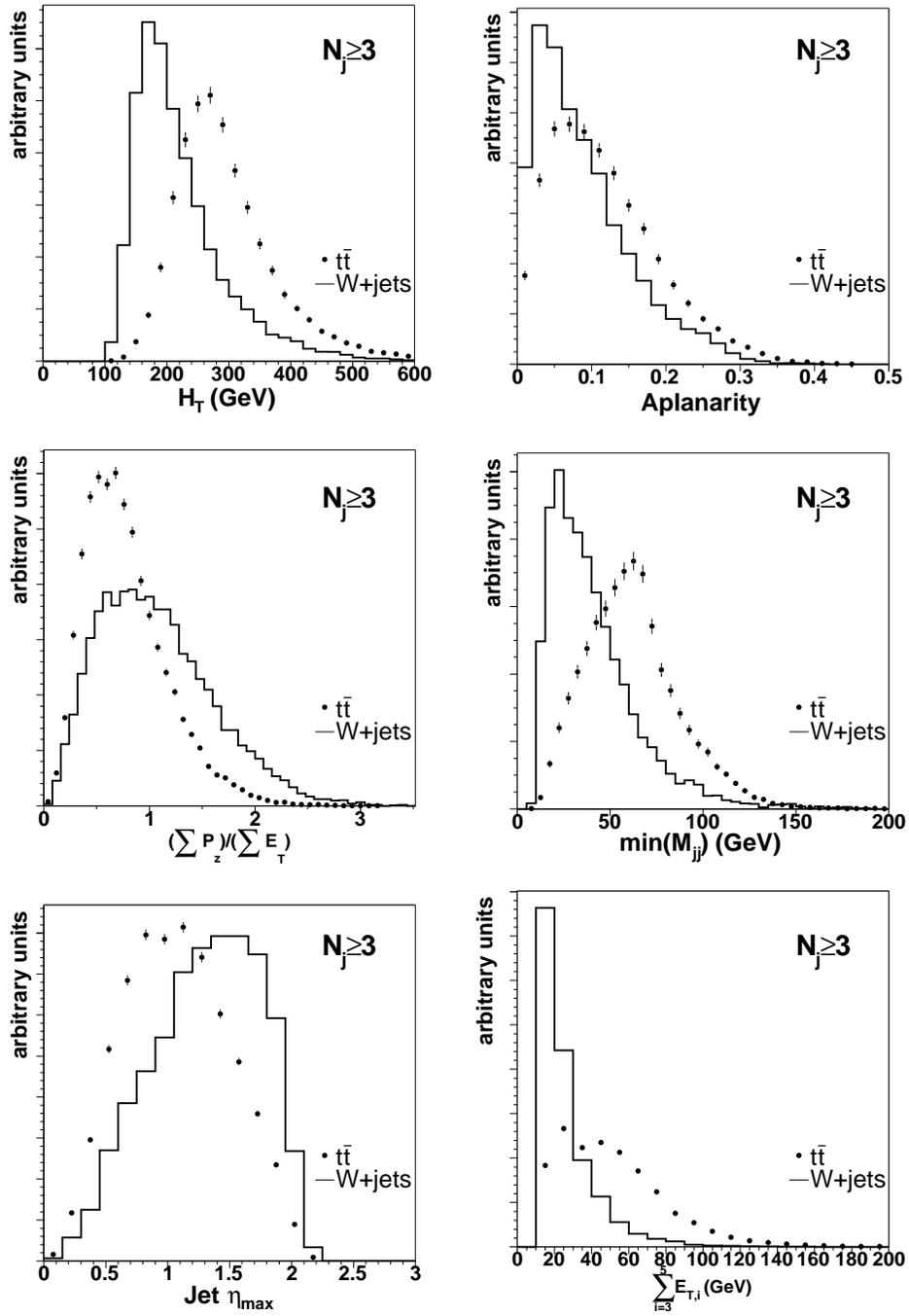


Figure 5.2: Comparison plots of kinematic variables using ALPGEN $W+3p$ and PYTHIA $t\bar{t}$ for a variety of kinematic variables in the reconstructed $W+\geq 3$ jets mode. The two histograms are normalized to equal area.

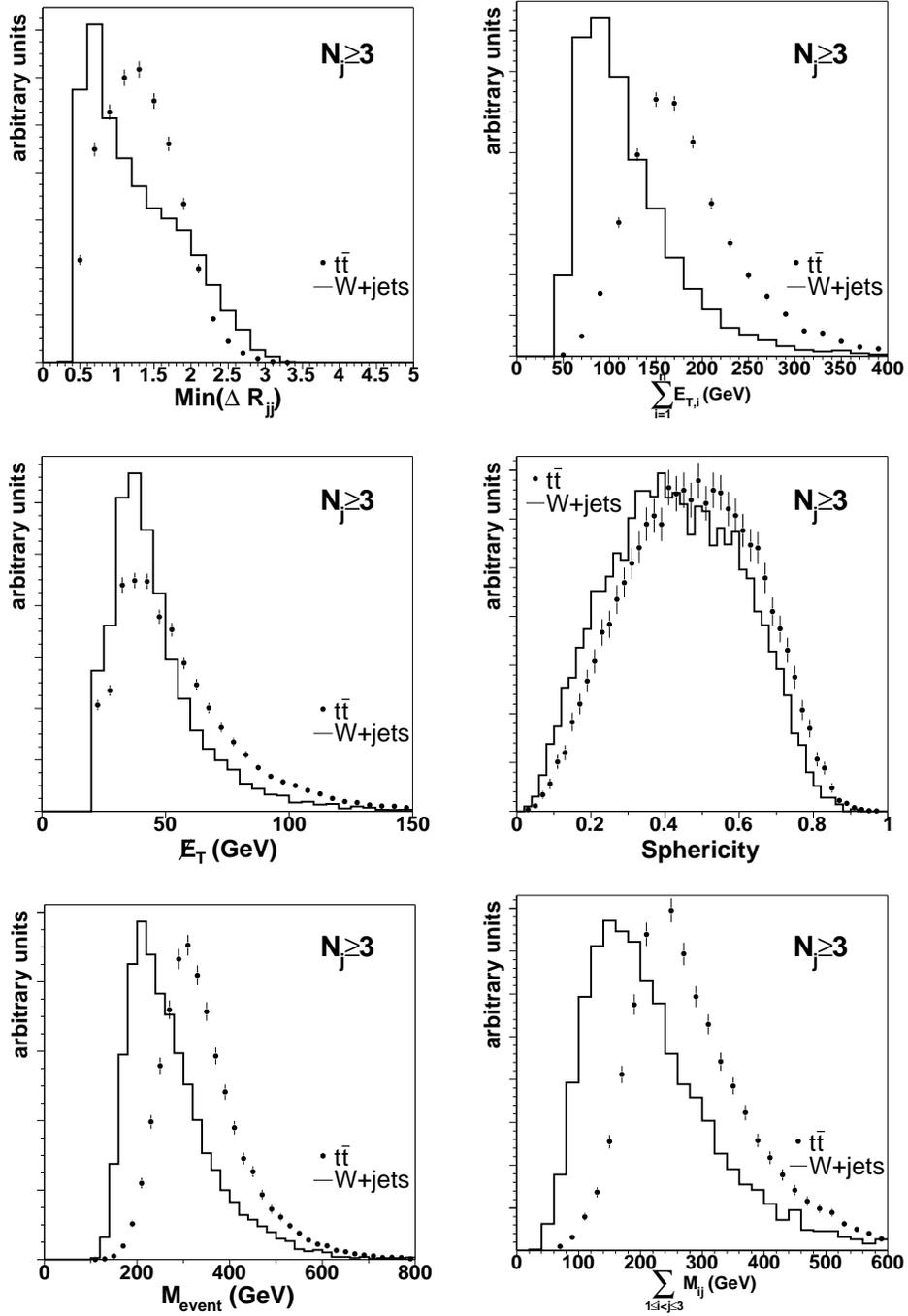


Figure 5.3: Comparison plots of kinematic variables using ALPGEN $W+3p$ and PYTHIA $t\bar{t}$ for a variety of kinematic variables in the reconstructed $W+\geq 3$ jets mode. The two histograms are normalized to equal area.

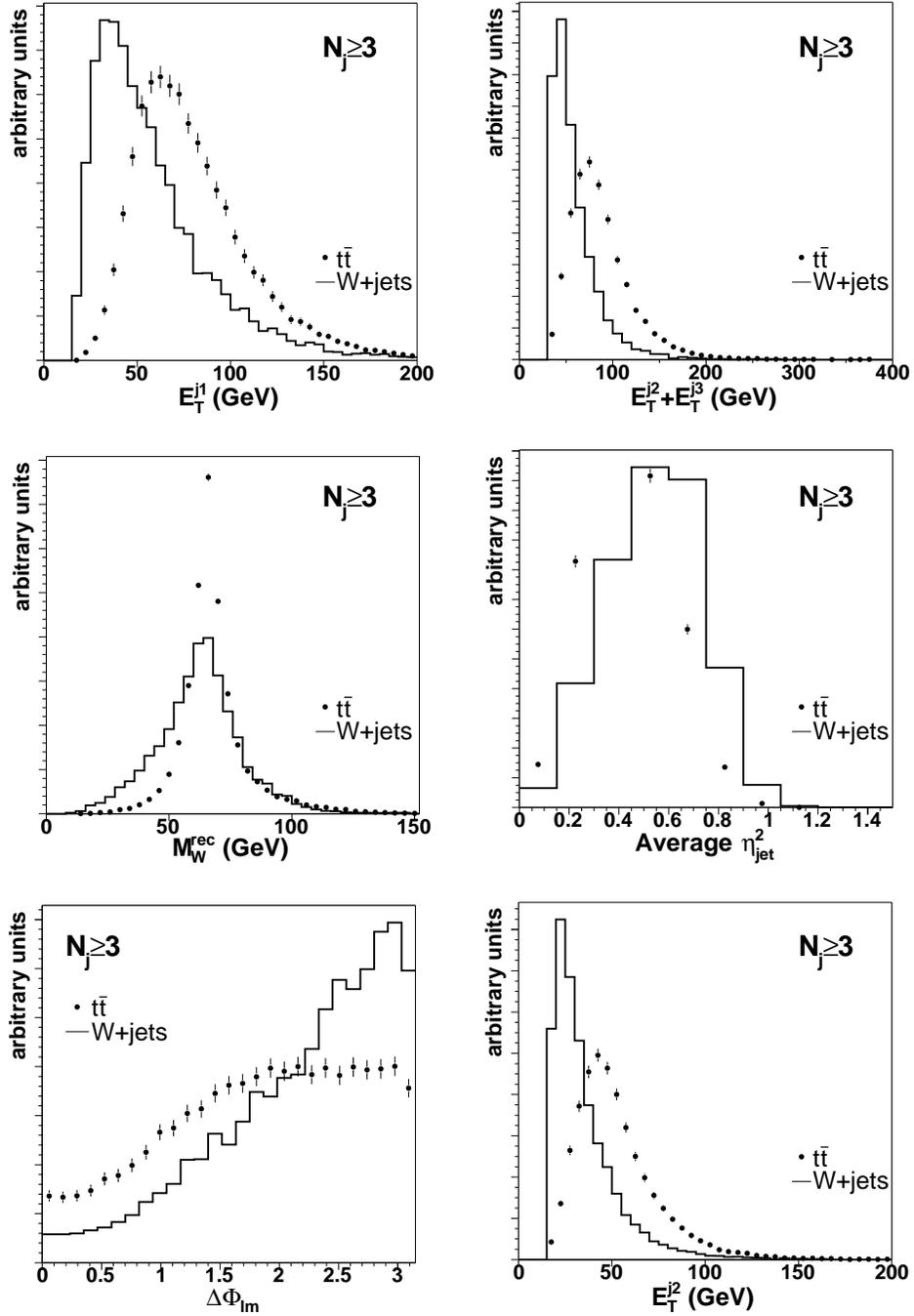


Figure 5.4: Comparison plots for some shape variables using ALPGEN $W+3p$ and PYTHIA $t\bar{t}$ for a variety of kinematic variables in the reconstructed $W + \geq 3$ jets mode. The two histograms are normalized to equal area.

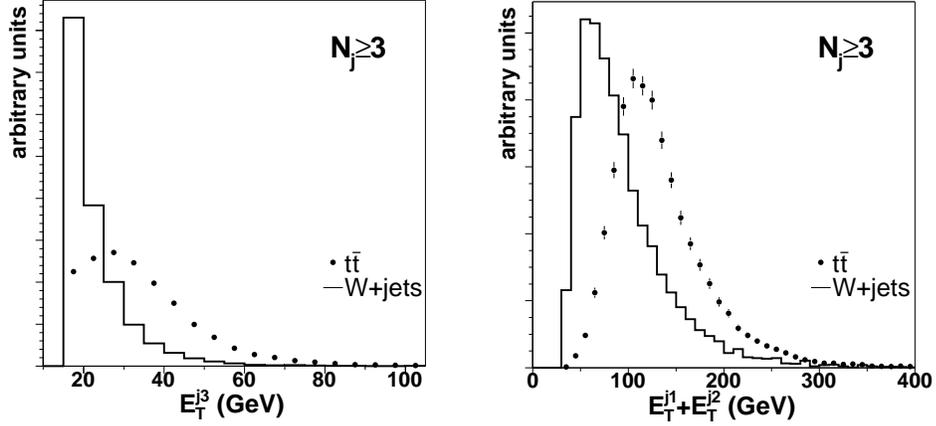


Figure 5.5: Comparison plots for some shape variables using ALPGEN $W+3p$ and PYTHIA $t\bar{t}$ for a variety of kinematic variables in the reconstructed $W+\geq 3$ jets mode. The two histograms are normalized to equal area.

5.6 Kinematic Variables

The main background to $t\bar{t}$ production in the “lepton+jets” channel comes from W production associated with hadronic jets. The contribution from QCD-fakes background, falls off rather fast with the missing transverse energy and generally is assumed to populate less energetic regions of the phase space. In selecting the discriminating variables and training the ANN we will try to achieve good separation between $t\bar{t}$ and $W+3, 4p$ simulated events. The other backgrounds will be taken into account in the final fit as well as the associated systematics.

Figure 5.2, 5.3, 5.4, 5.5 show a comparison in the shape of the distribution of various kinematic quantities for $t\bar{t}$ and $W+3p$ samples in events with three or more jets. The definitions for all 20 variables can be found in Table 5.5. A mass for the hadronically decaying W boson can be reconstructed as the dijet invariant mass

closest to 66 GeV out of the three highest E_T jets. The 66 GeV reference point was used instead of the world average for the W mas because full jet energy corrections are not performed in this analysis. Similar plots for events having four or more tight jets using $t\bar{t}$ and $W+4p$ MC would show the same trend. There is small loss of separation power in going from the $W+\geq 3$ jet mode to the $W+\geq 4$ jet mode but the expected signal to background fraction in the data sample will increase to ~ 0.9 . Besides the energy based variables, additional information can be extracted from event shape variables. Maximum jet rapidity, average jet rapidity, minimum dijet separation are all extracted from the three highest E_T jets. Since $t\bar{t}$ events are expected to be more spherical than the W +jets events we consider two topological variables: aplanarity and sphericity. These are constructed using the two lowest eigenvalues ($Q_1 \leq Q_2 \leq Q_3$) of the normalized momentum tensor:

$$\text{Aplanarity} = 3/2 Q_1 \tag{5.4}$$

$$\text{Sphericity} = 3/2 (Q_1 + Q_2) \tag{5.5}$$

The normalized momentum tensor is defined as:

$$\frac{\sum_i p_i^a p_i^b}{\sum_i p_i^2}, \tag{5.6}$$

where the a, b indices run over the three spatial directions and summation is taken over the five highest E_T jets, the lepton and the reconstructed transverse missing energy. No reconstruction of P_z , the z momentum component of the neutrino from the $W \rightarrow l\nu$ decay is attempted.

In order have a measure for the individual separation power for various variables, we looked at the expected statistical sensitivity from fitting the kinematic distribution

Variable	Definition
H_T	Scalar sum of transverse energies of all jets, lepton and \cancel{E}_T
Aplanarity	$3/2Q_1$
$\Sigma p_z / \Sigma E_T$	Ratio of total jet longitudinal momenta to total jet E_T
$\min(M_{jj})$	Minimum di-jet invariant mass for the three highest E_T jets
η_{max}	Maximum η for the three highest E_T jets
$\Sigma_{i=3}^5 E_{T,i}$	Sum E_T of third, fourth and fifth jets
$\min(\Delta R_{jj})$	Minimum di-jet separation in η and ϕ
$\Sigma_{i=1}^n E_{T,i}$	Sum E_T of all jets
\cancel{E}_T	Missing transverse energy
Sphericity	$3/2(Q_1 + Q_2)$
M_{event}	Invariant mass of jets, lepton and \cancel{E}_T
$M_{12} + M_{23} + M_{13}$	Sum of di-jet invariant masses for the three highest E_T jets
E_T^{j1}	E_T of jet with highest E_T
$E_T^{j2} + E_T^{j3}$	Sum of E_T of jets with second and third highest E_T
M_W^{rec}	Dijet invariant mass closest to 66.0 GeV.
$\Sigma \eta^2$	Sum of η^2 for the three highest E_T jets
$\Delta \Phi_{lm}$	Azimuthal angle between lepton and \cancel{E}_T
E_T^{j2}	E_T of jet with second highest E_T
E_T^{j3}	E_T of jet with third highest E_T
$E_T^{j1} + E_T^{j2}$	Sum of E_T of jets with first and second highest E_T

Table 5.5: The definition for all the kinematic variables considered in this analysis. The particular order for listing the variables was obtained from starting with the H_T and adding variables one by one in such order that performance in training a ANN on PHYTIA $t\bar{t}$ and ALPGEN W+3p was maximized during each step.

of 20 different single variables. The results for both $W+\geq 3$ and $W+\geq 4$ jets mode are presented in Figure 5.6 and 5.7 respectively. One can see that E_T based variables are better than the shape variables. The merit of the shape variables relative to the E_T based variables is that they are less dependent on the calorimeter calibration. They depend on the choice of Q^2 in MC simulation and the uncertainty on PDF's. Non of these variables shows to be biased in pseudo-experiments so in principle we could use all of these to measure a cross section. However one need to take into account that many of these variables are correlated so combining the measurement results is not a trivial operation.

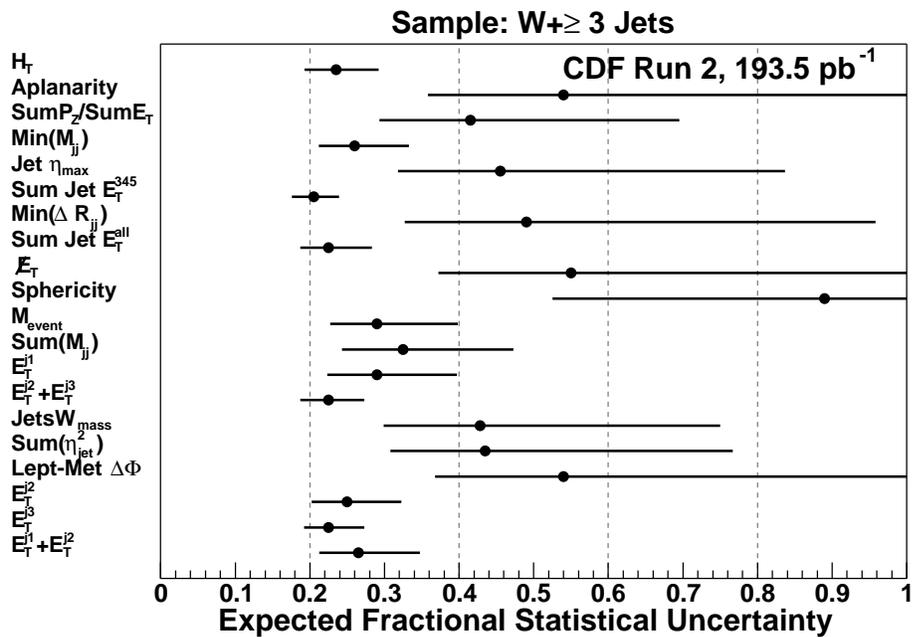


Figure 5.6: Expected statistical sensitivity from fitting variable distribution in the $W+\geq 3$ jets data sample. Points mark the median of the distribution of fit the uncertainty relative to the fitted $t\bar{t}$ fraction in pseudo-experiments. The error bars mark the 16-84 percentile interval in the relative uncertainty distribution. One should note that many of these variables are correlated.

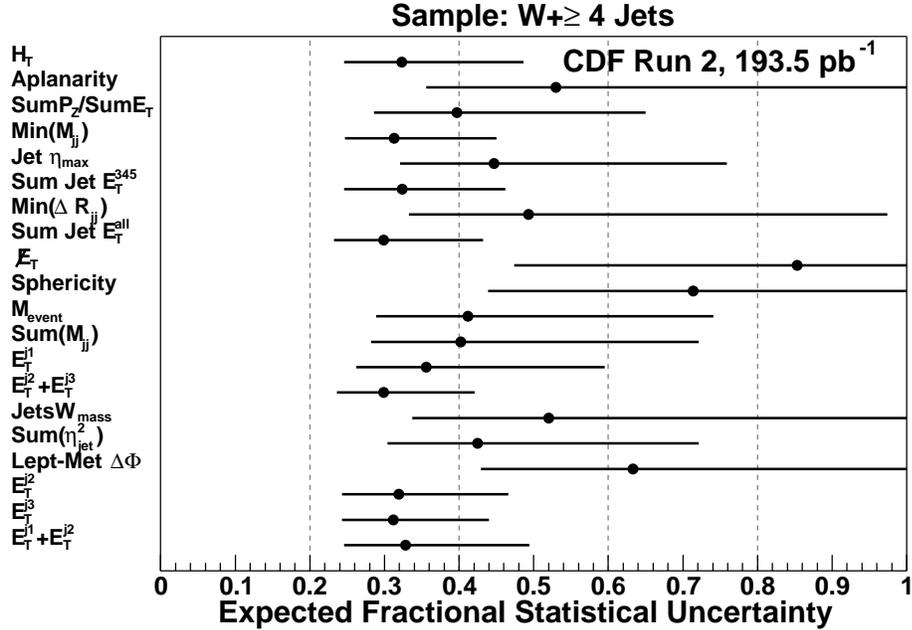


Figure 5.7: Expected statistical sensitivity from fitting variable distribution in the W+ \geq 4 jets data sample. Points mark the median of the distribution of fit the uncertainty relative to the fitted $t\bar{t}$ fraction in pseudo-experiments. The error bars mark the 16-84 percentile interval in the relative uncertainty distribution. One should note that many of these variables are correlated.

5.7 ANN Training and Testing

For the network training we use the JETNET [59] package, developed at CERN and the RootJetnet interface developed by Ohio-State group at CDF [81]. The network will have several input variables, one hidden layer and one continuous output unit in the range [0,1]. JETNET implements a variety of training algorithms. For this particular problem we get satisfactory results using the default back-propagation method with momentum added in order to improve learning stability at the end of

training. Learning is performed by modifying the network weights in order to minimize a mean square error function with a regularization term added in order to discourage unnecessary weights and avoid over-fitting, see Section 4.2.2 for details. This can be done by setting the pruning option in JETNET to be “on” ($MSTJN(21) = 1$).

We use the default JETNET learning rate $\eta = .001$ and momentum $\alpha = 0.5$. Ten different patterns are probed before an update of the weights is performed. In principle all these parameters as well as the training algorithm can be varied but we did not see any significant improvement in ANN performance. Signal events are trained for target 1 while background events are trained for target 0. When the input variables have different orders of magnitudes, sometimes in the learning process they need to be updated with different rates in order to achieve effective learning, which in turn requires more complicated learning algorithms. A simple method to avoid this is to normalize all input variables to a common scale. This is automatically done in the RootJetnet package, by dividing input variables by the first order of 10^n larger than the maximum value taken by the respective variable. For example: η_{max} having values smaller than 2.0, will be divided by 10 when transferred to the network input.

We are using PYTHIA for our default $t\bar{t}$ training sample. Training for the W +jets background is performed on ALPGEN $W+3p$ and $W+4p$ simulated samples for measurements in the $W+\geq 3$ and respectively $W+\geq 4$ jet samples. In the $W+\geq 3$ jets mode, we use 2×4000 events for training, 2×1000 events for testing and the rest of 3200 $W+3p$ (28000 $t\bar{t}$) events as a statistically independent validation sample which we use to check the quality of the training. In the $W+\geq 4$ jets mode, we use 2×2400 events for training, 2×1000 events for testing and the rest of 1600 $W+4p$ (14000 $t\bar{t}$) events for validation.

A short description of the training procedure follows here. In a first step the network is trained for a duration of 400 epochs. Since training is an iterative process, an epoch is defined as reading and learning being performed once upon the whole available training sample. The epoch for which the ANN has the lowest error when tested on a statistically independent sample is chosen as the reference stopping point. Then we repeat the training up the selected stopping epoch. In the end an statistically independent validation sample is used to check the quality training. Figure 5.8 compares the NN performance on a training and testing sample as a function of the number training epochs. One can see that the network learns quite fast, in about 100-200 epochs. Over-training effects, visible where the network error decreases in the training sample (red dots) but not in the testing sample (blue dots) are evident after a few hundred epochs. In the final ANNs used to fit the data, the actual stopping epoch was 126 for the $W+\geq 3$ jet mode and 180 for the $W+\geq 4$ jet mode network. One should note that for a different separation problem or when using a smaller training sample, overtraining effects can be far more pronounced after 100-200 epochs.

For the our final trained two networks, the ANN output for PYTHIA $t\bar{t}$ compared with ALPGEN $W+3(4)p$ is shown in Figure 5.9.

5.7.1 Choosing an Optimum Network Configuration

One has considerable liberty when choosing a particular combination of variables as inputs to the network. Ideally we would like to choose such variables which help define regions of the phase space where signal and background events are distributed differently. Figure 5.10 shows the average correlation coefficients for some 26 different kinematic variables in $W+3p$ (left) and $t\bar{t}$ (right) simulated events. There does not

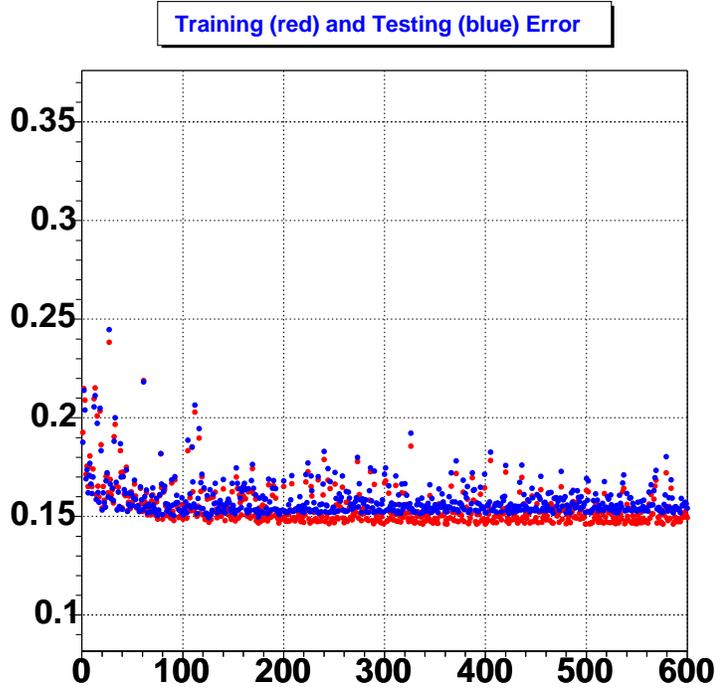


Figure 5.8: Typical variation of the training and testing ANN error with the number of training epochs.

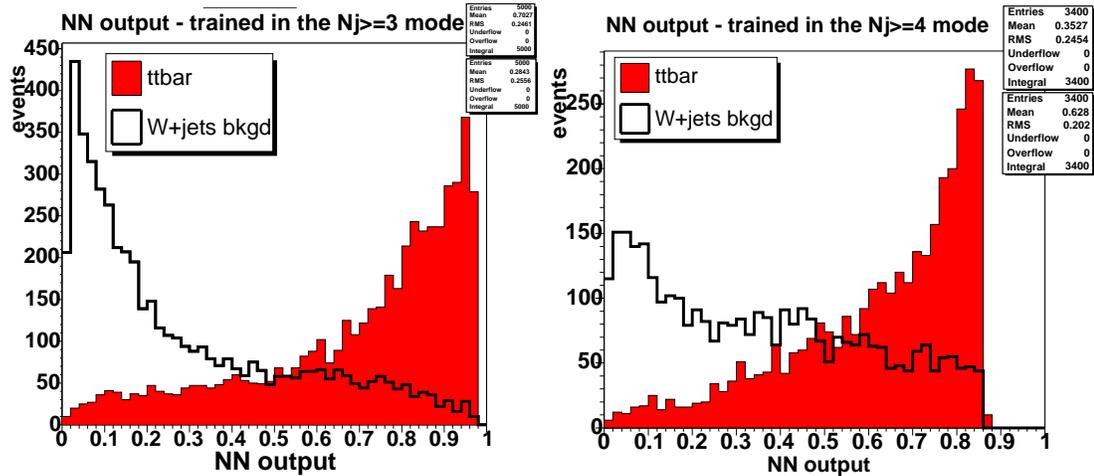


Figure 5.9: The output for ANN's trained in the $W+\geq 3(4)$ jets mode using PYTHIA $t\bar{t}$ and ALPGEN $W+3(4)p$ MC.

seem to be significant differences in terms of correlation coefficients between signal and background.

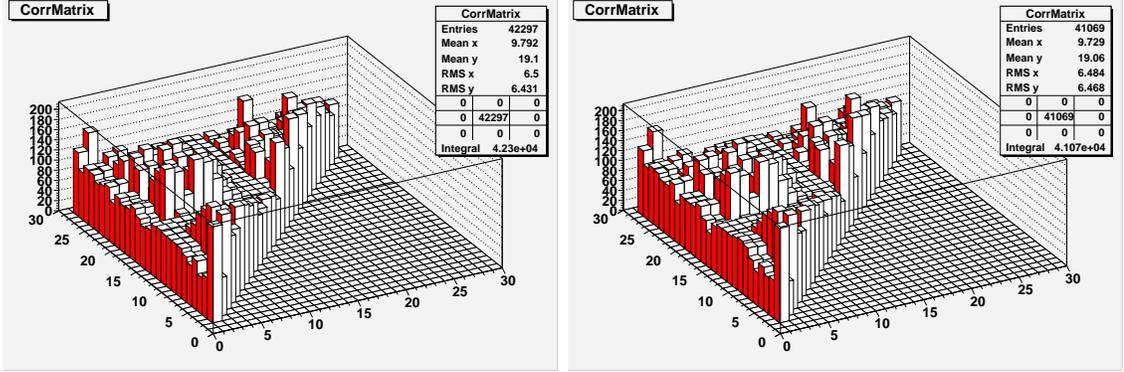


Figure 5.10: Average correlations coefficients for some 26 different kinematic variables in $W+3p$ (left) and $t\bar{t}$ (right) simulated events. The ordering of variables in the two plots is the same.

We have also tried to perform a quasi-exhaustive search hoping that maybe there is one combination of variables that performs much better than the rest. We trained a separate network for each combination of 2,3 variables out a total of 24 variables. For 4 to 8 input network configurations we gradually restricted the number of available variables to choose from down to 18. We also did a partial search for combinations of 17,18 and 20 input variables. Since performing pseudo-experiments for each individual ANN would take a long time, we monitor as a figure of merit the performance for the training, defined by the fraction of correct classifications in a test sample. A $t\bar{t}$ event is correctly classified if ANN output for this event is greater than 0.5 while a $W+jets$ event is correctly classified if ANN output is less than 0.5. The conclusion of this study was that there is no single combination of variables that performs much

better than the rest. For each given number of input variables, there are a number of slightly different combinations producing comparatively good results. We also note that there is an overall improvement in performance while the number of input variables increases.

Finally, out of the top performers we picked 3-4 different combinations and did pseudo-experiments in order to estimate the effect of fitting the ANN output. We use the shapes of the ANN output and generate 1500 pseudo-experiments using a combination of 19% $t\bar{t}$ events, 6.3% QCD and 74.7% $W+3p$ simulated events. The average number events in each pseudo-experiment corresponded to a 195^{-1}pb data sample. Figure 5.11 (top), shows the fit average fractional uncertainty for different choices of input combinations as a function of the number of input nodes. The number corresponding to 1-input is obtained by fitting the H_T shape alone. It is obvious that the more information is added in the better the fit performance. The overall gain with respect to a single variable kinematic fit is about 35%.

Systematic uncertainties are approached in a similar manner. The ANN is trained only once but we use two ANN output shapes, one from the nominal MC and the other from the systematic-shifted MC sample. We generate pseudo-experiments using the shifted MC, and do the fit twice using the regular and the shifted shapes. We take the systematic to be the average of the difference between the two fit results. In Figure 5.11 (bottom) a plot for the estimated systematic uncertainties²⁴ is presented. Again, numbers improve with adding more input variables. While for a given number of hidden variables the spread in fit uncertainty is generally small, the systematics seem to be more dependent on the particular choice in the combination

²⁴Systematic uncertainties are discussed in Section 5.9.

input variables. One should keep in mind that an $(n+1)$ -input ANN sometimes is quite different from the adjacent n -input configuration.

What we find is that a 7-input ANN is close enough to the overall best. When choosing a ANN architecture, simplicity may not be a stringent requirement [60], however we choose here a 7 input network as the minimal configuration yielding good performance. In future when more statistics in the data will allow a better validation of the W+jets simulation, a more complicated network or even a tree of multiple networks might be used to improve the measurement.

5.8 Comparisons Between Data and Simulation

The seven input variables we use are: the total transverse energy, H_T , the event aplanarity, minimum dijet mass, maximum jet rapidity, minimum dijet separation, the sum of transverse energy for all jets beyond the third jet, the ratio between jet longitudinal momenta and the jet transverse energy. This analysis relies on MC to simulate the background and signal properties. It is important to test our simulated events against the real data recorded by the detector. We do not have a high statistics top control sample but the data sample is large enough to test our W+jets model.

In order to validate the W+jets background modeling we looked at a top depleted region of our data sample: the $W + 3$ jet exclusive sample. Figure 5.12 shows the distributions for \cancel{E}_T , leading jet E_T , H_T and the other ANN input variables for $W + 3$ exclusive jet events compared to the prediction from ALPGEN+HERWIG, $W+3p$ MC, multi-jet background and PYTHIA $t\bar{t}$. For a 6.7 pb cross section we expect $\sim 10\%$ contribution from $t\bar{t}$ in this region. We add this contribution to overall shape, as well as the expected contributions from QCD and the smaller

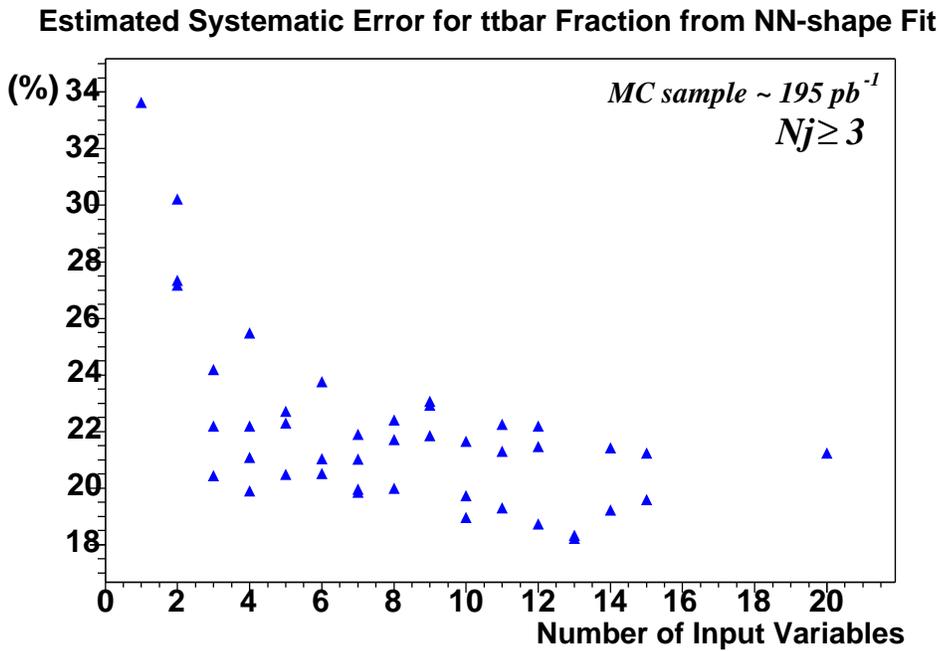
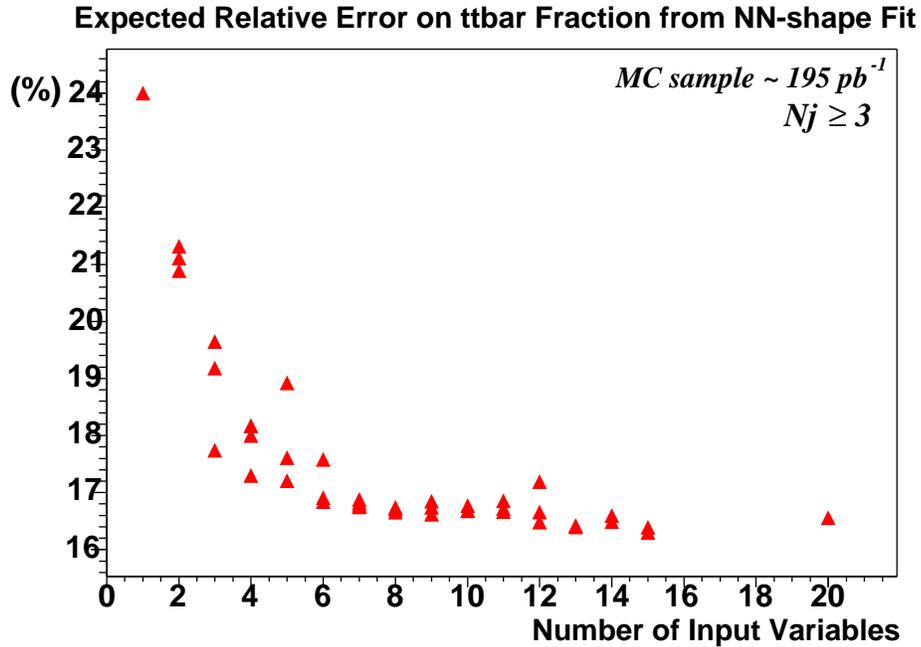


Figure 5.11: Predicted statistical and systematic uncertainties relative to a measured $t\bar{t}$ fraction for fitting the ANN output in the 3 or more jets mode. The average number events in pseudo-experiments corresponds to a 195^{-1}pb data sample. Statistical error bars, not shown here, should be slightly smaller in size than the triangles in the figure.

background process: $Wbb2p$, $WW1p$, $W\tau\nu3p$, $WZ0p$, $Z2p$, single-top. The MC model histogram is normalized to the number of events in the data sample. For a quantitative evaluation of the compatibility between the two shapes, each plot has attached the KS test result. The spectrum for the KS test results is expected to be flat if the shape of the smaller statistics data histogram were drawn from the same distribution as our MC model. The numbers in Figure 5.12 suggest good overall agreement between data and the MC simulation.

Variable correlations are also important for a multivariate approach. We use the following definition for an event by event correlation between two generic variables x and y :

$$\text{corr}(x, y) = \frac{(x - \bar{x}) \cdot (y - \bar{y})}{(\Delta x \cdot \Delta y)^{1/2}}, \quad (5.7)$$

where \bar{x} is the average and $\Delta x = \overline{(x - \bar{x})^2}$ for the distribution in the x variable. Comparisons between a simulated model and data in the $W + 3$ jets (exclusive) events are presented in Figure: 5.13, 5.14, 5.15 and 5.16. The model here is a combination of 10% $t\bar{t}$ and 90% W +jets simulated events. For a quantitative evaluation of the compatibility between the two shapes, each plot has attached the KS test result. The KS test numbers suggest in general good agreement between data and the MC model.

5.9 Systematic uncertainties

The cross section calculated from Equation: 5.2 is influenced by a series of systematic uncertainties. Uncertainties affect the luminosity, the signal efficiency and the ANN shape. The uncertainty on the measured luminosity is 5.9% [75].

Some systematics have impact on both the signal efficiency term and the ANN shape. When these effects modify the $t\bar{t}$ cross section in the same direction we add

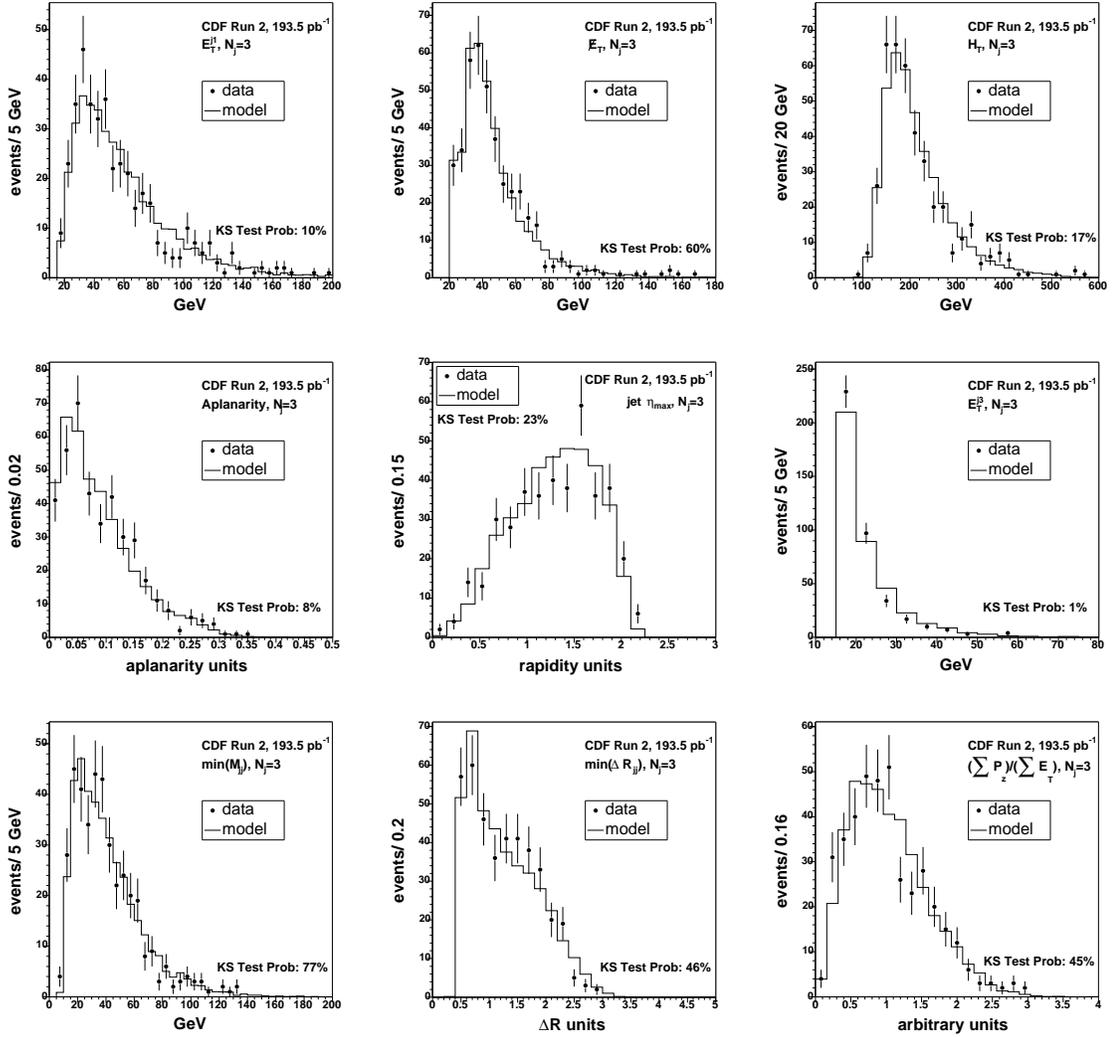


Figure 5.12: The leading jet E_T , \cancel{E}_T and the ANN input variables distributions in W+3 jets (exclusive) events compared to the prediction from ALPGEN+HERWIG MC, multi-jet background and PYTHIA $t\bar{t}$ MC. These are fit results, the MC model histogram is normalized to the number of events in the data sample. The KS test statistic is shown as a measure for the compatibility between the two shapes.

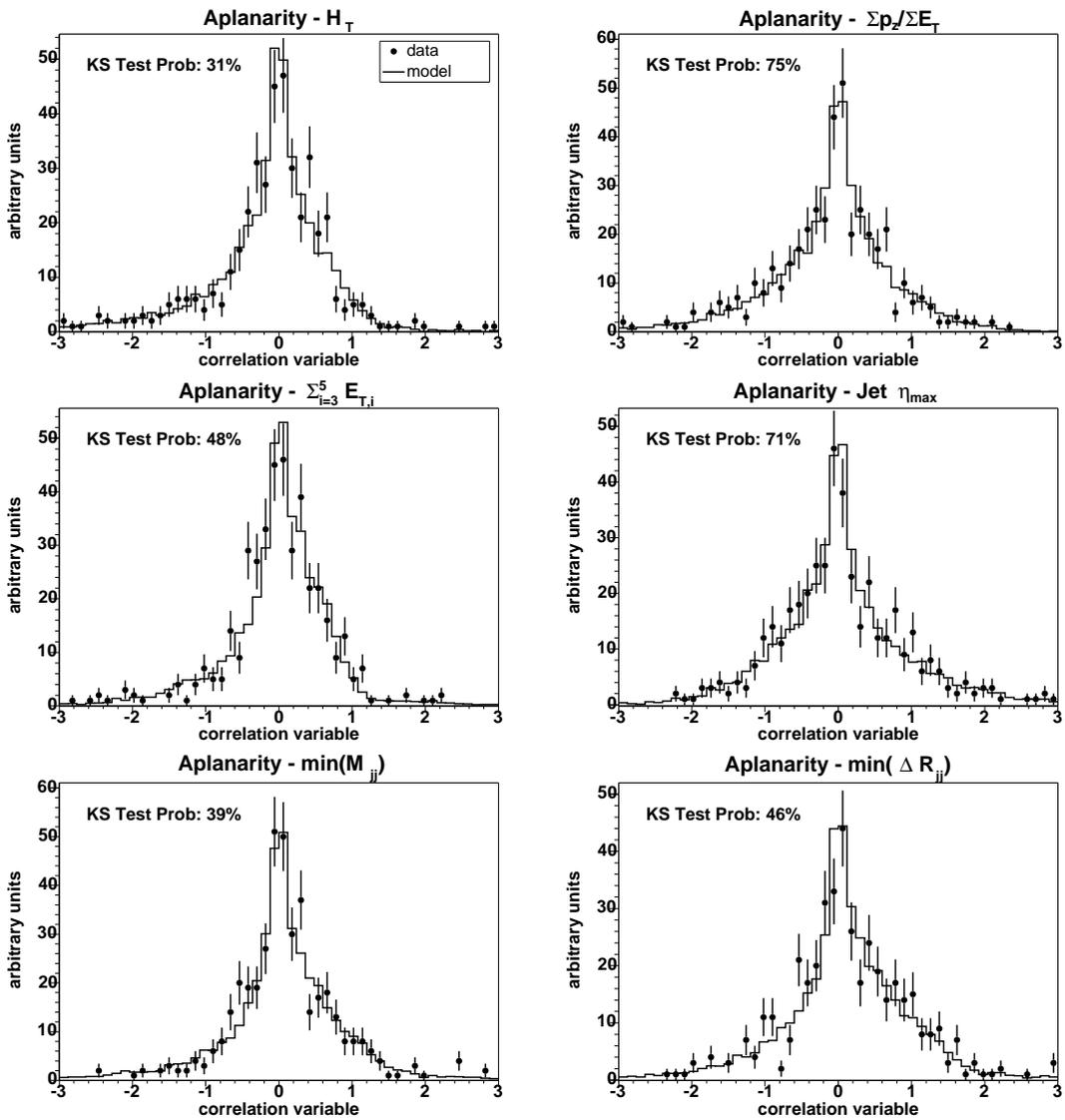


Figure 5.13: Distribution of the correlation coefficients in data compared to a mix of 10% PYTHIA $t\bar{t}$ and 90% ALPGEN $W+3p$ events. Plots are in the $W+3$ jets exclusive mode. Histograms are normalized to equal area. For a quantitative evaluation of the compatibility between the two shapes, each plot has attached the KS test result.

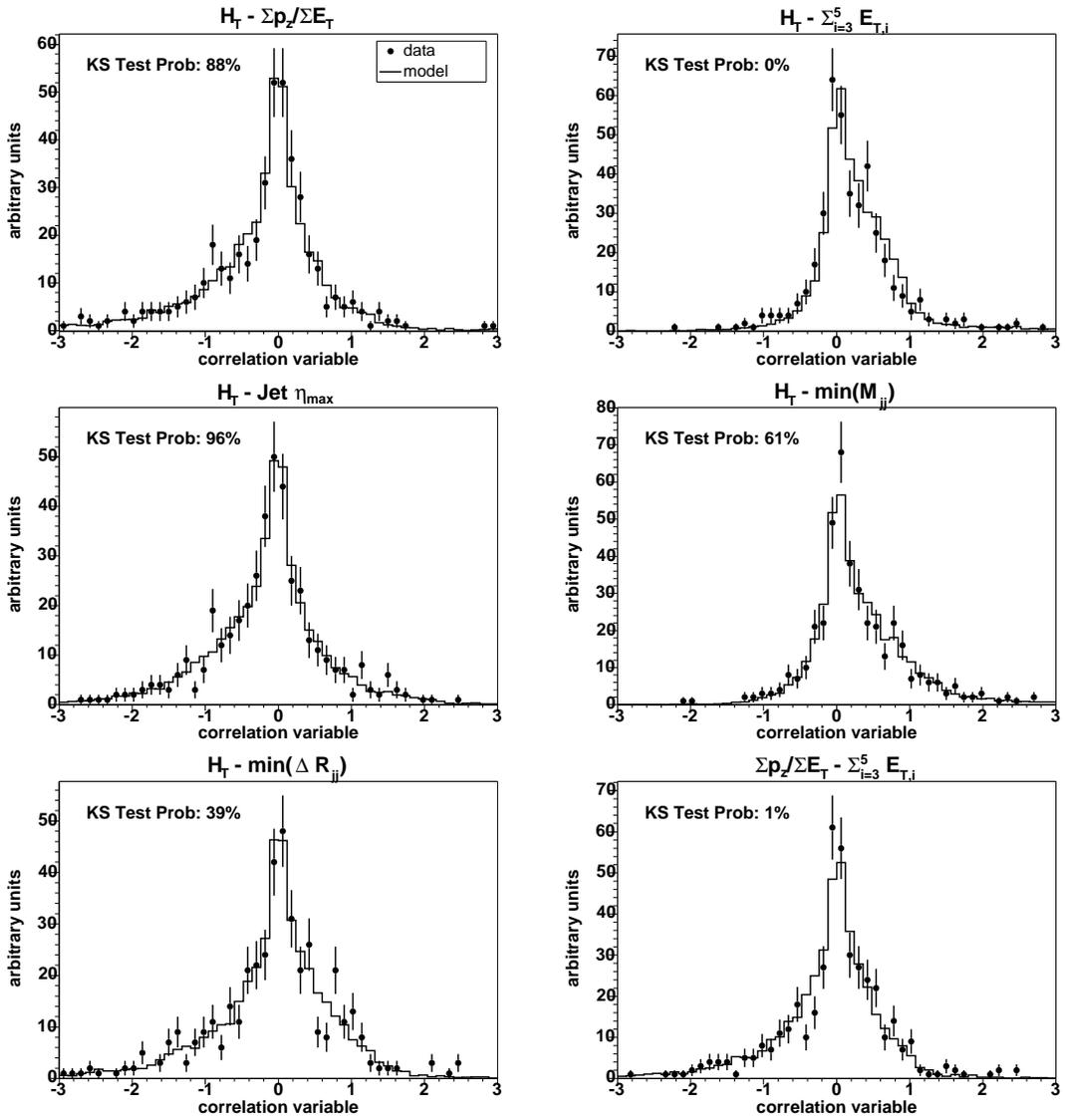


Figure 5.14: Distribution of the correlation coefficients in data compared to mix of 10% PYTHIA $t\bar{t}$ and 90% ALPGEN $W+3p$ events. Plots are in the $W+3$ jets exclusive mode. Histograms are normalized to equal area. For a quantitative evaluation of the compatibility between the two shapes, each plot has attached the KS test result.

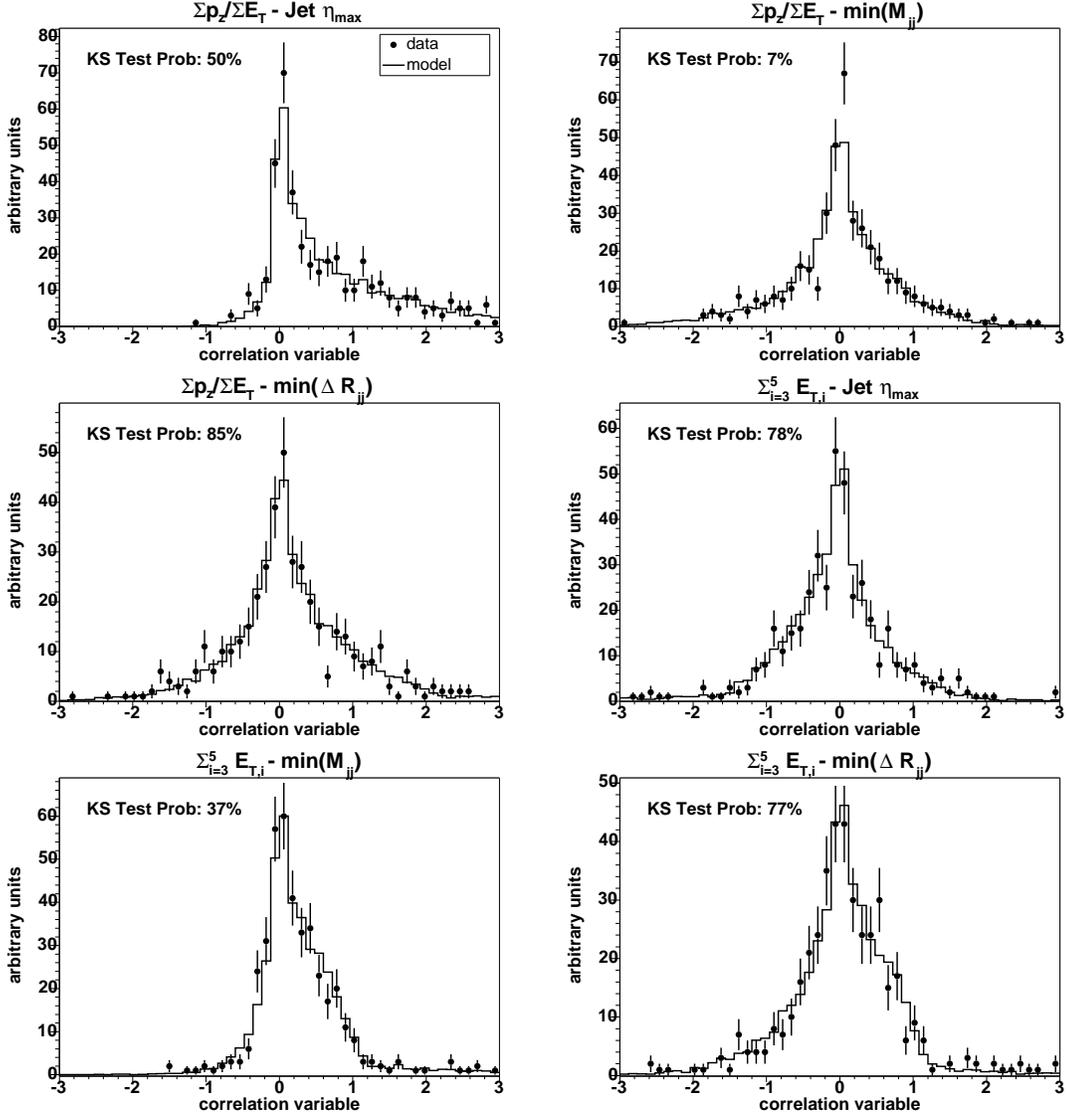


Figure 5.15: Distribution of the correlation coefficients in data compared to mix of 10% PYTHIA $t\bar{t}$ and 90% ALPGEN $W+3p$ events. Plots are in the $W+3$ jets exclusive mode. two histograms are normalized to equal area. For a quantitative evaluation of the compatibility between the two shapes, each plot has attached the KS test result.

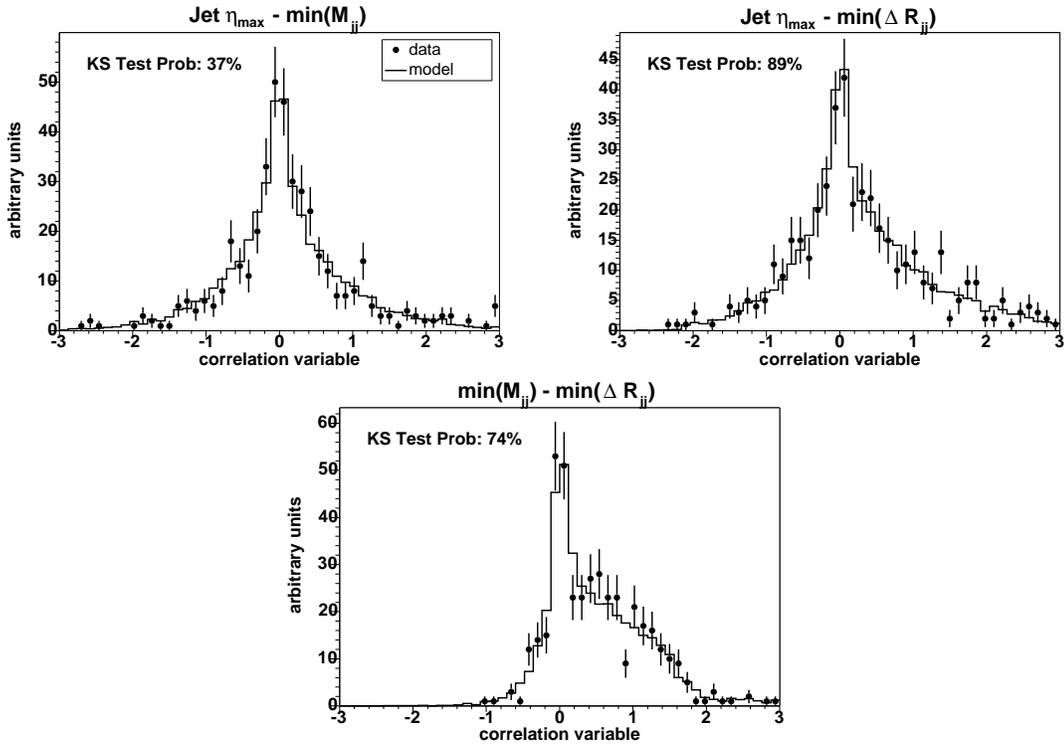


Figure 5.16: Distribution of the correlation coefficients in data compared to mix of 10% PYTHIA $t\bar{t}$ and 90% ALPGEN $W+3p$ events. Plots are in the $W+3$ jets exclusive mode. Histograms are normalized to equal area. For a quantitative evaluation of the compatibility between the two shapes, each plot has attached the KS test result.

them linearly as they are fully correlated. Numbers originating from different systematic sources will be added in quadrature for the final result. The systematic uncertainties on the $t\bar{t}$ fraction for measurement in the $W+\geq 3$ and $W+\geq 4$ jets sample are presented in Table 5.7 and Table 5.8 respectively.

The jet energy uncertainty is a major component of the overall systematic. We expect, after applying all 6 levels of correction to recover, on average, from the energy measured in a certain portion of the calorimeter the energy of the original parton. The jet correction systematics deal with uncertainties associated with each step of the correction procedure. These are actually uncertainties on how “wrong” the energy of a jet was measured in a particular portion of the detector. This is why, while we choose to correct²⁵ jet energy only up to the Level-4, jet energy systematics need to be calculated for the all 6 levels of correction. Acceptance terms are calculated and listed as the relative variation in the number of events. Shape systematics are calculated using pseudo-experiments. There are two ways to do this: one is to vary jet energy corrections according to each systematic effect at a time and add the calculated systematics from all effects in quadrature; the other is to vary jet energy corrections according to all systematics once with an amplitude calculated by adding all effects in quadrature and calculate only one systematic number. We found similar results using either method.

Systematics associated with the multijet fakes should account for both the modeling and our estimation of their contribution to the data samples. For calculating the modeling systematic we use the conversion electron sample as an alternative model. Uncertainties associated with the fraction of multijet events are calculated by varying

²⁵See section 3.1.1 for a discussion of the jet energy corrections.

the 6.3% contribution in the fit with $\pm_{50}^{+100}\%$ and taking half the difference in the fit results.

The electroweak backgrounds are not included in the training sample but will be included for the fit in the final $W+jets$ shape. The total contributions of these backgrounds to the data sample is estimated to be 11%. The leading order theory cross section values (see Table 5.1), were used to fix the contribution of each component relative to the dominant $W+3p$ process. Using pseudo-experiments we estimate the impact on the fit of including versus not including these and take half of the difference (2%) as a systematic.

The generator systematic is obtained from comparing PYTHIA and HERWIG simulated $t\bar{t}$ samples. The systematic for the acceptance calculation is 1.4%. For the systematics related on the amount of ISR and FSR we considered the following MC samples recommended by the CDF Top Group [77]:

- ttopbe(ISR less): Lambda QCD = 100 MeV, $K_{factor}=2.0$ for ISR evolution.
- ttopce(ISR more): Lambda QCD = 384 MeV, $K_{factor}=0.5$ for ISR evolution.
- ttopde(FSR less): Lambda QCD = 100 MeV, $K_{factor}=2.0$ for FSR PS.
- ttopee(FSR more): Lambda QCD = 384 MeV, $K_{factor}=0.5$ for FSR PS.

In addition, a 5% systematic due to the uncertainty on lepton identification efficiency scale factors for data compared to MC has been included. Details on how this uncertainty was evaluated can be found in [78].

For a systematic associated with our $W+jets$ model we considered several MC samples generated at various Q^2 . The nominal ALPGEN MC was generated at $Q^2 =$

$M_W^2 + \sum_i p_{T,i}^2$ where $p_{T,i}$ is the transverse momentum of the i -th parton. systematic effects on $t\bar{t}$ fraction for fits using H_T shape using a number of different samples are shown in Table 5.6. Based on these number we chose $Q^2 = 4M_W^2$ as the alternative MC sample for calculating this systematic. The results for ANN fits are given in tables 5.7 and tables 5.8.

Q^2	$W+\geq 3$	$W+\geq 4$
$4 * M_W^2$	-24.0%	-33.5%
$0.25 * M_W^2$	-25.0%	-37.0%
$\sum_i P t_i^2$	+4.3%	-0.9%
M_W^2	-26.5%	-
$M_W^2 + P t_W^2$	- 9.5%	-

Table 5.6: systematic effects on $t\bar{t}$ contribution extracted from fits using the H_T shape.

PDF systematics were calculated according to a method proposed in [79] which allows one to calculate such systematics independent of the ISR and FSR systematics. For the standard PYTHIA $t\bar{t}$ MC set generated using the CTEQ5L PDF library, each event can be re-weighted using various PDF libraries:

$$EventWeight = \frac{f1(x1, Q)_{NewPDF} * f2(x2, Q)_{NewPDF}}{f1(x1, Q)_{CTEQ5L} * f2(x2, Q)_{CTEQ5L}}, \quad (5.8)$$

using the incoming partons x_i momentum fractions and the event Q calculated using HEPG information. Here we compare MRST72 to MRST75 and the default CTEQ6M to 2×20 different eigenvalue variations of CTEQ6M (ISET = 101,140). We add in quadrature contributions for each systematic variation.

Effect	signal efficiency (%)	Shape (%)	Total (%)
Jet E_T Scale	4.7	12.2	16.9
W+jets Q^2 Scale	-	10.2	10.2
QCD fraction	-	0.6	0.6
QCD shape	-	1.1	1.1
Other EWK	-	2.0	2.0
$t\bar{t}$ PDF	1.5	2.9	4.4
$t\bar{t}$ ISR	2.1	1.9	3.0
$t\bar{t}$ FSR	1.7	1.0	2.7
$t\bar{t}$ generator	1.4	0.3	1.7
Lepton ID/trigger	2.0	-	2.0
Lepton Isolation	5.0	-	5.0
Luminosity	-	-	5.9
Total	-	-	22.3

Table 5.7: Table for systematic uncertainties in the $W+\geq 3$ jets mode. The overall uncertainty is obtained by adding in quadrature the individual effects.

Effect	signal efficiency (%)	Shape (%)	Total (%)
Jet E_T Scale	10.1	9.0	19.1
W+jets Q^2 Scale	-	16.0	16.0
QCD fraction	-	0.6	0.6
QCD shape	-	2.3	2.3
Other EWK	-	0.3	0.3
$t\bar{t}$ PDF	2.4	2.3	4.7
$t\bar{t}$ ISR	2.4	0.7	3.1
$t\bar{t}$ FSR	2.1	1.0	3.1
$t\bar{t}$ generator	1.4	0.2	1.6
Lepton ID/trigger	2.0	-	2.0
Lepton Isolation	5.0	-	5.0
Luminosity	-	-	5.9
Total	-	-	27.1

Table 5.8: Table for systematic uncertainties in the $W+\geq 4$ jets mode. The overall uncertainty is obtained by adding in quadrature the individual effects.

5.10 The Fitting Technique

The overall systematic uncertainty for a measurement using ANN fits in the $W+\geq 3$ jets data sample was estimated to be 22.3%. A similar measurement in the $W+\geq 4$ jets data sample would have a 27.1% estimated systematic uncertainty. Statistical sensitivity studies using pseudo-experiments in the $W+\geq 3$ jets data sample predict a statistical error in the range 15-19% for 68% of data-sized experiments, with a median at 16.5%. In the $W+\geq 4$ jets data sample we expect a statistical error in the 19-28% range with a median of 23%. Considering both the systematic and the predicted statistical uncertainties we conclude that the best measurement results will be obtained for ANN fits in the $W+\geq 3$ jets data sample.

In order to test the fitting method, Figure 5.17 shows the average cross section measured in 2500 pseudo-experiments as a function of generated top cross section. No bias in results from fitting the ANN shape is evident from this plot. In Figure 5.18 the fit pull distribution from pseudo-experiments is shown for measurements in the $W+\geq 3$ and $W+\geq 4$ jets sample. In both cases the shape of the histogram is consistent with being a Gaussian with RMS equal to 1.

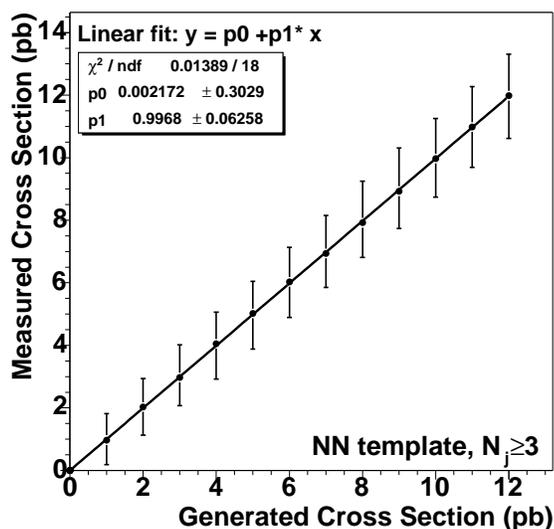


Figure 5.17: Test of the fitting technique under different signal hypotheses for the $t\bar{t}$ cross section. The error bars mark the 16-84 percentile interval in the distribution of measured cross section in pseudo-experiments.

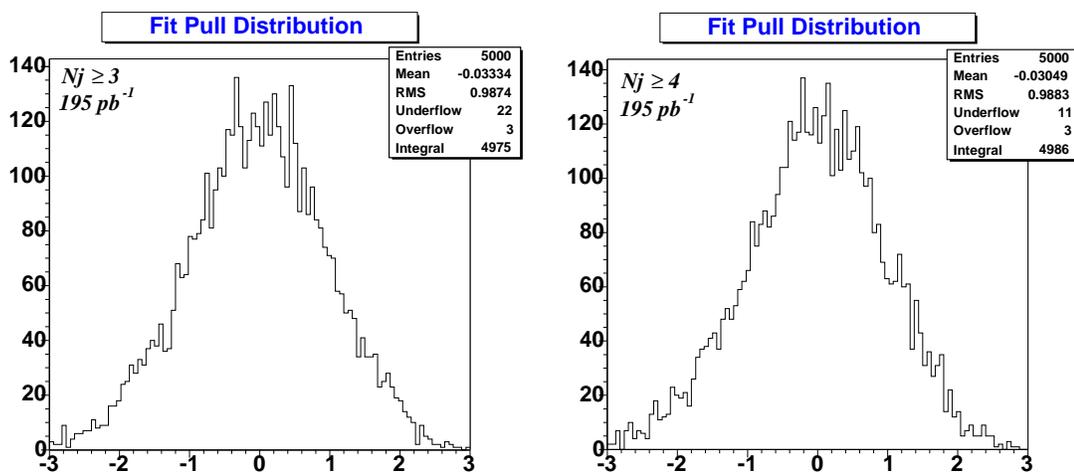


Figure 5.18: Fit pull distribution from pseudo-experiments for fits using the ANN's trained in the $W+\geq 3$ and $W+\geq 4$ jets mode.

5.11 Cross Section Calculation

5.11.1 Results for $M_{top}=175 \text{ GeV}/c^2$.

The final fit of the data using the $t\bar{t}$ W+jets and QCD background ANN templates to the $W+\geq 3$ jets data samples are presented in Figure 5.19. In order to give an idea for the quality of the fit, the uncertainty returned by the fitter is compared to pseudo-experiments.

A priori our best result is for the ANN fit in the 3 or more jets sample. We have repeated the analysis in the $W+\geq 4$ jet sample, where 118 events pass the event selection criteria and the expected S/B fraction is 4 times larger. The final fit of data in the $W+\geq 4$ data sample is presented in Figure 5.19. The same ANN configuration was used for both fits but each ANN was trained using the appropriate MC samples: ALPGEN $W+3p$ for the $W+\geq 3$ jets sample and ALPGEN $W+4p$ for the $W+\geq 4$ jets sample.

We extract a $t\bar{t}$ contribution to these data samples of:

$$f_{t\bar{t}} = 0.176 \pm 0.030 (W+\geq 3),$$

$$f_{t\bar{t}} = 0.473 \pm 0.100 (W+\geq 4),$$

where the uncertainty is statistical only and we have assumed a top mass of $175 \text{ GeV}/c^2$.

The measured top pair production cross section is then determined using Equation 5.2:

$$\sigma_{t\bar{t}} = 6.6 \pm 1.1 \pm 1.5 \text{ pb } (W+\geq 3),$$

$$\sigma_{t\bar{t}} = 7.5 \pm 1.6 \pm 2.0 \text{ pb } (W+\geq 4),$$

where the uncertainties are statistical and respective systematic. These numbers should be compared to the latest theoretical prediction of $6.7_{-0.88}^{+0.71} \text{ pb}$ [23].

Several cross-checks of this result have been done. Figure 5.20 shows the measured $t\bar{t}$ cross-section for a single variable fit to all 20 variables listed in Table 5.5 in both the $W+\geq 3$ and the $W+\geq 4$ jets data sample. *A priori* our best results for single variable fits are expected from measurements in the $W+\geq 3$ data sample. In Figure 5.20 the vertical line marks the expected $t\bar{t}$ fraction for a 6.7 pb cross section. Results look consistent within errors with the ANN fit result. One should keep in mind that many of the 20 variables we used are correlated and consequently the fit results are not expected to be completely independent.

5.11.2 Results as a Function of the Assumed Top Mass.

So far we presented results assuming a top mass of $175 \text{ GeV}/c^2$. In Table 5.9 cross section results for several top masses are presented. The change in the signal efficiency with respect to the central value of $175 \text{ GeV}/c^2$ is from a linear fit to HERWIG samples. The uncertainties listed in the Table 5.9 are statistical only.

In Figure 5.21 is shown a plot comparing $t\bar{t}$ cross section results with the theoretical predictions [23, 26] for a range of top masses. The error bars represent the statistical uncertainties. For the point at $175 \text{ GeV}/c^2$, systematic uncertainties are also shown with black, error bars.

5.11.3 Consistency Checks

Assuming that top indeed decays to a W boson and a b quark according to the SM expectation, a method to increase the signal to background ratio to $\sim 3/1$ is to select only b-tagged events. Figure 5.22 shows a comparison between the data sample tagged with the SECVTX algorithm [82] and simulated $t\bar{t}$ MC events. This is not a fit, instead the two histograms are normalized to equal area. The ANN output for the

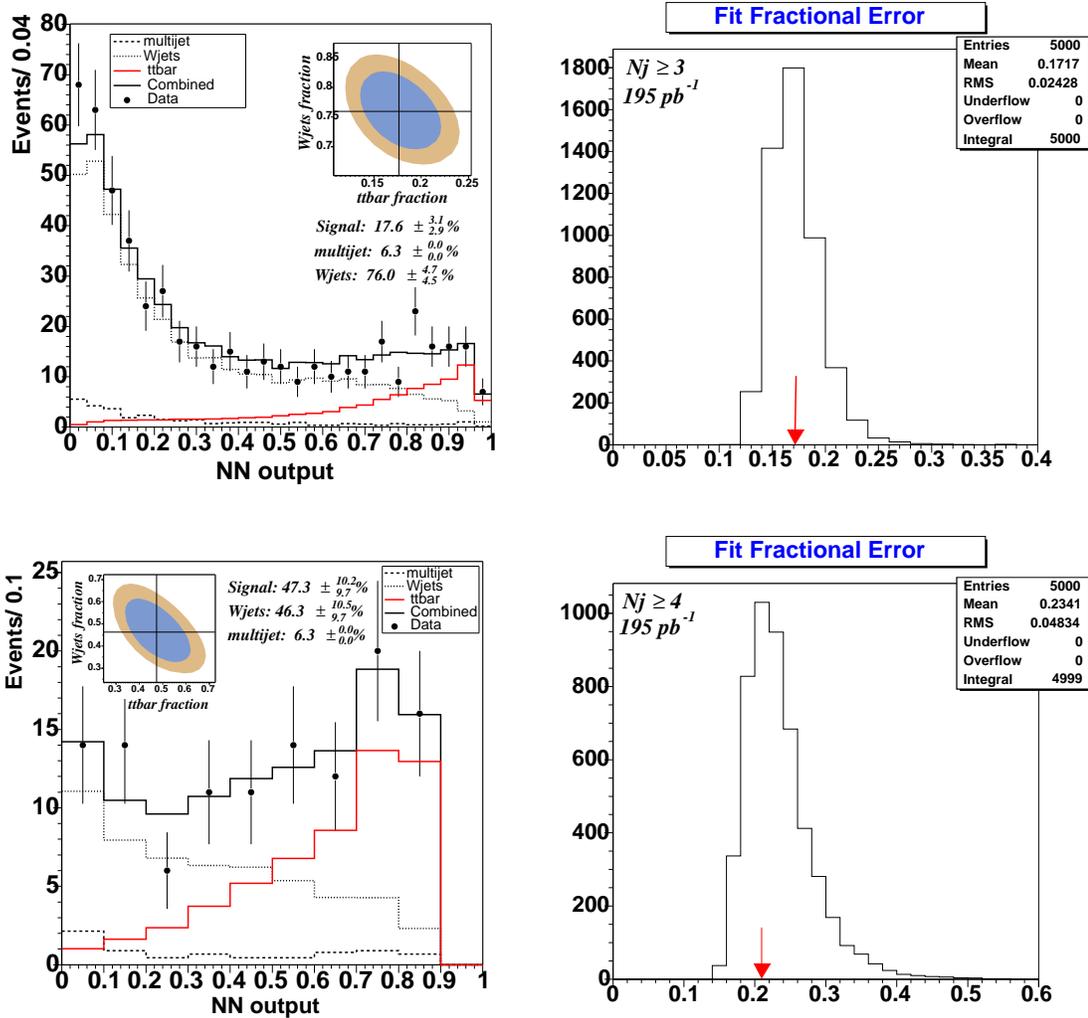


Figure 5.19: Fit to the data using the $t\bar{t}$ W+jets and QCD background ANN templates in the $W+\geq 3$ (top) and $W+\geq 4$ (bottom) jets samples assuming a top mass of $175 \text{ GeV}/c^2$. The relative uncertainty returned by the fitter is compared to relative uncertainty distribution in pseudo-experiments (right).

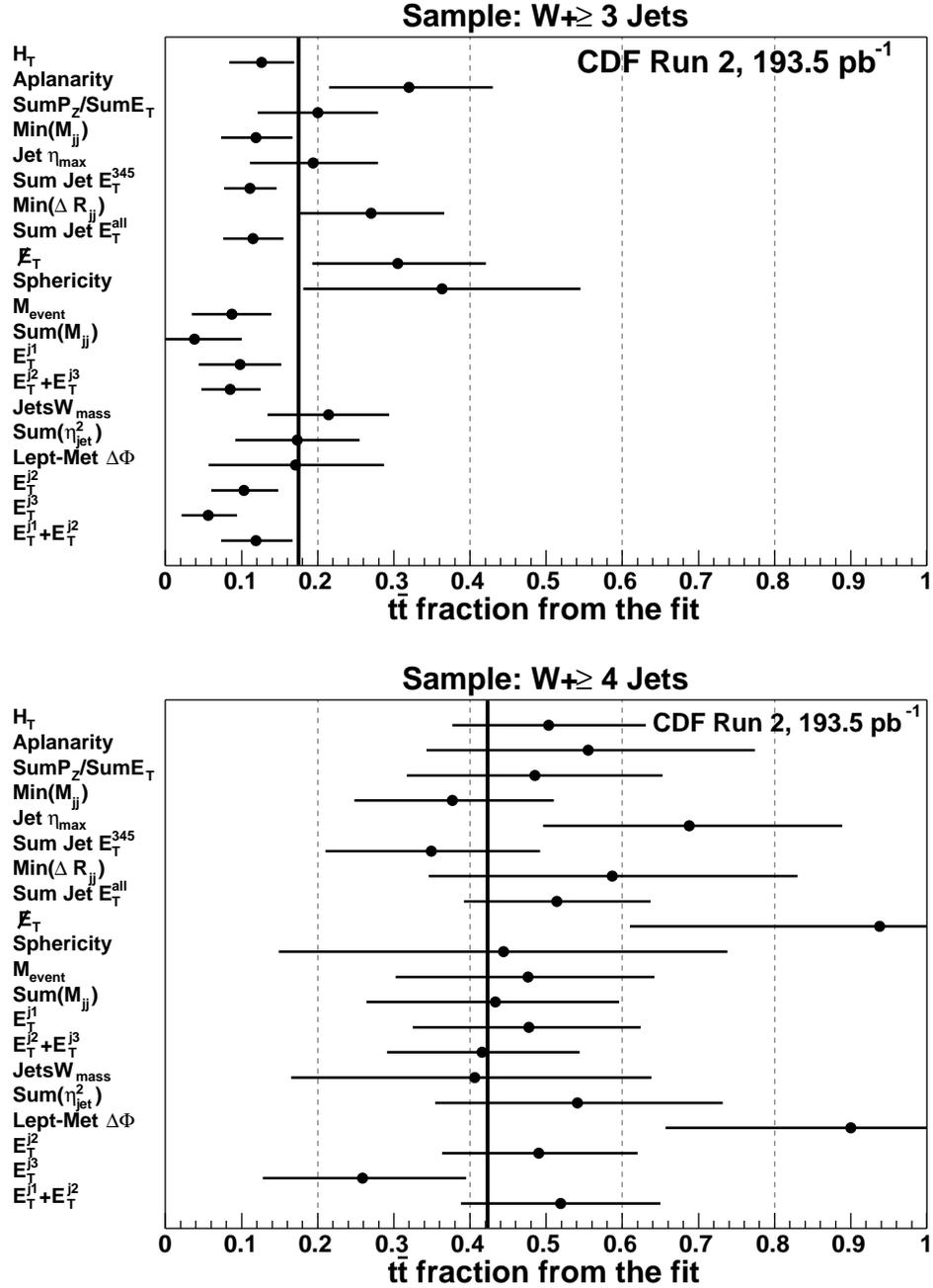


Figure 5.20: Measured $t\bar{t}$ fraction in the $W+\geq 3$ jets data sample (top) and $W+\geq 4$ jets data sample (bottom) for 20 different single variables assuming a top mass of $175 \text{ GeV}/c^2$. The uncertainty shown here is statistical only, the vertical line shows the expected fraction for a theory motivated $6.7 \text{ pb } t\bar{t}$ cross section.

Generated Top Mass	Relative Acceptance	$t\bar{t}$ fit fraction (%)	σ (pb)
160	0.934	19.6±3.4	7.9±1.3
165	0.956	18.9±3.2	7.5±1.3
170	0.978	18.1±3.1	7.0±1.2
175	1.000	17.5±3.0	6.6±1.1
180	1.022	16.9±2.9	6.3±1.1
185	1.044	16.2±2.8	5.9±1.0
190	1.066	16.0±2.8	5.7±1.0

Table 5.9: The $t\bar{t}$ production cross section measured in the $W+\geq 3$ jets data sample for fits to HERWIG $t\bar{t}$ MC generated at different top masses. The uncertainties shown here are statistical only.

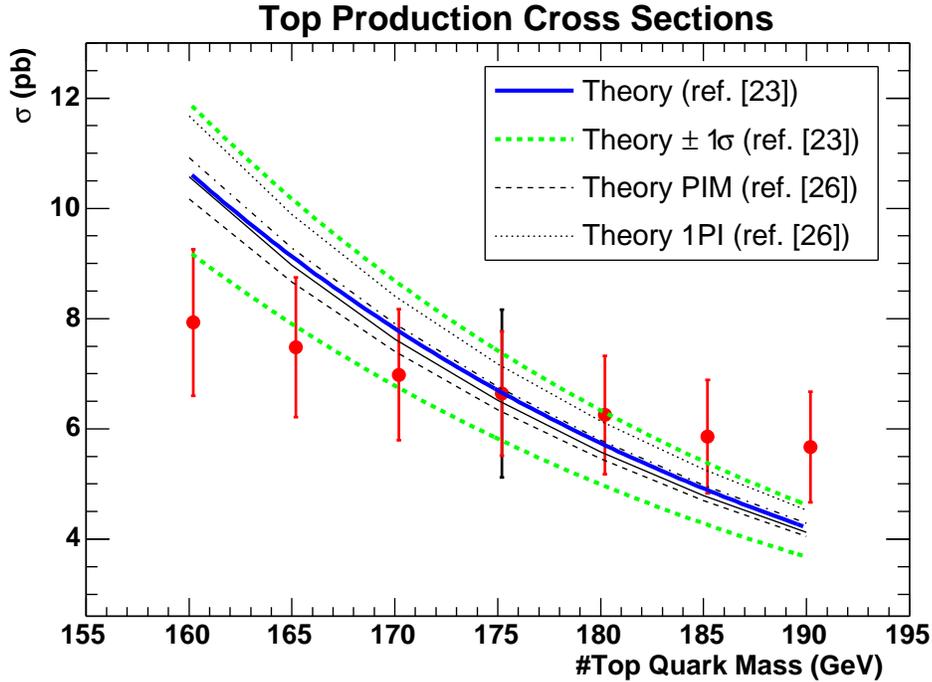


Figure 5.21: The $t\bar{t}$ production cross section in the $W+\geq 3$ jets data sample from fits to HERWIG $t\bar{t}$ MC samples generated at different top masses. The uncertainties shown are statistical, except for the 175 GeV point where the systematic uncertainties are shown in black.

data does tend to cluster near 1, as we would expect for a top-enriched sample. From an opposite point of view we can argue that our ANN verifies that the kinematics of the b-tagged events is indeed very top-like.

A cross-check in a top depleted region of the phase space can be done by looking at the 3-jets exclusive sample where we expect $\sim 10\%$ contribution from $t\bar{t}$. In Figure 5.23 the ANN output for data in the 3 jet exclusive mode is compared to the prediction from ALPGEN+HERWIG, W+jets, multi-jet background and PYTHIA $t\bar{t}$. No tagging requirement was made here. Again, this is not a fit, instead the two histograms are normalized to equal area. The KS test statistics suggests good agreement between data and the MC simulation.

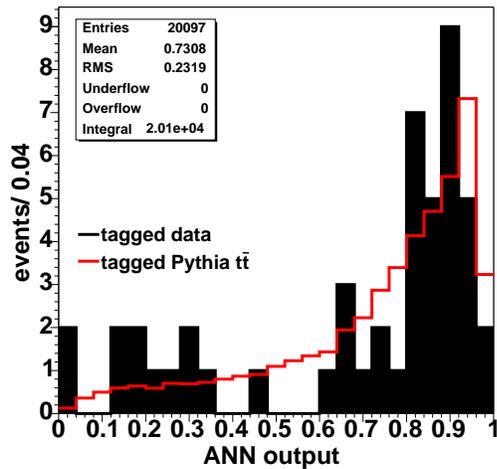


Figure 5.22: The SECVTX b-tagged data compared to the tagged PYTHIA $t\bar{t}$ MC in the $W+\geq 3$ jets sample. This is not a fit, instead the two histograms are normalized to equal area.

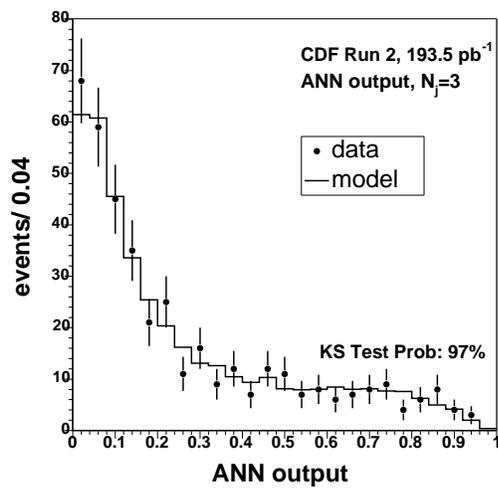


Figure 5.23: Comparison between data and MC simulation ANN output in $W+\geq 3$ jets exclusive mode. No b-tagging requirement was made here. This is not a fit, instead the two histograms are normalized to equal area.

CHAPTER 6

SUMMARY AND CONCLUSIONS

We present a measurement of the top pair production cross section in $p\bar{p}$ collisions at 1.96 TeV, from a data sample collected at CDF between March 2002 and September 2003 with an integrated luminosity of 193.5 pb^{-1} . At these energies, the top quarks are produced mostly via strong interactions in top anti-top pairs ($t\bar{t}$). The top quark has an extremely short lifetime and according to the Standard Model decays with $\sim 100\%$ probability into a b quark and a W boson. In the “lepton+jets” channel, the signal from top pair production is detected for events where one of the W bosons decays hadronically in two quarks, which we see as jets in the detector, and the other W decays to a charged lepton and a neutrino. For clean identification, here we require that the charged lepton must be an electron or muon of either charge. The presence of a neutrino in the event is inferred from an imbalance in the transverse energy.

In order to bring the signal to background ratio to manageable levels, measurements in this channel traditionally use precision tracking information to identify at least one secondary vertex produced in the decay of a long lived b hadron. An alternative approach was taken for this analysis. Because of the large mass of the top quark, $t\bar{t}$ events tend to be more spherical and more energetic than most background processes which otherwise mimic the $t\bar{t}$ signature in the “lepton+jets” channel. Monte Carlo

simulation are used to model the kinematics of background and $t\bar{t}$ events. We find a number of energy based and event shape variables that can be used to statistically discriminate between signal and background events. A neural network technique is employed to combine multiple variables in order to enhance signal separation. Such a measurement takes advantage of a larger data sample than the b-tagging based analyses and achieves a comparable level of precision.

A binned likelihood fit to the neural network output distribution for a 519 event $W+\geq 3$ jets data sample yields a $17.6 \pm 3.0(stat)\%$ fraction of $t\bar{t}$ events. The inclusive top pair production cross section is measured to be $\sigma_{t\bar{t}} = 6.6 \pm 1.1(stat) \pm 1.5(syst)$ pb. As a cross check a measurement was also made in the $W+\geq 4$ jets mode where the $t\bar{t}$ cross section is measured to be $7.5 \pm 1.6(stat) \pm 2.0(syst)$ pb. These numbers should be compared to the latest theoretical prediction of $6.7^{+0.71}_{-0.88}$ pb [23].

The cross section results are consistent with a large range of top masses. The current result is dominated by systematic uncertainties, somewhat larger than the 15% tolerance on $\sigma_{t\bar{t}}$ calculated from theory. Future improvements in the jet energy calibration at CDF and in the MC simulation of the W+jets backgrounds are expected to further reduce systematic uncertainties, while much larger data samples will improve the statistical precision.

BIBLIOGRAPHY

- [1] S. Glashow, Nucl. Pys. **22**, 579 (1961).
- [2] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).
- [3] A. Salam, *Proc. 8th Nobel Symp., W. Svartholm*, Almquist and Wiksell, Stockholm 1968.
- [4] D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973).
- [5] H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).
- [6] *Particle Data Book*, Phys. Rev. D **66**, 435 (2002).
- [7] *Particle Data Book*, Phys. Rev. D **66**, 383 (2002).
- [8] G. Kane, *Modern Elementary Particle Physics*, Addison-Wesley (1987).
- [9] B. Stugu, Nucl. Phys. Proc. Suppl. **76**, 123-132 (1999).
- [10] S. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D **2**, 1585 (1970).
- [11] C. Campagnari, M. Franklin, Rev. Mod. Phys. **69**, 137 (1997).
- [12] *Particle Data Book*, Phys. Rev. D **66**, 113 (2002).
- [13] F. Abe et al. (CDF Collaboration), Phys. Rev. Lett. **74**, 2626 (1995).
- [14] S. Abachi et al. (DØ Collaboration), Phys. Rev. Lett. **74**, 2632 (1995).
- [15] ALEPH, DELPHI, L3 and OPAL Collaborations, CERN-EP/2003-011 (2003).
- [16] R. D. McKeown, P. Vogel, Phys. Rep. **394**, 315 (2004).
- [17] *Particle Data Book*, Phys. Rev. D **66**, 173 (2002).
- [18] E. Laenen, J. Smith, W.L. van Neerven, arXiv: hep-ph/9310233.
- [19] E. L. Berger, H. Contopanagos, arXiv: hep-ph/9507363.

- [20] S. Catani, M. L. Mangano, P. Nason, L. Trentadue, arXiv: hep-ph/9602208.
- [21] R. Bonciani, S. Catani, M. L. Mangano, P. Nason, arXiv: hep-ph/9801375.
- [22] G. Altareli, Phys. Rep. **81**, 1 (1982).
- [23] M. Cacciari et al, arXiv: hep-ph/0303085.
- [24] A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, arXiv: hep-ph/0211080.
- [25] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. Nadolsky and W. K. Tung, JHEP **0207**, 012 (2002), (arXiv: hep-ph/9912439).
- [26] N. Kidonakis and R. Vogt, Phys. Rev. D **68**, 114014 (2003).
- [27] CDF Collaboration, F. Abe et al., Phys. Rev. Lett. **74**, 2626 (1995),
DØ Collaboration, S. Abachi et al., Phys. Rev. Lett. **74**, 2632 (1995).
- [28] A. Sill et al., Nucl. Instrum. Meth. **A 447**, 1 (2000),
A. Affolder et al., Nucl. Instrum. Meth. **A 453**, 84 (2000),
C.S. Hill, Nucl. Instrum. Meth. **A 530**, 1 (2004).
- [29] J. Thom, D. Glenzinski, M. Herndon, C. J. Lin, A. Yagil, *Determination of the Run II COT Tracking Efficiency using the W-No-Track Sample*: CDF Internal Note 6866.
- [30] T. Affolder et al., Nucl. Instrum. Meth. **A 526**, 249 (2004)
- [31] E. J. Thomson et al., IEEE Trans. Nucl. Sci. **49**, 1063 (2002).
- [32] W. Ashamanskas et al., Nucl. Instrum. Meth. **A 447**, 218 (2000).
- [33] L. Balka et al., NIMPR **A 267**, 272 (1988).
- [34] G. Apollinari, P. de Barbo and M. Mishina, *Proc. of the 4th Int. Conf. on Calorimetry in High Energy Physics*, World Scientific (1994) 200-225.
- [35] S. Kuhlmann et al., Nucl. Instrum. Meth. **A 518**,39 (2004).
- [36] K. Byrum, et. al., NIM **A 268**, 46 (1988),
G. Ascoli, et. al., NIM **A 268**, 33 (1988),
T. Dorigo et al., Nucl. Instrum. Meth., **A 461**, 560 (2001).
- [37] B. L. Winer, Int. J. Mod. Phys. **A**, 1169 (2001).
- [38] V. D. Barger, R. J. N. Phillips, *Collider Physics*, Addison-Wesley 1987.

- [39] B. Flaughner, J. Mueller, *A guide to JETCLU, The CDF jet cluster algorithm*: CDF Internal Note 1814.
- [40] G. Latino, Talk at the DPF-Meeting, Williamsburg May 24, 2002.
- [41] K. Burkett, Joao Guimaraes da Costa, *Development of a CTVMFT-based primary vertex finder PrimeVtxFinder*: CDF Internal Note 6430.
- [42] D. Acosta et al., submitted to Phys. Rev. Lett. (2004).
- [43] D. Acosta et al., to be submitted to Phys. Rev. D (2004).
- [44] CDFSim Webpage:
http://www-cdf.fnal.gov/cdfsims/cdfsims_main.html.
- [45] T. Sjöstrand, Report No. CERN-TH-6488/92 (1992).
- [46] G. Marchesini et al., Comput. Phys. Commun. **67**, 465 (1992).
- [47] M. L. Mangano et al., arXiv: hep-ph/0206293.
- [48] C. Peterson et al., Phys. Rev. D **27**, 105 (1983).
- [49] P. Avery, K. Read and G. Trahern, Cornell Internal Note CSN-212 (1985).
- [50] E. R. Kandel, J. H. Schwartz, T. M. Jessell, *Principles of Neural Science*, Elsevier, New York (1991).
- [51] W. S. McCulloch and W. Pitts, *A Logical Calculus of Ideas Immanent in Nervous Activity*, Bulletin of Mathematical Biophysics **5**, 115-133, (1943).
- [52] J. Hertz, A. Krogh, R. G. Palmer, *Introduction to the theory of Neural Computation*, Addison-Wesley (1991).
- [53] J. J. Hopfield, Rev. Mod. Phys. **72**, 431 (1999).
- [54] B. D. Ripley, *Pattern recognition and Neural Networks*, Cambridge University Press (1996).
- [55] S. J. Roberts, W. Penny, Sensor Review, Vol. 17 No. 1, 64-70 (1997).
- [56] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Reading Addison Wesley (1989).
- [57] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, Science, 220, 671 (1983).
- [58] M. D. Richard and R. P. Lippmann, Neural Computation, 3(4), 461-483 (1991).

- [59] C. Peterson, T. Rönngvaldsson, L. Lönnblad, *JETNET 3.0 - A Versatile Artificial Neural Network Package*, CERN-TH-7135/95 (1995).
- [60] S. Lawrence, C. L. Giles, A. C. Tsoi, AAI-97, AAI Press, Menlo Park California, 540-545 (1997).
- [61] V. N. Vapnik, and A. Y. Chervonenkis, *Theory of Probability and its Applications* **16**, 264-280 (1971).
- [62] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer (1995).
- [63] N. Tishby, E. Levin, S. A. Solla, *Consistent inference of probabilities in layered networks: predictions & generalization*, Proc. Int. Joint Conf. Neural Networks, vol. II, 1989.
- [64] V. Kurkova, In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 501–502. MIT Press, Cambridge, Massachusetts, (1995).
- [65] A. B. Nobel and A. Dembo, *Statistics and Probability Letters*, vol.17, 169-172 (1993).
- [66] F. Maltoni, T. Stelzer, arXiv: hep-ph/0208156.
- [67] S. Catani et al., arXiv: hep-ph/0109231.
- [68] S. Mrenna and P. Richardson, arXiv: hep-ph/0312274.
- [69] J. Campbell, R. K. Ellis, arXiv: hep-ph/0202176.
- [70] W. T. Giele and E. W. N. Glover, arXiv: hep-ph/0402152.
- [71] C. Friberg, T. Sjostrand, arXiv: hep-ph/9906316.
- [72] S. Frixione, B. R. Webber, arXiv: hep-ph/0204244, se also The MC@NLO webpage: <http://www.hep.phy.cam.ac.uk/theory/webber/MCatNLO>
- [73] M. A. Dobbs et al., arXiv: hep-ph/0403045.
- [74] E. Halkiadakis, C. Hays, M. Tecchio and W. Yao, *A Conversion Removal Algorithm for the 2003 Winter Coferences*: CDF Internal Note 6250.
- [75] Using the total $p\bar{p}$ cross section value of 60.7 ± 2.4 mb, S. Klimenko, J. Konigsberg and T. M. Liss, FERMILAB-FN-0741 (2003).
- [76] K. Bloom, K. Ebina, E. Thomson, K. Yorita, *Event Selection and Detection Efficiency for Winter 2004 Lepton+Jets Analyses*: CDF Internal Note 6844.

- [77] Un-ki Yang, *ISR studies on the Drell-Yann for top pair production*, Presentation at April APS Meeting, Denver CO, May 1-5, 2004.
- [78] K. Bloom, J. Conway, R. Erbacher, T. Maruyama, M. Shochet, *High pT Lepton ID Efficiency Scale Factor Studies in 4.11.1*: CDF Internal Note 6858.
- [79] S. Miller, Joint Physics meeting March 19 and April 2, 2004.
http://cdfreh0.grid.umich.edu/~miller/pdf/pdf_acceptance.html
- [80] MINUIT Webpage:
<http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html>
- [81] C. Ciobanu, R. Hughes, P. Koehn, C. Neu, B. Winer, *A ROOT Interface to JETNET*: CDF Internal Note 5434.
- [82] CDF II Collaboration, *Measurement of the t - t bar Cross Section using Lepton + Jets Events with Secondary Vertex Tagging*, to be submitted to Phys. Rev. D.