A Neural Networks Search for Single Top Quark Production in CDF Run I Data

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Cătălin Irinel Ciobanu, B.S., M.S.

* * * * *

The Ohio State University

2002

Dissertation Committee:

Prof. Richard E. Hughes, Adviser

Prof. Brian L. Winer

Prof. Richard Furnstahl

Prof. Samir Mathur

Approved by

Adviser
Department of Physics

© Copyright by

Cătălin Irinel Ciobanu

2002

ABSTRACT

In the proton-antiproton collisions at the Fermilab Tevatron collider, individual top quarks are expected to be produced by an electroweak mechanism referred to as single top production. We present here a Neural Networks method of searching for single top quarks in the 106 pb^{-1} of data collected by the Collider Detector at Fermilab (CDF) from 1992-1996. The prospects for observing single top in the current run at the Tevatron are also discussed.

In searching for single top, improvement to signal to noise can be obtained by requiring a charged lepton and missing energy from the W decay, as well as a B-tagged jet. We conduct our search in the $W + 1, 2$, and 3 jets channels, where we impose additional selections to further reduce background contributions from QCD multijet (W+jets) processes and top pair production $t\bar{t}$ respectively. Specifically, in the $W + 1$ jet channel we require exactly one additional jet with $E_T > 8\ GeV$ and $|\eta| < 2.4$, while in the $W + 3$ jets channel we require no such additional jets.

After applying these selection cuts to Monte Carlo samples, we are able to estimate the Run I expected contributions as: 4.2 single top events, 43.3 ± 8.4 W+jets events, and 7.4 ± 2.2 $t\bar{t}$ events. The signal purity is therefore 8%, which makes the Neural Networks approach particularly suitable for the single top analysis. To distinguish between signal and background we select seven event variables with good separating power. These variables are: the transverse energies of the leading two jets, the lepton,

and the neutrino: E_T^{jet1}, E_T^{jet2}, E_T^{lepton}, $\not\!\!E_T$, the transverse momentum of the leading two-jet system P_T^{jj}, the total transverse energy in the event H_T, and the product $Q \times \eta$ between the primary lepton charge and the pseudorapidity of the highest-E_T non B-tagged jet. The Neural Networks used is a three-output perceptron with one hidden layer trained on Monte Carlo generated events. We use the Monte Carlo signal and backgrounds output templates to estimate the composition of the 64-event Run I dataset passing our above selection. By using a simple binned-likelihood function with Gaussian background constraints to fit the data, we obtain the following contributions:

$$n_{signal} = 23.6 \pm 7.7, \quad n_{QCD} = 36.2 \pm 6.2, \quad n_{t\bar{t}} = 7.6 \pm 2.0 \quad \text{(events)}$$

We note that the fit result is approximately 2.5 standard deviations away from the expected value of 4.2 signal events. We further use this result to extract a 95% C.L. upper limit on single top cross section of 24.4 pb, roughly 10 times higher than the Standard Model prediction of 2.43 pb. For Run II, we expect a single top cross section measurement at 4σ precision level with roughly 2 fb^{-1} of data.

Dedicated to my parents Marioara and Tudor

ACKNOWLEDGMENTS

My first thanks go to Richard Hughes and Brian Winer, my advisers, who are wonderful people, good friends, and first-class scientists. Thank you for always supporting me in so many ways.

Next, I thank our OSU team at Fermilab: Phil Koehn, Evelyn Thomson, Chris Neu and Jong-young Chung, for providing quick help whenever I needed it. Special thanks go to Carlos Sánchez, a true *genus*. His simple yet brilliant idea of introducing the basketball game in Smith 2025 has generated countless hours of joy to almost everyone in the lab.

I would also like to thank my CDF collaborators, especially Pierre Savard, Sarah Wolinski, and Luc Demortier for the help with the tools used in my analysis, and Rob Roser and Mark Kruse for their encouragement and support. Thanks to the accelerator staff for making Run I a success.

Finally, I thank Luisa and Robert for the love and joy we share everyday.

****

VITA

12 January 1974	Born – Caracal, Romania
June 1997	B.S. in Physics, University of Bucharest, Bucharest, Romania
September 1997–September 1998	University Fellow, The Ohio State University, Department of Physics, Columbus, OH
December 1998	M.S. in Physics, The Ohio State University, Columbus, OH
October 1998–December 2000	Graduate Research Associate, Department of Physics, The Ohio State University, Columbus, OH
January 2001–December 2001	Presidential Fellow, Department of Physics, The Ohio State University, Columbus, OH
January 2002–September 2002	Graduate Research Associate, Department of Physics, The Ohio State University, Columbus, OH

PUBLICATIONS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The CDF Collaboration, "Measurement of the decay amplitudes of $B^0 \to J/\psi\ K^{*0}$ and $B_S^0 \to J/\psi\ \phi$ decays", *Phys. Rev. Lett.* 85, 4668 (2000)

The CDF Collaboration, "Dijet production by double Pomeron exchange at the Fermilab Tevatron", *Phys. Rev. Lett.* 85, 4215 (2000)

The CDF Collaboration, "Direct measurement of the W boson width in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 85, 3347 (2000)

The CDF Collaboration, "Measurement of J/ψ and ψ_{2S} polarization in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 85, 2886 (2000)

The CDF Collaboration, "Search for new particles decaying to top anti-top in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 85, 2062 (2000)

The CDF Collaboration, "Search for second and third generation leptoquarks including production via technicolor interactions in $p\bar{p}$ collisions $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 85, 2056 (2000)

The CDF Collaboration, "Limits on light gravitino production and new processes with large missing transverse energy in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 85, 1378 (2000)

The CDF Collaboration, "Search for scalar top and scalar bottom quarks in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 84, 5704 (2000)

The CDF Collaboration, "Search for scalar top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 84, 5273 (2000)

The CDF Collaboration, "Diffractive dijets with a leading anti-proton in $p\bar{p}$ collisions at $\sqrt{s} = 1800\ GeV$", *Phys. Rev. Lett.* 84, 5043 (2000)

The CDF Collaboration, "Production of $\Upsilon(1S)$ Mesons from χ_b decays in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 84, 2094 (2000)

The CDF Collaboration, "Measurement of b-quark fragmentation fractions in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 84, 1663 (2000)

The CDF Collaboration, "Search for color singlet technicolor particles in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 84, 1110 (2000)

The CDF Collaboration, "Search for a fourth-generation quark more massive than the Z^0 boson in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 84, 835 (2000)

The CDF Collaboration, "Transverse momentum and total cross section of e^+e^- pairs in the Z-boson region from $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. Lett.* 84, 845 (2000)

The CDF Collaboration, "Measurement of the helicity of W bosons in top quark decays", *Phys. Rev. Lett.* 84, 216 (2000)

The CDF Collaboration, "Observation of diffractive b-quark production at the Fermilab Tevatron", *Phys. Rev. Lett.* 84, 232 (2000)

The CDF Collaboration, "Search for the flavor-changing neutral current decays $B^+ \to \mu^+\mu^-K^+$ and $B^0 \to \mu^+\mu^-K^{*0}$", *Phys. Rev. Lett.* 83, 3378 (1999)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The CDF Collaboration, "Search for the charged Higgs boson in the decays of top quark pairs in the $e\tau$ and $\mu\tau$ channels at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 62, 012004 (2000)

The CDF Collaboration, "Measurement of the differential dijet mass cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$", *Phys. Rev. D* 61, 091101 (2000)

The CDF Collaboration, "Measurement of $\sin 2\beta$ from $B \to J/\psi K_S^0$ with the CDF detector", *Phys. Rev. D* 61, 072005 (2000)

The CDF Collaboration, "Measurement of the B^0 $\bar{B}^0$ oscillation frequency using $l^- D^{*+}$ pairs and lepton flavor tags", *Phys. Rev. D* 60, 112004 (1999)

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

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

C. Ciobanu, "Neural networks techniques for a Run II search of single top production at CDF", *Int. J. Mod. Phys. A*, Vol. 16, Suppl. 1A, 389 (2001)

FIELDS OF STUDY

Major Field: Physics

TABLE OF CONTENTS

Page

Abstract ii

Dedication iv

Acknowledgments v

Vita vi

List of Tables xvi

List of Figures xviii

Chapters:

1. Introduction 1

   1.1 The Standard Model 1
   1.2 Overview of the Analysis 4

2. Theoretical Perspective 7

   2.1 The W^* Channel. 8
   2.2 The W-gluon Channel. 10
   2.3 The Implications of Measuring Single Top 13

3. The CDF Experiment at the Tevatron 16

   3.1 The Tevatron 16
   3.2 The CDF Detector 18
      3.2.1 Tracking Detectors 19
      3.2.2 Calorimeters 23

3.2.3 Muon Systems 26
3.2.4 Event Triggers 28

4. Neural Networks 31

4.1 Introduction to Artificial Neural Networks 31
4.2 Multi-layered Perceptrons 34
4.3 Connection to Bayesian Statistics 37

5. Event Selection and Monte Carlo Samples 42

5.1 Initial Selection 42
5.1.1 Electron Identification 43
5.1.2 Muon Identification 45
5.1.3 Missing Transverse Energy $\not\!\!E_T$ 47
5.1.4 Z Boson Decays and $t\bar{t}$ Dilepton Events Removal 48
5.1.5 Jet Identification 49
5.2 Monte Carlo Simulations 52
5.2.1 Monte Carlo Acceptances and Run I Expectations 55

6. Measuring the Single Top Production Cross Section 60

6.1 Identifying Single Top with Neural Networks 60
6.1.1 Input Variables 61
6.1.2 Neural Network Output 68
6.1.3 Extracting the Signal by Fitting the Neural Network Output 70
6.2 A priori Study of Run I Single Top Cross Section 72
6.2.1 Setting the a priori Statistical Upper Limit 75
6.2.2 Folding the Systematic Uncertainties into the Confidence Limit 75
6.3 Single Top Cross Section Results from the Run I Data 83
6.3.1 A posteriori Limit on Single Top Cross Section 85
6.3.2 A posteriori Limit Interval on Single Top Cross Section . . . 88
6.3.3 Comparison with Other Single Top Analyses 89
6.4 Run II Prospects for Observing Single Top 92
6.5 Summary of the Results 94

7. Conclusions 95

Appendices:

A. Event Ratios in the Training Set 97

B. Neural Network Performance Indicators 100

C. Neural Network used for Single Top Identification 102

D. Calculating the Individual Confidence Limit μ_{95} 105

D.1 Standard Method 105
D.2 Alternate Method 106
D.2.1 Bayesian Approach to Setting a Limit Revisited 106
D.2.2 Alternate Method for Deriving the Likelihood 107
D.3 Equivalence between Methods 110

E. Neural Net Output for Run I Data 111

Bibliography 114

LIST OF TABLES

Table **Page**

3.1 Acceleration stages for protons and antiprotons. 17

3.2 Typical instantaneous, best instantaneous, and integrated luminosities during Run I at the Tevatron (integrated L corresponds to the CDF interaction region). 18

3.3 The physical properties for the central and endwall calorimeters. In the last row, the symbol $\oplus$ indicates that the constant term is to be added in quadrature to the resolution. 25

3.4 The physical properties for the plug and forward calorimeters. 26

5.1 Level 2 trigger requirements for primary muons. 46

5.2 Quality cuts designed to identify the secondary lepton from Z-decays. 48

5.3 Numbers of events passing the selections described in Sections 5.1 and 5.2.1. A total of N_{gen} events were generated with PYTHIA 6.129a for each Monte Carlo sample. 56

5.4 Run I expected numbers of W + 2 jets events passing the initial cuts of Section 5.1 (from [52]). 57

5.5 Predicted numbers of non-top background events in Run I data after the initial cuts of Section 5.1 have been imposed (from [51]). 57

5.6 N_{j8} cut efficiencies for signal and background. These numbers can be derived from the fourth and fifth rows below each process listed in Table 5.3. 58

5.7 Run I expectations for signal and backgrounds after the final selection. 59

6.1 Averaged performance P and error E for the best combinations of N variables. The last column lists the number of hidden nodes H for each case. 62

6.2 Systematic uncertainties for effects modifying both event rates and NN output shapes. The uncertainties are expressed as fractions of the theoretical prediction of $\mu_{SM} = 4.2$ signal events. As explained in the text, we average the two Jet E_T Scale entries, and we use the larger Top Mass systematic (corresponding to $M_{top} = 180\ GeV$). 78

6.3 Systematic uncertainties for effects modifying event rates only. 81

6.4 Run I observed and expected numbers of events. 83

6.5 Upper limits on single top cross section from CDF [53] and DØ [76, 75] analyses of Run I data. Our result is listed as CDF-3. 91

6.6 W+2 jets signal and background events expected in Run IIA (first row) and Run IIB (second row). 92

6.7 The most important results of our single top analysis. 94

C.1 Connection weights between the input and the hidden layers. 102

C.2 Connection weights between the output and the hidden layers (left table), and thresholds for the output and hidden nodes (right table). . 103

C.3 NN classification rates for Monte Carlo signal and background events. 104

D.1 Notation equivalence between Section 4.3 and Section D.2.1. 106

E.1 Run I W+1 jet events passing the final selection of Chapter 5. 111

E.2 Run I W+2 jet events passing the final selection of Chapter 5. 112

E.3 Run I W+3 jet events passing the final selection of Chapter 5. 113

LIST OF FIGURES

Figure **Page**

2.1 Single top production at the Tevatron: (a) LO diagram for W^* single top, (b)-(c) NLO corrections to W^*, (d) LO W-gluon single top, (e)-(f) NLO corrections of order $1/\ln(m_t^2/m_b^2)$, (g)-(k) NLO corrections of order α_s. Note that for the W^* channel (a)-(c) the W boson has $q^2 > (m_t + m_b)^2$, while for the W-gluon channel (d)-(k) the W has $q^2 < 0$. 11

3.1 The accelerator complex at Fermilab. 17

3.2 A quarter view of the CDF showing the locations of the different detector subcomponents. 20

3.3 Schematic view of an SVX barrel. 21

3.4 Transverse view the Central Tracking Chamber (CTC) which shows the alternating superlayers of axial and stereo cells. 22

3.5 A central calorimeter wedge. There are 24 wedges covering $\Delta\phi = 15°$ on either side of the transverse plane $z = 0$. 24

3.6 The location of the CMU system. Inside each calorimeter wedge there is a 2.4° ϕ interval not covered by the CMU. 28

4.1 A single nerve cell, or *neuron*, photographed by scanning electron microscopy. The inset shows a microscopy image of a biological neural network. 32

4.2 McCulloch-Pitts model: the connections between neurons are associated real numbers, or "weights", which are adjusted to implement a given task. (a) An artificial network of 5 neurons with 4 input terminals. In our notation, this is a 4-3-2 two-layered *perceptron*. (b) The McCulloch-Pitts neuron − seen as a binary threshold unit, where the weighted inputs are summed; the output is 1 if this sum exceeds a threshold μ_i, or 0 if not. 33

4.3 The two statistical approaches to classification. This example assumes three distinct classes of one-dimensional patterns. In both graphs, the horizontal axis represents the variable x that specifies the input patterns. As the three PDF's overlap, the classification can only be done in terms of probabilities. 39

5.1 Schematic view of an event containing a secondary vertex shown in the transverse $(r - \phi)$ view. The primary vertex is the point where the $p\bar{p}$ interaction occurs. The lines denoted 1, 2, and 3 are charged-particle tracks reconstructed in the SVX which yield the secondary vertex. The impact parameter for track 1 is denoted by d_1. L_{xy} is the 2-dimensional decay distance to the secondary vertex measured in the $r - \phi$ plane (adapted from [47]). 50

5.2 Representative Feynman diagrams for the main backgrounds to single top production. The $t\bar{t}$ background (a-d) dominates in the region of high energy and high jet multiplicity. At the other end of the spectrum (low energy and low jet multiplicity), the $Wb\bar{b}$ (e), $Wc\bar{c}$, and Wc (f) represent the most significant background. 53

5.3 N_{j8} versus N_j distribution for signal and background events passing the initial selection of Section 5.1. We note that most of the signal resides in the $N_j = 2$ channel. To suppress the non-top (QCD) background we require $N_{j8} = 1$ for events with $N_j = 1$. We reduce the $t\bar{t}$ background by imposing $N_{j8} = 0$ for events with $N_j = 3$ (see Section 5.2.1). . . . 54

6.1 Performance variation as training evolves (left). Learning saturation begins after epoch 100, and the net stabilizes after approximately 500 epochs. In estimating the "average" performance, we discard the first 100 epochs. The 400 P values corresponding to epochs $100 - 500$ are histogrammed (right). The average performance is 59.4%, while the RMS of 0.1% measures the variation of P from epoch to epoch. . . . 63

6.2 Averaged performance and error for the best combinations of N variables as functions of the number of hidden nodes H. See Table 6.1 for the list of the input variables for each value of N. 64

6.3 Monte Carlo distributions for the input variables. Left plot shows the two signal processes W^* (open) and W-gluon (shaded). Right plot shows the QCD background (shaded), signal mixed according to Table 5.7 (open), and $t\bar{t}$ background (lighter shade). For comparison purposes, all histograms are normalized to unit area. 65

6.4 Monte Carlo distributions for the input variables. Left plots show the two signal processes W^* (open) and W-gluon (shaded). Right plots show the QCD background (shaded), signal mixed according to Table 5.7 (open), and $t\bar{t}$ background (lighter shade). For comparison purposes, all histograms are normalized to unit area (*cont'd*). 66

6.5 Monte Carlo distributions for the input variables. Left plots show the two signal processes W^* (open) and W-gluon (shaded). Right plots show the QCD background (shaded), signal mixed according to Table 5.7 (open), and $t\bar{t}$ background (lighter shade). For comparison purposes, all histograms are normalized to unit area (*cont'd*). 67

6.6 Neural Network output (O_1, O_2, O_3) for Monte Carlo signal and background events. As required in the training, the QCD background events tend to cluster around the (1,0,0) point, while the $t\bar{t}$ events lie close to the (0,0,1) corner. The signal is more spreadout inside the output cube, with most of the events located in the region close to the (0,1,0) corner. 69

6.7 Output sum for Monte Carlo signal and background events. For each Monte Carlo sample, the corresponding $O_1 + O_2 + O_3$ distribution has a mean close to 1.0, and a small variance. This is a reflection of the normalization condition for the posteriori probabilities (see Section 4.2). 69

6.8 Graphical representation of the projection mapping $(O_1, O_2, O_3) \Rightarrow (x, y)$. We expect the dotted area to be little populated, as events in this region would have to have: $O_2 < O_1$, O_3. However, since our Network is mostly E_T-based, the signal should in general be situated in between the two backgrounds (Figures 6.3, 6.4, 6.5). Figure 6.9 shows that this is indeed the case. 71

6.9 Monte Carlo signal and background 2-dimensional output distributions resulted from projecting all output points onto the plane of equation $O_1 + O_2 + O_3 = 1$. 71

6.10 Results from 10000 pseudo-experiments. The 88 pseudo-experiments not shown have $n_{signal} < -23.0$, or $n_{QCD} < 2.0$ (events). 74

6.11 The distribution of the individual μ_{95}'s. The 95% C.L. limit is 16.65 signal events, or equivalently 9.63 pb. 76

6.12 The distribution of the individual μ_{95}'s after smearing the likelihood functions. The 95% C.L. limit is 18.14 signal events, or equivalently 10.50 pb. 83

6.13 Output sum for Run I data events ($left$). Run I data 2-dimensional distribution resulted from projecting the 64 output points onto the plane of equation $O_1 + O_2 + O_3 = 1$ ($right$). 84

6.14 Run I data likelihood function $\mathcal{L}(\mu_s|23.61)$. The arrow marks the 95% C.L. upper limit of $\mu_{95} = 38.25$ single top events (22.13 pb). 85

6.15 Jet E_T scale systematics as functions of μ_s. The circles correspond to a $+1\sigma$ increase in jets' transverse energies, the squares correspond to a -1σ decrease, while the triangles represent the average and define $\Delta S^{Jet\ E_T}$. 87

6.16 The Feldman-Cousins confidence belt. The confidence intervals $[\mu_1, \mu_2]$ are defined by the intersection of the belt with a vertical line drawn through the Run I fitted contribution of 23.61 signal events. The black, light grey, and dark grey horizontal lines represent the $[n_1, n_2]$ intervals corresponding to α = 95%, 90%, and 68.26% respectively. The circle with error bars located in the lower left corner represents the theoretically predicted value $\mu_s = 4.2$ events. 90

6.17 Fractional uncertainty in the single top cross section measurement versus integrated luminosity in Run II. A 3σ effect is expected to be observed with 1.2 fb^{-1} of data. No systematic uncertainties are included. 93

D.1 The likelihood function $\mathcal{L}(\mu_s|5.19)$. The corresponding confidence limit is $\mu_{95} = 17.22$ events. 109

D.2 The smeared likelihood function $\mathcal{L}_{smear}(\mu_s|5.19)$. The corresponding confidence limit is $\mu_{95} = 18.64$ events. 109

CHAPTER 1

INTRODUCTION

1.1 The Standard Model

What is the World made of? This is a fundamental question that scientists have tried to answer over the centuries. Tremendous progress into understanding the structure of matter has been made since the beginning of the last century, when the quantum theory of matter was founded. Today, particle physicists everywhere dedicate their efforts to studying the basic constituents of matter and the forces acting among them. There are four known types of interactions in the Universe: the strong, the electromagnetic, the weak, and the gravitational interaction. The Quantum Chromodynamics (QCD) describes the strong interaction, while the electromagnetic and the weak forces were unified in the framework of Electroweak Model [1, 2, 3] developed by Glashow, Weinberg, and Salam in the late 1960's. At microscopic level, the fundamental particles generate a gravitational field which is orders of magnitude weaker than its electromagnetic counterpart, and can therefore be neglected[1]. QCD and the Electroweak Model form the *Standard Model*, currently the best-established theory to describe the behavior of matter at fundamental level. In the Standard Model, the fundamental particles are fermions (spin $^1/_2$ particles) and bosons (spin 1 particles).

[1]For cosmological distance scales however, gravity is the most important force.

The fermions are grouped in three generations of quarks and leptons:

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} \quad quarks$$

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix} \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix} \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix} \quad leptons$$

The electron e, muon μ, and tau τ are massive, have negative electric charge (-1), and interact both electromagnetically with their antiparticle counterparts e^+, μ^+, τ^+, and weakly with their corresponding neutrinos ν_e, ν_μ, and ν_τ. Being electrically neutral, the neutrinos experience only weak interactions. Recent neutrino oscillation experiments indicate that the neutrinos posses a small rest mass [4]. The charge properties of quarks are different from those of leptons. First, quarks' electric charge has fractional value: $^2/_3$ for u, c, t, and $-^1/_3$ for d, s, b. Second, quarks also posses color charge red R, green G, or blue B. The quarks bind in colorless states (e.g. RGB, $\overline{RGB}$, $R\overline{R}$) to form the hadrons. It has been observed experimentally that mixing between quark generations occurs via weak interactions. Cross-generational coupling was first introduced by Cabbibo [5], and further developed by Glashow, Iliopulos, and Maiani [6] to fully explain the mixing between the first two generations. The extension to three generations is due to Kobayashi and Maskawa [7], who described quark mixing in terms of a 3×3 unitary matrix. This matrix is referred to as the Cabibbo-Kobayashi-Maskawa (CKM) matrix, and is defined by the following relation:

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

where the left (right) hand side column-matrix represents the eigenstates of the weak (strong) interactions. By convention, the lower elements of the quark doublets (*i.e.* d, s, b) were used in expressing the mixing. It is the V_{tb} element of the CKM matrix that we attempt to probe in the analysis to be presented in this document.

The Standard Model bosons are: the 8 gluons g which carry the strong force, the photon γ that mediates the electromagnetic interaction, and the W^+, W^-, and Z^0 which carry the charged and neutral weak forces. The Standard Model also predicts the existence of the Higgs boson, a spin 0 neutral particle which provides mass to the fundamental particles. The exchange of the Higgs boson would represent the fifth type of interaction in Nature. To date, no experimental evidence of the Higgs boson exists, although a significant excess consistent with the production of a 115 GeV/c^2 mass[2] Higgs boson has been recently reported by the ALEPH Collaboration [8].

We note that all existing data concerning elementary particles are consistent with the Standard Model, down to distance scales as small as 10^{-18} m. However, a major drawback of this theory is that it contains a number of arbitrary parameters that must be measured and put into the theory by hand. For this reason, many physicists believe that the Standard Model is a low energy approximation to a more fundamental theory, and that clues to this theory should appear at collider energies of several TeV's or more [9].

[2] A GeV is a unit for energy: 1 GeV = 0.001 TeV = $1.6 \cdot 10^{-10}$ J. Here c is the speed of light: c = $3 \cdot 10^8$ m/s. A GeV/c^2 is therefore a mass unit: 1 GeV/c^2 = $1.8 \cdot 10^{-27}$ kg.

1.2 Overview of the Analysis

This paper presents a search for electroweakly-produced single top quarks in the proton-antiproton collisions at the Tevatron. As already mentioned, this measurement can be used to extract one of the Standard Model parameters, namely the CKM matrix element V_{tb}, corresponding to the coupling between the top and the bottom quarks. In Chapter 2 we describe the two processes W^* and W-gluon that are expected to produce single top quarks at the Tevatron. We also discuss the implications of an experimental measurement of the single top cross section, both within and outside the Standard Model framework.

Chapter 3 presents a brief overview of the accelerator complex at Fermilab, and the Collider Detector at Fermilab (CDF) experimental setup.

Chapter 4 contains an introduction to the Artificial Neural Network (ANN) technique. In the beginning of this chapter we introduce the standard terminology and give a brief history of the ANN field of research. We then focus on a particular class of ANN's called *perceptrons*, and explain how these can be used in feature recognition problems. Finally, we pick a simple example to show how multi-class separation can be achieved using a multi-output perceptron, and illustrate the relation between perceptron's outputs and the Bayesian (posteriori) class probabilities.

The event selection for this analysis is detailed in Chapter 5. Single top identification starts from understanding its specific final state signatures. The final state of the W^* single top channel consists of one t quark and one b quark. The top quark decays into a W boson and a b quark almost exclusively [10], and therefore the W^* final state features two b quarks and the decay products of the W. Similarly, the W-gluon channel is characterized by two b quarks and the W decay products, plus

an additional light quark jet (u, d). As explained in Section 5.1.5, *jets* are sprays of hadrons produced by a fast-moving quark or gluon.

Consequently, we will be looking for events having one, two or three *good* jets, that is jets with high transverse energy $E_T > 15\ GeV$ and pseudorapidity[3] $|\eta| < 2.0$. We ask that at least one of the good jets have come from a B-hadron, as measured by the silicon tracking component of the CDF detector. A very distinctive signal is provided by the leptonic decays of the W boson. We accordingly demand the presence of a lepton and a neutrino[4] with a high transverse momentum $P_T > 20\ GeV/c$. The main backgrounds surviving these *cuts* are the QCD multijet W-production (W+jets), and top pair-production $t\bar{t}$. To reduce these backgrounds even further, we impose extra requirements on the number of *additional* jets in the event. These are jets having $E_T > 8\ GeV$, $|\eta| < 2.4$, and were not previously counted as *good* jets. Specifically, QCD rejection is significantly improved by requiring the presence of one additional jet in the $W + 1$ jet channel, while $t\bar{t}$ suppression is achieved by requiring no additional jets in the $W + 3$ jets events. After applying all these selection cuts to Monte Carlo-generated samples, we are able to estimate the Run I expected contributions as: 4.2 single top events, 43.3 W+jets events, and 7.4 $t\bar{t}$ events.

The low signal-to-background ratio (≈ 0.08) makes the ANN technique the ideal approach to single top recognition. ANN's allow us to combine the information from many different variables while naturally accounting for the correlations among these. Single top identification with ANN's is described in Chapter 6. We first present the

[3]Pseudorapidity is defined as: $\eta = -\ln \tan(\theta/2)$, where θ is the polar angle using the z vertex and the proton beam axis. For a detailed explanation of these quantities, see Chapter 3.

[4]The neutrino does not interact in the detector. Its presence is inferred from the amount of missing transverse energy in the event: $\overrightarrow{\not{E}_T} \equiv \overrightarrow{E_T}(\nu) = -\sum_i \overrightarrow{E_T}(i)$.

method we developed for choosing an appropriate input variable set. The best signal-background separation is obtained for the following set of input variables: E_T^{jet1}, E_T^{jet2}, E_T^{lepton}, $\not{E}_T$, H_T, P_T^{jj}, and $Q \times \eta$. Here, $jet1$ and $jet2$ are the highest-E_T jets in the event, H_T is the total transverse energy, jj refers to the $jet1 - jet2$ system, and $Q \times \eta$ is the product between the charge of the lepton and the pseudorapidity of the highest-E_T non-b jet. Next, we proceed to construct a perceptron whose task is to recognize the single top events in the data. To train the perceptron we use Monte Carlo-simulated signal and background events. Once the network is trained, we test the sensitivity of our ANN technique on fake-Run I experiments, or *pseudo-experiments*. A pseudo-experiment does not contain real data events; instead, Monte Carlo signal and background events are combined in the proportion that we *expect* to see in the Run I data. A measure of the sensitivity is the *a priori* 95% C.L. limit on single top production, which we find to be 10.5 *pb*, including the systematic effects. This result is 22% lower than the the limit obtained in the previous CDF search for the (W^*, W-gluon) combined single top signal, and 4.4 times higher than the theoretical cross section $\sigma_{single\ top} = 2.4\ pb$. Further, we give the *a posteriori* single top results obtained from the data collected during Run I. The ANN output fit to the data yields a signal content of 23.6 ± 7.7 events, which represents a 2.5 σ excess over the expected value of 4.2 signal events. We set an a posteriori upper limit on single top cross section of 24.4 *pb*. To close Chapter 6, we discuss the perspectives for observing single top in Run II at the Tevatron. The conclusions of this analysis are presented in Chapter 7.

CHAPTER 2

THEORETICAL PERSPECTIVE

The existence of the top quark was experimentally established in 1995 at Fermilab by the CDF [11] and DØ [12]. The top quark was observed in the strong production channel, or $t\bar{t}$, via quark-antiquark annihilation $q\bar{q} \to t\bar{t}$ and gluon-gluon fusion $gg \to t\bar{t}$. However, the Standard Model also allows for the top quark to be produced along with a b quark through electroweak interactions. This electroweak mechanism is referred to as *single top* production, because the final state features one top quark - as opposed to two in the $t\bar{t}$ case. There are three processes expected to generate single top quarks: (a) s-channel production $q\bar{q'} \to t\bar{b}$, usually called W^* production, (b) t-channel production $qg \to t\bar{b}q'$, called W-gluon fusion production, and (c) associated production with a W boson $bg \to tW$, known as the Wt production. The Wt process proceeds via gluon-b interaction, which makes the cross section negligible at the Tevatron[5] (approximately 7% of the total single top cross section). The representative Feynman diagrams for the W^* and W-gluon channels are shown in Figure 2.1(a, e). The next two sections present a summary of the theoretical calculations for W^* and W-gluon single top. The implications of an experimental measurement of the single top cross section are discussed in the third and final section of this chapter.

[5]This channel will account for about 30% of the single top cross section at the Large Hadron Collider at CERN.

2.1 The W^* Channel.

A high-energy proton beam may be usefully regarded as an unseparated, broad-band beam of quarks, antiquarks, and gluons [13]. The parton model is extremely accurate for hard scattering phenomena, where the momentum transfer is much larger than the QCD scale. The essence of the parton model is to regard the colliding proton as a collection of quasi-free partons which share its momentum. A proton of momentum P is envisaged as being made up of partons carrying longitudinal momenta $x_i P$, and negligible transverse momenta. In this model, the W^* process has a leading order cross section given by:

$$\sigma(p\bar{p} \to t\bar{b}+X) = \sum_{i,j} \int_0^1 \mathrm{d}x_1 \int_0^1 \mathrm{d}x_2 \, [q_i(x_1,\mu)\bar{q}_j(x_2,\mu)+\bar{q}_i(x_1,\mu)q_j(x_2,\mu)]\cdot\hat{\sigma}_{ij}(\hat{s}) \quad (2.1)$$

where $\hat{\sigma}_{ij}(\hat{s})$ refers to the W^* diagram shown in Figure 2.1(a), and has the following analytical expression [14]:

$$\hat{\sigma}_{ij}(\hat{s}) = \frac{\pi\alpha_W^2 |V_{ij}|^2 |V_{tb}|^2 \sqrt{q_0^2 - m_t^2}}{2(\hat{s} - m_W^2)^2} \left(q_0 - \frac{m_t^2 + 2q_0^2}{3\sqrt{\hat{s}}} \right), \quad q_0 \equiv \frac{\hat{s} + m_t^2 - m_b^2}{2\sqrt{\hat{s}}}. \quad (2.2)$$

The definitions of the quantities appearing in the previous two equations are:

- $q_i(x_1,\mu)\mathrm{d}x_1$ is the probability of finding the constituent quark i with momentum $x_1 P$ in the proton. The structure function $\bar{q}_j(x_2,\mu)$ has a similar interpretation in antiproton. The argument μ is the particular factorization scale where the structure functions are evaluated.

- $\hat{s}$ is the square of the center-of-mass energy of the (i, j) parton system, and is related to the $p\bar{p}$ center-of-mass energy S by $\hat{s} = x_1 x_2 S$.

- $\alpha_W = g^2/4\pi \equiv \sqrt{2}G_\mu m_W^2/\pi$. Equivalently, $\alpha_W = \alpha/\sin^2\theta_W$, where α is the fine structure constant, and θ_W is the weak mixing (Weinberg) angle.

- m_W, m_t, and m_b are the rest masses of the W-boson, top quark, and bottom quark respectively.

- V_{ij} and V_{tb} are the CKM matrix elements corresponding to the coupling between initial state quarks (i, j), and final state (t, b) respectively.

The integral in Eqn. (2.1) can be performed numerically using $m_t = 175\ GeV$, $\mu = M_{tb}$ (the mass of the top-bottom quark pair), CTEQ5L structure functions, and $\sqrt{S} = 1.8\ TeV$ to obtain $\sigma_{LO}(p\bar{p} \to t\bar{b} + X) = 0.49\ pb$ [15].

The next-to-leading-order corrections were evaluated by Smith and Willenbrock in [16]. The most important corrections to the $O(\alpha_W^2)$ leading order cross section are the QCD corrections of $O(\alpha_s)$, processes involving gluon radiation either in the initial or in the final state[6] (initial state gluon examples are given Figures 2.1(b,c)). By averaging the results found separately using two different sets of parton distribution functions (CTEQ3L,3M and MRS(A')), the authors find:

$$\sigma_{NLO}(p\bar{p} \to t\bar{b} + X) = 0.73 \pm 0.04\, pb, \quad \text{at } \sqrt{S} = 1.8\ TeV. \tag{2.3}$$

The systematic sources of uncertainty in the above σ_{NLO} are:

– the choices for the μ and μ_R, introducing an uncertainty of approximately 4%.

– the choice for a particular set of PDF's, together with the average taken in obtaining (2.3), which are assigned a 6% uncertainty.

[6] In [16] both the factorization scale μ and the renormalization scale μ_R for α_s are set to $\sqrt{q^2}$.

– the uncertainty in the top quark mass m_t taken to be 6 GeV, which is estimated to generate a 15% uncertainty.

Very importantly, we note that only the PDF uncertainty of 6% is quoted in the result (2.3). Adding all three systematic effects in quadrature yields:

$$\sigma_{NLO}(p\bar{p} \to t\bar{b} + X) = 0.73 \pm 0.12\ pb.$$

The higher order QCD corrections introduce a cross section uncertainty at the level of ±4%, included in the above result via the μ scale uncertainty.

2.2 The W-gluon Channel.

The W-gluon fusion channel is in some ways similar to the W^* single top. However, in the W^* process the excited W boson is timelike with $q^2 > (m_t + m_b)^2$, whereas the W-gluon fusion involves a spacelike W with $q^2 < 0$. These two production channels are therefore complementary since they probe the charged-current interaction of the top quark in different regions of q^2 [17]. We can interpret the W-gluon diagram shown in Fig. 2.1(d) as top quark production from a W boson fusing with a b quark which is in the proton sea, with the sea generated by gluons splitting into $b\bar{b}$ pairs [18, 19]. The W-gluon cross section calculation starts with this leading-order diagram, and using a b quark sea distribution function. Unlike the light-quark distribution functions, the b distribution function can be calculated perturbatively, as it involves energies of order m_b and larger; however, the gluon and light quark distribution functions are inputs to this calculation. The LO cross section of Figure 2.1(d) is given by [18]:

$$\frac{d\hat{\sigma}}{d\hat{t}} = \frac{\pi\alpha_W^2 |V_{ij}|^2 |V_{tb}|^2}{4} \frac{\hat{s} - m_t^2}{(\hat{s} - m_b^2)(\hat{t} - m_W^2)^2} \tag{2.4}$$

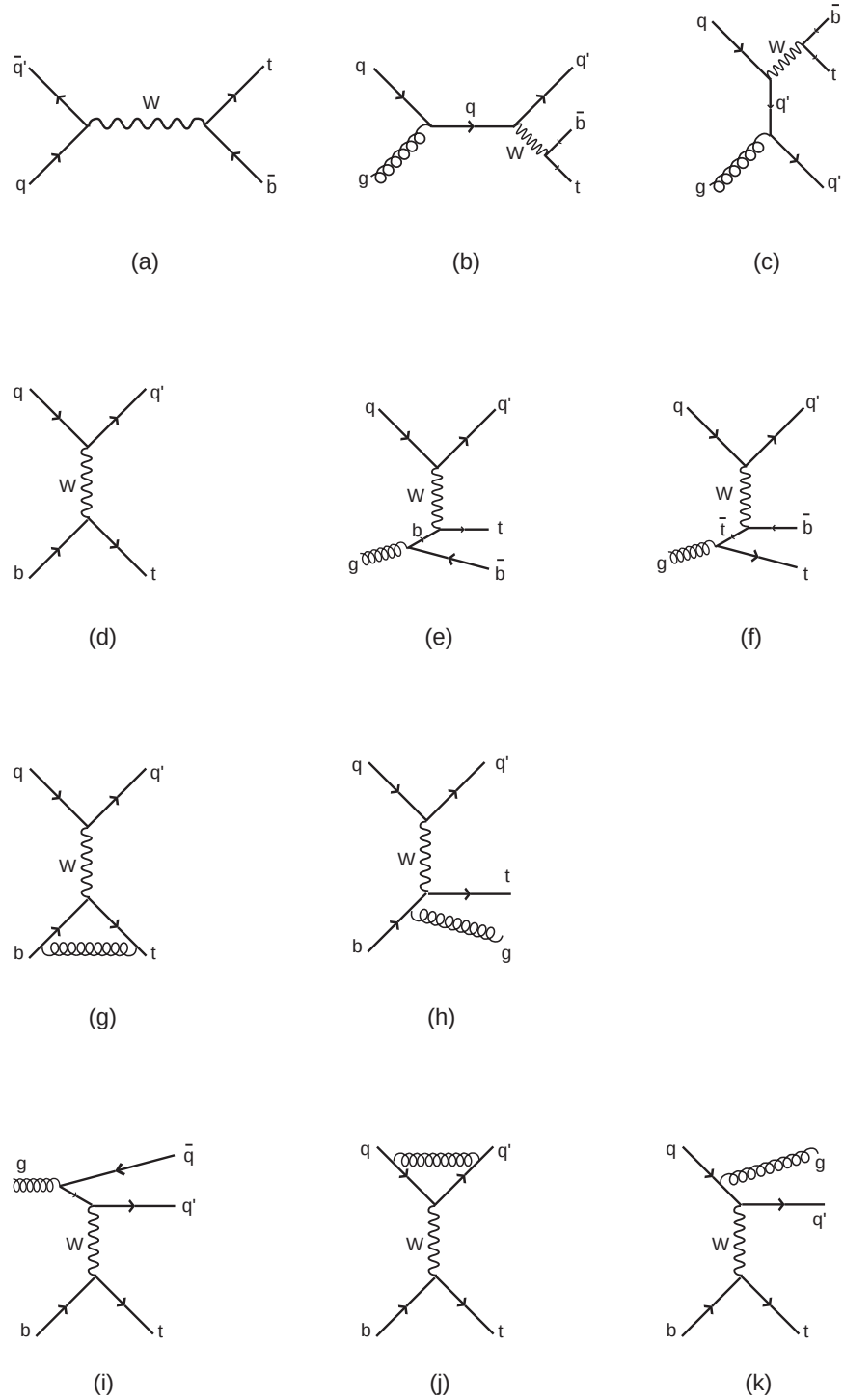

Figure 2.1: Single top production at the Tevatron: (a) LO diagram for W^* single top, (b)-(c) NLO corrections to W^*, (d) LO W-gluon single top, (e)-(f) NLO corrections of order $1/\ln(m_t^2/m_b^2)$, (g)-(k) NLO corrections of order α_s. Note that for the W^* channel (a)-(c) the W boson has $q^2 > (m_t + m_b)^2$, while for the W-gluon channel (d)-(k) the W has $q^2 < 0$.

for a quark-quark or antiquark-antiquark collision, and:

$$\frac{d\hat{\sigma}}{d\hat{t}} = \frac{\pi\alpha_W^2|V_{ij}|^2|V_{tb}|^2}{4}\frac{\hat{u}-m_b^2}{(\hat{s}-m_b^2)^2}\frac{\hat{u}-m_t^2}{(\hat{t}-m_W^2)^2} \tag{2.5}$$

for a quark-antiquark collision, where

$$\hat{t} = -\frac{\hat{s}}{2}\left[1-\frac{m_b^2}{\hat{s}}\right]\left[1-\frac{m_t^2}{\hat{s}}\right](1-z) \tag{2.6}$$

and $\hat{s}+\hat{t}+\hat{u} = m_t^2 + m_b^2$. In Eqn. (2.6) z is the cosine of the scattering angle in the parton-parton center-of-mass frame. To compute the leading-order cross section [19], the scales are chosen $\mu^2 = Q^2$ for the light-quark pdf's, and $\mu^2 = Q^2 + m_t^2$ for the b-quark pdf respectively, where $Q^2 = -q_W^2$. Calculations performed in the $\overline{MS}$ scheme using the CTEQ4M pdf set, $\sqrt{S} = 1.8\ TeV$, and $m_t = 175\ GeV$ yield: $\sigma_{LO}(p\bar{p} \to qt\bar{b} + X) \approx 1.87\ pb$.

As shown in reference [19], there are two independent next-to-leading-order corrections: of order $1/\ln(m_t^2/m_b^2)$, and of order α_s respectively. The diagrams corresponding to the two types of corrections are shown in Figures 2.1(e-f), and Figures 2.1(g-k) respectively. Using the same assumptions as for the leading-order calculation, the NLO result is [20]:

$$\sigma_{NLO}(p\bar{p} \to qt\bar{b} + X) = 1.70 \pm 0.09\ pb, \quad \text{at } \sqrt{S} = 1.8\ TeV. \tag{2.7}$$

In this expression, the 0.09 pb uncertainty is associated with varying the μ scale in the b pdf (5%). It does not incorporate systematics from the pdf (10%), top mass (9%), and the uncertainty associated with the gluon distribution function, which would reflect itself in an uncertainty in the b pdf. As in the previous section, we can add the systematic uncertainties in quadrature to obtain:

$$\sigma_{NLO}(p\bar{p} \to qt\bar{b} + X) = 1.70 \pm 0.24\ pb.$$

The effect of the higher order corrections is estimated to be $\pm 5\%$ of the NLO cross section, and was included in the previous σ_{NLO} result via the μ scale uncertainty.

2.3 The Implications of Measuring Single Top

An experimental measurement of the single top cross section is important for a number of reasons.

First, the single top *production* cross section[7] depends on the Cabbibo-Kobayashi-Maskawa (CKM) matrix element V_{tb}, one of the few parameters of the Standard Model not yet directly measured. Assuming three generations, the unitarity of the CKM matrix implies that $|V_{tb}|$ is close to unity [21]: $0.9990 < |V_{tb}| < 0.9993$ at 90% confidence level. In the first attempt to directly probe V_{tb}, the CDF collaboration [10] measured the ratio $|V_{tb}|^2/(|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2) = 0.94^{+0.31}_{-0.24}$. However, if there are more than three quark generations, $|V_{tb}|$ could in principle be any positive number less than unity. In this case, the experimental single top cross section coupled with an accurate theoretical calculation may provide the best direct estimate of $|V_{tb}|$ [17], independently of the assumption of the number of generations. Obtaining $|V_{tb}|$ in this manner presents both experimental and theoretical challenges. Experimentally, the main obstacle is isolating the signal from the large underlying background in order to accurately measure the single top rate. On the theory side, a key issue is the precise estimation of the gluon distribution function. Assuming that this is known to a reasonable degree of accuracy[8], then V_{tb} can be treated as a free parameter in the

[7] This is in contrast to the $t\bar{t}$ channel, where V_{tb} appears in the *decay* portion of the diagrams.

[8] Matching or exceeding the experimental accuracy.

NLO cross section calculation [9], and subsequently found by equating this result with the experimentally measured value.

Another important reason for studying single top is that this is a significant background to other processes, such as the Higgs boson production $q\bar{q}' \to WH$, with $H \to b\bar{b}$. This is the main decay channel for an intermediate mass Higgs: $90 < M_H < 130\ GeV$. Therefore understanding single top will be crucial for all future searches for the Higgs boson at the Tevatron or the LHC at CERN.

Beyond the Standard Model, single top could probe a variety of new phenomena, among which we mention the following:

(a) the existence of a new charged vector boson W' which would enhance the measured single top cross section. The W' boson is predicted by the left-right symmetric extensions of the Standard Model, where the the massive right-handed W' suppresses the $V + A$ charged current interaction $W' \to t\bar{b}$ [22].

(b) anomalous Wtb couplings [23, 24, 25]. As calculated in [24], such couplings can considerably enhance the Standard Model single top cross section at the Run II Tevatron.

(c) flavor changing neutral currents (FCNC) couplings $\bar{t}ug$ and $\bar{t}cg$. If these couplings exist, they would lead to a significantly larger single top cross section, as calculated in [26].

[9] The previous cross section results (2.3), (2.7) were calculated in the assumption that $V_{tb} = 1$.

(d) CP-violation effects within the Minimal Supersymmetric Standard Model (MSSM) [27]. Depending on the squark masses (particularly the stop) the W^* cross section can feature an $t\bar{b}$ - $\bar{t}b$ asymmetry of the order 1% at the Tevatron with 30 fb^{-1} of data.

(e) Kaluza-Klein excited W-boson within the MSSM. The Kaluza-Klein excitations can couple to the Standard Model quarks. As a result, the W^* single top cross section can be reduced by almost 25% for a mass of the W_{KK} of $\sim 1\ TeV$ [28].

To close this chapter, we summarize the most important aspects of a single top cross section measurement: within the Standard Model, it probes the V_{tb} coupling and impacts the future searches for the Higgs boson. Outside the Standard Model, this measurement has the potential to validate or invalidate several exotic phenomena such as the existence of heavy gauge bosons, flavor changing neutral currents, or supersymmetric extensions of the Standard Model.

CHAPTER 3

THE CDF EXPERIMENT AT THE TEVATRON

Fermi National Accelerator Laboratory (Fermilab), is a particle-physics experimental facility built in 1967 in Batavia, Illinois. Its largest component, the Tevatron, is currently the highest-energy accelerator in the world. The Tevatron will continue to dominate the arena of high-energy physics until 2007, when the Large Hadron Collider at CERN is scheduled to begin operations. Throughout this document, the main data taking periods at the Tevatron are referred to as *Run I* (1992-1995) and *Run II* (started 2001). The focus of the analysis presented here is the dataset accumulated during Run I by the Collider Detector at Fermilab (CDF) experiment. Consequently, in what follows we will describe the accelerator and the detector components as they were during Run I[10].

3.1 The Tevatron

The Tevatron is a proton-antiproton superconducting collider which for Run I ran at an energy level of $\sqrt{S} = 1.8\ TeV$. To attain this energy, the proton and antiproton beams go through the acceleration chain shown in Figure 3.1 and briefly described in Table 3.1. For protons, the acceleration process starts with the production of H^- ions

[10]For Run II, there have been a large number of improvements and additions to the experiment, as part of the upgrade following the 1995 shutdown.

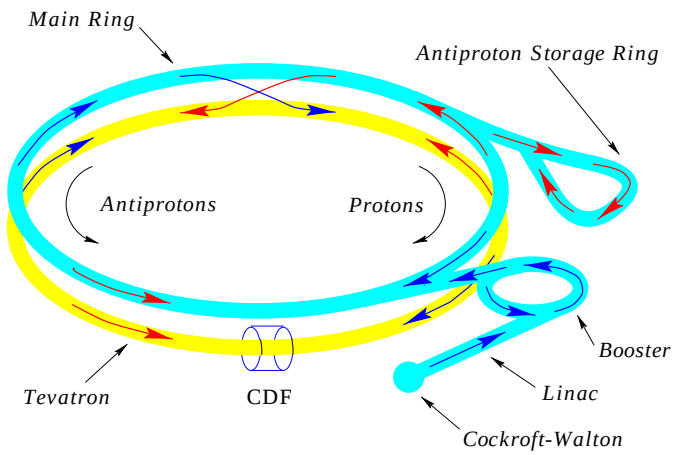

Figure 3.1: The accelerator complex at Fermilab.

Accelerator Component	Particle	Max. Energy	Approximate Size
Cockcroft-Walton	H^-	750 keV	few meters
Linac	H^-	200 MeV	$L = 150\ m$
Booster	p	8 GeV	$R = 75\ m$
Debuncher/Accumulator	$\bar{p}$	8 GeV	$\langle R \rangle = 90\ m$
Main Ring	p, $\bar{p}$	150 GeV	$R = 1\ km$
Tevatron	p, $\bar{p}$	900 GeV	$R = 1\ km$

Table 3.1: Acceleration stages for protons and antiprotons.

by ionizing Hydrogen gas. The H^- ions are first accelerated in a Cockcroft-Walton electrostatic accelerator followed by a linear accelerator (the Linac). The ions further pass through a carbon foil that strips off the electrons and permits the protons to pass through. From here, proton acceleration is done in three stages, corresponding to each of the three circular accelerators: Booster, Main Ring, and Tevatron. The protons are injected into the Tevatron as individual bunches. In Run I, the Tevatron

Period	Typical Instantaneous L	Best Instantaneous L	Integrated L
Run IA	$0.5 \times 10^{31}\ cm^{-2}\ s^{-1}$	$0.9 \times 10^{31}\ cm^{-2}\ s^{-1}$	$19.65\ pb^{-1}$
Run IB	$1.6 \times 10^{31}\ cm^{-2}\ s^{-1}$	$2.8 \times 10^{31}\ cm^{-2}\ s^{-1}$	$86.34\ pb^{-1}$

Table 3.2: Typical instantaneous, best instantaneous, and integrated luminosities during Run I at the Tevatron (integrated L corresponds to the CDF interaction region).

operated in a six-bunch regime, with the six proton bunches spaced evenly across the ring (3.5 μs temporal separation between successive bunches).

The antiprotons are made by colliding a proton beam from the Main Ring with a Tungsten target. Among the various particles created in this collision are the antiprotons, which are subsequently passed to the Debuncher and next to the Accumulator. When a large number of antiprotons ($\approx 10^{12}$) are collected in the Accumulator, they are grouped in a bunch to be injected first into the Main Ring and then into the Tevatron, similar to the protons but revolving in the opposite direction. Table 3.2 shows the various luminosity values during Run IA ($1992 - 1993$) and Run IB ($1994 - 1995$) respectively. Further information about the Tevatron at Fermilab can be found in reference [29].

3.2 The CDF Detector

The two detectors for $p\bar{p}$ collisions at Fermilab are CDF and DØ. Each of these detectors is located at one of the two high luminosity interaction regions (B0, and D0 respectively) along the Tevatron ring. CDF is a multi-purpose detector designed for good lepton (e, μ), photon, and jet identification. It is cylindrically symmetric, and also forward-backward symmetric with respect to the the transverse plane which

passes through its center. The CDF coordinate system defines the z direction along the direction of the proton beam, θ as the polar angle, and ϕ as the azimuthal angle. Instead of the polar angle θ, it is customary to use the pseudorapidity η expressed via the transformation: $\eta = -\ln[\tan(\theta/2)]$. CDF is composed of an array of individual detectors, each of which is designed for a specific type of measurement. According to their functionality, the various sub-detectors that make up CDF can be divided in three main categories: tracking chambers, calorimeters, and muon chambers. The tracking detectors are the closest to the interaction region. They are situated inside a 1.5 m radius superconducting solenoid which is co-axial with the beam pipe. The 1.4 T magnetic field generated by the solenoid causes a charged particle to follow a curved path, or *track*, whose radius can be used to infer particle's transverse momentum. Surrounding the solenoid are the calorimeters, which determine the energies of the outgoing particles. Finally, the outermost detectors are the muon chambers which measure the energies and positions of the muons that were not absorbed in the calorimeters. In the following section we will take a closer look at each of these three detector sub-systems. For the complete description of the detector, we point the reader to reference [30].

3.2.1 Tracking Detectors

There are three tracking components, all located inside the 1.4 T solenoid (Figure 3.2): the Silicon Vertex Detector (SVX), the Vertex Time Projection Chamber (VTX), and the Central Tracking Chamber (CTC).

The SVX surrounds the 1.9 cm radius beam pipe and lies within $2.86 < r < 7.87$ cm radial zone. It consists of two identical barrels located on either side of the

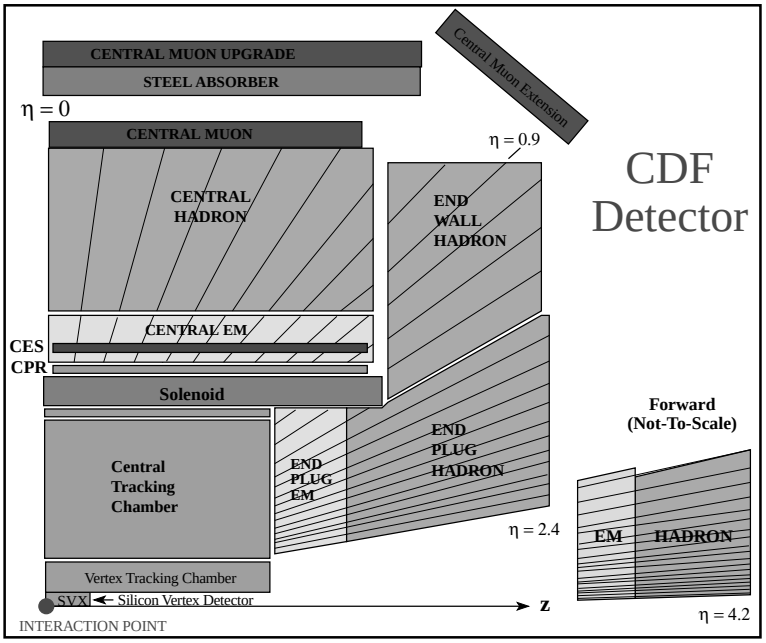

Figure 3.2: A quarter view of the CDF showing the locations of the different detector subcomponents.

$z = 0$ point, giving a total active length of 51 cm and a coverage of $|\eta| < 1.9$. The $p\bar{p}$ collisions occur within a range of roughly 60 cm along the z axis, which brings the SVX track acceptance to about 60%. Figure 3.3 shows one SVX barrel. In the azimuthal direction it features 12 wedges of 30° each, with the silicon strip detectors arranged in 4 concentric layers. In a given layer, the silicon strip detectors are bonded longitudinally in groups of three called *ladders*, for a total length of 25.5 cm. The total number of ladders is 96, and within each ladder the number of silicon strips is (depending on the layer): 256, 384, 512, 768 for layers 1 to 4 respectively. This gives a total of 46080 SVX channels which are read by 360 readout chips, or one chip per every 128 channels. Even though the read out is done is parallel and in sparse mode, the SVX readout time is still relatively long, about 2 ms per event. The role of the

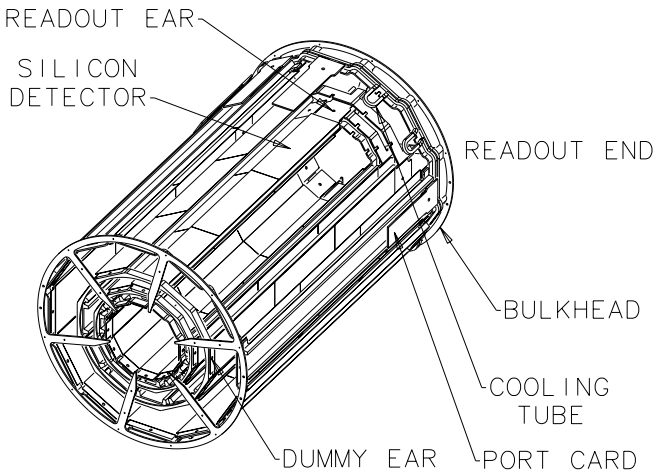

Figure 3.3: Schematic view of an SVX barrel.

SVX detector is to provide $r - \phi$ information (2-D tracking in the transverse plane) about the secondary vertices from b quark decays which have a mean decay length of $\approx 0.5\ mm$. The identification of b quark jets, or $B - tagging$, has played a major role in the discovery of the top quark [11] and impacts almost every aspect of top quark physics, including the search for single top presented here.

Surrounding the SVX is the VTX time-proportional drift chamber. In the radial direction it extends out to 21 cm, while longitudinally features 8 modules with a combined length of 2.8 m. Each module is segmented azimuthally into 8 wedges, and contains a 50%-50% mixture of argon-ethane gas. It has a central high voltage grid which divides the module into two 15.25 cm drift regions. The ionization electrons drift to the two endcaps containing sense wires running perpendicular to the

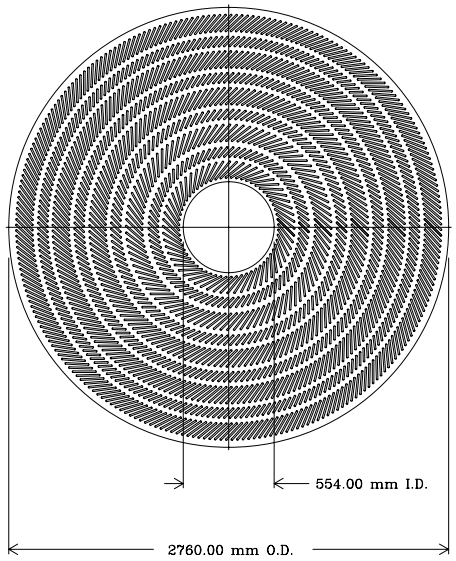

Figure 3.4: Transverse view the Central Tracking Chamber (CTC) which shows the alternating superlayers of axial and stereo cells.

longitudinal plane bisecting each wedge. The arrival times of the electrons give the z coordinate of the track, while the wire location specifies the particular r. The role of the VTX is to get the positions of the primary vertices by fitting the tracks in the $r - z$ plane. This information is subsequently used in the 3-dimensional CTC track reconstruction. In the multiple-interaction events, the VTX also helps establish which interaction a primary vertex and associated tracks belong to. The pseudorapidity coverage of the VTX is $|\eta| < 3.5$, and the primary vertex z resolution is $1 - 2$ mm. The last component of the tracking system is the CTC, which is a 3.2 m long cylindrical drift chamber whose inner and outer radii are 30.9, and 132.0 cm respectively. It covers a pseudorapidity range of $|\eta| < 1.0$. The CTC consists of 84 layers of sense wires which are grouped in 9 superlayers, numbered 0 to 8. There are two

types of superlayers, alternating in radial direction: *axial* superlayers (0,2,4,6, and 8) providing $r-\phi$ track information, and *stereo* superlayers (1,3,5, and 7) giving tracking information in the $r-z$ plane. Figure 3.4 presents a transverse view of the CTC, where each of the cells contains either 12 axial or 6 stereo wires and is tilted by 45° with respect to the radial direction. The tilt ensures that the electrons are drifting perpendicularly to the radial direction, which simplifies track reconstruction. The axial wires are strung parallel to the z direction, while the stereo wires are rotated by $\pm 3°$ relative to the beamline. The total number of sense wires in the chamber is 6156, and each wire has a design resolution of $\approx 200\mu m$ $(r-\phi)$, and 4 mm $(r-z)$. The CTC is filled with a gas mixture (argon/ethane/alcohol as 49.6% : 49.6% : 0.8%) in which the maximum drift time of the electrons to the wires is less than 800 ns. The track reconstruction in the CTC consists in fitting the hit wire pattern to a helix. The momentum of a charged particle traversing the CTC is determined from the reconstructed trajectory radius and has a resolution of: $\delta P_T/P_T = 0.002\ GeV^{-1}c \times P_T$. When information from the SVX is included in the reconstruction, the resolution becomes: $\delta P_T/P_T = 0.001\ GeV^{-1}c \times P_T$.

More detailed information about the SVX, VTX, and CTC, can be found in references [31], [32], and [33] respectively.

3.2.2 Calorimeters

The CDF calorimeters, located outside the solenoid, are segmented into towers (in $\eta-\phi$ space) which point back to the center of the detector (Figure 3.2). The *central* calorimeter extends over the range $|\eta| < 1.1$ and each of its towers corresponds to a solid angle $\Delta\phi = 15°$ and $\Delta\eta = 0.1$. The region $1.0 < |\eta| < 2.4$ is covered by the *plug*

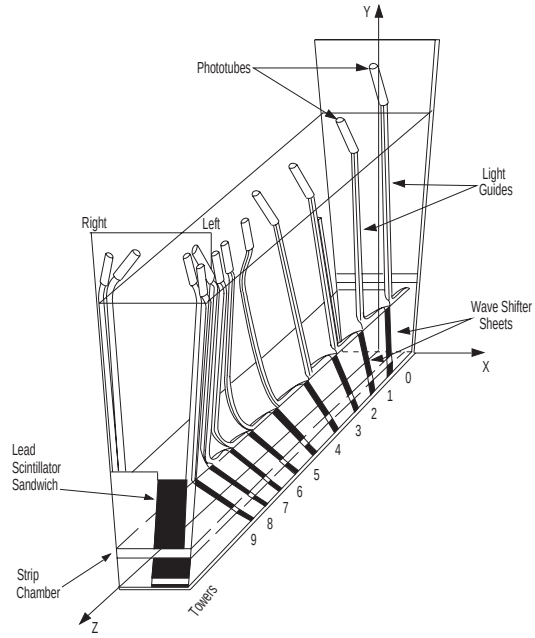

Figure 3.5: A central calorimeter wedge. There are 24 wedges covering $\Delta\phi = 15°$ on either side of the transverse plane $z = 0$.

calorimeters with a segmentation of $\Delta\phi = 5°$ and $\Delta\eta = 0.09$. Finally, the *forward* calorimeters match the pseudorapidity interval $2.4 < |\eta| < 4.2$, with a segmentation of $\Delta\phi = 5°$ and $\Delta\eta = 0.1$. We will describe each calorimetry system in what follows.

The central calorimeter system is composed of the central electromagnetic calorimeters (CEM), and central hadronic calorimeters (CHA). It is structured as 48 individual wedges each containing 10 towers. Figure 3.5 presents one of the central calorimeter wedges, where the lower and upper regions are occupied by the CEM and CHA calorimeters respectively. Towers 0 to 8 define the central region, and towers 6 to 8 share their hadronic portion with the endwall calorimeter (WHA) as shown in Figure 3.2. The relevant physical characteristics of the central calorimeters are summarized

	CEM	CHA	WHA
Coverage ($\|\eta\|$)	$0-1.1$	$0-0.9$	$0.7-1.3$
Tower Size ($\Delta\eta \times \Delta\phi$)	$0.1 \times 15°$	$0.1 \times 15°$	$0.1 \times 15°$
Module Length	250 cm	250 cm	100 cm
Module Width	$15°$	$15°$	80 cm
Number of Modules	48	48	48
Active Medium	polystyrene scintillator	acrylic scintillator	acrylic scintillator
Thickness	5 mm	10 mm	10 mm
Number of Layers	31	32	15
Absorber	Pb	Fe	Fe
Thickness	3 mm	25 mm	51 mm
Number of Layers	30	32	15
Energy Resolution ($\sigma(E)/E(GeV)$)	$13.7\%/\sqrt{E_T} \oplus 2\%$	$50\%/\sqrt{E_T} \oplus 3\%$	$75\%/\sqrt{E_T} \oplus 4\%$

Table 3.3: The physical properties for the central and endwall calorimeters. In the last row, the symbol $\oplus$ indicates that the constant term is to be added in quadrature to the resolution.

in Table 3.3. A detailed description of the CEM, CHA, and WHA calorimeters is given in references [34], [35]. The CEM is constructed of alternating layers of lead and scintillator, and has an effective thickness of 18 *radiation legths*. Similarly, the CHA is composed of alternating layers of iron and scintillator, with a thickness of 4.7 *attenuation lengths*. Besides CEM and CHA, each wedge contains a proportional strip chamber (CES) which measures the electromagnetic showers produced by the incident electrons or photons. The CES is positioned inside the CEM at roughly 6 radiation lengths - where the shower activity is expected to peak. Finally, a set of proportional tubes called CPR or the *central pre-radiator* are placed in the central region between the solenoid and CEM. The role of CPR is to detect the development of the early showers produced by electrons as they traverse the solenoid coil.

	PEM	PHA	FEM	FHA
Coverage ($\|\eta\|$)	$1.1 - 2.4$	$1.3 - 2.4$	$2.2 - 4.2$	$2.3 - 4.2$
Tower Size ($\Delta\eta \times \Delta\phi$)	$0.1 \times 5^\circ$	$0.1 \times 5^\circ$	$0.1 \times 5^\circ$	$0.1 \times 5^\circ$
Number of Layers	34	20	30	27
Active Medium	Proportional Tube Chambers with Cathode Pad Readout			
Tube Size	$7 \times 7\ mm^2$	$14 \times 8\ mm^2$	$10 \times 7\ mm^2$	$15 \times 10\ mm^2$
Absorber	Pb	Fe	96% Pb, 4% Sb	Fe
Thickness	$2.7\ mm$	$51\ mm$	$4.8\ mm$	$51\ mm$
Energy Resolution ($\sigma(E)/E(GeV)$)	$22\%/\sqrt{E_T} \oplus 2\%$	$90\%/\sqrt{E_T} \oplus 4\%$	$26\%/\sqrt{E_T} \oplus 2\%$	$137\%/\sqrt{E_T} \oplus 4\%$

Table 3.4: The physical properties for the plug and forward calorimeters.

In the reminder of this subsection we will very briefly review the plug and forward calorimeters. These are used to determine the energies of the forward jets (and indirectly to extract the missing energy $\not{E}_T$ in an event). As in the case of the central calorimetry, both the plug and the forward calorimeters feature electromagnetic and hadronic sections: PEM, FEM, and PHA, FHA respectively. These calorimeters use a 50%-50% argon-ethane gas mixture as their active medium. Proportional tube arrays containing this gas are interleaved with layers of lead and steel, as summarized in Table 3.4. PEM and FEM have thicknesses of 18-21 and 25 radiation lengths respectively, while their hadronic counterparts PHA and FHA have 5.7 and 7.7 attenuation lengths respectively.

3.2.3 Muon Systems

Most of the charged particles exiting the calorimeters are high energy muons ($P_T > 1.5\ GeV/c$). At CDF, muon detection is accomplished by using single wire, rectangular drift tubes. The muon system is composed of three separate muon detectors, all situated in the central pseudorapidity region: the central muon detector

(CMU), the central muon upgrade (CMP), and the central muon extension (CMX). The CMU [36] offers a coverage of $|\eta| < 0.6$, and is located inside the central calorimeter wedges radially outside of the CHA (Figure 3.6). The coverage in ϕ is about 85% of the full 2π range due to the $2.4°$ gaps between detector modules located in adjacent wedges. In each wedge the CMU is segmented into three towers, each consisting of 4 radial layers of 4 drift cells. In turn, each drift cell is filled with a mixture of 49.6% argon, 49.6% ethane, and 0.8% ethanol gas, and has a sense wire running parallel to the z direction (2260 mm). To resolve the ambiguity as to which side of the wires a given track has passed, the upper and lower layers are shifted relatively by 2 mm in the azimuthal direction. The $r - \phi$ track coordinates are inferred from the locations of the hit wires and the differences in the arrival times of the drift electrons. The track z information is obtained by comparing the pulse heights at each end of the sense wires. The CMU resolution is 250 μm in the transverse plane, and 1.2 mm longitudinally. To further reduce the number of hadrons outside the CMU, CDF uses a 60 cm - thick layer of steel provided by two side walls and the return magnet yoke. Behind this steel layer there is the CMP which consists of 4 layers of drift cells. CMP has essentially the same η range as the CMU, and covers most of the CMU gaps. The muon coverage was extended in the pseudorapidity range of $0.6 < |\eta| < 1.0$ with the addition of the CMX system. CMX is structured as four conical arches of drift tubes covering 71% of the solid angle in the $0.6 < |\eta| < 1.0$ interval. CMX inner and outer surfaces feature scintillator planes called CSX which provide timing information used in the CMX muon trigger as described in the following section.

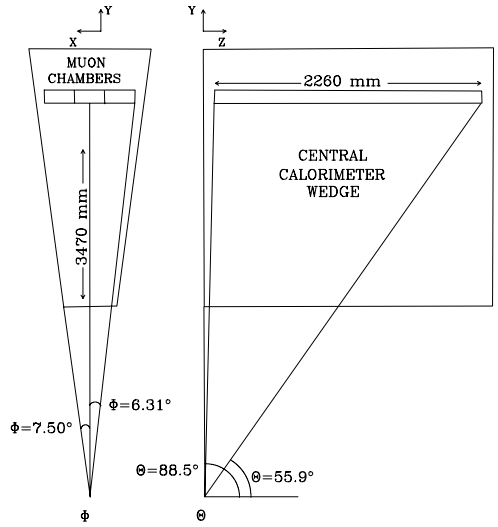

Figure 3.6: The location of the CMU system. Inside each calorimeter wedge there is a 2.4° ϕ interval not covered by the CMU.

3.2.4 Event Triggers

In the Run IB, the $p\bar{p}$ collisions were occurring at a rate of 280 kHz, which considerably exceeds the maximum event storage rate, or the rate of writing out the events to tape ($\sim$ 10 Hz). Consequently, CDF devised a three-level trigger system which rejects most of the events, while retaining only those which are potentially useful for various physics analyses. Events passing the requirements of the Level 1 trigger are passed to Level 2; the events passing Level 2 are allowed to propagate to Level 3 which impose further requirements before the events are ready to be written to tape. Typical acceptance rates for the three trigger levels are of the order 1:300, 1:50, and 1:5 for Level 1, Level 2, and Level 3 respectively. Level 1 and Level 2 triggers are implemented in hardware, while Level 3 is a software trigger.

The Level 1 does not have any dead time as the decision of accepting an event is made in less than 3.5 μs. It is based on identification of energy clusters in the calorimeters or muon tracks in the muon chambers. The Level 1 decision is a logical OR of the following calorimetry conditions: for inclusive electrons and jets, a single tower above a threshold of 8 GeV for CEM, 11 GeV for PEM, 12 GeV for CHA, 51 GeV for PHA, FEM, and FHA. These are further OR-ed with the muon conditions, which first require hits on two radially aligned sense wires in the CMU or CMX systems. The P_T of the muon is obtained by first measuring the slope of the track (given by the difference between the electron drift times) and then extrapolating back to the CTC to get track's deflection in the magnetic field. The events having muons with $P_T > 6\ GeV/c$ are accepted if the CMU track segment matches with hits in the CMP, or if the CMU hits are in time with the hadronic calorimeter TDC's (this rejects the cosmic muon events). The last muon trigger requirement demands a $P_T > 10\ GeV/c$ track segment in coincidence with the CSX response and the hadron TDC's.

The Level 2 decision time is approximately 20 μs during which further beam crossings are ignored by the detector (dead time of a few percent). The decision to accept events is made based on calorimeter clusters, high-P_T tracks in the CTC, and muon candidates. A hardware calorimeter cluster finder outputs a list of clusters by searching for a seed tower above a specific threshold accompanied by neighboring towers exceeding a second lower threshold. The process is repeated until no new seed towers are found, and for each cluster found, the E_T is calculated along with average values for η and ϕ. Track information is provided by the Central Fast Tracker [37], or CFT, a highly efficient hardware processor that uses hits to reconstruct high

momentum tracks in the transverse plane with a resolution of $\delta P_T/P_T^2 = 3.5\%$. To form a muon candidate at Level 2, the muon track segments in the CMU, CMP, and CMX that triggered Level 1 are required to have a matching CFT track. Similarly, the Level 2 central electron candidates are required to have energy clusters in the CEM with $E_T > 9\ GeV$ matched to CFT tracks with $P_T > 9.2\ GeV/c$ and a ratio of CHA/WHA to CEM cluster energies less than 0.125. The electromagnetic clusters are formed by demanding a seed CEM tower over 9 GeV and possible adjacent towers exceeding 7 GeV.

If the Level 2 trigger passes the event, the whole detector is read out, and the information is fed to the Level 3 which consists of FORTRAN 77 filter algorithms run on a farm of Silicon Graphics processors. The detector readout time is 3 ms which on average results in 10% dead-time. The trigger software reconstructs events using a simplified version of the $offline$ analysis code. Electrons and muons from Level 2 have their tracks fully reconstructed in 3 dimensions, and are required to match a reconstructed electromagnetic cluster or a muon stub. Once an event is accepted by Level 3, it will be written to 8 mm tape as part of a particular data $stream$.

CHAPTER 4

NEURAL NETWORKS

4.1 Introduction to Artificial Neural Networks

Artificial Neural Networks (ANN) are inspired from the structure of biological neural networks and their functioning [38, 39]. Though extremely primitive from a neurophysiological point of view, the ANN models are still valuable for gaining insight into how the human brain works. As we will see, just using a simple mathematical function to approximate the complex nerve cell allows us to reproduce the essential *collective* behavior of a large network of interacting cells.

The human brain contains approximately 10^{11} nerve cells, or *neurons*. Figure 4.1 shows an image of single neuron, and a network of such cells (inset). The *cell body* receives an input from the surrounding cells by means of electric signals (ions flow) arriving through the *dendrites*. If the input signals exceed a threshold value −the *resting* potential of the cell body− then an electric discharge takes place via the *axon* which is in turn connected to the dendrites of other cells. The discharge process is customarily referred to as "firing". In 1943, McCulloch and Pitts [40] developed the first mathematical approach for modeling the networks of neurons. In this model, a neuron is seen as a binary threshold unit, and its connections to the surrounding

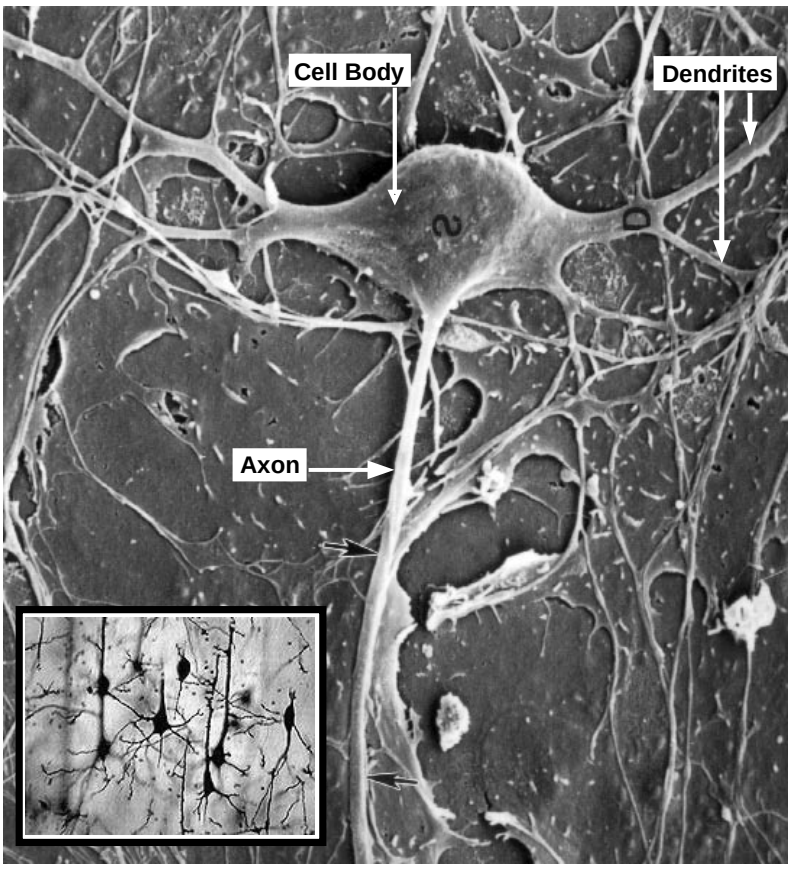

Figure 4.1: A single nerve cell, or *neuron*, photographed by scanning electron microscopy. The inset shows a microscopy image of a biological neural network.

neurons are represented by real numbers, called *weights*. Figure 4.2(b) presents a graphical picture of the McCulloch - Pitts model, where the idealized neuron i performs the following calculation:

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

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

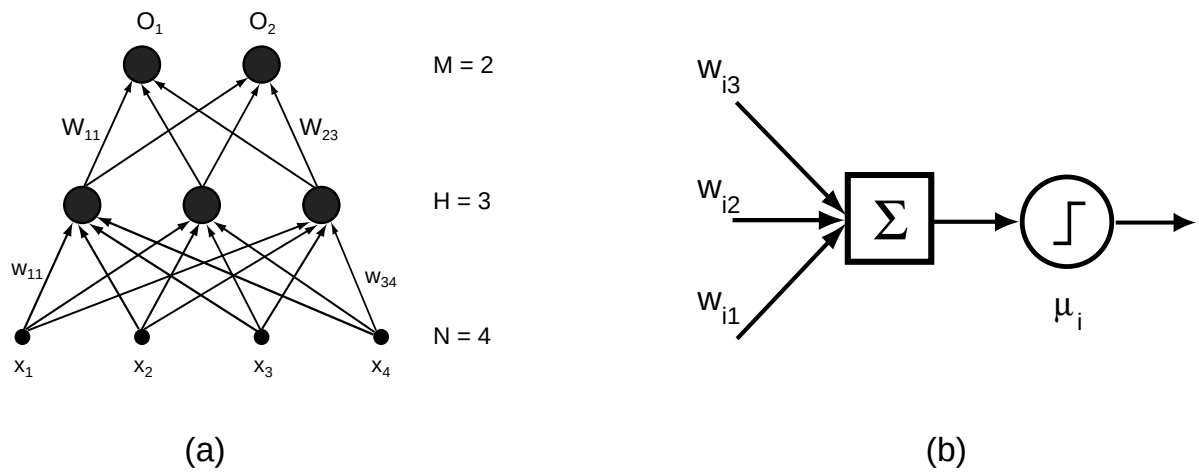

Figure 4.2: McCulloch-Pitts model: the connections between neurons are associated real numbers, or "weights", which are adjusted to implement a given task. (a) An artificial network of 5 neurons with 4 input terminals. In our notation, this is a 4-3-2 two-layered *perceptron*. (b) The McCulloch-Pitts neuron − seen as a binary threshold unit, where the weighted inputs are summed; the output is 1 if this sum exceeds a threshold μ_i, or 0 if not.

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

The following terms are conventionally used when referring to ANN modeling:

- **node, unit** for the neuron
- **architecture** for the configuration of nodes
- **activation function g(x)** for the unit calculation

In Eqn. (4.1) the activation function $g(x)$ was the Heaviside step function $\Theta(x)$. In general, continuous s-shaped, or "sigmoid" functions are used, such as the hyperbolic tangent $g(x) = \tanh(x)$. There are two different kinds of ANN architectures: feed-forward and feed-back. In a feed-forward network (see Figure 4.2(a)), signal

processing starts from a set of input terminals whose only role is to feed input information to the network. After this can come one or more intermediate layers of nodes[11], followed by a final output layer where the result of the calculation is read off. In contrast, feed-back networks process information bidirectionally, and the computation continues until a fixed point is reached, similar to the equilibrium state in a statistical mechanics system. In high-energy physics applications, feed-forward networks are used for pattern recognition problems such as signal/background separation, while feed-back architectures are used for optimization problems as in the case of track-finding algorithms.

4.2 Multi-layered Perceptrons

Neural networks learn from examples [41]. This statement is obviously true for the brain, but artificial networks also adapt their "synaptic" weights to a set of examples. In what is called the *learning phase*, the weights are changed iteratively until the network acquires some ability to generalize, to make predictions for input information not seen before. Learning in artificial systems has first been demonstrated in 1962 by Rosenblatt [42] who devised a special class of feed-forward networks called *perceptrons*. In a perceptron, the nodes are organized into layers and the unidimensional connections run only between nodes in consecutive layers. For example, Figure 4.2(a) presents a 4-3-2 two-layer[12] perceptron. In what follows we will illustrate how multi-layered perceptrons (MLP's) can be used in feature recognition (or pattern classification) situations.

[11]These are usually referred to as *hidden* nodes, because they have no connection to the outside world, neither input nor output.

[12]The bottom (input) layer is not counted because it contains no neurons. As mentioned in the text, its only role is to feed the input information to the network.

In a general classification problem, one wants to categorize the input data into classes based on a set of distinctive features. This task can always be cast in the form of asking a particular output pattern $\overrightarrow{T^k}$ in response to the kth input pattern $\overrightarrow{x^k}$. That is, one wants the *actual* output pattern $\overrightarrow{O^k}$ to be equal to the *target* pattern $\overrightarrow{T^k}$:

$$\overrightarrow{O^k} = \overrightarrow{T^k} \qquad \text{(desired)} \tag{4.2}$$

The dimension N of the input space gives the number of input terminals needed in the input layer. The number M of output nodes is directly related to the number of different classes (categories). As we will explain in the next section, it is particularly advantageous to use one output node for each data category. Given the notations of Figure 4.2(a), the output of the network $\overrightarrow{O^k}$ depends on the input pattern $\overrightarrow{x^k}$ as:

$$O_m^k = g\left(-\mu_m + \sum_{h=1}^{H} W_{mh} \cdot g\left(-\mu_h + \sum_{n=1}^{N} w_{hn} \cdot x_n^k\right)\right), \quad \text{for } m = 1, \cdots, M \tag{4.3}$$

In the above equation, O_m^k is the mth component of output pattern $\overrightarrow{O^k} = (O_1^k, \cdots, O_M^k)$; x_n^k is the nth component of input pattern $\overrightarrow{x^k} = (x_1^k, \cdots, x_N^k)$; μ_m, μ_h are the thresholds for the mth output node and h^{th} hidden node respectively; H is the total number of nodes in the hidden layer.

There are a total of $H \cdot (M+N) + H + M$ free parameters in Eqn. (4.3): the weight connections W_{mh}, w_{hn}, and the thresholds μ_m, μ_h. In adjusting these parameters, one needs to have a list or *training set* of correct input/output pairs as examples: $(\overrightarrow{x^k}, \overrightarrow{T^k})$ with $k = 1, \cdots, N_{train}$. To achieve the desired condition of Eqn. (4.2), one minimizes an error function, as for example:

$$E^k(\overrightarrow{W}, \overrightarrow{w}, \overrightarrow{\mu}) = \frac{1}{2}|\overrightarrow{T^k} - \overrightarrow{O^k}|^2 = \frac{1}{2}\sum_{m=1}^{M}(T_m^k - O_m^k)^2 \tag{4.4}$$

If $g(x)$ is a differentiable function (e.g. $\tanh(x)$), then $E^k(\overrightarrow{W}, \overrightarrow{w}, \overrightarrow{\mu})$ is also differentiable with respect to every weight and threshold, therefore we can use a gradient descent algorithm to find the appropriate parameters. Specifically, after feeding training pattern $\overrightarrow{x^k}$ to the net, we adjust the weights by:

$$\Delta W_{mh} = -\eta\frac{\partial E^k}{\partial W_{mh}}, \quad \text{and} \quad \Delta w_{hn} = -\eta\frac{\partial E^k}{\partial w_{hn}}. \tag{4.5}$$

The parameter η is called the *learning rate*. Similar equations can be written for the thresholds. The partial derivatives can be easily worked out starting from Eqn. (4.4) and Eqn. (4.3). In calculating these derivatives, one finds that the changes in the input-hidden weights Δw_{hn} depend on the hidden-output weight changes ΔW_{mh}. Therefore the errors appear to propagate backwards in the network, reason for which this algorithm is called $back - propagation$. One can apply this weight-updating procedure for each and all training examples $(\overrightarrow{x^k}, \overrightarrow{T^k})$ [13].

Besides training, the other important aspect of learning is the *generalization*. After exhausting the training set, the weights are "frozen", and the generalization performance of the network is tested on a separate set of N_{gen} patterns: $(\overrightarrow{x^j}, \overrightarrow{T^j})$, $j = 1, \cdots, N_{gen}$. The weights are not updated at this stage. The generalization performance is usually lower than the performance on the training set, although for very large data sets the two performances can be approximately equal. A good indicator of the performance is, for example, the mean error:

$$E_{train} = \frac{1}{N_{train}}\sum_{k=1}^{N_{train}} E^k, \quad \text{and} \quad E_{gen} = \frac{1}{N_{gen}}\sum_{j=1}^{N_{gen}} E^j \tag{4.6}$$

[13] We say that an *epoch* has passed after all training patterns were presented to the network once.

where E^k, E^j are defined by Eqn. (4.4). Obviously, a high performance implies a low error and vice-versa. If the generalization performance is unsatisfactory, one should execute the training procedure again. Depending on the complexity of the network and also on the task involved, there can be as many as a few thousand epochs until the weights converge to the optimal set of values.

The role of the generalization set $(\overrightarrow{x^j}, \overrightarrow{T^j})$ is to help us establish when to stop training. There is therefore a dependence of the networks' parameters on the generalization set, even if the latter was not used in updating the weights. For this reason, a third set of examples $(\overrightarrow{x^l}, \overrightarrow{T^l})$, independent of the training and generalization sets, is often used to gauge the expected performance of the net for data not seen before. This third set is called the *validation* set, and its performance is defined as in Eqn. (4.6).

4.3 Connection to Bayesian Statistics

In the previous section we have shown how neural networks can be used for classification. In order to understand the connection between perceptrons and Bayesian probabilities, it is useful to consider the problem of pattern classification in the larger statistical context.

The purpose of a pattern classifier is to assign every input pattern $(x_1, x_2, \cdots, x_N) \equiv \overrightarrow{x}$ to one of a small number M of discrete classes or groups $C_1, C_2, \cdots, C_M$ [43]. Each input pattern $\overrightarrow{x}$ is fed into a classifier which calculates M output numbers $(O_1, O_2, \cdots, O_M) \equiv \overrightarrow{O}$ and based on these values assigns the pattern to one of the M classes[14]. The two major statistical approaches to classification are to use *probability*

[14]We note again that this discussion applies but is not limited to neural networks.

distribution functions or PDF's, and *posteriori* or *Bayes probabilities*. These approaches differ in the statistical quantity their outputs model, and in the procedure used for training. We will illustrate both methods through the following example taken from [44]. In the conclusion of this chapter we will discuss the similarity between this example and our single top identification task.

The problem. Consider the simplest case $N = 1$, where an input pattern is defined by one real number x. Further, let us assume that we have $M = 3$ categories of patterns. Figure 4.3(A) shows the three probability distribution functions $\mathcal{P}(x|C_1)$, $\mathcal{P}(x|C_2)$, $\mathcal{P}(x|C_3)$ of x for classes C_1, C_2, C_3 respectively. All three PDF's are unit variance, gaussian distributions and differ only in their means. For a **given** class C_i, the probability that the input value is in the $[x,\, x + \mathrm{d}x]$ interval is:

$$\mathrm{d}P_i(x) = \mathcal{P}(x|C_i)\,\mathrm{d}x, \quad \text{with} \quad \int_{-\infty}^{\infty} \mathcal{P}(x|C_i)\,\mathrm{d}x = 1, \quad \text{for } i = 1, 2, 3. \tag{4.7}$$

As already stated however, we are interested in deriving class probabilities **given** a certain value of x[15]. This are called Bayes probabilities, and are denoted by $P(C_i|x)$. We point out that these are not probability densities, as can be seen from the normalization condition:

$$P(C_1|x) + P(C_2|x) + P(C_3|x) = 1, \quad \text{for all } x \in \mathbf{R}. \tag{4.8}$$

The solution. The PDF's $\mathcal{P}(x|C_i)$ and the posteriori probabilities $P(C_i|x)$ are related via the Bayes theorem:

$$P(C_i|x) \cdot \mathrm{d}P(x) = \mathcal{P}(x|C_i)\,\mathrm{d}x \cdot P(C_i), \quad \text{for } i = 1, 2, 3 \text{ and all } x \in \mathbf{R}. \tag{4.9}$$

[15]Since the three PDF's overlap, an input pattern x can not be assigned with 100% certainty to any one class. The classification task reduces therefore to calculating the *probabilities* that pattern x originated from class C_1, C_2, or C_3.

Here, $\mathrm{d}P(x) = \mathcal{P}(x)\,\mathrm{d}x$ is the unconditional probability that the input is situated in the $[x, x + \mathrm{d}x]$ interval, and $P(C_i)$ is the *a priori* or prior class probability of class C_i. More precisely, $P(C_i)$ is the probability that a random selection of a new input pattern will come from class C_i, and can be obtained by measuring the relative frequency of occurrence of patterns from different classes. Solving Eqn. (4.9) for $P(C_i|x)$, we obtain:

$$P(C_i|x) = \frac{\mathcal{P}(x|C_i) \cdot P(C_i)}{\mathcal{P}(x)} \tag{4.10}$$

We note that $\mathcal{P}(x)$ does not play any role in the classification process, because it

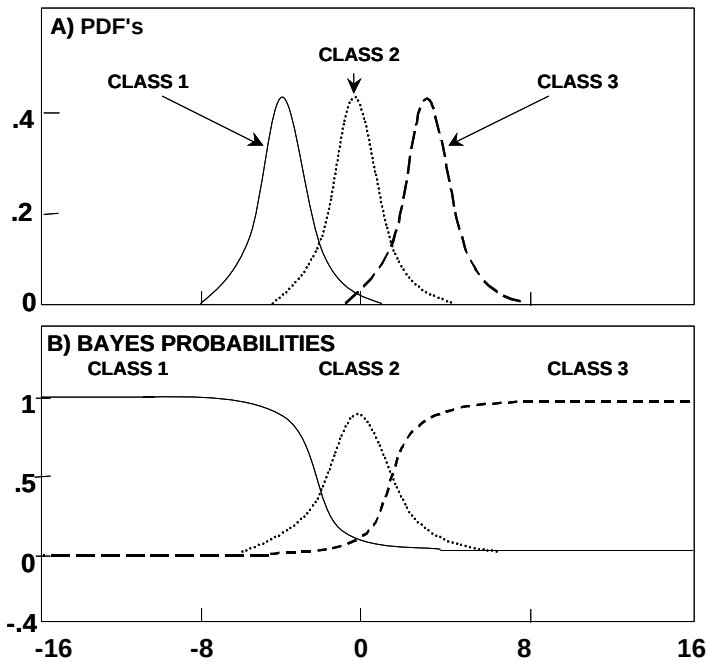

Figure 4.3: The two statistical approaches to classification. This example assumes three distinct classes of one-dimensional patterns. In both graphs, the horizontal axis represents the variable x that specifies the input patterns. As the three PDF's overlap, the classification can only be done in terms of probabilities.

is the same for the three classes. It is customary to treat $\mathcal{P}(x)$ as a constant factor, which we can calculate from the normalization condition (4.8). Based on expression (4.10), we can compute the three Bayesian $P(C_i|x)$ as function of the input x. These are plotted in Figure 4.3(B), where we assumed that the prior class probabilities are equal: $P(C_1) = P(C_2) = P(C_3) = 1/3$.

Practical implementation. Conventional PDF classifiers do not estimate posterior probabilities directly. Instead, they estimate the PDF's and the prior probabilities for each class, i.e. the right hand side of Eqn. (4.10). The most common PDF's used for PDF classifiers are gaussian distributions whose means and variances for each class are estimated from the training data using simple non-iterative algorithms. Examples of this type of classifiers are linear discriminants, as for example Fisher discriminants in the case $M = 2$.

In contrast, Bayesian classifiers estimate posterior probabilities directly, i.e. left hand side of Eqn. (4.10). Richard and Lippmann [44] have shown that perceptrons are such classifiers, when the network has one output for each class, target (desired) outputs are $1\ of\ M$ (the output corresponding to the correct class is 1, all other outputs are 0), and a squared-error function (4.4) is used. When posterior probabilities are estimated accurately, the M output node values will sum up to unity, and these outputs can be treated as probabilities.

Analogy to single top. This simple example exhibits the essential features of our single top search. As it will be described in the next chapter, we attempt to classify patterns (events) into three classes corresponding to the three physical processes: QCD, single top, and top pair production. The patterns are characterized by $N = 7$ real numbers, which are final state particles' energies, momenta, angles

(or combinations of these) in the event. The individual PDF's for these seven inputs have a significant overlap among the three classes, similar to Figure 4.3(A). We will use a 3-output perceptron, where in the training we turn on one output for each corresponding process. To train the network we use Monte Carlo generated events. Further, we assume no prior knowledge with respect to classes: $P(C_1) = P(C_2) = P(C_3) = 1/3$, therefore we train the network with equal number of events from each class[16]. The actual class probabilities, or process rates, are subsequently found by comparing (fitting) the output distribution of the entire Run I dataset to the individual Monte Carlo output distributions corresponding to each process. This is precisely how we extract the single top cross section from the Run I data.

[16]Appendix A gives a more detailed discussion on the *prior* class probabilities.

CHAPTER 5

EVENT SELECTION AND MONTE CARLO SAMPLES

The data studied for this analysis were collected with the CDF detector during Run I, and correspond to an integrated luminosity of $106.0 \pm 4.1\ pb^{-1}$. For both single top production channels $W-$gluon and W^*, the final state features a W boson and two b-quarks. To improve signal resolution, we select only the events in which the W decayed leptonically. We are therefore looking for events having a high transverse momentum lepton and a large amount of missing transverse energy $\not\!\!E_T$ from the neutrino. We require at least one high transverse energy jet from a B hadron, and reject dilepton events from Z boson or $t\bar{t}$ decays ($Z \to ll$ or $t\bar{t} \to ll\nu\nu jj$). After this initial selection, we use a jet counting variable N_{j8} to further reduce the backgrounds. The final section of this chapter discusses the Monte Carlo acceptances for signal and backgrounds, along with the expected Run I contributions.

5.1 Initial Selection

We present here the selection criteria for electrons, muons, missing energy, and jets. We do not explicitly consider τ leptons here, as their hadronic decay channel makes the identification considerably more difficult.

5.1.1 Electron Identification

Electrons deposit most of their energy in the electromagnetic calorimeter, and the shower is usually contained within a small cluster of towers (one or two towers in the CEM). In addition, an electron has a CTC track which points towards the cluster. At the Level 1 trigger stage, a CEM cluster is classified as an electron candidate if it has $E_T > 8\ GeV$. At Level 2, an electron candidate is defined by a CEM cluster with $E_T > 16\ GeV$ which matches a CTC track with $P_T > 12\ GeV/c$, or a CEM cluster with $E_T > 16\ GeV$ and at least 20 GeV of missing transverse energy. Additional cuts are applied at Level 3, which are looser versions of the final selection cuts given in the list below. The variables used to identify electrons, along with the specific cuts, are:

- E_T - transverse energy in the CEM cluster. Require: $E_T > 20\ GeV$.
- E/P - the ratio of electromagnetic calorimeter energy E to the track momentum P. Require: $E/P < 1.8$.
- HAD/EM - the ratio of the energy HAD deposited in the hadronic calorimeters to the energy EM deposited in the electromagnetic calorimeter (within the electron cluster). Require: $HAD/EM < 0.05$.
- L_{shr} - the lateral shower profile for electrons, which compares the energy in adjacent CEM towers to the seed tower for the cluster. L_{shr} is defined as:

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

  where the sum is over the towers adjacent to the seed tower, E_i^{obs} is the electromagnetic energy measured in tower i, E_i^{exp} is the energy expected from test

beam electrons, $0.14\sqrt{E}$ is the CEM energy resolution, and $\sigma_{E_i^{exp}}$ is the uncertainty in E_i^{exp}. We require: $L_{shr} < 0.2$.

- $|\Delta x|$ and $|\Delta z|$ - the separation between the extrapolated CTC track position and the CES chamber position in the $r-\phi$ plane, and $r-z$ plane respectively. Require: $|\Delta x| < 1.5\ cm$, and $|\Delta z| < 3.0\ cm$.

- χ^2_{strip} - the χ^2 resulted from the comparison of the CES shower profile in the z direction between the electron candidate and test beam electrons. We require: $\chi^2_{strip} < 10$.

- z-vertex match - the distance along the beam axis z between the primary vertex (interaction point) and the reconstructed track. If there is more than one vertex in the event, the distance to the closest vertex is used. The closest vertex is required to have $|z| < 60\ cm$ measured from the $z = 0$ point. We require: z-vertex match $< 5.0\ cm$

- Iso - electron isolation. $Iso = (E_T^{cone} - E_T)/E_T$, where E_T^{cone} is the sum of the transverse energies in towers within a cone of $R = 0.4$[17] around the track direction, and E_T is the electron cluster transverse energy. We require: $Iso < 0.1$.

- Fiducial cut on the electron. The electron is required to be in the fiducial volume, which covers 84% of the solid angle in the region $|\eta| < 1.0$. With this requirement we seek to avoid the regions where the detection response is not accurate (cracks, or inter-module spacing).

[17] R is defined by: $R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$

The combined efficiency for these cuts was measured using electrons from Z decays, and is approximately 82% without including the isolation cut. The process $Z \rightarrow e^+e^-$ was used because of its very distinct signatures; the electrons coming from W decays are expected to have similar kinematic characteristics. In addition to these requirements, an algorithm is applied to remove the photon conversion electrons [45], with an efficiency of 91%.

5.1.2 Muon Identification

Depending on the muon system subcomponent they traverse, the muons are classified as: (a) CMU-CMP muons is they are reconstructed in both CMU and CMP chambers, (b) CMU-only or CMP-only if the muon is reconstructed in only one of the CMU and CMP chambers, and (c) CMX muons if they pass through the CMX detector. The Level 1 muon trigger requires a pair of hits in two parallel muon drift tubes (within a certain time window). The hits in CMU, CMP, and CMX are then linked into track segments, or *stubs*. At Level 2, the muon candidates are defined by a CTC track with $P_T > 12\ GeV/c$ pointing to within 5° of a Level 1 muon stub. The Level 2 selection criteria are given in Table 5.1 [45]. The Level 2 triggers having a very high event rate are *prescaled*, which means that only 1 out of every n triggers is accepted. The Level 3 muon trigger runs full offline reconstruction of muon stubs and a simplified version of the offline tracking code. It requires that the distance $|\Delta x|$ between the extrapolated CTC track and the muon stub be less than 10 cm for CMU-only or CMU-CMP, 25 cm for CMX muons, and 40 cm for CMP-only muons. Finally, the following conditions are imposed for muon candidates passing the Level 3:

Triggers	Prescaled
CMU-only muons	
1) $\not\!\!E_T > 35\ GeV$ and two jets with $E_T > 3\ GeV$	NO
2) CTC track with $P_T > 12\ GeV/c$ matched to CMU stub	YES
3) CTC track with $P_T > 12\ GeV/c$ matched to CMU stub *and* one jet with $E_T > 15\ GeV$	NO
CMU-CMP muons	
1) $\not\!\!E_T > 35\ GeV$ and two jets with $E_T > 3\ GeV$	NO
2) CTC track with $P_T > 12\ GeV/c$ matched to CMU and CMP stubs	NO
3) CTC track with $P_T > 12\ GeV/c$ matched to CMU and CMP stubs *and* one jet with $E_T > 15\ GeV$	NO
CMP-only muons	
1) $\not\!\!E_T > 35\ GeV$ and two jets with $E_T > 3\ GeV$	NO
CMX muons	
1) $\not\!\!E_T > 35\ GeV$ and two jets with $E_T > 3\ GeV$	NO
2) CTC track with $P_T > 12\ GeV/c$ matched to CMX stub	YES
3) CTC track with $P_T > 12\ GeV/c$ matched to CMX stub *and* one jet with $E_T > 15\ GeV$	YES

Table 5.1: Level 2 trigger requirements for primary muons.

- P_T - transverse momentum of the muon track in the CTC. Require: $P_T > 20\ GeV/c$.

- EM, HAD - the energy deposition in the electromagnetic, and hadronic calorimeters. As muons are minimum ionizing particles, we require: $EM < 2\ GeV$, $HAD < 6\ GeV$, and also $EM + HAD > 0.1\ GeV$.

- d_0 - the impact parameter. This is defined as the distance of closest approach between the reconstructed muon track and the beam axis in the $r - \phi$ plane. Require: $d_0 < 3\ mm$.

- $|\Delta x|$ - the track-stub separation. Require: $|\Delta x| < 2\ cm$ for CMU muons, and $|\Delta x| < 5\ cm$ for CMP or CMX muons.

- z-vertex match (defined in the previous subsection). Require: z-vertex match $< 5.0\ cm$. Also require that the closest vertex to have $|z| < 60\ cm$ measured from the $z = 0$ point.

- Iso - muon isolation. $Iso = (E_T^{cone} - E_T^{tower})/P_T$, where E_T^{cone} is the sum of the transverse energies in towers within a cone of $R = 0.4$ around the track direction, E_T^{tower} is the transverse energy measured in the tower associated with the muon track, and P_T is the transverse momentum of the muon track. We require: $Iso < 0.1$.

- Fiducial cut on the muon. The muons are required to be in the fiducial volume within a pseudorapidity range of $|\eta| < 1.2$.

Similar to the electrons, muon cut efficiencies are extracted from the Z boson decays $Z \to \mu^+\mu^-$. Excluding the isolation cut, the combined cut efficiencies are 91% for CMX muons, 90% for CMU-only muons, 88% for CMP-only muons, and 94% for the CMU-CMP muons.

5.1.3 Missing Transverse Energy $\not\!\!E_T$

The neutrinos originating from W boson decays do not interact in the detector. Their presence however can be inferred indirectly by calculating the total transverse momentum in the event [18]. The latter should be equal to zero, as the incoming proton and antiproton were moving in the longitudinal direction. The deviation from zero is caused by the undetected neutrino(s), and is referred to as the missing transverse energy. The so called raw missing transverse energy $\not\!\!E_T{}^{raw}$ is the negative of the

[18] It is customary to use energy instead of momentum, as this is what the calorimeters measure. In the high-energy limit, the two quantities are equivalent up to the constant c factor.

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

Table 5.2: Quality cuts designed to identify the secondary lepton from Z-decays.

vector sum of all transverse energy in the calorimeter towers with $|\eta| < 3.6$. The thresholds for each calorimeter tower included in the sum are: 0.1 GeV for CEM, CHA, and WHA, 0.3 GeV for PEM, 0.5 GeV for PHA and FEM, and 0.8 GeV for FHA towers. If the primary lepton is a muon, the missing energy $\not\!\!E_T$ is obtained from $\not\!\!E_T{}^{raw}$ by vectorially adding the P_T of the muon track and subtracting the E_T of the calorimeter tower associated with the muon. If the primary lepton is an electron, then $\not\!\!E_T = \not\!\!E_T{}^{raw}$. We do not include jet corrections in calculating $\not\!\!E_T$. For our analysis, we select events having $\not\!\!E_T > 20\ GeV$.

5.1.4 Z Boson Decays and $t\bar{t}$ Dilepton Events Removal

We remove events consistent with Z boson decays: $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$. On of the two leptons resulting from the decay is identified by the selection cuts in Sections 5.1.1, 5.1.2. A Z-decay candidate is first required to have a secondary lepton passing the cuts listed in Table 5.2 [45]. Second, we demand that the invariant mass of the ee or $\mu\mu$ satisfy: $75 < M_{ee},\ M_{\mu\mu} < 105\ GeV/c^2$. If both these conditions are met, the event is rejected.

For $t\bar{t}$ background suppression, dilepton events are also removed. These are events having two opposite-sign leptons (e or μ) with $P_T > 20\ GeV/c$, missing transverse

energy $\not\!\!E_T > 25\ GeV$, and two jets with $E_T > 20\ GeV$ [46]. In addition, we reject events with an isolated track with $P_T > 15\ GeV/c$ and charge opposite to that of the primary lepton. The track isolation requirement is $\sum P_T < 2\ GeV/c$ in a cone of radius $R = 0.4$ around the lepton candidate track.

5.1.5 Jet Identification

Colored quarks and gluons can be regarded as free during a hard collision. Shortly afterwards ($\sim 10^{-24}\ s$) the color forces will organize them into colorless hadrons in a process known as hadronization. A quark or gluon therefore exits the interaction region as a narrow spray of hadrons, or *jet*. Jets will deposit their energy in a localized cluster of calorimeter towers. Cluster identification starts from *seed* towers featuring a transverse energy deposition E_T of 3 GeV or more. Neighboring towers with E_T above 1 GeV are further added to the cluster. An energy weighted centroid is calculated for the cluster, and only towers within a cone of radius $R = 0.4$ about the centroid are considered in estimating jet energy. The *raw* energy E_T^{raw} is defined as the transverse energy sum within the 0.4 radius cone. E_T^{raw} is in general different from the true energy of the original parton, for the following possible reasons:

- A fraction of jet energy may be deposited outside the cone (*out-of-cone* losses).
- Extra energy in the calorimeters due to multiple hard interactions, and/or soft interactions between the spectator partons in the event (*underlying event*).
- Calorimeters non-linearity and response at the boundaries between modules.
- The low momentum tracks in the jet that curl and fail to exit the CTC.
- Muons and neutrinos in the jet which are not detected in the calorimeters.

For this analysis, we use the raw E_T values for the jets, therefore we do not correct for the effects listed above. We select only events having one, two, or three jets with $E_T^{raw} > 15\ GeV$ and $|\eta| < 2.0$.

Secondary Vertex Tagging of B Jets.

Secondary Vertex Tagging (SVX) allows the identification of the jets produced by b quarks. This method exploits the long life time of B hadrons[19], which implies that they can travel up to $1-2\ cm$ before their decay. The spatial location of a B decay (*secondary* vertex) is therefore measurably displaced from the the $p\bar{p}$ interaction point (*primary* vertex). This is illustrated in Figure 5.1.

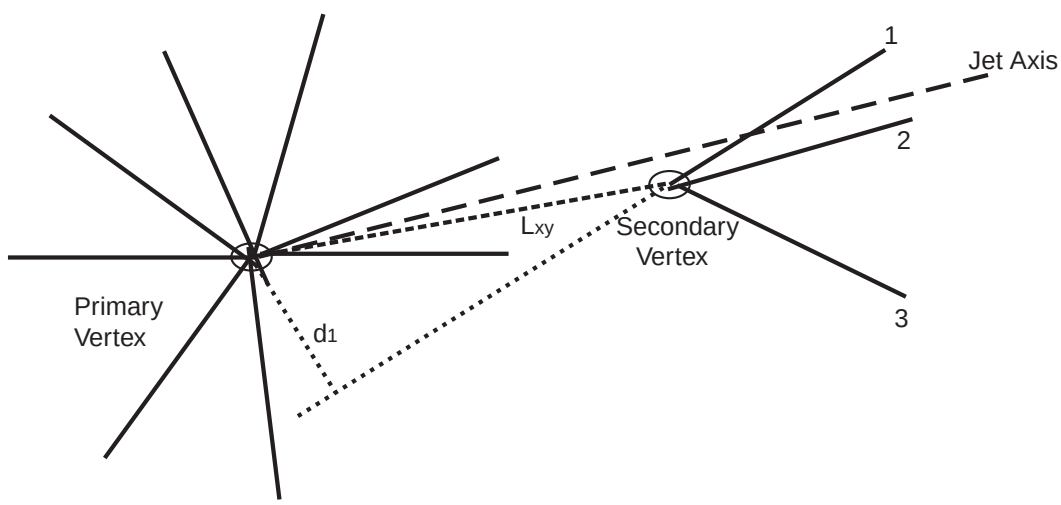

Figure 5.1: Schematic view of an event containing a secondary vertex shown in the transverse ($r-\phi$) view. The primary vertex is the point where the $p\bar{p}$ interaction occurs. The lines denoted 1, 2, and 3 are charged-particle tracks reconstructed in the SVX which yield the secondary vertex. The impact parameter for track 1 is denoted by d_1. L_{xy} is the 2-dimensional decay distance to the secondary vertex measured in the $r-\phi$ plane (adapted from [47]).

[19]The mean life time in the rest frame is $\tau_0 = 1.65 \cdot 10^{-12}\ s$ ($c\tau_0 = 0.5\ mm$). A 15 GeV b quark has a Lorentz factor $\gamma \approx 4$.

The primary vertex in each event is found by a weighted fit of the SVX tracks and the VTX z event vertex position. An iterative search removes tracks from the fit which have large impact parameters. The impact parameter d is the distance of the closest approach of a track to the primary vertex in the $r-\phi$ (transverse) plane. The uncertainty in the fitted primary vertex coordinates in the transverse plane depends on the number of tracks and the event topology, and ranges between 6 and 36 μm. If there are multiple primary interactions in the event, then the primary vertex is chosen to be the one with the greatest total P_T of associated tracks. All tracks used in the vertex fit and in the analysis are required to extrapolate to within 5 cm of this vertex in the z direction.

The $B-$tagging algorithm is applied to groups of two or more SVX tracks associated with jets having $E_T^{raw} > 15\ GeV$ and $|\eta| < 2.0$. An SVX track is associated with a jet if the opening angle between track direction and the jet direction (given by the calorimeter) is less than 35°. These tracks are also required to have $P_T > 2\ GeV/c$ and impact parameter significance $|d|/\sigma_d \geq 3$, where σ_d is the uncertainty on d. A 2-dimensional vertex constrained fit is performed for the tracks, which yields the transverse displacement L_{xy} and its uncertainty $\sigma_{L_{xy}}$. A jet is tagged as a $B-$jet if the secondary vertex has significance $|L_{xy}|/\sigma_{L_{xy}} > 3.0$, and $L_{xy} > 0$. The sign of L_{xy} is given by the dot product of the L_{xy} direction and the direction of the vector sum of the tracks' momenta.

We finally note that not all B-tags identified by this algorithm come from B hadrons. Displaced vertices are also produced by other long-lived particles, as c, τ, K_S^0, Λ etc. In addition, "fake" secondary vertices may result from the reconstruction of mismeasured tracks.

5.2 Monte Carlo Simulations

Monte Carlo is a numerical method of sampling the probability distribution functions (pdf's) for a physical system by using sequences of random numbers. How can Monte Carlo computer simulations be applied to high energy physics analyses?

First, let us review the main stages in our experiment: **(1)** the Tevatron accelerator produces $p\bar{p}$ interactions, **(2)** the CDF detector measures the various particles resulted from the interaction and writes interesting events to tape **(3)** recorded events are reconstructed, i.e. the electronic signals from the detector subsystems are translated in a setup of charged tracks, jets, etc., and **(4)** the final stage - physics analysis.

In the "Monte Carlo world" the similar stages are: **(1')** the Monte Carlo *generator* produces a user-specified type of events e.g. W^*, $t\bar{t}$, etc., **(2')** the Monte Carlo *detector simulation* models the subsequent interactions and decays within the detector, and its output has the same format as the real data recorded to tape, **(3')**, **(4')** same as **(3)**, **(4)** above. The purpose of Monte Carlo generation and simulation is therefore to model detector's response to certain type (physical process) of events.

The program we use for Monte Carlo generation of events is PYTHIA 6.129a [48]. This program implements a variety of tree-level Standard Model processes of interest at a hadron collider. The parton-shower approach is employed for initial and final state radiation, and string fragmentation [49] is used to model parton hadronization. The default set of parton distribution functions is GRV 94L. B-hadron decays are modeled with the CLEOMC software [50]. Detector simulation is performed with the QFL' package developed by the CDF collaboration.

The processes we generated for our analysis are:

- electroweak single top production $q\bar{q}' \to t\bar{b}$, $qg \to t\bar{b}q'$
- strong top pair production: $q\bar{q} \to t\bar{t}$, $gg \to t\bar{t}$ (see Figure 5.2)
- $W + jets$ QCD production: $q\bar{q}' \to gW$, $qg \to q'W$ (see Figure 5.2)

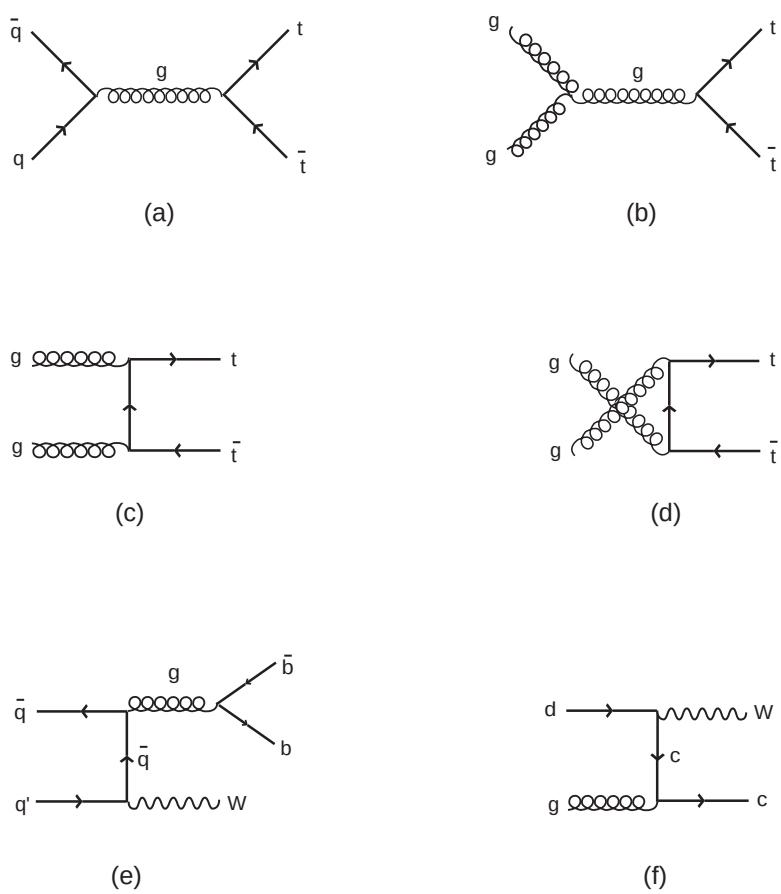

Figure 5.2: Representative Feynman diagrams for the main backgrounds to single top production. The $t\bar{t}$ background (a-d) dominates in the region of high energy and high jet multiplicity. At the other end of the spectrum (low energy and low jet multiplicity), the $Wb\bar{b}$ (e), $Wc\bar{c}$, and Wc (f) represent the most significant background.

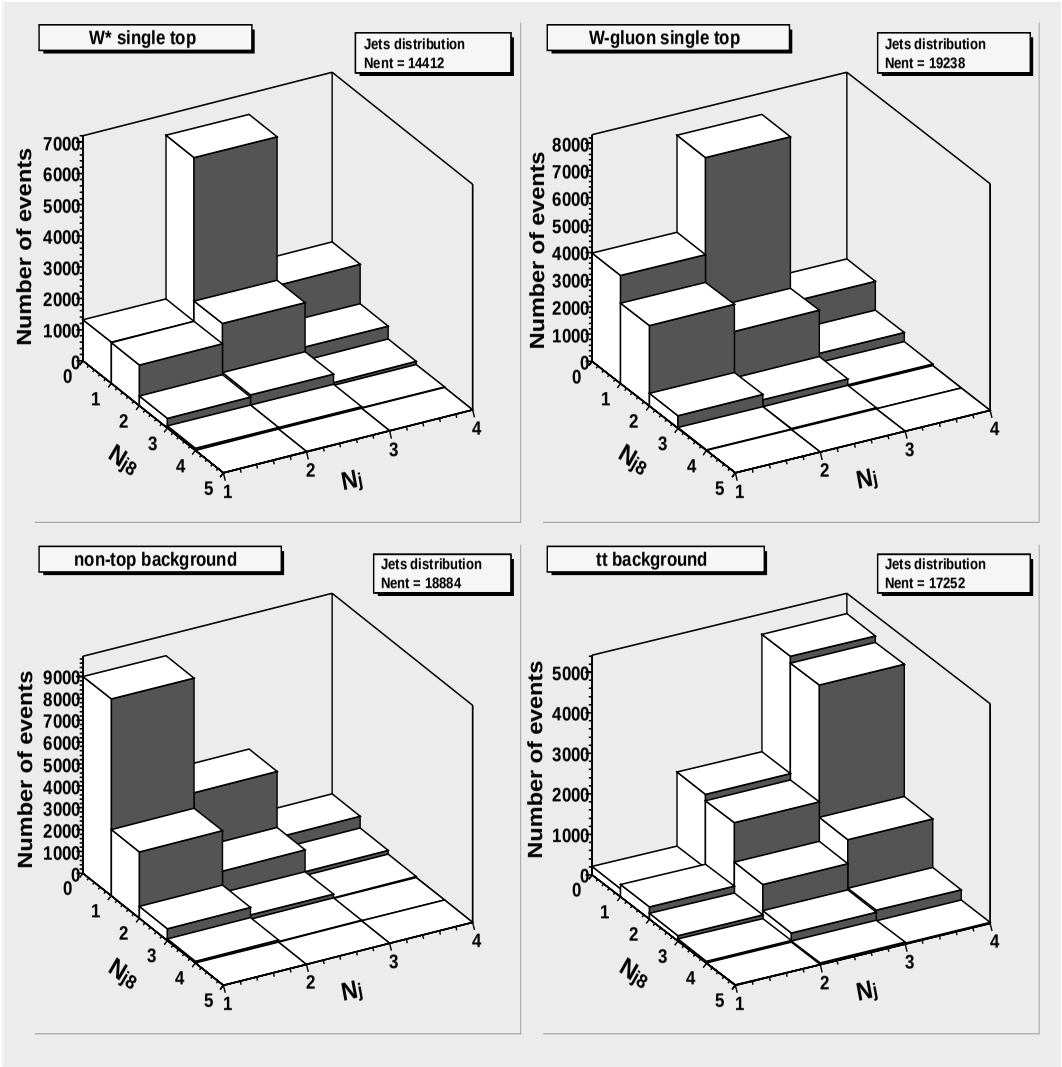

Figure 5.3: N_{j8} versus N_j distribution for signal and background events passing the initial selection of Section 5.1. We note that most of the signal resides in the $N_j = 2$ channel. To suppress the non-top (QCD) background we require $N_{j8} = 1$ for events with $N_j = 1$. We reduce the $t\bar{t}$ background by imposing $N_{j8} = 0$ for events with $N_j = 3$ (see Section 5.2.1).

The main $W + jets$ QCD processes are $Wb\bar{b}$, $Wc\bar{c}$, and Wc production. However, we used the $W + jets$ QCD sample as a model for the entire non-top background (see Table 5.5 for a list of all non-top contributors). We have checked that the mistags are well modeled by our Monte Carlo sample, and that the W+1 jet Run I data is also in agreement with our model.

5.2.1 Monte Carlo Acceptances and Run I Expectations

Let us first introduce the terminology which will be used in what follows:

- *Initial selection* - the ensemble of selection requirements listed in Section 5.1
- N_j - the number of jets with $E_T^{raw} > 15\ GeV$ and $|\eta| < 2.0$ in an event
- N_{j8} - the number of jets with $E_T^{raw} > 8\ GeV$ and $|\eta| < 2.4$, which were not previously counted in the above N_j jets
- $W + N\ jets$ event - an event which satisfies the initial selection requirements, and has a total of $N = N_j$ jets with $E_T^{raw} > 15\ GeV$ and $|\eta| < 2.0$
- N_{j8} *selection* - the N_{j8} requirements defined as:
  - in the W+1 jet channel ($N_j = 1$), require $N_{j8} = 1$
  - in the W+3 jet channel ($N_j = 3$), require $N_{j8} = 0$

  We will justify the N_{j8} selection later in this Section.
- *Final selection* - the initial selection plus the N_{j8} selection. This is the tightest and final set of selection criteria we will use in our analysis.

W^* sample: $N_{gen} = 0.5$ million events			
Cuts Imposed	$W + 1$ jet	$W + 2$ jets	$W + 3$ jets
Lepton and $\not\!\!E_T$	11040	20497	5522
AND ≥ 1 B-tag	3057	9091	2492
AND Z and dilepton removal	3011	8951	2450
AND N_{j8} selection	1308	8951	1785
$W-$gluon sample: $N_{gen} = 0.905$ million events			
Cuts Imposed	$W + 1$ jet	$W + 2$ jets	$W + 3$ jets
Lepton and $\not\!\!E_T$	31493	33480	5732
AND ≥ 1 B-tag	7570	10070	1999
AND Z and dilepton removal	7451	9838	1949
AND N_{j8} selection	2958	9838	1466
QCD (non-top) sample: $N_{gen} = 37.85$ million events			
Cuts Imposed	$W + 1$ jet	$W + 2$ jets	$W + 3$ jets
Lepton and $\not\!\!E_T$	13661	5633	1129
AND ≥ 1 B-tag	12822	5296	1051
AND Z and dilepton removal	12663	5192	1029
AND N_{j8} selection	3032	5192	756
$t\bar{t}$ sample: $N_{gen} = 0.6$ million events			
Cuts Imposed	$W + 1$ jet	$W + 2$ jets	$W + 3$ jets
Lepton and $\not\!\!E_T$	4865	19961	29930
AND ≥ 1 B-tag	1406	8134	13408
AND Z and dilepton removal	712	5076	11464
AND N_{j8} selection	304	5076	4912

Table 5.3: Numbers of events passing the selections described in Sections 5.1 and 5.2.1. A total of N_{gen} events were generated with PYTHIA 6.129a for each Monte Carlo sample.

The acceptances after making various cuts on the Monte Carlo samples can be extracted from Table 5.3. In this Table, the initial selection corresponds to the fourth row below each listed process. The fifth and last row presents the numbers of events which pass the final selection.

	W^*	W-gluon	$t\bar{t}$
Number of W + 2 jet events	1.2 ± 0.3	1.8 ± 0.5	3.7 ± 1.1

Table 5.4: Run I expected numbers of W + 2 jets events passing the initial cuts of Section 5.1 (from [52]).

Source	$W+1$ jet	$W+2$ jets	$W+3$ jets
Mistags	18 ± 7	6.3 ± 2.6	1.6 ± 0.6
$Wb\bar{b} + Wc\bar{c}$	18 ± 7	9.0 ± 2.5	1.7 ± 0.5
Wc	17 ± 6	4.3 ± 1.5	0.6 ± 0.2
Z+hf	1.5 ± 0.5	0.7 ± 0.3	0.2 ± 0.1
Diboson	1 ± 0.4	1.4 ± 0.5	0.3 ± 0.1
Non-W	9 ± 4	2.3 ± 1.0	0.8 ± 0.4
Total Events	65 ± 13	24 ± 4.5	5.2 ± 1.1

Table 5.5: Predicted numbers of non-top background events in Run I data after the initial cuts of Section 5.1 have been imposed (from [51]).

N_{j8} Selection and Run I Expected Contributions

The Run I expected contributions in the W+2 jets channel at CDF have been studied previously [51, 52, 53]. Table 5.4 gives these expectations for the top quark processes, derived by normalizing Monte Carlo contributions to the theoretical predictions for single top and $t\bar{t}$. The numbers in this Table include the various data/simulation ratios as the lepton trigger and identification scale, B-tag efficiency scale, and muon trigger prescaling. Table 5.5 gives the W+1,2,3 jets expected numbers of events in Run I for the non-top processes[20]. The non-top backgrounds were normalized to the data content of the W + 1 jet channel, with no B-tag requirement.

[20]In this document we often refer to the non-top background as "QCD". We note however that in addition to $Wb\bar{b}$, $Wc\bar{c}$, and Wc events, the non-top backgrounds also include mistags, Z+heavy flavor jets, diboson, and other non-W processes [51].

	W^* signal	W-gluon signal	QCD (non-top)	$t\bar{t}$
$\epsilon_{N_{j8}}(W + 1 \text{ jet})$	43.4%	39.7%	23.9%	42.7%
$\epsilon_{N_{j8}}(W + 3 \text{ jet})$	72.9%	75.2%	73.5%	42.8%
$\epsilon_{N_{j8}}(W + 1,2,3 \text{ jet})$	83.6%	74.1%	47.6%	59.7%

Table 5.6: N_{j8} cut efficiencies for signal and background. These numbers can be derived from the fourth and fifth rows below each process listed in Table 5.3.

We notice that for Run I, in the W+2 jets channel, single top contribution is considerably smaller than the combined backgrounds contribution. This is even more true for the W+1 jet and W+3 jets channels. We investigated the possibility of reducing the backgrounds by exploiting jet multiplicity distributions. Figure 5.3 presents the distributions of N_j versus N_{j8} for signal and backgrounds. An optimization study shows that optimal signal to background ratio is obtained by selecting events that satisfy the following criteria:

- in the W+1 jet channel ($N_j = 1$), require $N_{j8} = 1$
- in the W+3 jet channel ($N_j = 3$), require $N_{j8} = 0$

Note that because both W^* and $W-$gluon processes peak in the $N_j = 2$ channel, there is no N_{j8} cut for the W+2 jets events. For the $W + 1$ jet events, the N_{j8} cut efficiency $\epsilon_{N_{j8}}(W + 1 \text{ jet})$ is defined as the ratio between the number of events passing the final cut and the number of events passing the initial cut. Similarly, we can define $\epsilon_{N_{j8}}(W + 3 \text{ jets})$ and $\epsilon_{N_{j8}}(W + 1,2,3 \text{ jet})$ for the $W + 3$ jets channel, and the combined $W + 1, 2, 3$ jet channels respectively. These efficiencies can be derived from Table 5.3, and are given in Table 5.6.

We are now in position to calculate the Run I expected numbers of events passing the final selections. For the top samples, the W+1, 3 jets contributions (before N_{j8}

Process	$W+1$ jet	W+2 jets	W+3 jets	Total events
W^*	0.2 ± 0.04	1.2 ± 0.3	0.2 ± 0.1	1.6 ± 0.4
W-gluon	0.5 ± 0.2	1.8 ± 0.5	0.3 ± 0.1	2.6 ± 0.7
QCD (non-top)	15.6 ± 3.1	24.0 ± 4.5	3.8 ± 0.8	43.3 ± 8.4
$t\bar{t}$	0.2 ± 0.1	3.7 ± 1.1	3.6 ± 1.1	7.4 ± 2.2

Table 5.7: Run I expectations for signal and backgrounds after the final selection.

cuts) are obtained by normalizing to the W + 2 jets predictions given in Table 5.4. The normalization factors can be extracted from Table 5.3. We then factor in the N_{j8} cut efficiencies (Table 5.6) to arrive at the values listed in Table 5.7. To illustrate how this is done, we give the W+1 jet calculation for the W^* process below:

$$1.2 \cdot \frac{3011}{8951} \cdot 43.4\% = 0.175 \approx 0.2 \text{ (events)}$$

To find the amount of non-top background expected in Run I, we start from the W + 1, 2, and 3 jets predictions given in Table 5.5. Next, we take into account the N_{j8} efficiencies for the W + 1 jet channel, and W + 3 jets channel respectively (Table 5.6). The resulting non-top Run I contributions are given in Table 5.7.

We close this Section with a brief remark on single top identification. According to Table 5.7, we should expect 4.2 single top events and 50.7 background events in the Run I dataset. The signal to background ratio is therefore 0.08, which implies that signal identification is very challenging. In this case, Artificial Neural Networks appear more suitable than a cut-based analysis, as they use a greater amount of information. To better understand how small the 0.08 value is, let us consider the $t\bar{t}$ production whose cross section is $\sigma_{t\bar{t}} = 2.7 \cdot \sigma_{single\ top}$. In the lepton+jets channel, where the top quark was originally discovered, the signal to background ratio is 2.6, more than 30 times higher than our single top value.

CHAPTER 6

MEASURING THE SINGLE TOP PRODUCTION CROSS SECTION

The analysis presented in this Chapter starts from the final selection requirements presented in Chapter 5. In Section 6.1, the Neural Networks technique for single top event identification is explained. The expected (*a priori*) sensitivity of this technique for the Run I data is discussed in Section 6.2. The results obtained from the Run I dataset are presented in Section 6.3. Section 6.4 contains the future prospects for observing single top in Run II at the Tevatron. We close this Chapter with a brief summary of the results obtained.

6.1 Identifying Single Top with Neural Networks

ANN's offer a simple way of combining the information from many kinematic and geometry variables, which in principle increases our potential to resolve the signal in the data. Our Neural Network is a two-layer perceptron (see Figure 4.2) with standard activation function: $g(x) = 0.5 \cdot [1 + \tanh(x)]$. For training, we use JETNET v3.5 [54] subroutines interfaced to ROOT [55] via the Root_Jetnet [56] package. The weights and thresholds are updated according to the Manhattan Back-Propagation algorithm with the default parameters. Our method for selecting the optimal set

of seven variables for the input layer is explained in Section 6.1.1. The left plots in Figures 6.3, 6.4, 6.5 show the distributions of the seven variables for W^* events and W-gluon events. These distributions are very similar, with the exception of $Q \times \eta$. To make up for this difference, we train the network with W^* and W-gluon in the proportion expected in Run I (Table 5.7). It is important to note that, for all six kinematic variables the two background components QCD and $t\bar{t}$ peak on different sides of the signal, as shown by the right plots in Figures 6.3, 6.4, 6.5. It is this major difference between the two backgrounds that requires the use of three output categories: single top signal (W^* and W-gluon), QCD background, and $t\bar{t}$ background. For training we use 10000 events from each class, or 30000 total. We require an output of (1,0,0) for the QCD events, (0,1,0) for the single top events, and (0,0,1) for the $t\bar{t}$ events respectively. As discussed in Section 4.3, the outputs have a straightforward interpretation: each one of them estimates the a posteriori Bayes probability for the corresponding class.

6.1.1 Input Variables

To distinguish between signal and background, a set of variables with good separation power is needed. Our procedure for choosing the input set is described below.

First, we form a large set of input variables based on previous single top studies [52, 57, 58, 59], and also on a visual study of the individual variable distributions for signal and backgrounds. The following 18 variables were found to have some separation potential: E_T^{jet1}, E_T^{jet2}, E_T^{lepton}, $\not{E}_T$, H_T, $\sqrt{\hat{s}}$, $M^{l\nu b}$, M^{jj}, P_T^{jj}, η^{jj}, η^{jet1}, η^{jet2}, $Q \times \eta$, $cos(\widehat{lq})$, R_{min}, N_j, N_{j8}, N_{b-tags}. Here, $jet1$ and $jet2$ are the leading jets in the event, H_T is the total transverse energy, $l\nu b$ refers to the lepton, neutrino, and

leading b-jet system, jj refers to the $jet1 - jet2$ system, $Q \times \eta$ is the product between the charge of the lepton and the pseudorapidity of the highest-E_T untagged jet (q), $\widehat{lq}$ is the angle between the direction of the lepton and that of the q jet, R_{min} is the minimum separation $\sqrt{(\delta\eta)^2 + (\delta\phi)^2}$ among all possible pairs of jets in the event.

N	Best combination $\overrightarrow{X}$	mean $P(\%)$	mean E	H
1	H_T	59.4	0.524	5
2	H_T,E_T^{jet1}	61.5	0.500	4
3	H_T,E_T^{jet1},N_j	62.7	0.487	6
4	H_T,E_T^{jet1},E_T^{jet2},E_T^{lepton}	64.5	0.471	9
5	H_T,E_T^{jet1},E_T^{jet2},E_T^{lepton},$\not{E}_T$	66.5	0.453	11
6	H_T,E_T^{jet1},E_T^{jet2},E_T^{lepton},$\not{E}_T$,P_T^{jj}	67.3	0.444	10
7	H_T,E_T^{jet1},E_T^{jet2},E_T^{lepton},$\not{E}_T$,P_T^{jj},$Q \times \eta$	67.9	0.436	19
8	H_T,E_T^{jet1},E_T^{jet2},E_T^{lepton},$\not{E}_T$,P_T^{jj},M^{jj},M^{lnb}	68.1	0.434	12
9	H_T,E_T^{jet1},E_T^{jet2},E_T^{lepton},$\not{E}_T$, P_T^{jj},M^{jj},M^{lnb},$Q \times \eta$	68.3	0.433	16
18	$A\ L\ L\quad V\ A\ R\ I\ A\ B\ L\ E\ S$	68.2	0.449	21

Table 6.1: Averaged performance P and error E for the best combinations of N variables. The last column lists the number of hidden nodes H for each case.

Second, we proceed to investigate a large number of combinations $\overrightarrow{X}$ of N variables that can be drawn from the 18-variable set[21]. To be able to compare the relative performances of two sets of input variables $\overrightarrow{X_1}$ and $\overrightarrow{X_2}$, we use two quantities: the performance P and the error E. The performance is defined as the rate of correct classifications, while the error measures how far the output is from the desired value. The exact definitions of P and E are given in Appendix B. We note that the two quantities are correlated; the better a network is, the higher the performance, and the lower the error. To facilitate the $\overrightarrow{X_1}$ vs $\overrightarrow{X_2}$ comparison, we have to deal with the variations of P and E from epoch to epoch. These variations are usually of the order

[21] For $N > 9$, only combinations containing the core $\{H_T, E_T^{jet1}, E_T^{jet2}, E_T^{lepton}, \not{E}_T, P_T^{jj}\}$ were tested.

of a few percent or less. Our solution to account for these variations is to average P and E over 400 training epochs *after* the learning begins to saturate. For example, Figure 6.1 shows the epoch-to-epoch variation of P for the case $\overrightarrow{X} = H_T$ and number of hidden nodes $H = 5$. We will use the mean of this histogram to plot the point marked by an arrow in Figure 6.2, and the RMS of the histogram to plot the error bars on this point. The exact procedure is repeated for different values of H.

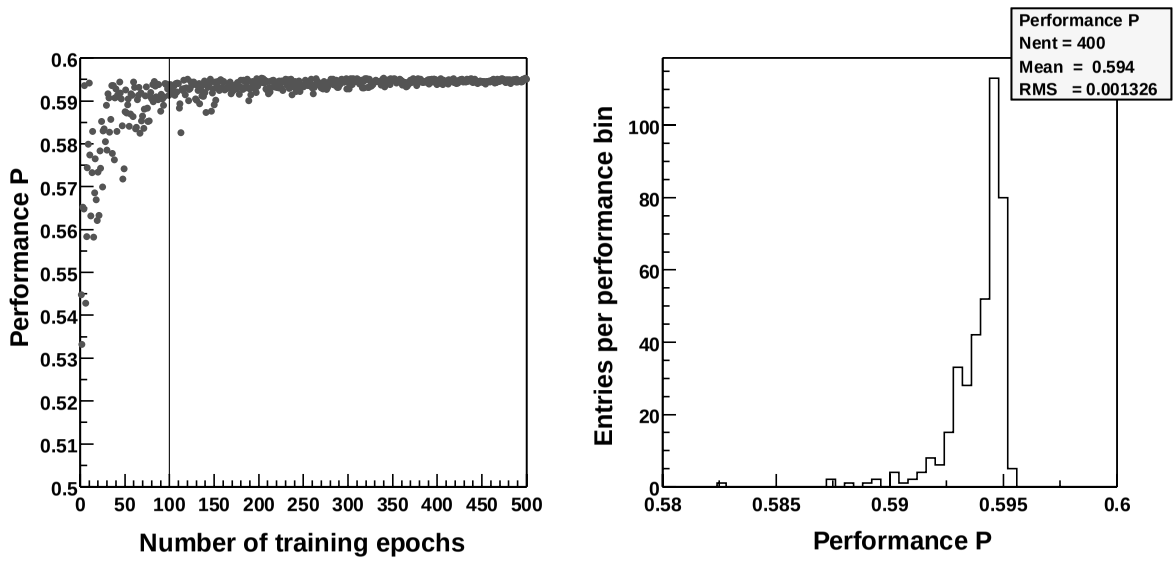

Figure 6.1: Performance variation as training evolves (left). Learning saturation begins after epoch 100, and the net stabilizes after approximately 500 epochs. In estimating the "average" performance, we discard the first 100 epochs. The 400 P values corresponding to epochs $100 - 500$ are histogrammed (right). The average performance is 59.4%, while the RMS of 0.1% measures the variation of P from epoch to epoch.

Table 6.1 presents the best (lowest E, highest P) combinations $\overrightarrow{X}$ of N variables that can be extracted from the 18-variable set. For these combinations, the dependence of P and E on H is shown in Figure 6.2. We note that using the entire set of N

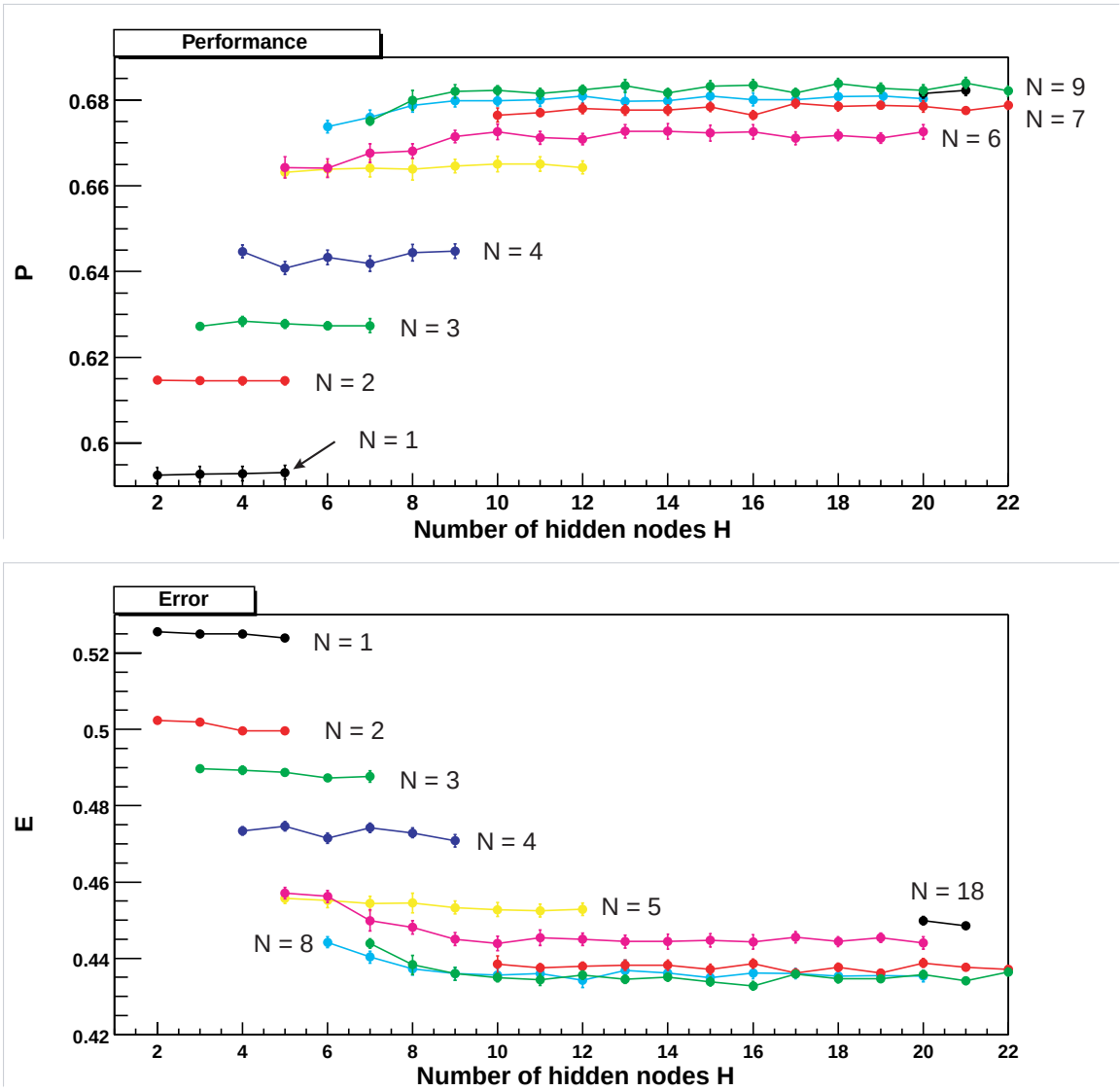

Figure 6.2: Averaged performance and error for the best combinations of N variables as functions of the number of hidden nodes H. See Table 6.1 for the list of the input variables for each value of N.

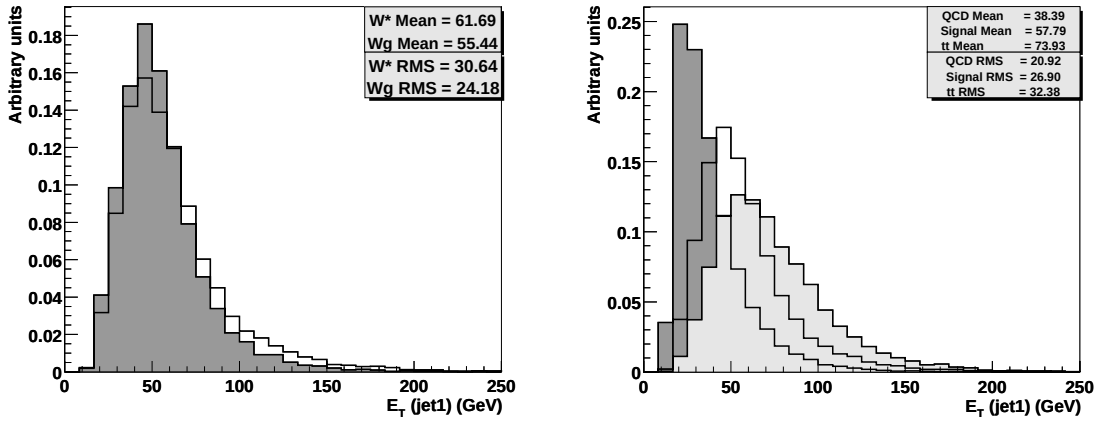

Figure 6.3: Monte Carlo distributions for the input variables. Left plot shows the two signal processes W^* (open) and W-gluon (shaded). Right plot shows the QCD background (shaded), signal mixed according to Table 5.7 (open), and $t\bar{t}$ background (lighter shade). For comparison purposes, all histograms are normalized to unit area.

$= 18$ variables results in less signal-background separation than that obtained for $N = 7, 8, 9$. In principle, this should not happen, as the amount of information is larger for $N = 18$. In practice however, the training sample is limited, while the space that must be searched for minimization is considerably expanded, making it more difficult to find a good minimum. Further, by performing a study of the systematic uncertainties for the cases $N = 7, 9$ we found that the a priori limit is in fact better for $N = 7$. In conclusion, we select for the input layer the following seven variables: E_T^{jet1}, E_T^{jet2}, E_T^{lepton}, $\not{E}_T$, H_T, P_T^{jj}, $Q \times \eta$. The corresponding distributions for Monte Carlo signal and backgrounds are shown in Figures 6.3, 6.4, 6.5. Note the significant overlap among the three distributions shown in the right plots of these Figures, which makes single top identification a challenging task.

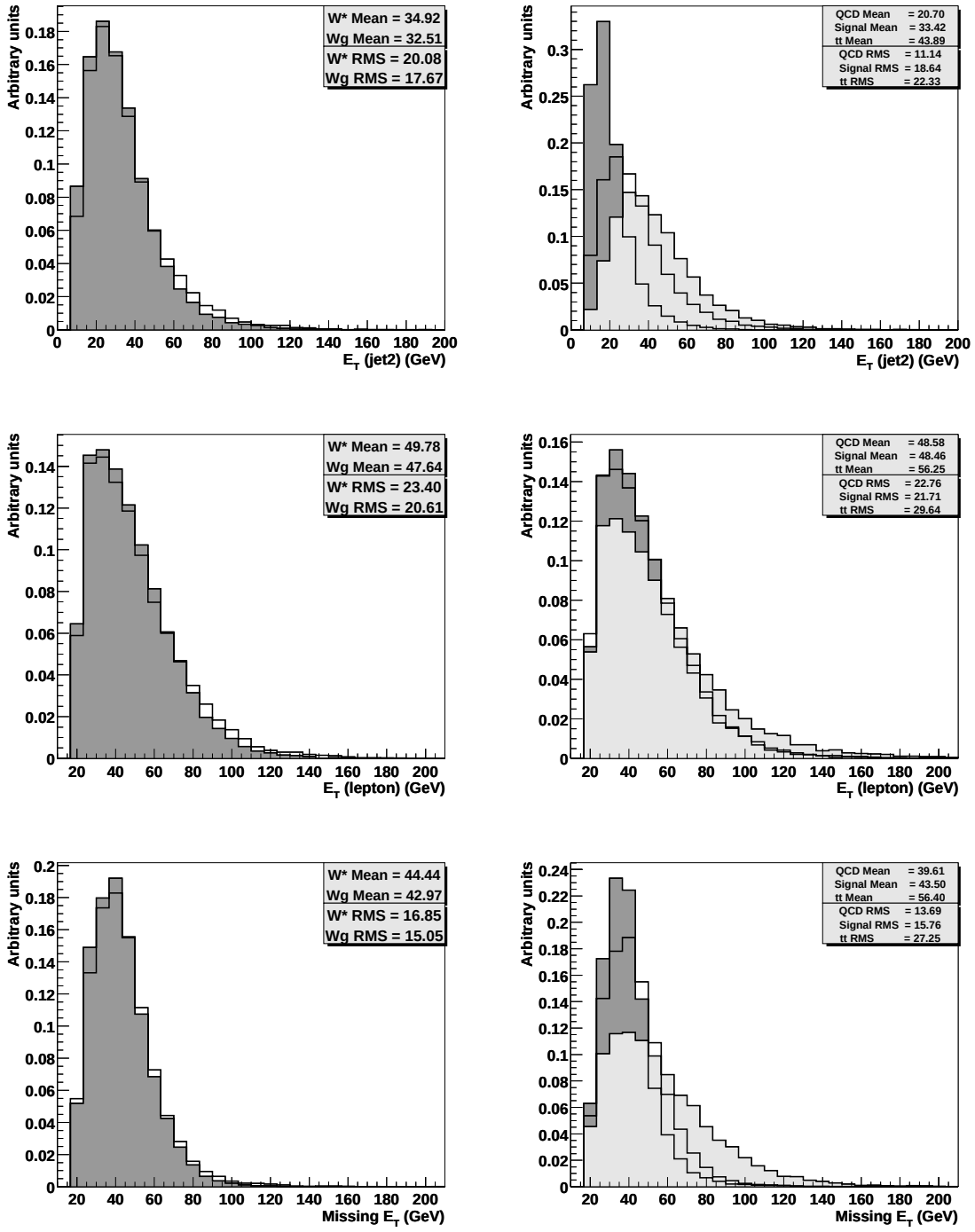

Figure 6.4: Monte Carlo distributions for the input variables. Left plots show the two signal processes W^* (open) and W-gluon (shaded). Right plots show the QCD background (shaded), signal mixed according to Table 5.7 (open), and $t\bar{t}$ background (lighter shade). For comparison purposes, all histograms are normalized to unit area (*cont'd*).

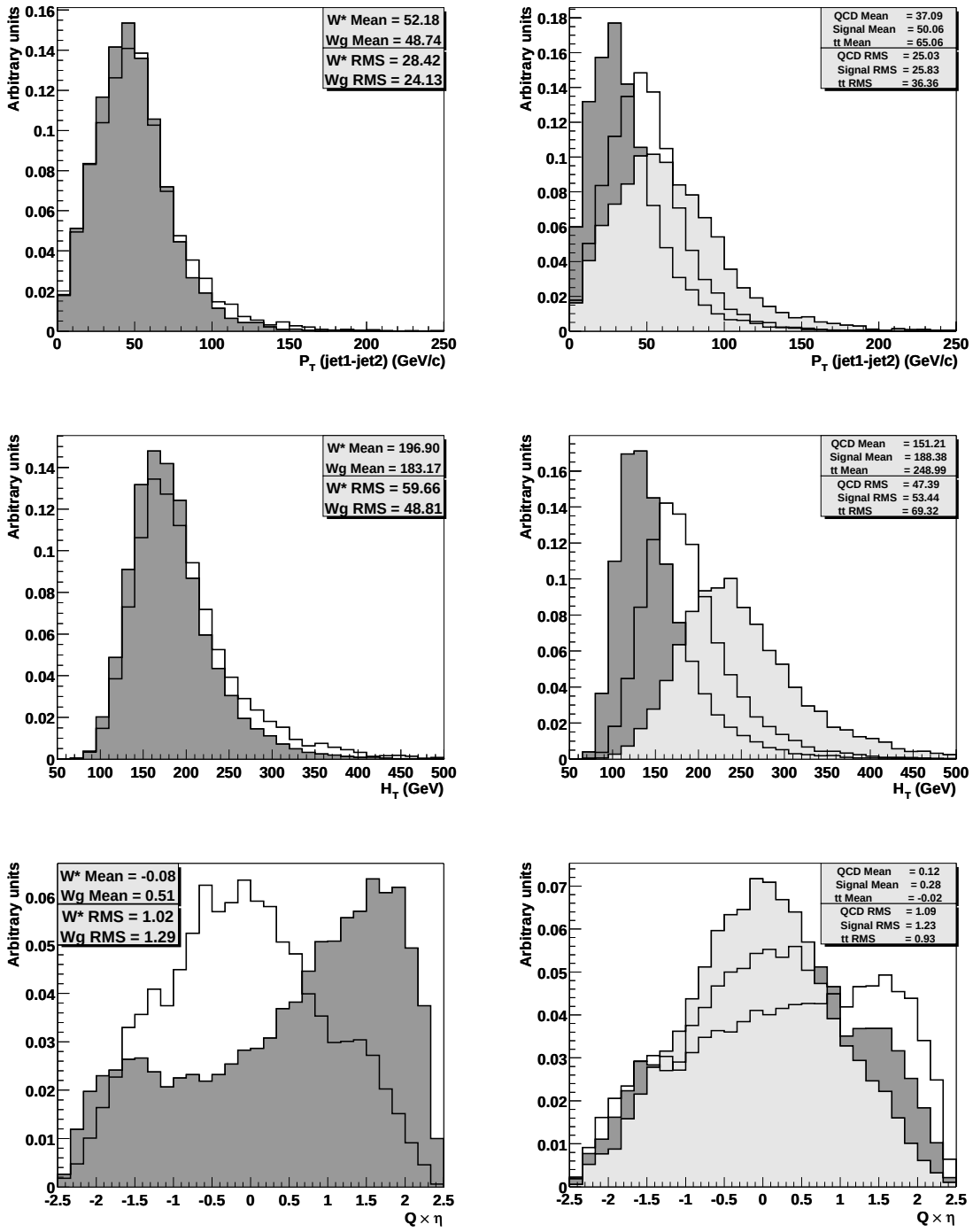

Figure 6.5: Monte Carlo distributions for the input variables. Left plots show the two signal processes W^* (open) and W-gluon (shaded). Right plots show the QCD background (shaded), signal mixed according to Table 5.7 (open), and $t\bar{t}$ background (lighter shade). For comparison purposes, all histograms are normalized to unit area (*cont'd*).

6.1.2 Neural Network Output

As mentioned previously, to train the network we require the outputs to be one for the correct class, and zero for the other two classes. The seven input variables require seven nodes in the input layer. Within a range of 7 to 30 nodes in the hidden layer, we found the generalization performance to vary slightly, with 17 nodes providing the best classification. Our Neural Network is therefore a 7-17-3 perceptron with 190 parameters adjusted by training. We stopped the training procedure at epoch 124, when the generalization performance was optimal. Appendix C presents the network obtained, which we will use to derive our Run I single top results.

Using the standard activation function implies that each output takes values in the (0,1) interval, therefore the output space is a cube of side one. After the training is completed, we run the three Monte Carlo samples[22] through the trained net to get the 3-dimensional distributions presented in Figure 6.6.

We note that the output points (O_1, O_2, O_3) are not spread throughout the cube, but rather tend to lie in the same plane. The reason for this, as mentioned in Section 4.3, is that the outputs estimate the a posteriori Bayes probabilities, which means that $O_1+O_2+O_3 \approx 1.0$. But $O_1+O_2+O_3=1.0$ is the equation of the plane determined by the three target output points, and this explains the planar appearance mentioned above. From Figure 6.7 we see that the output sum distributions for signal and background events have means of approximately 1.0 with small variances, consequences of the successful training process. The output sum being equal to one essentially reduces the output space to two dimensions. It is therefore natural to

[22]Our signal sample is obtained by mixing Monte Carlo W^* and W-gluon events in the Run I expected proportion of Table 5.7.

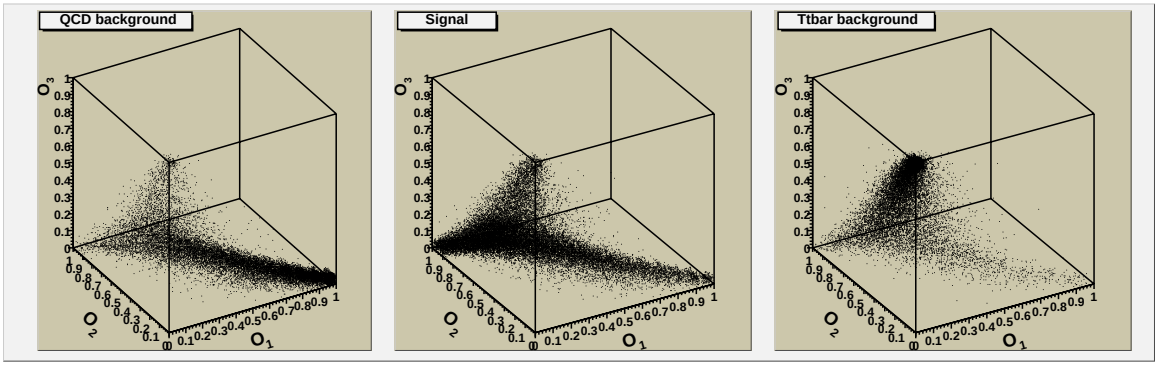

Figure 6.6: Neural Network output (O_1, O_2, O_3) for Monte Carlo signal and background events. As required in the training, the QCD background events tend to cluster around the (1,0,0) point, while the $t\bar{t}$ events lie close to the (0,0,1) corner. The signal is more spreadout inside the output cube, with most of the events located in the region close to the (0,1,0) corner.

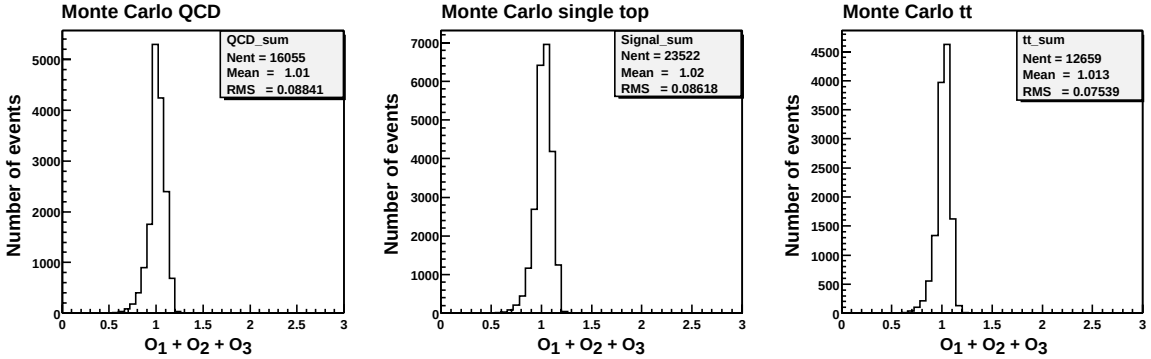

Figure 6.7: Output sum for Monte Carlo signal and background events. For each Monte Carlo sample, the corresponding $O_1 + O_2 + O_3$ distribution has a mean close to 1.0, and a small variance. This is a reflection of the normalization condition for the posteriori probabilities (see Section 4.2).

consider projecting the 3-dimensional output distributions onto the O_1+O_2+O_3=1.0 plane. For an arbitrary event, the projection mapping (O_1, O_2, O_3)⇒(x, y) is given by:

$$x = -\frac{1}{2} \cdot (a - \sqrt{a^2 + 4b^2 - 4ab}), \qquad y = \frac{\sqrt{3}}{2} \cdot (a + \frac{1}{3} \cdot \sqrt{a^2 + 4b^2 - 4ab}) \tag{6.1}$$

$$\text{where } a = (1 + O_1 - O_3)/\sqrt{2}, \text{ and } b = (1 + O_2 - O_3)/\sqrt{2}$$

A graphical representation of the projection mapping is presented in Figure 6.8. The resulting 2-dimensional distributions for the three Monte Carlo samples are shown in Figure 6.9. In the next section, we will describe how these 2-dimensional templates are used to extract the signal from a sample containing a mixture of signal and background events.

6.1.3 Extracting the Signal by Fitting the Neural Network Output

In what follows we will illustrate how the signal can be extracted from a dataset.

The problem. Suppose we are given a mixed sample of single top, QCD, and $t\bar{t}$ events, whose exact composition is unknown. We want to find out the signal content of this sample, by using the Neural Network developed in the previous Section.

The solution. First, we run the sample through the network, and obtain a 2-dimensional distribution (x, y) as described in Section 6.1.2. For convenience, let us use the same binning as in Figure 6.9, where the number of bins is $N_{bins} = 20^2 = 400$. Further, we fit this (x, y) distribution as the weighted sum of the three Monte Carlo templates shown in Figure 6.9. The fit is performed by maximizing the following standard binned likelihood function [60]:

$$\mathcal{L}(n_{signal}, n_{QCD}, n_{t\bar{t}}) = \prod_{i=1}^{N_{bins}} \frac{e^{-n_i} \cdot n_i^{d_i}}{d_i!} \tag{6.2}$$

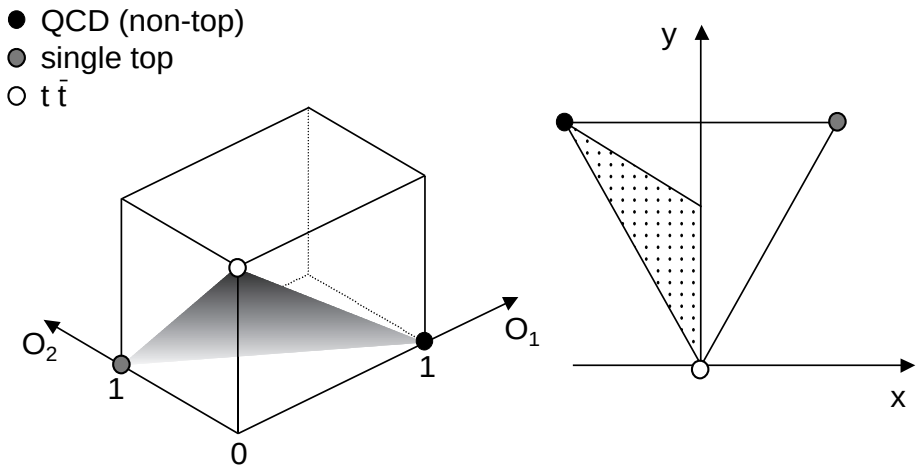

Figure 6.8: Graphical representation of the projection mapping $(O_1, O_2, O_3) \Rightarrow (x, y)$. We expect the dotted area to be little populated, as events in this region would have to have: $O_2 < O_1$, O_3. However, since our Network is mostly E_T-based, the signal should in general be situated in between the two backgrounds (Figures 6.3, 6.4, 6.5). Figure 6.9 shows that this is indeed the case.

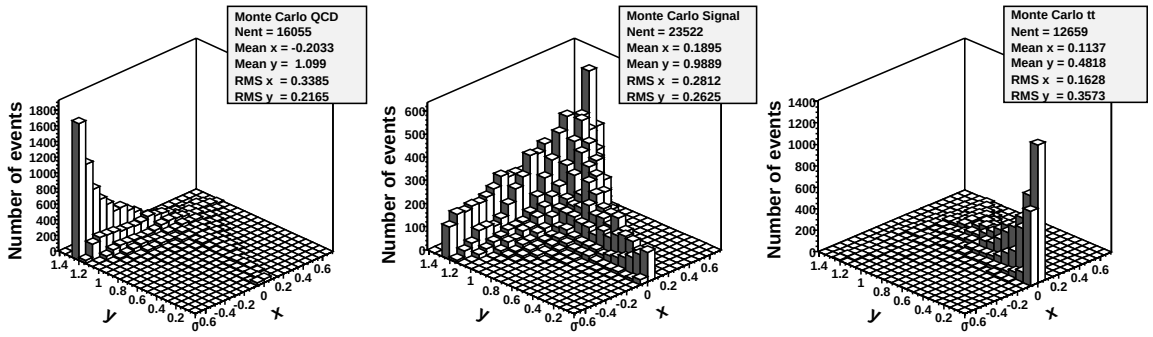

Figure 6.9: Monte Carlo signal and background 2-dimensional output distributions resulted from projecting all output points onto the plane of equation $O_1+O_2+O_3 = 1$.

The quantities appearing in expression (6.2) are defined below:

- n_{signal}, n_{QCD}, $n_{t\bar{t}}$ are the parameters of the fit, representing the numbers of single top, QCD, and $t\bar{t}$ events in the sample

- $n_i = n_{signal} \cdot f_{signal,i} + n_{QCD} \cdot f_{QCD,i} + n_{t\bar{t}} \cdot f_{t\bar{t},i}$ is the *expected*[23] number of events in the i-th bin of the (x, y) output space. $f_{signal,i}$, $f_{QCD,i}$, $f_{t\bar{t},i}$ are the probabilities that a Monte Carlo single top, QCD, or $t\bar{t}$ event appears in the i-th bin. These probabilities are derived by simply normalizing the 2-dimensional distributions of Figure 6.9.

- d_i is the number of events in the sample that populate the i-th bin, often called the *observed* number.

In conclusion, the signal and background contents of the sample are given by the values n_{signal}, n_{QCD}, $n_{t\bar{t}}$ which maximize the likelihood in Eqn. (6.2). We will refer to these values as the *fitted* numbers of signal and background events.

6.2 A priori Study of Run I Single Top Cross Section

Before even running on the Run I data, we can ask ourselves how sensitive the NN method is expected to be. To calculate this *a priori* sensitivity we use Monte Carlo events to mimic the real data, starting from the contributions of Table 5.7.

The first step is constructing fake-Run I datasets, or pseudo-experiments. A pseudo-experiment is composed of N_{signal}, N_{QCD}, and $N_{t\bar{t}}$ Monte Carlo signal, QCD, and $t\bar{t}$ events respectively. We generate the values of N_{signal}, N_{QCD}, and $N_{t\bar{t}}$ according to the following recipe:

[23]This is the total number of events in the i-th bin that we would expect based on our Monte Carlo samples.

- Start from the Run I expected contributions: $\mu_{signal} = 4.2$, $\mu_{QCD} = 43.3$, and $\mu_{t\bar{t}} = 7.4$ (events), with the uncertainties: $\sigma_{QCD} = 8.4$, and $\sigma_{t\bar{t}} = 2.2$ (events).

- Generate random numbers from Gauss distributions: $N'_{QCD} = \text{Gauss}(\mu_{QCD}, \sigma_{QCD})$, and $N'_{t\bar{t}} = \text{Gauss}(\mu_{t\bar{t}}, \sigma_{t\bar{t}})$.

- Generate random numbers from Poisson distributions: $N_{signal} = \text{Poisson}(\mu_{signal})$, $N_{QCD} = \text{Poisson}(N'_{QCD})$, and $N_{t\bar{t}} = \text{Poisson}(N'_{t\bar{t}})$.

Next, we randomly draw N_{signal}, N_{QCD}, $N_{t\bar{t}}$ events from our Monte Carlo samples to form the pseudo-experiment. Note that we do not constrain the sum $N_{signal} + N_{QCD} + N_{t\bar{t}}$ to the number of events observed in the Run I data. To get the signal content of this fake-Run I dataset we apply the procedure described in the Section 6.1.3. The only difference is that we introduce a likelihood constraint factor $\mathcal{L}_{constr}(n_{QCD}, n_{t\bar{t}})$ in Eqn. (6.2) [53]. This factor accounts for the fact that we already have backgrounds estimates which have been measured in separate analyses: $\mu_{QCD} \pm \sigma_{QCD}$ and $\mu_{t\bar{t}} \pm \sigma_{t\bar{t}}$. Our likelihood function is therefore $\mathcal{L}' = \mathcal{L}_{constrain} \cdot \mathcal{L}$, or explicitly:

$$\mathcal{L}'(n_{signal}, n_{QCD}, n_{t\bar{t}}) = \frac{\mathrm{e}^{-\frac{(n_{QCD}-\mu_{QCD})^2}{2\sigma^2_{QCD}}}}{\sqrt{2\pi}\sigma_{QCD}} \cdot \frac{\mathrm{e}^{-\frac{(n_{t\bar{t}}-\mu_{t\bar{t}})^2}{2\sigma^2_{t\bar{t}}}}}{\sqrt{2\pi}\sigma_{t\bar{t}}} \prod_{i=1}^{N_{bins}} \frac{e^{-n_i} \cdot n_i^{d_i}}{d_i!} \tag{6.3}$$

We generate and fit 10000 pseudo-experiments in this manner.

If the likelihood fit is working properly, then the fitted values n_{signal}, n_{QCD}, $n_{t\bar{t}}$ should be (on the average) close to the original values used in generating each pseudo-experiment: μ_{signal}, μ_{QCD}, $\mu_{t\bar{t}}$. Figure 6.10 shows that this is precisely the case.

To further test our fitting method, we calculate the signal and backgrounds *pulls* for each-pseudo-experiment. We define the pulls as:

$$P_{signal} = \frac{(n_{signal} - \mu_{signal})}{e_{signal}}, \quad P_{QCD} = \frac{(n_{QCD} - N'_{QCD})}{e_{QCD}}, \quad P_{t\bar{t}} = \frac{(n_{t\bar{t}} - N'_{t\bar{t}})}{e_{t\bar{t}}} \tag{6.4}$$

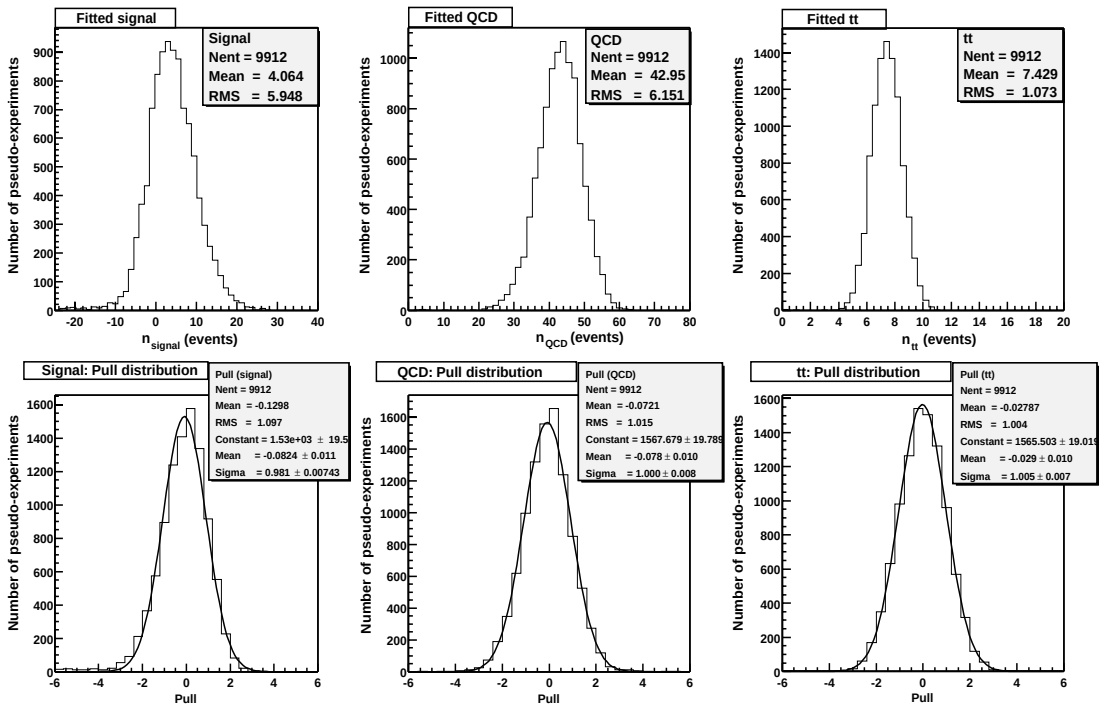

Figure 6.10: Results from 10000 pseudo-experiments. The 88 pseudo-experiments not shown have $n_{signal} < -23.0$, or $n_{QCD} < 2.0$ (events).

where e_{signal}, e_{QCD}, and $e_{t\bar{t}}$ are the errors returned by the fit. We use these definitions for the pulls, because n_{signal}, n_{QCD}, and $n_{t\bar{t}}$ estimate[24] the corresponding Poisson means: μ_{signal}, N'_{QCD}, and $N'_{t\bar{t}}$ (rather than the actual amounts N_{signal}, N_{QCD}, $N_{t\bar{t}}$). If our fitting procedure is unbiased and returns the correct errors, then the pulls distributions should be centered on zero and have unit standard deviation. Figure 6.10 demonstrates that our fitter satisfies both of these conditions.

Note. In less than one percent of the 10000 experiments, the likelihood fit did not converge. This is due to the low values of μ_{signal} and $\mu_{t\bar{t}}$, which can lead to problematic fluctuations as for example $N_{signal} = N_{t\bar{t}} = 0$. To eliminate these cases,

[24]For an explanation of this, we point the reader to Appendix C of Ref. [57].

we demand that the fitted contributions satisfy: $n_{signal} > -23.0$, $n_{QCD} > 2.0$, and $n_{t\bar{t}} > 0.0$ (hence 9912 entries instead of 10000).

Finally, Figure 6.10 shows that the percent uncertainty in the fitted number of signal events is approximately 146%, which determines us to continue this a priori study by setting an upper limit on the Run I single top production.

6.2.1 Setting the a priori Statistical Upper Limit

Our Bayesian approach to setting the a priori 95% C.L. confidence limit is described in Refs. [52, 57, 61]. We use the set of pseudo-experiments constructed previously, and we extract a limit μ_{95} from each individual pseudo-experiment. For example, if a given pseudo-experiment has the fitted signal contribution of n_0 events, then the limit μ_{95} is defined by the following equation:

$$\int_0^{\mu_{95}} \mathcal{L}(\mu_s|n_0)\mathrm{d}\mu_s = 0.95 \cdot \int_0^{\infty} \mathcal{L}(\mu_s|n_0)\mathrm{d}\mu_s \tag{6.5}$$

where $\mathcal{L}(\mu_s|n_0)$ is the likelihood that the true signal mean is μ_s, given the observed n_0. In Appendix D we describe how $\mathcal{L}(\mu_s|n_0)$ can be obtained. We solve Eqn. (6.5) for each and every one of the 9912 pseudo-experiments, histogram μ_{95}, and define our 95% C.L. on single top production as the mean of the μ_{95} distribution (Figure 6.11). We conclude that the statistical a priori 95% C.L. limit on single top production is 16.65 events, or equivalently 9.63 *pb*.

6.2.2 Folding the Systematic Uncertainties into the Confidence Limit

Following the procedure of Ref. [64], we divide the systematic uncertainties into two groups. The first group consists of those systematics that can modify both the shape of the NN output templates as well as the number of events accepted. The

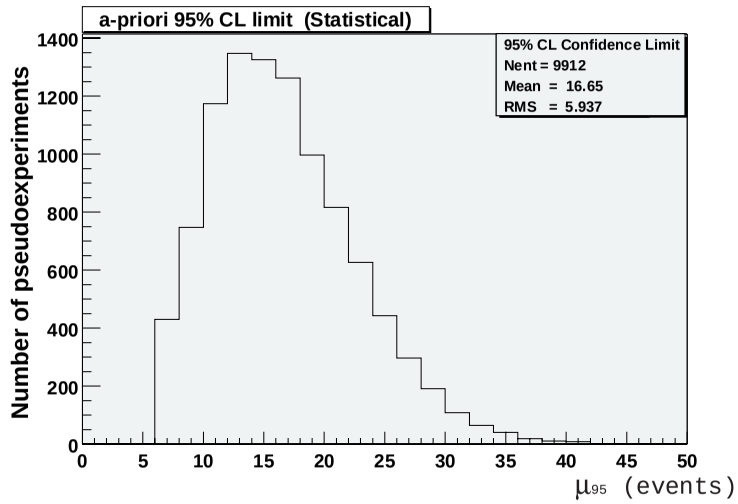

Figure 6.11: The distribution of the individual μ_{95}'s. The 95% C.L. limit is 16.65 signal events, or equivalently 9.63 *pb*.

systematic uncertainty associated with these effects is denoted by ΔS. The second group of systematic uncertainties contains those effects that alter only the rate of accepted events, and not the shape of the NN output distributions. The systematic uncertainty due to these effects is expressed fractionally and it is denoted by ΔR. We will describe both groups and explain how the effects are incorporated into the 95% C.L. limit on single top cross section.

Systematic uncertainties impacting both event acceptance rates and shapes of the NN output distributions

Let ΔS^i be the contribution of the i-th effect to ΔS. To derive ΔS^i, we use the following method:

- Generate new signal and backgrounds Monte Carlo samples where the i-th systematic effect was shifted by one standard deviation or a reasonable approximation of this.

- Compute the new acceptance rates, and also get the new 2-dimensional NN output templates by running the new Monte Carlo samples through the Neural Network[25]. We will use these templates only for pseudo-experiment generation, and not for fitting (see below).

- Using the newly calculated acceptances and NN output distributions, construct 10000 pseudo-experiments as described in Section 6.2, and fit each of these to the standard (unshifted) set of NN output templates. Let $\langle n^i_{signal} \rangle$ denote the mean of the distribution of the fitted number of signal events n^i_{signal}.

- Define ΔS^i as the difference $|\langle n^i_{signal} \rangle - \mu^i_{signal}|$, or half this difference for certain effects (ISR, FSR).

The systematic effects in this group are summarized below. For each of these we explain how we generate the shifted Monte Carlo samples and how we define ΔS^i.

- **Jet E_T scale**

  The jet E_T systematic uncertainty is attributed to (a) calorimeter response and (b) fragmentation-related effects [65]. In the first case, the parameters that describe the calorimeter's response to incident electrons, photons, and pions have uncertainties due to finite statistics and assumptions which are made. Additional uncertainty comes from the non-linear response of the calorimeters,

[25]For the systematics study we use the *same* 7-17-3 Neural Network constructed in Section 6.1.2.

Systematic Effect i	$\Delta S^i/\mu_{SM}$
Jet E_T Scale (+1σ)	0.503
Jet E_T Scale (-1σ)	0.264
Initial State Radiation	0.061
Final State Radiation	0.116
Parton Distribution Function	0.041
Top Mass 170 GeV	0.028
Top Mass 180 GeV	0.042
Signal Generator	0.033
Background Generator	0.043
Sum in quadrature ΔS	0.413

Table 6.2: Systematic uncertainties for effects modifying both event rates and NN output shapes. The uncertainties are expressed as fractions of the theoretical prediction of $\mu_{SM} = 4.2$ signal events. As explained in the text, we average the two Jet E_T Scale entries, and we use the larger Top Mass systematic (corresponding to $M_{top} = 180\ GeV$).

and energy deposition in cracks between calorimeter cells. In the second case, the uncertainty in the fragmentation parameters comes from the modeling of the tracking efficiency in jets, and the level of agreement between simulation and data. The contributions to the jet E_T uncertainty from these sources are evaluated by shifting the corresponding parameters by $+1$ and -1 standard deviations, and calculating the resulting shift in jets' energies. We apply these shifts to both signal and backgrounds samples. $\Delta S^{Jet\ E_T}$ is taken to be the average $(|\Delta S^{+1\sigma}| + |\Delta S^{-1\sigma}|)/2$.

- **Initial State Radiation** (ISR)

  The QCD radiation ($e.g.$ gluon emission) originating from the incoming partons is referred to as *initial state radiation*, and increases the jet content of an event. To study the associated systematic uncertainty, we turn off the initial

state radiation in PYTHIA and generate new signal and backgrounds samples. The uncertainty ΔS^{ISR} is taken to be one half of the shift in the mean fitted signal contribution.

- **Final State Radiation (FSR)**

  The final state radiation is defined as the QCD radiation from the outgoing partons, and leads to extra jets in the final state. To isolate the effect of FSR, we start from Monte Carlo samples generated with the initial state radiation turned off. From these noISR samples we select the subset noISR−noFSR of events in which every jet matches to a final state parton within a (η, ϕ) distance of 0.4. Since we did not store the generator level information for QCD events, we only consider the noISR, noISR−noFSR shifted signal and $t\bar{t}$ samples. We take ΔS^{FSR} to be half the difference: $(\Delta S^{ISR,FSR} - \Delta S^{ISR})/2$.

- **Parton Distribution Function**

  Our Monte Carlo samples used to measure the single top cross section are generated with the GRV 94L [66] set of parton distribution functions (PDF), which is the default choice in PYTHIA 6.129a. Other PDF choices are available as well, as for example CTEQ 3L [67]. To evaluate the associated systematic uncertainty, we generate new signal and backgrounds samples using the CTEQ 3L functions. ΔS^{PDF} is defined as the shift in the mean signal fitted contribution.

- **Top quark mass**

  The top mass result measured by CDF is: $M_{top} = 176.1 \pm 5.1\ GeV/c^2$ [65]. We therefore change the default top quark mass $M_{top} = 175\ GeV$ to $170\ GeV$ and

180 GeV respectively, and generate new W^*, W-gluon, and $t\bar{t}$ samples. We take $\Delta S^{M_{top}}$ to be the larger of the differences ΔS^{170} and ΔS^{180}.

- **Signal Generator**

  To estimate this systematic effect, we use W^* and W-gluon generated with HERWIG [68] instead of PYTHIA. The two generators are similar for the most part. There are however a few differences, among which we note the hadronization approach (*cluster* model in HERWIG versus *string* model in PYTHIA) and the underlying event modeling (HERWIG model is based on UA5 Collaboration data study [69], while PYTHIA uses the multiple interactions model described in Ref. [70]). We define $\Delta S^{Sig.Gen.}$ as the shift in the mean signal fitted contribution.

- **Background Generator**

  To estimate this systematic effect, we use $Wb\bar{b}$ and $t\bar{t}$ samples generated in HERWIG instead of PYTHIA. To model the QCD (non-top) background we mix HERWIG $Wb\bar{b}$ with PYTHIA $Wc\bar{c}$ and Wc. The mixing proportion was derived from our standard PYTHIA QCD sample. As in the previous cases, we define $\Delta S^{Bck.Gen.}$ as the shift in the mean signal fitted contribution.

The systematic uncertainties divided by the theory prediction are given in Table 6.2.

Systematic uncertainties effecting event acceptance rates only

Let ΔR^i be the contribution of the i-th effect to ΔR. To calculate ΔR^i we shift the i-th effect by one standard deviation, and recompute signal acceptance A^i. We define the uncertainty as the relative difference: $(A^i - A^0)/A^0 \equiv \Delta R^i$, where A^0 is

Systematic Effect i	$(A^i - A^0)/A^0$
Trigger and Lepton ID	0.10
B-tag Efficiency	0.10
Luminosity	0.04
Sum in quadrature ΔR	0.15

Table 6.3: Systematic uncertainties for effects modifying event rates only.

the acceptance in the standard case (*i.e.* before shifting the systematic effect). The rate uncertainties are listed in Table 6.3. The 10% systematic associated with lepton triggering and identification was derived in Ref. [71]. The B-tag systematic comes from the B-tag efficiency scale factor of 1.0 ± 0.1 [57]. Finally, the uncertainty on luminosity of 4.1% was taken from Ref. [72].

Incorporating Systematic Uncertainties in the a priori Limit

To include the effect of the systematic uncertainties in the calculation of the a priori limit, we follow the method of Ref. [52]. We use the same 9912 pseudo-experiments as in Section 6.2.1, and we extract a limit μ_{95} from each individual pseudo-experiment. The calculation is similar to the one performed in the case of the statistical a priori limit. The difference is, however, that before solving Eqn. (6.5) we smear the likelihood function $\mathcal{L}(\mu_s|n_0)$. The smeared likelihood $\mathcal{L}_{smear}(\mu_s|n_0)$ is obtained by executing the following procedure a large number (10^6) of times:

- Draw a random number r with probability density $\mathcal{L}(\mu_s|n_0)$; only positive values are allowed;
- Draw a Gaussian random number g_1 with mean 0 and width 1 and such that $1 + g_1\Delta R > 0$, where $\Delta R = 0.15$ was obtained in the previous section;

- Draw a Gaussian random number g_2 with mean 0 and width 1;

- Set $r' = r/(1 + g_1\Delta R) + g_2\Delta S$, where $\Delta S = 1.73$ events was obtained in the previous section[26]. The second term has a straightforward interpretation, while the first term is justified below. If the acceptance changes from A_0 to A', we have $n' = n_0 \cdot A'/A_0$, and the likelihood shifts to:

  $$\mathcal{L}'(\mu_s|n') = \mathcal{L}'\left(\mu_s\middle|n_0\frac{A'}{A_0}\right) = \mathcal{L}\left(\mu_s\frac{A'}{A_0}\middle|n_0\right)$$

  We can also express this in terms of probabilities. If r and r' are random numbers drawn from $\mathcal{L}'(\mu|n')$ and $\mathcal{L}(\mu|n_0)$, then:

  $$P(r' < \mu_s) = P(r < \mu_s A'/A_0) = P(rA_0/A' < \mu_s)$$

  which explains the ratio $A_0/A' = 1/(1 + g_1\Delta R)$ appearing in the r' definition;

- Histogram r'.

The resulting histogram defines $\mathcal{L}_{smear}(\mu_s|n_0)$. To find μ_{95} we have to solve the following equation, which replaces Eqn. (6.5):

$$\int_0^{\mu_{95}} \mathcal{L}_{smear}(\mu_s|n_0)\mathrm{d}\mu_s = 0.95 \cdot \int_0^{\infty} \mathcal{L}_{smear}(\mu_s|n_0)\mathrm{d}\mu_s \tag{6.6}$$

An example of likelihood smearing is given in Appendix D. Solving Eqn. (6.6) for each pseudo-experiment yields the μ_{95} distribution shown in Figure 6.12. We conclude that the a priori limit with systematics included is 18.14 signal events, or equivalently:

$$\sigma_{single\ top} < 10.5\ pb \text{ at } 95\% \text{ C.L. (a priori, including systematics)} \tag{6.7}$$

[26]Note that ΔS listed in Table 6.2 is expressed as a fraction of the theoretical prediction of 4.2 signal events.

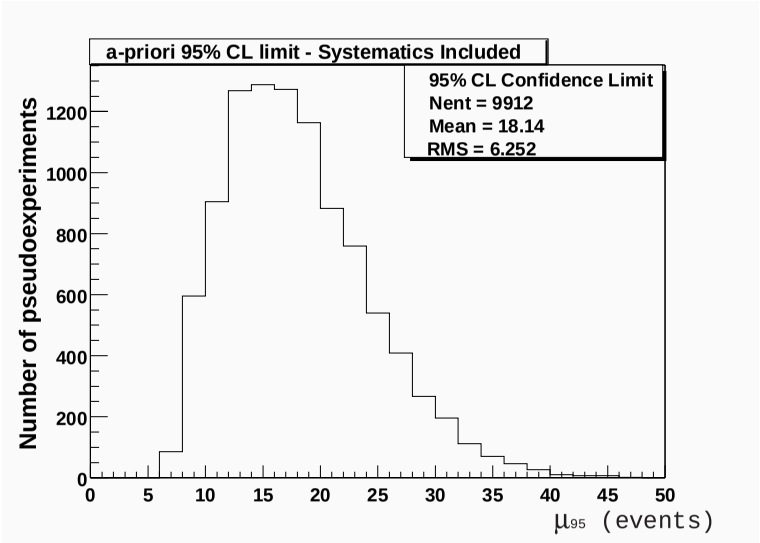

Figure 6.12: The distribution of the individual μ_{95}'s after smearing the likelihood functions. The 95% C.L. limit is 18.14 signal events, or equivalently 10.50 *pb*.

6.3 Single Top Cross Section Results from the Run I Data

We have analyzed the data samples collected during Run IA and Run IB. A Run and Event list with the events passing the final selection of Chapter 5 is provided in Appendix E. Table 6.4 shows the numbers of events predicted and observed in each of the three W+jets channels.

	$W+1$ jet	$W+2$ jets	$W+3$ jets	Total events
Observed in Run IA	1	6	0	7
Observed in Run IB	13	35	9	57
Total Observed	14	41	9	64
Expected	16.6	30.7	7.8	54.9

Table 6.4: Run I observed and expected numbers of events.

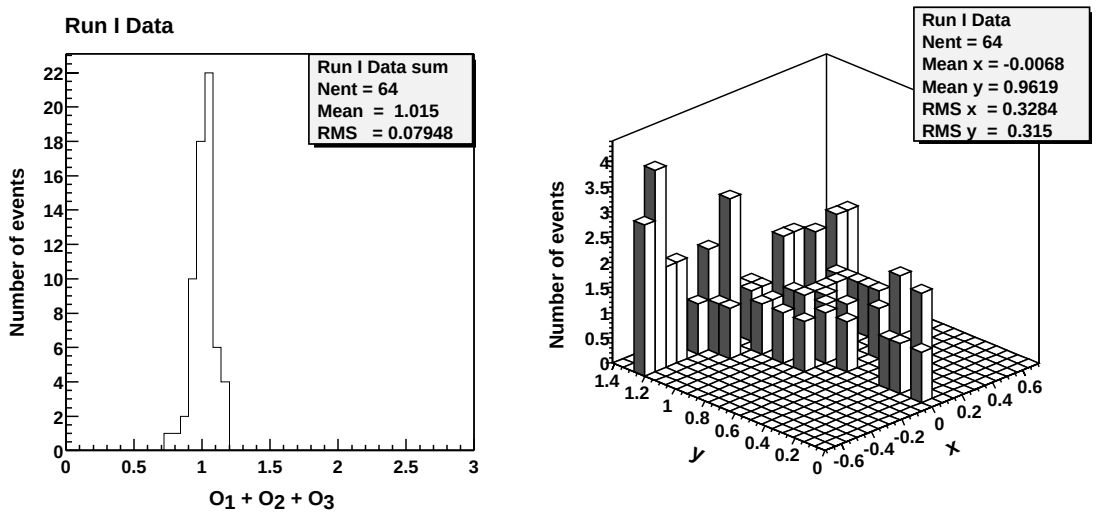

Figure 6.13: Output sum for Run I data events ($left$). Run I data 2-dimensional distribution resulted from projecting the 64 output points onto the plane of equation $O_1 + O_2 + O_3 = 1$ ($right$).

We feed the 64 Run I events to our 7-17-3 Neural Network. The outputs for all events are given in Appendix E. Figure 6.13 shows the (x, y) projections of the 64 output points, along with a histogram of the output sum. We note that the sum of the three outputs has a mean of approximately 1.0 and a small variance, similar to the Monte Carlo distributions of Figure 6.7.

To extract the signal content of the Run I dataset, we fit the (x, y) distribution of Figure 6.13 as the weighted sum of the three (x, y) Monte Carlo templates of Figure 6.9. We use the likelihood function defined in Eqn. (6.3). The result of the fit is:

$$n_{signal} = 23.61 \pm 7.70, \quad n_{QCD} = 36.18 \pm 6.24, \quad n_{t\bar{t}} = 7.59 \pm 2.00 \quad (events) \tag{6.8}$$

In cross section units, the signal content of Eqn. (6.8) corresponds to:

$$\sigma_{single\ top} = 13.66 \pm 4.46\ pb \tag{6.9}$$

not including the systematic uncertainties.

6.3.1 A posteriori Limit on Single Top Cross Section

Statistical a posteriori Limit

To obtain the 95% C.L. statistical limit, we use the Bayesian technique presented in Section 6.2. The likelihood function $\mathcal{L}(\mu_s|23.61)$ corresponding to the Run I data fit result is shown in Figure 6.14. By solving the following equation for μ_{95}:

$$\int_0^{\mu_{95}} \mathcal{L}(\mu_s|23.61)\mathrm{d}\mu_s = 0.95 \cdot \int_0^{\infty} \mathcal{L}(\mu_s|23.61)\mathrm{d}\mu_s \tag{6.10}$$

we extract an upper limit $\mu_{95} = 38.25$ single top events, or equivalently 22.13 pb. Next,

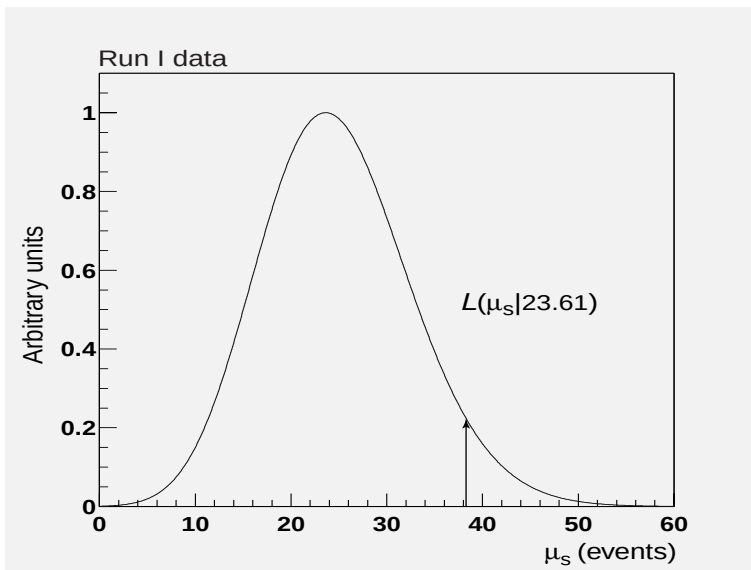

Figure 6.14: Run I data likelihood function $\mathcal{L}(\mu_s|23.61)$. The arrow marks the 95% C.L. upper limit of $\mu_{95} = 38.25$ single top events (22.13 pb).

we perform a frequentist check of this upper limit. We generate pseudo-experiments using a signal mean of $\mu_s = \mu_{95} = 38.25$ events and default backgrounds, and then

extract the fraction of pseudo-experiments returning a fitted signal contribution less or equal to the Run I data result of 23.61 events. We find this fraction to be 3.8%, which indicates that our Bayesian-calculated limit is in agreement (slightly conservative) with the frequentist point of view.

Including Systematic Uncertainties in the a posteriori Limit

The definitions of the different ΔS and ΔR uncertainties were given in Section 6.2.2, and their values are listed in Tables 6.2 and 6.3.

We point out that the individual ΔS values of Table 6.2 were calculated for the Run I *expected* single top contribution $\mu_{SM} = 4.2$ events. However, if we want to get a limit from the data we must decouple ourselves from the theoretical prediction, and consider a range of possible μ_s values. We therefore have to investigate how the systematics ΔS depend on the particular μ_s, which requires varying μ_s in constructing the pseudo-experiments. From Table 6.2 we see that the dominant systematic is associated with the jet energy scale[27] and accounts for approximately 93% of the total ΔS. We studied the dependence $\Delta S^{Jet\ E_T}(\mu_s)$ by running 10000 pseudo-experiments for each of the 14 values of μ_s shown in Figure 6.15, while background contributions are kept constant. The circles represent the $\Delta S^{+1\sigma}$ systematics obtained by increasing the jet transverse energy scale by one standard deviation, the squares are $\Delta S^{-1\sigma}$ systematics corresponding to a decrease in the jet E_T scale of one standard deviation, and the triangles represent the combined systematics, defined as the average $\Delta S^{Jet\ E_T} = (|\Delta S^{+1\sigma}| + |\Delta S^{-1\sigma}|)/2$. We notice that $\Delta S^{Jet\ E_T}$ increases roughly quadratically with μ_s (the fit shown is a parabola). To account for the other

[27] Five of our seven input variables are affected by changing the jet E_T scale: $E_T(jet1)$, $E_T(jet2)$, $\not\!\!E_T$, H_T, and $P_T(jj)$.

systematic effects contributing to the total ΔS we make the assumption that their relative ratio to $\Delta S^{Jet\ E_T}$ does not significantly change as μ_s vary, or equivalently:

$$\Delta S(\mu_s) \approx 1.1 \cdot \Delta S^{Jet\ E_T}(\mu_s) \quad \text{for all } \mu_s. \tag{6.11}$$

As we will see, this assumption should yield correct results, as the confidence limit is dominated by the statistical uncertainties and the impact of $\Delta S(\mu_s)$ on the limits is relatively small. For example, we found the limit to vary by less than two events when changing the systematics from $\Delta S(0)$ to $\Delta S(50)$.

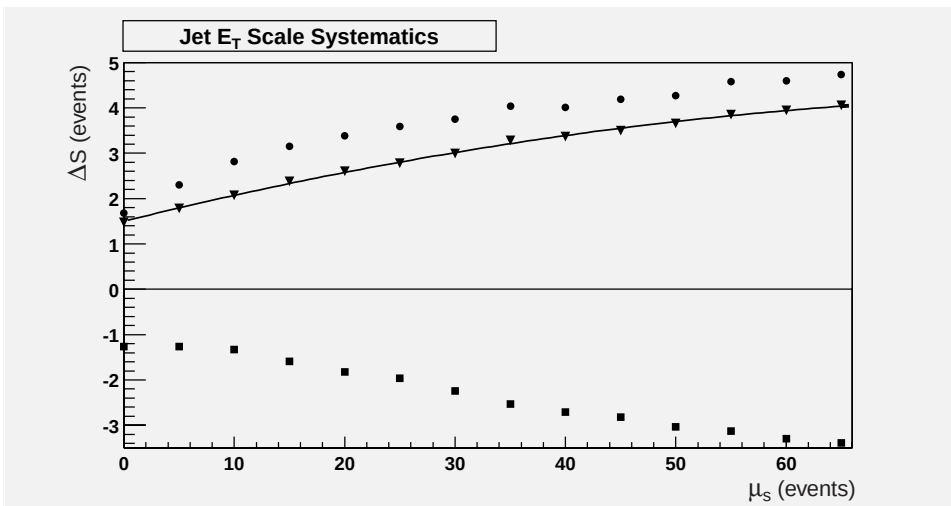

Figure 6.15: Jet E_T scale systematics as functions of μ_s. The circles correspond to a $+1\sigma$ increase in jets' transverse energies, the squares correspond to a -1σ decrease, while the triangles represent the average and define $\Delta S^{Jet\ E_T}$.

We smear the likelihood function $\mathcal{L}(\mu_s|23.61)$ using the recipe given in Section 6.2.2. For the systematics, we use $\Delta R = 0.15$, and $\Delta S = \Delta S(\mu_s = 23.61) = 3.02$ events. Note that we use the measured result $n_{signal} = 23.61 \pm 7.70$ as the estimate for the true signal mean which generated the Run I data set. After likelihood smearing,

we solve Eqn. (6.6) to obtain a 95% C.L. limit of 42.1 single top events, or in terms of cross section:

$$\sigma_{single\ top} < 24.4\ pb \text{ at } 95\% \text{ C.L. (a posteriori, including systematics)} \tag{6.12}$$

6.3.2 A posteriori Limit Interval on Single Top Cross Section

In Figure 6.14 we note that the likelihood function $\mathcal{L}(\mu_s|23.61)$ goes to zero in the vicinity of $\mu_s = 0$. It is natural therefore to consider deriving a confidence interval (upper and lower limits), rather than an upper limit only. In the frequentist approach, we can use the result (6.8) to derive a confidence interval via the Feldman-Cousins technique [62]. References [62, 73, 74] discuss the advantages of this method, among which we mention that it defines a proper coverage[28] of the parameter to be determined, and also makes a smooth transition between setting a limit and making an actual measurement.

Statistical Confidence Intervals

The construction of the Feldman-Cousins confidence belts starts from the PDF's $P(n_{signal}|\mu_s)$ (referred to as the "likelihood templates" in Appendix D.2.2). $P(n_{signal}|\mu_s)$ represents the probability of observing n_{signal} signal events given the "true" rate of μ_s signal events. For this study, we use the same likelihood templates as in Appendix D.2.2. Suppose we want to construct a 95% C.L. interval $[n_1, n_2]$ for a given value of μ_s: $\mu_s = \mu_0$. To do this, we compute the quantity $R(n_{signal})$ for the entire range [29]

[28] If μ denotes the single top cross-section, then correct coverage means that the 95% confidence intervals $[\mu_1, \mu_2]$ cover μ at the 95% confidence level (see Section II B. of Ref. [62]).

[29] In practice, we choose a discrete set of equally spaced n_{signal} values which span the relevant (non-zero probability) range. The spacing in n_{signal} is 0.2 events, which matches the spacing in μ_s (See Appendix D.2.2).

of n_{signal} values:

$$R(n_{signal}) = P(n_{signal}|\mu_0)/P(n_{signal}|\mu_{best}) \tag{6.13}$$

where μ_{best} is that value of μ_s which maximizes $P(n_{signal}|\mu_s)$ for the given n_{signal}. R determines the order in which values of n_{signal} are added to the acceptance interval at the particular value of $\mu_s = \mu_0$. This means that for the given μ_0, we find the α C.L. confidence interval $[n_1,\, n_2]$ by requiring:

$$R(n_1) = R(n_2) \;\text{ and }\; \int_{n_1}^{n_2} P(n_{signal}|\mu_0)\mathrm{d}n_{signal} = \alpha \cdot \int_{-\infty}^{\infty} P(n_{signal}|\mu_0)\mathrm{d}n_{signal} \tag{6.14}$$

We considered three values for α: 0.95, 0.90, and 0.6826 (one-sigma level). The n_1, n_2 form horizontal intervals which are plotted for each μ_s and shown in Figure 6.16.

We use the value found from the Run I data $n_{signal} = 23.61$ events to draw a vertical line in the $(n_{signal},\, \mu_s)$ two-dimensional space. The intersections of this vertical line with the band of horizontal intervals define the confidence interval for μ_s. The results expressed in numbers of events (not *pb*) are:

$$\begin{aligned}(\mu_1,\ \mu_2) &= (9.7,\ 41.1) \quad \text{at } 95\% \text{ C.L.}\\ (\mu_1,\ \mu_2) &= (11.5,\ 38.1) \quad \text{at } 90\% \text{ C.L.}\\ (\mu_1,\ \mu_2) &= (15.9,\ 32.5) \quad \text{at } 68.3\% \text{ C.L.}\end{aligned}$$

Note that no systematic uncertainties are included in the above results.

6.3.3 Comparison with Other Single Top Analyses

We will discuss here how the a posteriori results obtained in this Section compare with the other experimental searches for single top. We will only consider CDF and DØ analyses, as the Tevatron is at present the only accelerator with energy sufficient for producing top quarks.

We note that none of the analyses conclusively *observed* single top production. In general, the *observation* of a physical process requires a measurement of a signal

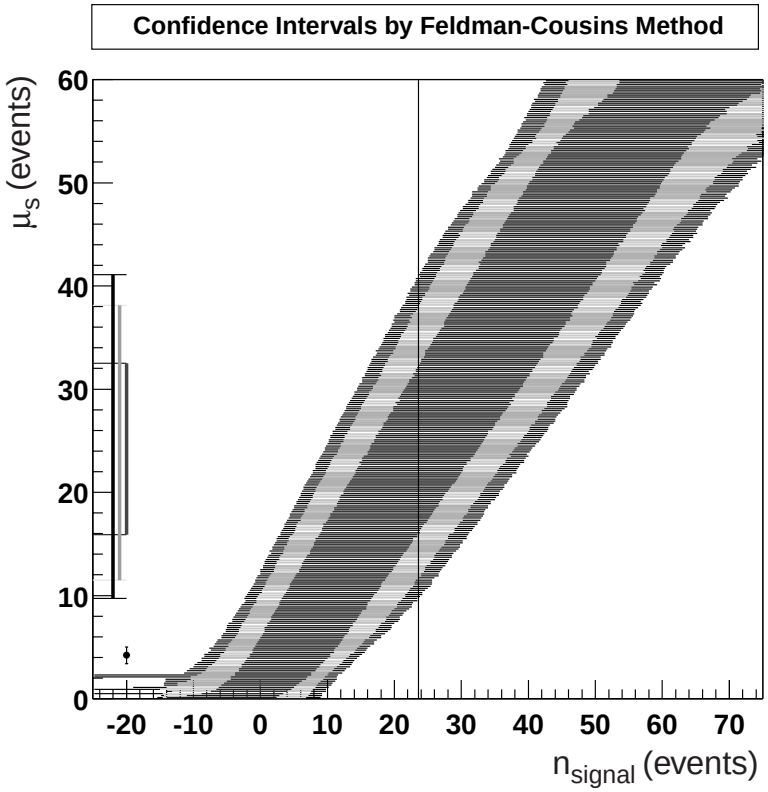

Figure 6.16: The Feldman-Cousins confidence belt. The confidence intervals $[\mu_1, \mu_2]$ are defined by the intersection of the belt with a vertical line drawn through the Run I fitted contribution of 23.61 signal events. The black, light grey, and dark grey horizontal lines represent the $[n_1, n_2]$ intervals corresponding to $\alpha = 95\%$, 90%, and 68.26% respectively. The circle with error bars located in the lower left corner represents the theoretically predicted value $\mu_s = 4.2$ events.

content at least five standard deviations away from zero. Therefore, upper limits on Run I single top cross section are set (see Table 6.5).

For the two DØ analyses, we note that the Neural Networks study finds significantly lower limits than the cut-based analysis. This is also true in the case of the a priori results in the CDF analyses, where our limit of 10.5 *pb* is 22% lower than that obtained in the combined search cut-based study CDF-2 [57]. This improvement

Analysis	W^* (95% C.L. limit)	W-gluon (95% C.L. limit)
CDF-1 (cut-based)	18 pb	13 pb
CDF-2 (cut-based)	14 pb	
CDF-3 (3-output ANN)	24 pb	
DØ-1 (cut-based)	39 pb	58 pb
DØ-2 (1-output ANN array)	17 pb	22 pb

Table 6.5: Upper limits on single top cross section from CDF [53] and DØ [76, 75] analyses of Run I data. Our result is listed as CDF-3.

comes primarily from two sources: (a) the N_{j8} cut selection yields a higher signal to background ratio than the $M_{l\nu b}$ cut used in CDF-2, and (b) the neural network combines information from seven variables which results in better signal separation. However, because we find a large excess of signal events, the posteriori limit increases accordingly, and in fact exceeds the cut-based CDF-2 result by 10 pb.

Further, it is worth pointing out that all CDF analyses, including the NN analysis presented here, have measured an excess in the number of Run I single top events. Below we list the fit results for each single top study, along with the corresponding signal expectations (in number of events):

CDF-1 ▷ W^* : $7.6^{+5.9}_{-4.8}$ expect 0.32 ± 0.07, W-gluon : $0.6^{+4.8}_{-4.0}$ expect 1.4 ± 0.3

CDF-2 ▷ W^* and W-gluon : 8.7 ± 7.8 expect 4.3 ± 0.6

CDF-3 ▷ W^* and W-gluon : 23.6 ± 7.7 expect 4.2 ± 0.9

In every case, the uncertainty in the fitted result is not sufficiently low to claim single top observation. We finally stress out that Table 6.5 or the above fit results can not be used to judge the sensitivity of the particular analyses. Sensitivity comparison only results from the a priori studies, where a large number of pseudo-experiments

Luminosity	Single top	QCD (non-top)	$t\bar{t}$
2 fb^{-1}	140	1040	340
10 fb^{-1}	700	5200	1700

Table 6.6: W+2 jets signal and background events expected in Run IIA (first row) and Run IIB (second row).

are performed. In this respect, we have shown that using the NN technique we are able to achieve a lower a priori limit than that of the cut-based approach [57].

6.4 Run II Prospects for Observing Single Top

Run II at the Tevatron started in March 2001. The target integrated luminosity at CDF in Run IIA is 2 fb^{-1}, almost 20 times greater than the Run I luminosity. In addition to getting more data, the $p\bar{p}$ center-of-mass energy $\sqrt{S} = 2.0\ TeV$ is also higher than in Run I ($\sqrt{S} = 1.8\ TeV$). This leads to an increase in the combined single top cross section by roughly 38% [77]. Finally, the upgraded CDF detector has a better overall acceptance than the Run I CDF detector. Taking into account these factors, the Run II-*expected* contributions in the $W + 2$ jets channel are listed in Table 6.6 from Ref. [78].

We studied single top production in the W+2 jet channel. Since the statistics are considerably better than those of Run I, we did not consider including the $W+1$, and $W+3$ jets channels and optimizing the search. Two more variables were added to the input layer of the Neural Network used for Run I: $M^{l\nu b}$ and M^{jj}, where $l\nu b$ refers to the lepton, neutrino, and leading b-jet system, while jj refers to the leading two jet system $jet1 - jet2$. We trained a 9-18-3 Neural Network as explained in Section 6.1. Further, we constructed Run IIA pseudo-experiments by mixing Monte Carlo events

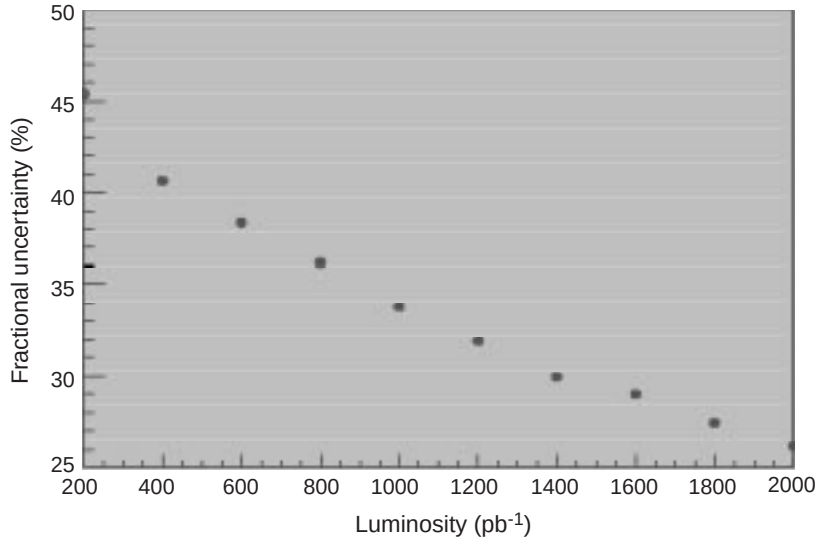

Figure 6.17: Fractional uncertainty in the single top cross section measurement versus integrated luminosity in Run II. A 3σ effect is expected to be observed with 1.2 fb^{-1} of data. No systematic uncertainties are included.

according to the contributions of Table 6.6. For 2 fb^{-1} of Run IIA, the fitted signal distribution has a mean of 140.0 events with a standard deviation of 36.7 events. The fractional uncertainty is therefore 26.2%, roughly corresponding to a 4σ single top cross section measurement. We can repeat this simple study for intermediate Run IIA luminosity values. Figure 6.17 shows the fractional uncertainty versus integrated luminosity within the range of $0.2 - 2$ fb^{-1}. All points in this graph assume the theoretically predicted value for single top cross section.

With the higher luminosity of Run IIB, we should be able to measure the single top cross section with an uncertainty better than 10% [79].

6.5 Summary of the Results

Table 6.7 gives a brief summary of the main results obtained in our analysis:

Run I a priori results
$\sigma_{single\ top} < 9.6\ pb$ at 95% C.L. (stat.)
$\sigma_{single\ top} < 10.5\ pb$ at 95% C.L. (stat. + syst.)

Run I a posteriori results
NN output fit result: $\sigma_{single\ top} = 13.7 \pm 4.5\ pb$
$\sigma_{single\ top} < 22.1\ pb$ at 95% C.L. (stat.)
$\sigma_{single\ top} < 24.4\ pb$ at 95% C.L. (stat. + syst.)
Cross section 95% C.L. interval: $[5.6,\ 23.8]$ (pb) (stat. only)

Run II expected results
With 2 fb^{-1} of data, roughly expect a 4σ measurement

Table 6.7: The most important results of our single top analysis.

CHAPTER 7

CONCLUSIONS

Measuring single top is an important test of the Standard Model, as its production cross section is directly proportional to $|V_{tb}|^2$, where V_{tb} represents the coupling between the top and the bottom quarks. In addition, it impacts the future searches for the Higgs boson, and has the potential to validate a number of phenomena outside of the Standard Model.

We have developed a Neural Networks technique to identify single top events. In this case the signal to noise is very small (8%), and therefore the NN approach is more suitable than a cut based analysis as it allows combining information from multiple event variables. Rather than using an array of networks [75], we devised a three-output perceptron with one intermediate layer, whose output nodes estimate signal and backgrounds probabilities. We searched for Standard Model single top production in the 106 pb^{-1} of data collected by the Collider Detector at Fermilab (CDF). Performing a Run I a priori study we found an expected 95% C.L. limit on single top cross section of 10.5 pb, roughly 22% lower than that obtained in the previous cut-based CDF analysis [57]. Using a sample of 64 $W + 1, 2, 3$ jets events from Run I, the fit to the Neural Network output yields a signal content of 23.6 ± 7.7 events, approximately 2.5 standard deviations away from the theoretically predicted

value of 4.2 events. We employed a fully Bayesian treatment to set an upper limit on single top production of 24.4 *pb* at 95% C.L., including systematic uncertainties.

A final comment regards the observed excess of single top events. Among the different possible sources, we note that the excess can be caused by a statistical fluctuation, or a slightly larger single top cross section which fluctuated high, or by other processes with similar final state. Our measurement does not have the precision necessary to claim single top observation. However, with the higher statistics of Run II, this method has the potential to yield conclusive results on single top production.

APPENDIX A

EVENT RATIOS IN THE TRAINING SET

A question that often arises regards the particular class proportions that should be used in the training phase. Should we use the relative proportions expected from theory and previous experimental studies? Or should we assume no prior knowledge about the rates for each class? The latter would imply using the same number of events from each class for training. It is definitely a sensible thing to do, especially when there are the process rates that we are trying to measure.

The first argument in the favor of training with equal proportions of training events is purely intuitive: we would like the network to recognize signal or the different backgrounds with similar precision. Take for example the case of single top, whose rate is approximately 10 times smaller than that of the $W+$ jets background. If we were to use 10 times more QCD events for training, than the single top would have very little role in the updating of the weights, essentially being treated as noise by the network.

But does the training sample resemble the real data? The second argument will show that there is a simple transformation which compensates for the differences in the compositions of the training and real data sets. It is important to clearly state the problem, because confusion can easily arise.

The problem. Let us study the example of section 4.3, where our training and testing set contained equal numbers of patterns from each class. This implies that in training/testing $P(C_1) = P(C_2) = P(C_3) = 1/3$. Let us denote NN_0 the perceptron trained with this set. Further, let us assume that in the real world the actual class priors are $P'(C_1), P'(C_2), P'(C_3)$. We want now to find a transformation $NN_0 \mapsto NN'$ that allows us to use the same training/testing set, but reproduce real data output. For example, when we present NN' with 60 training patterns[30], we want to get the same output distribution that NN_0 would produce for 60 real data events.

The solution. As argued in section 4.1 of [44], it turns out that we need not retrain the network with the correct priors $P'(C_i)$. The reason is that the class priors occur only as multiplicative terms in producing the outputs, according to Eqn. (4.10). We can therefore simply multiply the three output nodes by the correct $P'(C_i)$[31] obtained from fitting the Run I data. In this way, we essentially normalize the output distributions for each class i to the corresponding probability $P'(C_i)$ (see Eqn. (A.4)). A rigorous derivation of this transformation is immediate, as shown in what follows.

When the network NN_0 is trained with equal proportions, we have $P(C_1) = P(C_2) = P(C_3) = 1/3$, and therefore Eqn. (4.10) becomes:

$$O_i(x) \approx P(C_i|x) = \frac{\mathcal{P}(x|C_i)}{\mathcal{P}(x)} \cdot \frac{1}{3}, \quad \text{for } i = 1, 2, 3. \tag{A.1}$$

However, given that the actual proportions are in fact $P'(C_1), P'(C_2), P'(C_3)$, the new network NN' should output:

$$O'_i(x) \approx P'(C_i|x) = \frac{\mathcal{P}(x|C_i)}{\mathcal{P}(x)} \cdot P'(C_i), \quad \text{for } i = 1, 2, 3. \tag{A.2}$$

[30] We mean a random selection here; therefore ≈ 20 patterns from each class.

[31] More precisely, the factors should be $P'(C_i)/({}^1/_3)$, as it will be explained later in the text.

Therefore, NN' can be constructed by starting from NN_0 and merely transforming the output values as:

$$O'_i(x) = O_i(x) \cdot 3P'(C_i), \quad \text{for } i = 1, 2, 3. \tag{A.3}$$

What about the sum of the NN' outputs O'_1, O'_2, O'_3? To check this, we integrate Eqn. (A.2) with respect to x:

$$\int_{-\infty}^{\infty} O'_i(x)\mathcal{P}(x)dx \approx \int_{-\infty}^{\infty} \mathcal{P}(x|C_i) \cdot P'(C_i)dx = P'(C_i) \int_{-\infty}^{\infty} \mathcal{P}(x|C_i)dx = P'(C_i) \tag{A.4}$$

Using the above relation to sum over the class index i, we obtain:

$$\int_{-\infty}^{\infty} \left[O'_1(x) + O'_2(x) + O'_3(x)\right]\mathcal{P}(x)dx \approx P'(C_1) + P'(C_2) + P'(C_3) = 1 \tag{A.5}$$

which implies that the *mean* of the output sum is unity. We finally note that NN' was designed to and should be used only for samples containing equal numbers of events from each class. It should not be applied to real data sets, as this will imply accounting twice for the correct class priors.

The interpretation. There are therefore two ways to make the classification equivalent for the cases of real data and training datasets. One way is to fit the real data output as the sum of the three Monte Carlo distributions (one output distribution for each class), and then use the fit parameters $P'(C_1), P'(C_2), P'(C_3)$ when mixing training events to reproduce the real data. The second way is to keep equal proportions of events in the training sample, but transform the outputs of the network as in Eqn. (A.3), which guarantees that the real data output will be correctly modeled. Obviously, the two procedures are fully equivalent.

APPENDIX B

NEURAL NETWORK PERFORMANCE INDICATORS

To quantify how well a Neural Network performs, we use two indicators: performance P and error E.

The performance P is the fraction of events correctly *classified* by the network:

$$P = \frac{N_{correct}}{N_{events}} \tag{B.1}$$

Classification is realized by assigning event k to the class m having the least distance $|\overrightarrow{O^k} - \overrightarrow{T^m}|^2$ among all classes $m = 1, \ldots, M$. The notations of Section 4.2 are used, and $\overrightarrow{T^m}$ are the target output vectors for the $M = 3$ classes. For example, an event which produces the output $\overrightarrow{O} = (0.7, 0.4, 0.6)$ has the following distances:

$$\begin{aligned}
|\overrightarrow{O} - \overrightarrow{T^1}|^2 &= (0.7-1)^2 + (0.4-0)^2 + (0.6-0)^2 = 0.61 \\
|\overrightarrow{O} - \overrightarrow{T^2}|^2 &= (0.7-0)^2 + (0.4-1)^2 + (0.6-0)^2 = 1.21 \\
|\overrightarrow{O} - \overrightarrow{T^3}|^2 &= (0.7-0)^2 + (0.4-0)^2 + (0.6-1)^2 = 0.81
\end{aligned}$$

and is therefore assigned to class 1, which is the QCD (non-top) background. The classification is considered *correct* if the event was indeed originating from the Monte Carlo QCD sample.

Unlike the performance P, which finds the closest target output $\overrightarrow{T^m}$, the error E accounts for how close the correct target output $\overrightarrow{T^k}$ is relative to the actual output

$\overrightarrow{O^k}$. It is therefore a more accurate measure of the NN capability, being defined as:

$$E = \frac{1}{N_{events}} \cdot \sum_{k=1}^{N_{events}} |\overrightarrow{O^k} - \overrightarrow{T^k}|^2 = \frac{1}{N_{events}} \cdot \sum_{k=1}^{N_{events}} \sum_{m=1}^{3} [O_m^k - T_m^k]^2 \qquad \text{(B.2)}$$

where $\overrightarrow{T^k}$ is the target output corresponding to the correct class. Note that in calculating both P and E we assumed that the correct class for each event k is known. Therefore P and E can only be evaluated on Monte Carlo generated samples, and not on real data. We also point out that P and E are highly correlated; a low E implies a high P and vice versa.

Finally, we mention that the allowed ranges for P and E are (0,1) and (0,3) respectively. However, the minimum performance is obtained in the case of purely random classification. For our three-class problem we have: $P_{min} = {}^1/_3$ and $E_{max} = 1$.

APPENDIX C

NEURAL NETWORK USED FOR SINGLE TOP IDENTIFICATION

Tables C.1, C.2 present the set of 190 parameters which define the Neural Network developed for single top event recognition.

Node	E_T^{jet1}	E_T^{jet2}	E_T^{lepton}	$\not{E}_T$	P_T^{jj}	H_T	$Q \times \eta$
h_1	2.479	3.934	-0.287	-0.514	-1.905	-23.658	-2.829
h_2	1.171	-0.338	3.176	0.867	-1.059	-26.491	0.784
h_3	-2.848	-13.358	0.200	-3.323	-6.019	9.408	-1.465
h_4	-3.270	2.765	1.219	9.327	-0.894	26.917	0.340
h_5	-4.099	-8.271	2.922	-3.253	1.412	20.729	0.244
h_6	5.183	-0.433	5.322	3.412	0.895	-26.168	-1.249
h_7	0.989	16.437	-9.306	8.457	0.105	9.110	0.733
h_8	4.265	9.316	-0.784	0.545	-0.600	-19.234	3.610
h_9	9.805	1.502	-1.407	5.326	8.963	28.308	0.086
h_{10}	-2.933	-0.273	-0.788	3.150	-5.392	28.160	0.366
h_{11}	-19.387	-8.953	-6.598	-1.335	-13.492	-21.724	0.182
h_{12}	-4.774	-4.095	-0.968	-2.556	-4.620	21.391	-0.169
h_{13}	6.624	2.262	3.880	3.087	4.203	-26.995	-0.036
h_{14}	6.151	6.365	3.021	2.006	2.439	12.290	-5.623
h_{15}	-8.930	-17.259	13.602	-5.164	4.293	-27.734	0.265
h_{16}	-7.655	-5.274	-1.448	-6.974	-11.927	-3.571	-0.315
h_{17}	-1.622	3.247	-10.288	-17.986	7.680	21.025	0.108

Table C.1: Connection weights between the input and the hidden layers.

Node	O_1	O_2	O_3
h_1	-0.498	4.167	-6.182
h_2	-0.508	2.751	-4.688
h_3	-1.689	-3.433	4.240
h_4	-6.385	-9.119	9.889
h_5	-1.476	-2.423	6.314
h_6	1.518	3.386	-3.918
h_7	-1.521	0.334	0.157
h_8	-8.445	10.160	-10.425
h_9	-0.757	-5.279	2.496
h_{10}	-5.205	-5.343	11.729
h_{11}	13.109	-9.530	-13.340
h_{12}	-3.838	-1.722	5.537
h_{13}	3.089	4.025	-4.911
h_{14}	-1.311	-0.143	0.759
h_{15}	6.037	-7.432	-19.684
h_{16}	0.678	-3.403	-0.212
h_{17}	-7.466	-9.970	12.877

Node	Threshold
O_1	0.743
O_2	0.966
O_3	0.725
h_1	-0.823
h_2	-0.482
h_3	0.799
h_4	-2.131
h_5	-0.705
h_6	-0.159
h_7	0.737
h_8	-1.761
h_9	-1.755
h_{10}	-1.465
h_{11}	0.345
h_{12}	-1.490
h_{13}	0.431
h_{14}	-1.598
h_{15}	-1.235
h_{16}	1.212
h_{17}	-0.821

Table C.2: Connection weights between the output and the hidden layers (left table), and thresholds for the output and hidden nodes (right table).

The activation function for the hidden and output neurons is:

$$g(x) = \frac{1}{2} \cdot [1 + \tanh(x)] = \frac{1}{1 + \mathrm{e}^{-2x}} \tag{C.1}$$

Before we run an event through the network, the seven input variables are transformed as follows: $10^{-3} \cdot E_T^{jet1}(GeV)$, $10^{-3} \cdot E_T^{jet2}(GeV)$, $10^{-3} \cdot E_T^{lepton}(GeV)$, $10^{-3} \cdot \not{E}_T(GeV)$, $10^{-3} \cdot P_T^{jj}(GeV/c)$, $10^{-4} \cdot H_T(GeV)$, and $10^{-1} \cdot Q \times \eta$. This scaling is necessary to ensure that the argument of the exponential in Eqn. (C.1) stays of order $\mathcal{O}(1)$ so that the exponential does not "blow up".

	Classified as QCD	Classified as signal	Classified as $t\bar{t}$
QCD sample	65.8%	26.9%	7.3%
Signal sample	19.3%	67.0%	13.7%
$t\bar{t}$ sample	5.6%	24.1%	70.3%

Table C.3: NN classification rates for Monte Carlo signal and background events.

Table C.3 shows how Monte Carlo signal and background events are classified by our network, using the classification definition given in Appendix B. The overall correct classification rate is $P = 67.7\%$.

APPENDIX D

CALCULATING THE INDIVIDUAL CONFIDENCE LIMIT μ_{95}

The confidence limit μ_{95} for one pseudo-experiment is obtained by solving Eqn. (6.5):

$$\int_0^{\mu_{95}} \mathcal{L}(\mu_s|n_0)\mathrm{d}\mu_s = 0.95 \cdot \int_0^{\infty} \mathcal{L}(\mu_s|n_0)\mathrm{d}\mu_s$$

To solve this equation, we need to know the likelihood $\mathcal{L}(\mu_s|n_0)$. There are two Bayesian methods for obtaining $\mathcal{L}(\mu_s|n_0)$, which we will describe in what follows.

D.1 Standard Method

This simple procedure is the most widely used in upper limit analyses. It involves maximizing Eqn. (6.3) with respect to the background contributions n_{QCD}, $n_{t\bar{t}}$ only, while the signal contribution is fixed to the value $n_{signal} = \mu_s$. In mathematical translation, we maximize:

$$\mathcal{L}''(n_{QCD}, n_{t\bar{t}}) = \frac{\mathrm{e}^{-\frac{(n_{QCD}-\mu_{QCD})^2}{2\sigma^2_{QCD}}}}{\sqrt{2\pi}\sigma_{QCD}} \cdot \frac{\mathrm{e}^{-\frac{(n_{t\bar{t}}-\mu_{t\bar{t}})^2}{2\sigma^2_{t\bar{t}}}}}{\sqrt{2\pi}\sigma_{t\bar{t}}} \prod_{i=1}^{N_{bins}} \frac{e^{-n_i} \cdot n_i^{d_i}}{d_i!} \tag{D.1}$$

where $n_i = \mu_s \cdot f_{signal,i} + n_{QCD} \cdot f_{QCD,i} + n_{t\bar{t}} \cdot f_{t\bar{t},i}$ is the expected number of events in the i-th bin. We then set:

$$\mathcal{L}(\mu_s|n_0) = \mathcal{L}''(n_{QCD}, n_{t\bar{t}})$$

Section 4.3	Section D.2.1
Class C_i	Model μ_s
Input vector $\overrightarrow{x}$	Observed data n_0
PDF $\mathcal{P}(\overrightarrow{x}\|C_i)$	PDF $P(n_0\|\mu_s)$
Posteriori $P(C_i\|\overrightarrow{x})$	Likelihood $\mathcal{L}(\mu_s\|n_0)$

Table D.1: Notation equivalence between Section 4.3 and Section D.2.1.

Note that the dependence on μ_s and n_0 in the RHS of the above relation is given by the n_i and d_i appearing in Eqn. (6.3) defining $\mathcal{L}''$. By its very construction, the function $\mathcal{L}(\mu_s|n_0)$ will peak at $\mu_s = n_0$.

D.2 Alternate Method

There is another Bayesian procedure for obtaining $\mathcal{L}(\mu_s|n_0)$, which can be applied with minimal modifications to a frequentist study (*e.g.*, Feldman-Cousins intervals). The equivalence to the standard approach is described in Section D.3. Before presenting this alternate method, we go through the exercise of re-deriving Eqn. (6.5), with the purpose of introducing the terminology.

D.2.1 Bayesian Approach to Setting a Limit Revisited

We start by introducing a few basic concepts and notations in Bayesian statistics [62]. The concepts were briefly encountered in Section 4.3, but the notations were different. For reference, Table D.1 presents the analogy to the discrete-class ANN study of Section 4.3.

The problem. We wish to make an inference about the single top rate, whose true value μ_s is unknown. As in the case of the Run I data, suppose we do this by

making a single measurement (NN output fit) whose result is n_0 signal events. If the uncertainty in the fitted result n_0 is large, we want to derive an upper limit for μ_s.

The solution. Let us denote $P(n_0|\mu_s)$ the probability of measuring n_0 events given the true value μ_s. In the Bayesian formalism we associate a probability to each possible value of μ_s; these probabilities, or PDF's, represent the degree of belief about μ_s. To make inferences about the single top rate, one looks at the "posteriori" distribution, or likelihood $\mathcal{L}(\mu_s|n_0)$, which is the probability that the true rate is μ_s, given the observed n_0 events ("posteriori" = *after* we measured the n_0 events).

The posteriori $\mathcal{L}(\mu_s|n_0)$ can be used to construct Bayesian intervals $[\mu_1, \mu_2]$ corresponding to a confidence level α by requiring:

$$\int_{\mu_1}^{\mu_2} \mathcal{L}(\mu_s|n_0)d\mu_s = \alpha \cdot \int_0^{\infty} \mathcal{L}(\mu_s|n_0)d\mu_s \tag{D.2}$$

In the present study, we choose $\mu_1 = 0$, and denote $\mu_2 = \mu_\alpha$. We only investigate the case $\alpha = 95\%$, and refer to μ_{95} as the 95% *C.L. confidence limit*. Making these substitutions into the previous equation, we arrive at Eqn. (6.5).

D.2.2 Alternate Method for Deriving the Likelihood

In Section D.1 we have shown how $\mathcal{L}(\mu_s|n_0)$ can be obtained. In what follows we present the *alternate* method for constructing $\mathcal{L}(\mu_s|n_0)$ based on the Bayes theorem. This theorem states that the probability that an element is in both sets A and B is $P(A|B)P(B) = P(B|A)P(A)$. Applying this to our case, we get:

$$\mathcal{L}(\mu_s|n_0) = P(n_0|\mu_s)P(\mu_s)/P(n_0) \tag{D.3}$$

Typically, the denominator is just a normalization constant, and so the major issue is what to use for $P(\mu_s)$, which is called the "prior" distribution. Following the Particle

Data Group review of the Bayesian approach [63], we claim no prior knowledge, so that before the measurement all physically reasonable values of μ_s are equally probable: $P(\mu_s) = const.$ for $\mu_s \geq 0$. The previous equation becomes therefore:

$$\mathcal{L}(\mu_s|n_0) = constant \cdot P(n_0|\mu_s) \tag{D.4}$$

This is precisely how we obtain $\mathcal{L}(\mu_s|n_0)$ from the PDF's $P(n_0|\mu_s)$. We will explain next in which conditions the *constant* factor in Eqn. (D.4) can be ignored.

Implementation. We will describe the steps necessary to derive $\mathcal{L}(\mu_s|n_0)$ using the Alternate Method. In order to perform the integral in Eqn. (6.5) we need to know the function $\mathcal{L}(\mu_s|n_0)$ for $0 \leq \mu_s < \infty$. In practice, we derive $\mathcal{L}(\mu_s|n_0)$ for a discrete set of N values:

$$\mu_s = \mu_1,\ \mu_2,\ \mu_3,\ \ldots,\ \mu_N$$

For a given μ_i in this set, we construct the PDF $P(n_{signal}|\mu_i)$ by running a large number (*e.g.* 10000) pseudo-experiments starting with μ_i signal events, 43.3 ± 8.4 QCD events, and 7.4 ± 2.2 $t\bar{t}$ events respectively, as in Section 6.2. The resulting n_{signal} histogram of 10000 entries represents the un-normalized PDF $P(n_{signal}|\mu_i)$. We repeat this procedure for each value of μ_i ($i = 1, 2, \ldots, N$), and store the resulting N distributions $P(n_{signal}|\mu_i)$, to be referred to as "*likelihood templates*". For greater precision, we smooth these templates and perform a cubic spline fit for each.

Once this is done, we are ready to build the likelihood function $\mathcal{L}(\mu_s|n_0)$ corresponding to a given measurement n_0. From Eqn. (D.4), we have:

$$\mathcal{L}(\mu_i|n_0) = constant \cdot P(n_0|\mu_i), \quad i = 1,\ 2,\ \ldots,\ N \tag{D.5}$$

where $P(n_0|\mu_i)$ are extracted from the N likelihood templates $P(n_{signal}|\mu_i)$ by setting $n_{signal} = n_0$. We point out that since all likelihood templates were constructed by

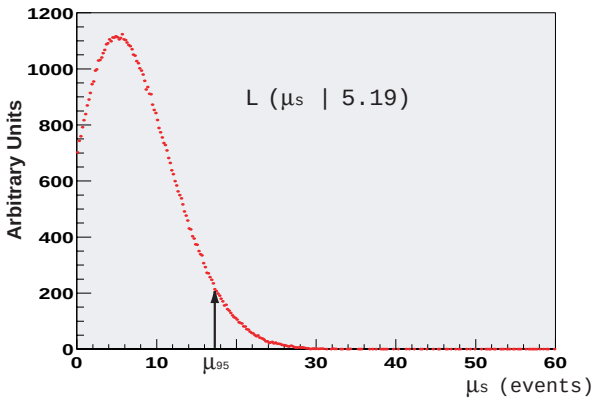

Figure D.1: The likelihood function $\mathcal{L}(\mu_s|5.19)$. The corresponding confidence limit is $\mu_{95} = 17.22$ events.

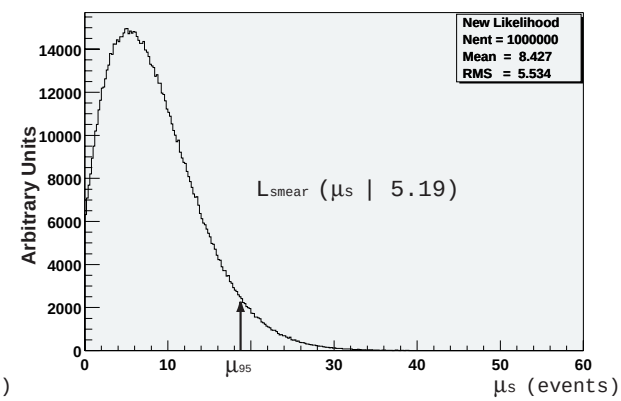

Figure D.2: The smeared likelihood function $\mathcal{L}_{smear}(\mu_s|5.19)$. The corresponding confidence limit is $\mu_{95} = 18.64$ events.

running the same number of pseudo-experiments, there is no need to normalize the individual $P(n_{signal}|\mu_i)$'s. Also, the overall normalization (including the *constant* factor) of the likelihood function $\mathcal{L}(\mu_s|n_0)$ is not important, since it factors out in Eqn. (6.5).

An example. Let us suppose that the fit to one of the 9912 Run I-like pseudo-experiments returns $n_0 = 5.19$ signal events. We look up this n_0 value into each of the N likelihood templates and read out the N values $P(5.19|\mu_i)$, which, as explained previously, are in fact the N likelihood points $\mathcal{L}(\mu_i|5.19)$ (via Eqn. (D.5) where we ignore the *constant* factor). We plot these values[32] as shown in Figure D.1. Next, we integrate the likelihood function $\mathcal{L}(\mu_s|5.19)$ of Figure D.1 and solve Eqn. (6.5) to get the limit: $\mu_{95} = 17.22$ events.

Let us use the same example to illustrate the smearing of the likelihood $\mathcal{L}(\mu_s|5.19)$. Starting from the systematic uncertainties values of $\Delta S = 1.73$ events and $\Delta R = 0.15$,

[32]We have selected $N = 300$ equally spaced μ_i values: $0 < \mu_i < 60$ (events).

we execute the smearing procedure given in Section 6.2.2 one million times. The new likelihood function is denoted by $\mathcal{L}_{smear}(\mu_s|5.19)$, has a higher mean and RMS than $\mathcal{L}(\mu_s|5.19)$, and is presented in Figure D.2. Integrating $\mathcal{L}_{smear}(\mu_s|5.19)$ to solve Eqn. (6.6) yields: $\mu_{95} = 18.64$ events.

D.3 Equivalence between Methods

The methods described in Sections D.1 and D.2.2 give very similar results. By performing a 10000 pseudo-experiment comparison, we found the a priori 95% C.L. limit to differ by 0.07 *pb*. However, we do point out that the Standard Method is theoretically more rigorous, as it calculates $\mathcal{L}(\mu_s|\text{data})$, rather than $\mathcal{L}(\mu_s|n_0)$. By $\mathcal{L}(\mu_s|\text{data})$ we understand:

$$\mathcal{L}(\mu_s|\text{data}) = \mathcal{L}(\mu_s|d_1, d_2, \ldots, d_{N_{bins}})$$

where d_i is the number of data events in the i-th bin of the (x, y) space.

In contrast, the Alternate Method yields $\mathcal{L}(\mu_s|n_0)$. One can imagine that there are multiple experiments $(d_1, d_2, \ldots, d_{N_{bins}})$ which are fitted the same amount of signal n_0. $\mathcal{L}(\mu_s|n_0)$ is therefore an *average* of $\mathcal{L}(\mu_s|\text{data})$ over all *data* sets having n_0 fitted signal events.

APPENDIX E

NEURAL NET OUTPUT FOR RUN I DATA

Tables E.1, E.2, E.3 present the Run and Event numbers for the Run I events passing the final selection requirements described in Chapter 5. The last three columns give the Neural Network output (O_1, O_2, O_3) for the events. As explained in Section 4.3, the values of O_1, O_2, O_3 estimate the Bayes probabilities that the event originates from a QCD, single top, and $t\bar{t}$ process respectively.

Run	Event	O_1 (QCD)	O_2 (signal)	O_3 ($t\bar{t}$)
42640	55874	0.848	0.206	0.010
58400	230496	0.975	0.031	0.002
58485	4226	0.991	0.013	0.001
63127	173868	0.751	0.292	0.008
64021	46988	0.840	0.180	0.001
64311	43078	0.377	0.544	0.202
64700	98497	0.295	0.817	0.007
64880	623030	0.891	0.100	0.004
65426	273116	0.433	0.548	0.112
67561	450608	0.889	0.163	0.004
67842	5949	0.933	0.090	0.005
68110	123757	0.934	0.088	0.003
68685	411056	0.988	0.013	0.001
69007	60885	0.933	0.053	0.000

Table E.1: Run I W+1 jet events passing the final selection of Chapter 5.

Run	Event	O_1 (QCD)	O_2 (signal)	O_3 ($t\bar{t}$)
40190	98182	0.256	0.642	0.174
41627	87219	0.303	0.667	0.193
45776	386857	0.024	0.135	0.897
46357	511399	0.639	0.314	0.039
47439	128290	0.708	0.346	0.044
47689	80060	0.058	0.279	0.729
60705	93795	0.610	0.402	0.039
60766	299452	0.261	0.591	0.037
61167	368226	0.012	0.563	0.407
61377	114526	0.508	0.534	0.030
63603	4029	0.226	0.450	0.141
63883	935	0.242	0.582	0.249
63946	43019	0.444	0.262	0.070
64126	52063	0.125	0.628	0.186
64916	499208	0.471	0.431	0.008
64997	46557	0.120	0.593	0.198
65022	34157	0.589	0.369	0.042
65298	907072	0.321	0.435	0.214
65384	266051	0.076	0.614	0.289
65470	4390	0.584	0.544	0.057
65741	654870	0.084	0.385	0.474
65750	106257	0.376	0.636	0.131
66103	190513	0.454	0.395	0.139
66103	563542	0.600	0.299	0.086
66103	743101	0.044	0.799	0.115
66412	121506	0.173	0.457	0.396
66518	203555	0.113	0.206	0.739
67692	420568	0.115	0.766	0.098
68044	53510	0.123	0.746	0.120
68231	157759	0.724	0.404	0.024
68374	364586	0.344	0.468	0.092
68423	3326	0.148	0.658	0.141
68464	275644	0.613	0.385	0.058
68593	37659	0.250	0.698	0.110
68637	225974	0.260	0.416	0.309
68774	150313	0.332	0.614	0.167
69498	36574	0.342	0.184	0.326
69520	136405	0.915	0.059	0.006
69683	21986	0.046	0.815	0.093
69709	173294	0.223	0.460	0.255
69761	157205	0.455	0.458	0.051

Table E.2: Run I W+2 jet events passing the final selection of Chapter 5.

Run	Event	O_1 (QCD)	O_2 (signal)	O_3 ($t\bar{t}$)
59124	31243	0.224	0.319	0.516
61074	103772	0.034	0.061	0.949
61548	284898	0.039	0.146	0.875
64934	416715	0.326	0.332	0.410
65025	152	0.079	0.289	0.706
65277	209495	0.420	0.264	0.302
67879	407958	0.051	0.554	0.333
67899	82457	0.125	0.173	0.810
70627	56836	0.069	0.179	0.824

Table E.3: Run I W+3 jet events passing the final selection of Chapter 5.

BIBLIOGRAPHY

[1] S. Glashow, Nucl. Phys. **22**, 579 (1961)

[2] S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967)

[3] A. Salam, Proc. 8th Nobel Symp., W. Svartholm, *ed.*, (Almquist and Wiksell, Stockholm 1968).

[4] The Super-Kamiokande Collaboration, Phys. Rev. Lett. 81, 1562 (1998)

[5] N. Cabibbo, Phys. Rev. Lett. 10, 531 (1963)

[6] S. Glashow, J. Iliopulos, and L. Maiani, Phys. Rev. D **2**, 1585 (1970)

[7] M. Kobayashi and T. Maskawa, Progr. Theor. Phys. **49**, 652 (1973)

[8] A. Heister *et al.*, Phys. Lett. B 526: 191-205 (2002)

[9] M. Gaillard, P. Grannis, and F. Sciulli, Rev. Mod. Phys. 71 (1999) S96-S111

[10] T. Affolder *et al.*, Phys. Rev. Lett. 86, 3233 (2001)

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

[12] S. Abachi *et al.*, Phys. Rev. Lett. 74, 2632 (1995)

[13] E. Eichten *et al.*, Rev. Mod. Phys. **56**, 579 (1984)

[14] S. Cortese and R. Petronzio, Phys. Lett. B **253**, 494 (1991)

[15] Z. Sullivan, private communication.

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

[17] T. Stelzer and S. Willenbrock, Phys. Lett. B **357**, 125 (1995)

[18] S. Willenbrock and D. Dicus, Phys. Rev. D **34**, 155 (1986)

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

[20] T. Stelzer, Z. Sullivan, and S. Willenbrock, Phys. Rev. D **58**, 094021 (1998)

[21] D.E. Groom *et al.*, The European Physical Journal **C15** (2000) 1 and 2001 off-year partial update for the 2002 edition available on the PDG WWW pages (URL: http://pdg.lbl.gov/)

[22] J. Pati and A. Salam, Phys. Rev. D **10**, 275 (1974)

[23] K. Whisnant *et al.*, Phys. Rev. D **56**, 467 (1997)

[24] A. Heinson, A. Belyaev, E. Boos, Phys. Rev. D **56**, 3114 (1997)

[25] E. Boos, L. Dudko, and T. Ohl, Eur. Phys. J. C **11**, 473 (1999)

[26] T. Han *et al.*, Phys. Rev. D **58**, 073008 (1998)

[27] S. Bar-Shalom, D. Atwood, and A. Soni, Phys. Rev. D **57**, 1495 (1998)

[28] A. Datta *et al.*, Phys. Lett. B **483**, 203 (2000)

[29] `http://www.fnal.gov/` contains a wealth of information regarding Fermilab.

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

[31] P. Azzi *et al.*, Nucl. Inst. Meth. Phys. Res. Sect. A **360**, 137 (1995)

[32] F. Snider *et al.*, Nucl. Inst. Meth. Phys. Res. Sect. A **268**, 75 (1988)

[33] F. Bedeschi *et al.*, Nucl. Inst. Meth. Phys. Res. Sect. A **268**, 50 (1988)

[34] L. Balka *et al.*, Nucl. Inst. Meth. Phys. Res. Sect. A **267**, 272 (1988)

[35] S. Bertolucci *et al.*, Nucl. Inst. Meth. Phys. Res. Sect. A **267**, 301 (1988)

[36] G. Ascoli *et al.*, Nucl. Inst. Meth. Phys. Res. Sect. A **268**, 33 (1988)

[37] G. Foster *et al.*, Nucl. Inst. Meth. Phys. Res. Sect. A **269**, 93 (1988)

[38] J. Hertz, A. Krogh and R. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, CA (1991)

[39] L. Lönnblad, C. Peterson and T. Rögnvaldsson, Nucl. Phys. B 349, 675-702 (1991)

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

[41] M. Opper and W. Kinzel's article in *Models of Neural Networks III*, E. Domany, J. van Hemmen, and K. Schulten, *eds.*, Springer-Verlag New York, NY (1996)

[42] F. Rosenblatt, *Principles of Neurodynamics* Spartan Books, New York, NY (1962)

[43] R. Lippmann's article in *From Statistics to Neural Networks*, V. Cherkassky, J. Friedman, and H. Wechsler, *eds.*, Springer-Verlag Berlin Heidelberg, Germany (1994)

[44] M. Richard and R. Lippmann, Neural Computation **3**, 461-483 (1991)

[45] K. Tollefson, *A Measurement of the Top Quark Mass*, Ph.D. thesis, University of Rochester (1997)

[46] M. Kruse, *Observation of Top Quark Pair Production in the Dilepton Decay Channel from proton-antiproton Collisions at* $\sqrt{S} = 1.8\ TeV$, Ph.D. thesis, Purdue University (1996)

[47] F. Abe *et al.*, Phys. Rev. D 50, 2966-3026 (1994)

[48] T. Sjöstrand, P. Edén, C. Friberg, L. Lönnblad, G. Miu, S. Mrenna and E. Norrbin, Computer Phys. Commun. 135 (2001)

[49] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, Phys. Rep. **97**, 31 (1983)

[50] P. Avery, K. Read, and G. Trahern, Cornell Internal Note CSN-212, unpublished (1985)

[51] P. Giromini, F. Ptohos and W. Yao, *Search for New Particles* $X \to b\bar{b}$ *Produced in W+2jet Sample*, CDF Internal Note 4112 (1997)

[52] T.Kikuchi L.Demortier S.Kim, CDF Note 5104: *Upper Limit on the Combined Cross Section of Single Top Quark Production*, CDF Internal Note 5104, unpublished (2001)

[53] D. Acosta *et al.*, Phys. Rev. D **65**, 091102 (2002)

[54] L. Lönnblad, C. Peterson, H. Pi, and T. Rögnvaldsson, Comput. Phys. Commun. 81, 185 (1994)

[55] R. Brun and F. Rademakers, Nucl. Inst. Meth. Phys. Res. Sect. A **389**, 81 (1997). See also `http://root.cern.ch/`

[56] C. Ciobanu, R. Hughes, P. Koehn, C. Neu, and B. Winer, *A ROOT Script Interface to JETNET*, CDF Internal Note 5434, unpublished (2000). The Root_Jetnet package can be downloaded from `http://cdfpc2.mps.ohio-state.edu/root_to_jetnet/rtj.html`

[57] S. Wolinski and P. Savard, *An Upper Limit on Standard Model Single Top Production in CDF Run I*, CDF Internal Note 5199, unpublished (2000)

[58] B. Abbott *et al.*, hep-ex/9907041 (1999)

[59] D. Carlson, Ph.D. thesis, hep-ph/9508278 (1995)

[60] See for example [21]. A more accurate treatment, which takes into account the finite size of the Monte Carlo samples, is given at: `http://wwwinfo.cern.ch/asdoc/hbook_html3/node132.html`

[61] R. Cousins, Am. J. Phys. 63, 398 (1995)

[62] G. Feldman and R. Cousins, Phys. Rev. D **57**, 3873 (1998)

[63] F. Solmitz, Annu. Rev. Nucl. Sci. **14**, 375-402 (1964)

[64] F. Abe *et al.*, Phys. Rev. D **43**, 664 (1991)

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

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

[67] H. Lai *et al.*, Phys. Rev. D **51**, 4763 (1995)

[68] G. Marchesini *et al.*, Comput. Phys. Commun. **67**, 465 (1992). We used HERWIG version 5.6.

[69] G. Alner *et al.*, Nucl. Phys. B 291, 445 (1987)

[70] T. Sjöstrand and M. van Zijl, Phys. Rev. D **36**, 2019 (1987)

[71] R. Hughes and T. Liss, *$t\bar{t}$ Cross Section for the SVX B-Tag Analysis*, CDF Internal Note 2992, unpublished (1995)

[72] D. Cronin-Hennessy and A. Beretvas, *Luminosity at CDF*, CDF Internal Note 4721 (1998)

[73] C. Blocker, *Comment On the Feldman and Cousins Method of Limits*, CDF Internal Note 4611, unpublished (1998)

[74] C. Hill, *Measurement of the production cross-section ratio $\sigma_{t\bar{t}j}/\sigma_{t\bar{t}X}$ in 106 ± 4.1 pb^{-1} of data collected by CDF during Run1A+1B of the Tevatron*, Ph. D. thesis, University of California, Davis (2001)

[75] V. Abazov *et al.*, Phys. Lett. B 517, 282 (2001)

[76] B. Abbot *et al.*, Phys. Rev. D **63**, 031101 (2001)

[77] B. Harris *et al.*, *The fully differential single-top-quark cross section in next-to-leading order QCD*, `hep-ph/0207055` (2002)

[78] S. Truitt, *Estimates of Single Top Yield and Signal-to-Background Ratio at CDF in Run II*, presentation given at Fermilab's Top Thinkshop (1998). `http://lutece.fnal.gov/thinkshop/default.html`

[79] C. Ciobanu, Int. J. Mod. Phys. A, Vol. 16, Suppl. 1A, 389 (2001)