Measurement of the Top Quark Mass with Neural Networks

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Carlos Andrés Sánchez, B.S., M.S.

* * * * *

The Ohio State University

2002

Dissertation Committee:

Prof. Richard E. Hughes, Adviser

Prof. Brian L. Winer

Prof. Tom Humanic

Prof. Stephen Pinsky

Approved by

Adviser
Department of Physics

© Copyright by

Carlos Andrés Sánchez

2002

ABSTRACT

At the Tevatron, protons and antiprotons collide with a center-of-mass energy of 1.8 TeV. In this energy range the dominant source of top quarks is the production of $t\bar{t}$ pairs via quark-antiquark annhilation and gluon-gluon fusion. We present our analyses to determine the mass of the top quark reconstructed through the "lepton+jets" decay channel in the 106 $\pm$ 4.1 pb^{-1} of data collected by the Collider Detector at Fermilab (CDF) from 1992-1996.

In the past, the top mass was obtained by comparing the observed kinematic features of top events to those predicted for different top quark masses. These distributions are known as *templates*. While any kinematic variable, which exhibits sensitivity to the top mass can be used to calculate the mass of the top quark, the lowest statistical uncertainty is achieved by reconstructing the top mass from the decay products of the $t\bar{t}$ pair. It was also observed that it is possible to obtain a better estimate of the top mass by fitting the different templates to a smooth function. Previously, we have used a combination of a Gaussian and a Gamma function to fit the distributions. In this analysis, we find that using a Neural Network (NN) to fit the distribution gives slightly better results, and that the NN fitting technique is applicable to any kinematic variable. Following this recipe the top mass is measured to be $177.9 \pm 4.7(stat.) \pm 4.6(syst.)$ GeV/c^2 when using the reconstructed mass, M_{rec}.

When we use the total transverse energy of the events, $\mathrm{H_T}$, the mass of the top quark is found to be $204.4 \pm 9.2(stat.) \pm 9.2(syst.)$ GeV/c^2.

As noted, there are different kinematic variables that can be used to calculate the top mass. A Neural Network provides a simple and elegant way of combining all of these variables which have mass information. The idea is that combining the information from more than one kinematic variable would result in a more accurate measurement of the top mass. Therefore, this NN based technique uses a combination of M_{rec}, H_T, the invariant mass of the $t\bar{t}$ system, $M_{t\bar{t}}$, and the sum of the P_T's of the two leading jets, $P_T(1) + P_T(2)$. These variables were chosen because they exhibit the greatest mass dependence. The Neural Networks attempt to classify the events as $t\bar{t}$ signal or background. For each event the NN provides a set of probabilities that it has come from any of the top masses used in this analysis, as well as background. Using this information we construct a discrete likelihood function from which the top mass is calculated to be $181.9 \pm 5.1(stat.) \pm 5.2(syst.)$ GeV/c^2.

Dedicado a mis abuelos

ACKNOWLEDGMENTS

I would like to thank Richard Hughes and Brian Winer for being the best advisers I could have asked for. I would have never finished my Ph.D. if it wasn't for you. Thanks to Catalin Ciobanu for making my life in the lab more enjoyable. I want to thank Phil Koehn, Evelyn Thomson and Jong-young Chung for all their help. Thanks to Chris Neu for all the times we spent together. Also, many thanks to all the people who have made the CDF experiment a success.

Outside the physics world, I would like to thank Michael for all the things that make him the "Man of the Year". Gracias a Nacho y Mercedes por su amistad y por no volverme completamente loco. Special thanks to Robyn for always making me laugh and for always being there for me.

Outside the U.S world, quiero darle las gracias a mi familia especialmente a mis padres y a mis hermanos Esther y Alvaro por todo su apoyo y su amor. Sin vosotros no seria nadie. Gracias a mis amigos de la juventud: Eduardo, Jorge, Sergio y Arturo por siempre estar ahi cuando mas lo he necesitado.

J & K's are on me.

TABLE OF CONTENTS

Page

Abstract ii

Dedication iv

Acknowledgments v

List of Tables x

List of Figures xiii

Chapters:

1. Introduction 1

2. Theory 3

   2.1 The Standard Model 3
   2.2 Constraining the Higgs Mass with the Top Mass 5
   2.3 Pair Production of Top Quarks 7
   2.4 Top Hadronization and Decay 9
   2.5 Decay Signature of the Lepton+jets Channel 10

3. The CDF Experiment at the Tevatron 12

   3.1 The Tevatron 12
   3.2 The CDF Detector 14
      3.2.1 Tracking Detectors 16
      3.2.2 Calorimeters 19
      3.2.3 Muon Systems 22
      3.2.4 Event Triggers 22

4. Introduction to Neural Networks 25

4.1 Historical Background 25
4.2 The Multi-layer Perceptron 28
4.3 Understanding Artificial Neural Network 30
4.3.1 The Perceptron Learning Rule 30
4.3.2 The Convergence Theorem 31
4.3.3 The Back-propagation Algorithm 32
4.4 The Universal Approximation Theorem 34
4.4.1 Neural Networks and Fourier Series 35
4.5 NN performance 35

5. Event Sample 38

5.1 High P_T Leptons 38
5.2 Missing Transverse Energy, $\not\!\!E_T$ 40
5.3 Jet Reconstruction 40
5.3.1 Secondary Vertex Tagging of B Jets 41
5.3.2 Soft Lepton Tagging of B Jets 41
5.4 Top Mass Candidate Sample 43
5.4.1 Mass Subsamples 44
5.5 Backgrounds 45
5.6 Monte Carlo Simulation 47
5.6.1 Top Samples 47
5.6.2 Background Sample 48

6. Mass Reconstruction 49

6.1 Jet Energy Corrections 49
6.2 The Unclustered Energy 51
6.3 Event Reconstruction 53
6.3.1 Combinatorics 55
6.3.2 Gluon radiation 56

7. Template-based Mass Analysis 58

7.1 Neural Network Template Parametrization 58
7.1.1 Fitting the M_{rec} Distributions 59
7.1.2 Fitting the H_T Distributions 62
7.2 Background Calculation 68
7.3 Likelihood Procedure 68
7.4 Results from Simulated Experiments 70

7.5 Combining the Top Mass Results 74

8. Standalone Neural Network Mass Analysis 76

8.1 Input Variables 76
8.2 Extracting M_{top} with a Neural Network 81
8.2.1 Neural Network Architecture 82
8.2.2 Training and Testing 82
8.2.3 Likelihood Procedure 84
8.2.4 Results from Pseudoexperiments 86

9. Systematic Uncertainties 88

9.1 Jet energy scale 88
9.2 Initial and final state radiation 90
9.3 b-Tagging 92
9.4 Parton distribution function 92
9.5 Monte Carlo generators 92
9.6 Summary of Systematic Errors 93

10. Measurement of the Gluon Content 94

10.1 Gluon Effect on the Top Mass 94
10.1.1 Gluon Fraction in the Monte Carlo 95
10.1.2 Pseudoexperiments 97
10.2 Measuring the Gluon Content 99
10.2.1 Neural Network Input Variables 99
10.2.2 Neural Network Training and Testing 100
10.2.3 Definition of the Likelihood 103
10.2.4 Pseudoexperiments 104
10.3 Measurement of the Gluon Fraction in the Run I Data 105
10.4 Systematic Errors 107
10.5 Application to the Mass Analysis 108

11. Run I Results 110

11.1 Summary of Monte Carlo Results 110
11.2 Template-Based Data Results 110
11.3 Comparison of Data to Expectations 115
11.4 Standalone NN Run I Data Results 117
11.5 Comparison of Data to Expectations 118

12. Measuring the Top Mass in Run II 120

12.1 Motivation 120
12.2 M_{rec} Vs. $\mathrm{H_T}$ Templates 121
12.3 Fitting Procedure 122
12.4 Run II Expectations 123

Appendices:

A. The Delta Rule 126

B. Back-propagation 128

C. Fitting with a Neural Network 131

D. VECBOS Background 134

D.1 Background Shape Effect on NN Template Method 135
D.2 Background Shape Effect on Standalone NN Analysis 135

E. Combining Measurements 138

E.1 Information 139

F. Generation of "Fake" Monte Carlo 142

Bibliography 145

LIST OF TABLES

Table **Page**

2.1 Results of several different $p\bar{p} \to t\bar{t}$ calculations, for $M_{top} = 175$ GeV and $\sqrt{s} = 1.8$ TeV. 8

2.2 Branching ratios for the three different $t\bar{t}$ decay modes. Here q stands for a u, d, c, or s, and ℓ stands for e and μ. τ leptonic decays are not included. 10

3.1 Acceleration stages for protons and antiprotons. 13

3.2 Typical instantaneous, best instantaneous, and integrated luminosities during Run I at the Tevatron (integrated L corresponds to the CDF interaction region). 14

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. 21

3.4 The physical properties for the plug and forward calorimeters. In the last row, the symbol $\oplus$ indicates that the constant term is to be added in quadrature to the resolution. 21

5.1 Subsamples used in the lepton+jets mass analysis and the expected signal to background ratio (S/B) for each. See Section 7.2 for background estimates for these subsamples. 45

6.1 Efficiency of the χ^2 cut for each of the mass subsamples. The probability of finding a solution with $\chi^2 < 10.0$ increases as the number of possible configurations increases. 56

7.1 We show the χ^2 per degree of freedom for each of the four subsamples. The results from both fitting techniques are given. 61

7.2 Number of background events for the four mass subsamples from various processes. 69

7.3 These are the results obtained by using the NN fitting technique and the GG fitting method for the reconstructed mass. We only show the results from the H_T NN templates since we were unable to obtain continuous templates with the previous Gaussian plus Gamma method. 71

8.1 We show the *slope* and RMS values for the single SVX-tagged events. 77

8.2 Neural Network architecture. The choice of variables depends solely on the network's performance. 82

9.1 Summary of the systematic uncertainties associated with each of the different Neural Network analysis methods described in this thesis. In parenthesis we give the (ISR/FSR) individual systematic errors. . . . 93

10.1 We show the results from the 1000 pseudoexperiments for the Single SVX subsample. In each of the pseudoexperiments, we draw 15 events from the templates, which we have constructed with the appropriate amount of gluon events, and we compare them to the default Monte Carlo templates. Background is not included. 98

10.2 We show the results from the 1000 pseudoexperiments for the Double SVX subsample. In each of the pseudoexperiments, we draw 5 events from the templates, which we have constructed with the appropriate amount of gluon events, and we compare them to the default Monte Carlo templates. Background is not included. 98

10.3 We show the expected number of events without the χ^2 cut. In parenthesis we show the expected number of events after the χ^2 cut. These numbers are taken from Table 7.2. 104

10.4 This is a summary of the Run I, and Run IIa pseudoexperiment results. 106

10.5 Median top mass and median statistical MINOS error results for the 2000 pseudoexperiments using Run I statistics. 109

11.1 Summary of the Monte Carlo results. The (gluon) description means that the templates were constructed with the gluon content measured in the data. 111

11.2 Results of applying the mass likelihood procedure using the GG parametrized M_{rec} templates to the four subsamples and for all subsamples combined. We show the results when the background fractions are constrained to their expected values (Table 7.2), and when they are not. The combined background fraction is the average of the x_b fit results weighted by the number of events in the subsamples. 112

11.3 Results of applying the mass likelihood procedure using the NN parametrized M_{rec} templates to the four subsamples and for all subsamples combined. We show the results when the background fractions are constrained to their expected values (Table 7.2), and when they are not. The combined background fraction is the average of the x_b fit results weighted by the number of events in the subsamples. 113

11.4 Results of applying the mass likelihood procedure using the $\mathrm{H_T}$ templates to the four subsamples and for all subsamples combined. We show the results when the background fractions are constrained to their expected values (Table 7.2), and when they are not. The combined background fraction is the average of the x_b fit results weighted by the number of events in the subsamples. 114

11.5 Results of applying the standalone NN method to the four subsamples and for all subsamples combined. We show the results when the background fractions are constrained to their expected values (Table 7.2), and when they are not. The combined background fraction is the average of the x_b fit results weighted by the number of events in the subsamples. 118

12.1 We show the χ^2 per degree of freedom for each of the four subsamples. The NN does an excellent job at approximating the 2-D surfaces. . . 122

D.1 We give K-S tests results for each of the VECBOS subsamples. We note that the SVX Single background is used for both SVX subsamples. 135

F.1 We show the correlation factor between M_{rec} and H_T for all the different Monte Carlo $t\bar{t}$ samples. 143

LIST OF FIGURES

Figure **Page**

2.1 Self-coupling loops contributing higher order quantum corrections. . . 6

2.2 Lowest order processes for $t\bar{t}$ pair production. 7

2.3 Lepton+jets $t\bar{t}$ event. One of the W bosons decays into a lepton+neutrino pair while the other one decays into a quark-antiquark pair. 11

3.1 The accelerator complex at Fermilab. 13

3.2 A quarter view of the CDF showing the locations of the different detector subcomponents. 15

3.3 Schematic view of an SVX barrel. 16

3.4 Transverse view the Central Tracking Chamber (CTC) which shows the alternating superlayers of axial and stereo cells. 18

3.5 A central calorimeter wedge. There are 24 wedges covering $\Delta\phi = 15°$ on either side of the transverse plane $z = 0$. 19

3.6 The location of the CMU system. Inside each calorimeter wedge there is a 2.4° ϕ interval not covered by the CMU. 23

4.1 A biological neuron. The inset shows the synapse which is a specialized connection between neurons. 26

4.2 (Left) McCulloch-Pitt model for the computational neuron. The different weights are either excitatory $w_i = +1$ or inhibitory $w_i = -1$. (Right) Simple graphical represention of a perceptron with one output and two inputs. 27

4.3 A multi-layer network with l layers of units. The information goes from the input units (left) to the output units (right). 29

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 [24]). 42

6.1 The $t\bar{t}$-specific corrections applied to jets according to available b-jet information. The curves show the fractional change to the corrected jet $\mathrm{P_T}$ after all the generic jet corrections have been applied. The curves are for: (A) jets from the decay of W bosons, (B) jets from all b quarks, (C) jets from b quarks containing an electron, and (D) jets from b quarks containing a muon. 52

7.1 We show a 2-D histogram of all the Single SVX mass templates as a function of top mass. The M_{rec} axis is normalized to be between 0 ($M_{rec} = 80$ GeV/c^2) and 1 ($M_{rec} = 380$ GeV/c^2). The M_{top} axis is also normalized to be between 0 ($M_{top} = 120$ GeV/c^2) and 1 ($M_{top} = 230$ GeV/c^2). 60

7.2 Reconstructed mass distributions for single SVX events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the GG templates (green) and the NN templates (blue) are also shown. 63

7.3 Reconstructed mass distributions for double SVX events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the GG templates (green) and the NN templates (blue) are also shown. 63

7.4 Reconstructed mass distributions for SLT events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the GG templates (green) and the NN templates (blue) are also shown. 64

7.5 Reconstructed mass distributions for events with no SLT or SVX tags in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the GG templates (green) and the NN templates (blue) are also shown. 64

7.6 $\mathrm{H_T}$ distributions for single SVX events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the NN templates are also shown. 65

7.7 $\mathrm{H_T}$ distributions for double SVX events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the NN templates are also shown. 65

7.8 $\mathrm{H_T}$ distributions for SLT events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the NN templates are also shown. 66

7.9 $\mathrm{H_T}$ distributions for events with no SLT or SVX tags in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the NN templates are also shown. 66

7.10 Reconstructed mass distributions for VECBOS background, which contains events generated with $Q^2 = \langle P_T \rangle^2$ and $Q^2 = \langle M_W \rangle^2$ (see Appendix D). The fits from the GG templates (green) and the NN templates (blue) are also shown. 67

7.11 $\mathrm{H_T}$ distributions for VECBOS background, which contains events generated with $Q^2 = \langle P_T \rangle^2$ and $Q^2 = \langle M_W \rangle^2$ (see Appendix D). The fits from the NN templates are also shown. 67

7.12 Results for 2000 pseudoexperiments using M_{rec} NN templates. (Left) Fitted top mass distribution. (Right) Pull distribution. Signal and background fractions as observed in the Run I data. 72

7.13 Results for 2000 pseudoexperiments using M_{rec} GG templates. (Left) Fitted top mass distribution. (Right) Pull distribution. Signal and background fractions as observed in the Run I data. 73

7.14 Results for 2000 pseudoexperiments using H_T NN templates. (Left) Fitted top mass distribution. (Right) Pull distribution. Signal and background fractions as observed in the Run I data. 73

7.15 Results from 2000 pseudoexperiments in which we combine the top mass measurements obtained using the M_{rec} templates and the H_T templates. The left histogram shows the top mass distribution, and the right histogram shows the pull distribution. 74

8.1 (Left) Mean and (Right) RMS, of the reconstructed mass distribution versus the generated top mass. 78

8.2 (Left) Mean and (Right) RMS, of the H_T distribution versus the generated top mass. 78

8.3 (Left) Mean and (Right) RMS, of the invariant $t\bar{t}$ mass distribution versus the generated top mass. 79

8.4 (Left) Mean and (Right) RMS, of the $P_T(1)+P_T(2)$ distribution versus the generated top mass. 79

8.5 (Left) Mean and (Right) RMS, of the missing E_T distribution versus the generated top mass. 80

8.6 (Left) Mean and (Right) RMS, of the lepton P_T distribution versus the generated top mass. 80

8.7 (Left) Sum over the output values of the Neural Network. (Right) Shows the training and testing error. The inset shows a close view of the testing error curve. When this error begins to increase the training is stopped. 84

8.8 (Left) The -log $\mathcal{L}$ distribution for a given Run I standalone NN pseudoexperiment. (Right) Close-up of the left pseudoexperiment. The black curve is the 3rd degree polynomial used to calculate the mass and uncertainties. The red line shows the point at which -log $\mathcal{L}$ changes by 0.5 units. 86

8.9 Results for 2000 pseudoexperiments using the standalone NN method. (Left) Fitted top mass distribution. (Right) Pull distribution. Signal and background fractions as observed in the Run I data. 87

9.1 M_{rec} distribution for Single SVX tagged events from the PYTHIA $t\bar{t}$ Monte Carlo simulation using an input top mass of 175 GeV/c^2 for events with (a) a -1σ shift in the jet momenta, (b) no shift in the jet momenta, and (c) a $+1\sigma$ shift in the jet momenta. 89

9.2 M_{top} distribution returned from the likelihood procedure from simulated experiments using the NN-fitted M_{rec} templates. The solid histogram shows the distribution when the jet E_T scale is shifted down by -1σ, and the dashed histogram shows the distribution when the jet E_T scale is shifted up $+1\sigma$. The systematic uncertainty is half the difference between the medians of the distributions. 90

10.1 Reconstructed top mass distributions for the single SVX Monte Carlo sample for (Left) non-gluon events and (Right) gluon events. 96

10.2 Reconstructed top mass distributions for the double SVX Monte Carlo sample for (Left) non-gluon events and (Right) gluon events. 96

10.3 The four input variables for the SVX Single Neural Network. (a) Di-jet invariant mass between the two most energetic untagged jets. (b) Di-jet invariant mass between the most and least energetic untagged jets. (c) χ^2 returned by the top mass kinematic fit. (d) Number of extra jets with $E_T > 8.0$ GeV and $|\eta| < 2.4$. Background is not included. 101

10.4 The three input variables for the SVX Double Neural Network. (a) Di-jet invariant mass between the two untagged jets. (b) χ^2 returned by the top mass kinematic fit. (c) Number of extra jets with $E_T > 8.0$ GeV and $|\eta| < 2.4$. Background is not included. 102

10.5 (Left) NN output distribution for the Single SVX subsample. (Right) NN output distributions for the Double SVX subsample. The single SVX-tagged background is used for both subsamples. 102

10.6 Results from 5000 pseudoexperiments for (Left) fraction of gluon events and (Right) fraction of non-gluon events. 106

10.7 Pull distribution from 5000 pseudoexperiments for (Left) fraction of gluon events and (Right) fraction of non-gluon events. 106

10.8 Reconstructed mass distribution for (Left) Single SVX events and (Right) Double SVX events. The gluon fraction in the Monte Carlo is set to that measured in the Run I data. The dashed line histogram has the standard MC gluon fraction ($\sim 50\%$). Signal and background fractions as observed in the Run I data. 107

11.1 Results from applying the likelihood procedure to the combined four subsamples using the NN-fitted M_{rec} templates. The figure shows the Run I data (histogram), fitted background and signal (shaded region), fitted signal (blue curve), and fitted background (red curve). 113

11.2 Results from applying the likelihood procedure to the combined four subsamples using the $\mathrm{H_T}$ templates. The figure shows the Run I data (histogram), fitted background and signal (shaded region), fitted signal (blue curve), and fitted background (red curve). 114

11.3 Distribution of the difference between the top mass obtained using the M_{rec} templates and the top mass obtained using the $\mathrm{H_T}$ templates from simulated experiments of $t\bar{t}$ (M_{top} = 175 GeV/c^2) plus background. Also shown is the difference obtained from the data sample. 115

11.4 Statistical uncertainties from pseudoexperiments of $t\bar{t}$ (M_{top} = 175 GeV/c^2) plus background using (Left) the NN M_{rec} templates, and (Right) the $\mathrm{H_T}$ templates. Also shown in the statistical uncertainty obtained from the data sample. 116

11.5 Distribution of the minimum value of -log $\mathcal{L}$ from pseudoexperiments of $t\bar{t}$ ($M_{top} = 175$ GeV/c^2) plus background using (Left) the M_{rec} templates, and (Right) the $\mathrm{H_T}$ templates. Also shown in the minimum -log $\mathcal{L}$ value obtained from the data sample. 117

11.6 Distribution of the correlation coefficient between the different variables used in the standalone NN mass analysis along with the correlation value seen in the data sample. (a) Correlation between M_{rec} and $\mathrm{H_T}$, (b) correlation between $P_T(1) + P_T(2)$ and $\mathrm{H_T}$, (c) correlation between $P_T(1) + P_T(2)$ and M_{rec} , (d) correlation between $P_T(1) + P_T(2)$ and $M_{t\bar{t}}$, (e) correlation between $\mathrm{H_T}$ and $M_{t\bar{t}}$, and (f) correlation between M_{rec} and $M_{t\bar{t}}$. 119

12.1 M_{rec} Vs. H_T surfaces generated from the functional form obtained by fitting the "fake" Monte Carlo Single SVX subsample. (a) Contour plot for M_{top} = 150 GeV/c^2. (b) Contour plot for M_{top} = 175 GeV/c^2. (c) Contour plot for M_{top} = 200 GeV/c^2. (d) Surface plot for M_{top} = 175 GeV/c^2. 123

12.2 We apply the different template-based analyses methods to 2000 simulated experiments. The graph shows the median error obtained from the pseudoexperiments containing 1, 2, 4, 6, and 10 times the Run I statistics. For the standalone NN method we only show the results from the Run I simulated experiments. 124

C.1 We show the M_{rec} distributions for three different top masses. All of the histograms have been normalized to the same number of events. . 132

D.1 Comparisons of the VECBOS background generated with two different values of Q^2 for SVX-tagged events for (a) reconstructed top mass, M_{rec} , (b) $\mathrm{H_T}$, (c) $t\bar{t}$ invariant mass, $M_{t\bar{t}}$, and (d) $P_T(1) + P_T(2)$. . . . 136

E.1 (Left) Shows $\sigma^2_{\bar{s}}(\rho)$. The function is not linear and it is equal to 0 when the $\rho = +1$ or -1. (Right) The Fisher information is shown as a function of ρ. The information is maximum when the measurements are 100% correlated (values not shown because they approach infinity). 140

F.1 (a) Lego histogram of our "fake" MC distribution for M_{top} = 175 GeV/c^2. (b) Contour plot of the same distribution. (c) Projection along the M_{rec} axis of our 2-D plot. The blue curve corresponds to the 1-D NN fit of the true MC for a top mass of 175 GeV/c^2. (d) Projection of the 2-D plot along the H_T axis. The blue curve corresponds to the 1-D NN fit of the true MC for a top mass of 175 GeV/c^2. 144

CHAPTER 1

INTRODUCTION

In 1995, the top quark was discovered at the Fermilab Tevatron. This was the culmination of nearly two decades of intense research at particle accelerators around the world. Also, it was a major triumph for the Standard Model of particle physics since it predicted the top quark existence. The top quark is by far the heaviest fundamental particle known, and this large mass has caused much excitement in the physics community. Since the top quark mass is close to the electroweak scale, it is believed that it plays a role in the breaking of the electroweak symmetry, and therefore the origin of the fermion masses.

Previous mass analyses have exploited the use of a single kinematic variable to extract the mass of the top quark. This thesis explores the use of Neural Networks to improve the top mass measurement. Neural Networks are applied in two distinct ways:

- Neural Networks are used as function approximators to obtain a functional form, which describes any distribution as a function of top mass.
- Neural Networks are used for event classification. This technique allows us to combine information from different variables.

The Standard Model is summarized in Chapter 2 with specific attention given to the production and decay of the top quark. Chapter 3 presents a description of the CDF detector, emphasizing the subsystems most important in this analysis. Artificial Neural Networks are introduced in Chapter 4. Chapter 5 describes the selection criteria used to obtain the event sample used in this analysis. The top mass reconstruction algorithm is presented in Chapter 6. Here, we also discuss the jet energy corrections employed in the mass analysis. The description of the standard template-based likelihood method used to extract the top mass is given in Chapter 7. We introduce a new Neural Network method to parametrized the M_{rec} templates. This technique is compared to the previous Gaussian + Gamma parametrization method. The NN fitting technique is also used to parametrize the $\mathrm{H_T}$ distributions. Chapter 8 presents the standalone Neural Network technique, which combines the mass information from multiple kinematic variables to extract the mass of the top quark. The systematic uncertainties are presented in Chapter 9. In Chapter 10, we describe a method to measure the gluon content in the Run I SVX-tagged data sample. The Run I results from the different analyses presented in this thesis are given in Chapter 11. Finally, in Chapter 12 we describe a powerful technique to measure the top mass using the large statistical samples of Run II.

CHAPTER 2

THEORY

All known particle physics phenomena are extremely well described within the Standard Model (SM) of elementary particles and their fundamental interactions. The SM provides a very elegant theoretical framework and it has successfully passed very precise tests in a variety of high-energy interactions. Although no significant discrepancies between experimental data and the Standard Model have yet been found, there are several critical issues which remain unsolved. For instance, the Standard Model does not predict the masses of the fundamental particles nor provides any information why these masses are so different. Furthermore, the Higgs mechanism by which the fundamental particles acquire their mass is completely artificial. Hence, physics beyond the Standard Model seems inevitable and it is possible that the top quark might be the window to that new physics.

2.1 The Standard Model

This is a simple sketch of the Standard Model of particle physics. There are several books available to the reader that give a much more thorough description [1]. The Standard Model is a quantum field theory that is based on the gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$. The first gauge group, $SU(3)_C$, corresponds to

the symmetry of the strong interactions, and the second and third gauge groups, $SU(2)_L \times U(1)_Y$, correspond to the symmetry of the electroweak interactions. The group symmetry of the electromagnetic interactions, $U(1)_{em}$, appears in the SM as a subgroup of $SU(2)_L \times U(1)_Y$ and it is in this sense that the weak and electromagnetic interactions are said to be unified.

In the SM there are two kinds of fundamental particles: *fermions* and *bosons*. Fermions are spin-1/2 particles, which are the constituents of matter, and they are divided into leptons (ℓ) and quarks (q). The known leptons are: the electron, e^-, the muon, μ^- and the τ^- with electric charge Q = -1; and the corresponding neutrinos ν_e, ν_μ, and ν_τ with electric charge Q = 0. The known quarks are of six different flavors: u, d, c, s, t, and b with fractional electric charge Q = $\frac{2}{3}$, $-\frac{1}{3}$, $\frac{2}{3}$, $-\frac{1}{3}$, $\frac{2}{3}$, and $-\frac{1}{3}$ respectively. Quarks also carry a *color* charge labelled red, green or blue for reference. We know that color is not seen in Nature and therefore elementary quarks must bind into colorless composite particles named hadrons. Fermions are organized in three families with identical properties except for mass. The particle content in each family is:

$$1^{st} \text{ family: } \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, e_R^-, \begin{pmatrix} u \\ d \end{pmatrix}_L, u_R, d_R$$

$$2^{nd} \text{ family: } \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \mu_R^-, \begin{pmatrix} c \\ s \end{pmatrix}_L, c_R, s_R$$

$$3^{rd} \text{ family: } \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L, \tau_R^-, \begin{pmatrix} t \\ b \end{pmatrix}_L, t_R, b_R$$

and their corresponding antiparticles.

The second kind of elementary particles, bosons, are the mediators of the fundamental forces. By leaving apart the gravitational interaction, all relevant interactions

in particle physics are known to be mediated by spin-1 bosons. The photon, γ, is the exchange particle in electromagnetism. The photon is chargeless, massless and it does not interact with itself. The strong force is mediated by eight different gluons. Gluons are massless, electrically neutral and carry color which means that they not only interact with quarks but also with themselves. The $W^{\pm}$ and Z particles mediate the weak interaction. The weak bosons are massive and they are self-interacting. The $W^{\pm}$ is charged with $Q = \pm 1$ and the Z is electrically neutral.

In the Standard Model, particles acquire their mass via the "Higgs mechanism". In order to accommodate electroweak and flavor symmetry breaking, we must artificially introduce a weak-isospin doublet of fundamental scalar field $\Phi = \binom{\phi^+}{\phi^0}$ with a potential of the form

$$V(\Phi^\dagger\Phi) = \mu^2(\Phi^\dagger\Phi) + |\lambda|(\Phi^\dagger\Phi)^2, \tag{2.1}$$

where λ is the self coupling of the scalar field. If we choose μ^2 to be negative, the electroweak symmetry is spontaneously broken when the field is expanded about its non-zero vacuum expectation value $v = \sqrt{-\mu^2/\lambda} = (G_F\sqrt{2})^{-\frac{1}{2}} = 246$ GeV. This value is referred to as the electroweak scale. The spontaneous symmetry breaking gives mass to the W^+, W^-, and Z^0 particles, and it also gives rise to a spin-0 (scalar) particle called the Higgs boson. Each quark and lepton has its own Yukawa coupling to the Higgs boson G_f and thus acquires a mass $m_f = G_f v/\sqrt{2}$, where f stands for fermion.

2.2 Constraining the Higgs Mass with the Top Mass

The mass of the top quark is one of the most important parameters of the Standard Model. It enters into calculations of higher-order (radiative) corrections which connect

several other Standard Model input parameters. Thus, radiative corrections to many electroweak processes depend on the masses of the top quark and the Higgs boson via loop diagrams such as those shown in Figure 2.1.

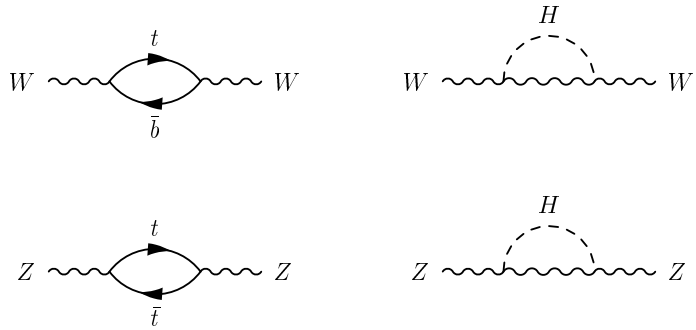

Figure 2.1: Self-coupling loops contributing higher order quantum corrections.

At one loop, for instance, the ρ parameter,

$$\rho = \frac{M_W^2}{M_Z^2(1-\sin^2\theta_W)} \equiv 1 + \Delta r, \tag{2.2}$$

which relates the W and Z boson masses and the weak angle, gets a radiative correction

$$\Delta r = \frac{3G_F}{8\pi^2\sqrt{2}} M_{top}^2 + \frac{\sqrt{2}G_F}{16\pi^2} M_W^2 \left[\frac{11}{3}\ln\left(\frac{M_H^2}{M_W^2}\right) + \ldots\right] + \ldots, \tag{2.3}$$

which is quadratic in the top mass. Note, however, that the dependence on the mass of the Higgs boson is only logarithmic. Therefore, the mass of the top quark is the dominant term in the corrections for electroweak processes. By measuring M_{top} very accurately, and using additional constraints from the large body of precision electroweak data we can test for consistency and predict unknowns in the Standard Model. One of such predictions is the mass of the Higgs boson, which can be constrained by the direct measurement of the W boson and the top quark along with other precision data.

2.3 Pair Production of Top Quarks

In $p\bar{p}$ collisions at a center-of-mass energy of 1.8 TeV, top quarks are expected to be produced by two distinct processes: $q\bar{q}$ annhilation and gluon-gluon (gg) fusion. Figure 2.2 shows the Feynman diagrams for these processes. At the Tevatron, the relative contributions from these processes are about 90% and 10% respectively. For a complete derivation of heavy quark production there are several reviews available to the reader [2]. Here, only the most important issues of these calculations will be presented.

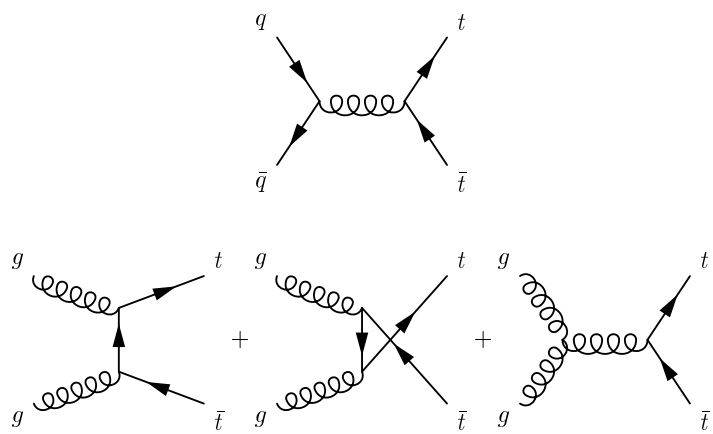

Figure 2.2: Lowest order processes for $t\bar{t}$ pair production.

The total cross section for the production of heavy quarks has the form,

$$\sigma(p\bar{p} \to t\bar{t}) = \sum_{i,j} \int dx_1 dx_2 f_i^p(x_1, \mu^2) f_j^{\bar{p}}(x_2, \mu^2) \hat{\sigma}(ij \to t\bar{t}; \hat{s}, \alpha_s(\mu^2), M_{top}), \qquad (2.4)$$

where the sum is over all partons: gluons, light quarks and antiquarks. This formula expresses the total cross section in terms of the parton-parton processes $ij \to t\bar{t}$. The parton distribution function f_i^p corresponds to the probability density of finding a parton with a given fraction of the proton momentum between x_1 and $x_1 + dx_1$.

Calculation	Type	Structure Function	$\sigma_{t\bar{t}}$
(1) Exact NLO[3, 4]	NLO only	MRSR2[5]	$4.87^{+0.30}_{-0.56}$ pb
(2) LSvN[6]	Resummed	MRSD′[7]	$4.94^{+0.71}_{-0.45}$ pb
(3) BC[8]	Resummed	CTEQ3[9]	$5.52^{+0.07}_{-0.42}$ pb
(4) BCMNT[4]	Resummed	MRSR2	$5.06^{+0.13}_{-0.36}$ pb

Table 2.1: Results of several different $p\bar{p} \to t\bar{t}$ calculations, for $M_{top} = 175$ GeV and $\sqrt{s} = 1.8$ TeV.

The function $f_j^{\bar{p}}$ has the same definition but it corresponds to the antiproton. The total short-distance cross section for the parton-parton subprocesses that intervene in the heavy quark production is $\hat{\sigma}$. The center-of mass energy of the i-j parton system is given by $\hat{s}$ and it is related to the $p\overline{p}$ center-of-mass energy by $\hat{s} = x_1 x_2 s$. The renormalization scale μ is the result of including higher order Feynman diagrams. If the calculation could be carried out to all orders then the dependence on μ would vanish. The strong coupling constant is given by α_s and M_{top} is the mass of the heavy quark.

The cross section for heavy quark production is rather sensitive to higher order corrections to the leading order (LO) calculation. Subsequently, several groups have calculated the complete next-to-leading order (NLO) cross section. For top quark production at the Tevatron, the NLO contribution for the gg process is roughly 70% that of the size of the LO term, and for the $q\bar{q}$ process it is about 20% that of the LO result. The theoretical cross section results are summarized in Table 2.1.

The $t\bar{t}$ cross section is small compared with other processes which take place at the Tevatron. Of the five trillion or so collisions that occurred at CDF during the entire Run I, one expects about 600 $t\bar{t}$ pairs to have been produced. Of these events,

only a small fraction are observed. Throughout the text, it will become evident why most of these events fail to be categorized as top events.

2.4 Top Hadronization and Decay

Within the Standard Model, the top quark decays via t $\rightarrow$ W$^+$b with a branching ratio close to unity. The decays t $\rightarrow$ Ws and t $\rightarrow$ Wd are also allowed but they are suppressed by factor of 10^{-3}-10^{-4} by the Cabibbo-Kobayashi-Maskawa (CKM) mixings [10]. The decay of the top quark can be written as,

$$\Gamma(t \rightarrow Wb) = \frac{G_F M_W^2}{8\pi\sqrt{2}} \frac{1}{M_{top}^2} V_{tb}^2 \left[\frac{(M_{top}^2 - m_b^2)^2}{M_W^2} + M_{top}^2 + m_b^2 - 2M_W^2 \right] \times 2k \tag{2.5}$$

where k denotes the W momentum in the t rest frame and it is,

$$k = \frac{\sqrt{(M_{top}^2 - (M_W + m_b)^2)(M_{top}^2 - (M_W - m_b)^2)}}{2M_{top}} \tag{2.6}$$

Let $M_W = 80.4$ GeV/c^2, $m_b = 5$ GeV/c^2, and the top quark mass $M_{top} = 175$ GeV/c^2. Substituting these numbers into the above equation gives the following top quark decay rate and lifetime:

$$\Gamma(t \rightarrow Wb) \approx 1.55 \text{ GeV} \Longrightarrow \tau_{\mathrm{t}} = \left(\frac{1}{\Gamma_{\mathrm{t}}}\right) \sim 4 \times 10^{-25} \text{ sec.} \tag{2.7}$$

At the Tevatron energies, the top quark decays before it has time to hadronize. This implies that the top quark can be treated as a free particle. Note however, that it is not only the large mass of the top quark which gives it its extremely short lifetime, but also the fact that it has a CKM-allowed t $\rightarrow$ b decay very close to unity. If this decay was more suppressed by the CKM mixings, the lifetime of the top quark would be longer.

A $t\bar{t}$ final state contains two t $\rightarrow$ Wb decays. The two b quarks will be observed as a spray or particles known as *jets*, and the W bosons will decay into a lepton-neutrino

Category	Decay Mode	Branching Ratio
Lepton+jets	$t\bar{t} \rightarrow (q\bar{q}'b)(\ell\nu b)$	24/81 (30%)
All hadronic	$t\bar{t} \rightarrow (q\bar{q}'b)(q\bar{q}'b)$	36/81 (44%)
Dilepton	$t\bar{t} \rightarrow (\ell\nu b)(\ell\nu b)$	4/81 (5%)

Table 2.2: Branching ratios for the three different $t\bar{t}$ decay modes. Here q stands for a u, d, c, or s, and ℓ stands for e and μ. τ leptonic decays are not included.

or a quark-antiquark pair. To a good approximation, the possible decays of the W boson are equally probable. However, one must remember to count each quark flavor three times since there are three different colors for each of them. Therefore, the probability that the W boson will decay into either of the three leptons is about 1/9, while the probability that it will decay into a quark-antiquark pair is 2/3. Since the W bosons decay independently of each other, the $t\bar{t}$ decays can be classified according to the W boson's decay. A summary of the different branching ratios of the possible decay channels is given in Table 2.2.

2.5 Decay Signature of the Lepton+jets Channel

Figure 2.3 shows an artist representation of the lepton+jets decay channel. From this picture we expect the final state of the $t\bar{t}$ decay to have the following:

- **An energetic lepton**, electron or muon, coming from the leptonic decay of the W boson. The W boson is a massive object so in general the lepton produced when it decays will have a high energy.

- **Large missing energy** due to the undetected neutrino from the leptonically decaying W boson. Neutrinos do not interact with the CDF detector so their presence is inferred by an imbalance in the total energy measured in the detector.

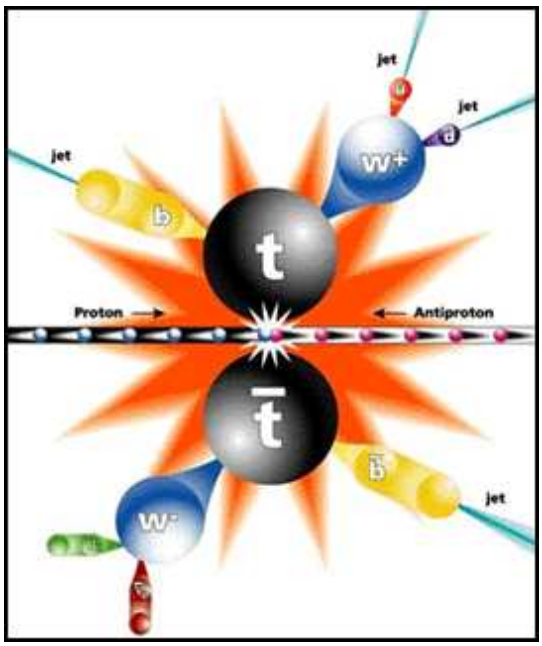

Figure 2.3: Lepton+jets $t\bar{t}$ event. One of the W bosons decays into a lepton+neutrino pair while the other one decays into a quark-antiquark pair.

- **Four jets**, two of them from the b quarks and two more from the two light quarks produced by the hadronic decay of the W boson. After a high energy collision, a quark or a gluon becomes free from the color-neutral object which contained it. As the partons move apart the energy required to separate them increases and $q\bar{q}$ pairs are created from the vacuum. These new quarks and the original ones recombine themselves to produce new hadrons. Thus, what emerges from the collision is a spray of colorless hadrons, or jet.

CHAPTER 3

THE CDF EXPERIMENT AT THE TEVATRON

Located in Batavia, Illinois, the Fermi National Accelerator Laboratory (Fermilab) is a particle physics experimental facility built in 1967 that is home to the Tevatron, the highest-energy particle accelerator in the world. The Tevatron will continue to dominate the world 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. In the material and observations that follow, we will describe the accelerator and the detector components as they were during Run I[1].

3.1 The Tevatron

At the Tevatron, protons and antiprotons collide with a $center - of - mass$ energy of 1.8 TeV. The proton and antiproton beam acquire their final energy by passing through the acceleration chain shown in Figure 3.1. The intermidiate energies attained at each step are given Table 3.1.

[1]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.

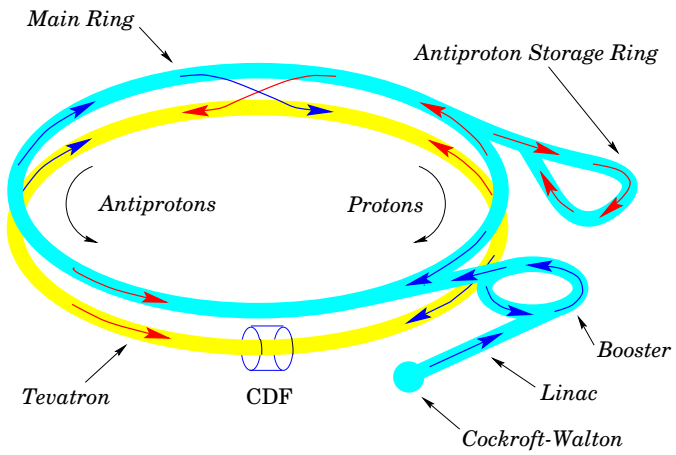

Figure 3.1: The accelerator complex at Fermilab.

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.

At the Cockroft-Walton accelerator, negatively charged hidrogen ions are first accelerated before being inserted in the 500-feet Linear accelerator (Linac) . Then, the H^- ions pass through a carbon foil which removes the electrons, leaving only the protons. From here, proton acceleration is done in three stages, corresponding to each of the three circular accelerators: Booster, Main Ring, and Tevatron. The antiprotons are made by colliding a proton beam from the Main Ring into a Tungsten target. The collisions in the target produce a wide range of secondary particles

Period	Typical Instantaneous L	Best Instantaneous L	Integrated L
Run IA	$0.5 \times 10^{31}\ cm^{-2}\ s^{-1}$	$0.9 \times 10^{31}\ cm^{-2}\ s^{-1}$	$19.6\ pb^{-1}$
Run IB	$1.6 \times 10^{31}\ cm^{-2}\ s^{-1}$	$2.8 \times 10^{31}\ cm^{-2}\ s^{-1}$	$86.3\ pb^{-1}$

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

incluiding many antiprotons. These antiprotons are selected and subsequently passed to the Debuncher where they are reduced by stochastic cooling and transferred to the Accumulator. When a sufficient number of antiprotons ($\approx 10^{12}$) has been collected, the antiprotons are reinjected into the Main Ring and then into the Tevatron, similar to the protons but revolving in the opposite direction. The bunches of protons and antiprotons cross every 3.5 μs. 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 [11].

3.2 The CDF Detector

The CDF detector is a general purpose detector designed for good lepton (e, μ), photon, and jet identification. It is cyllindrically symmetric, and also forward-backward symmetric with respect to the transverse plane, which passes throught the center of the detector. CDF employs a coordinate system in which the z-axis is along the beam line. In this system, we express the particle track coordinates in the detector by the pseudorapidity, η = -ln(tan$\frac{\theta}{2}$), and ϕ, where θ is the polar angle, and ϕ is the azimuthal angle. CDF is composed of an array of individual detectors, which are divided into three main categories: the tracking chambers, calorimeters,

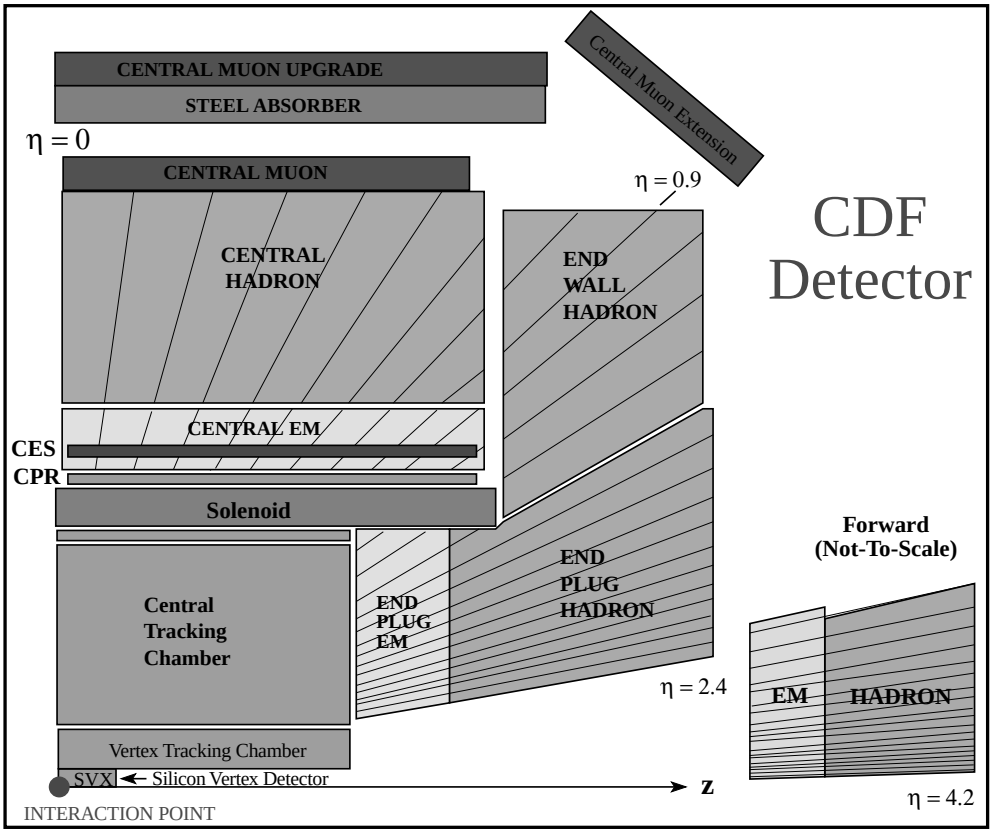

Figure 3.2: A quarter view of the CDF showing the locations of the different detector subcomponents.

and muon chambers. First, the tracking detectors are the closest to the interaction region. They are enclosed by a superconducting solenoid, which generates a 1.4 T magnetic field along the incident beam direction. This magnetic field allows precise momentum determination of charged particles in the central region. Surrounding the solenoid are the calorimeters that determine the energies of the outgoing particles. Finally, the outermost detectors are the muon chambers that measure the energies and positions of the muons that were not absorbed in the calorimeters. For the complete description of the detector, we point the reader to reference [12].

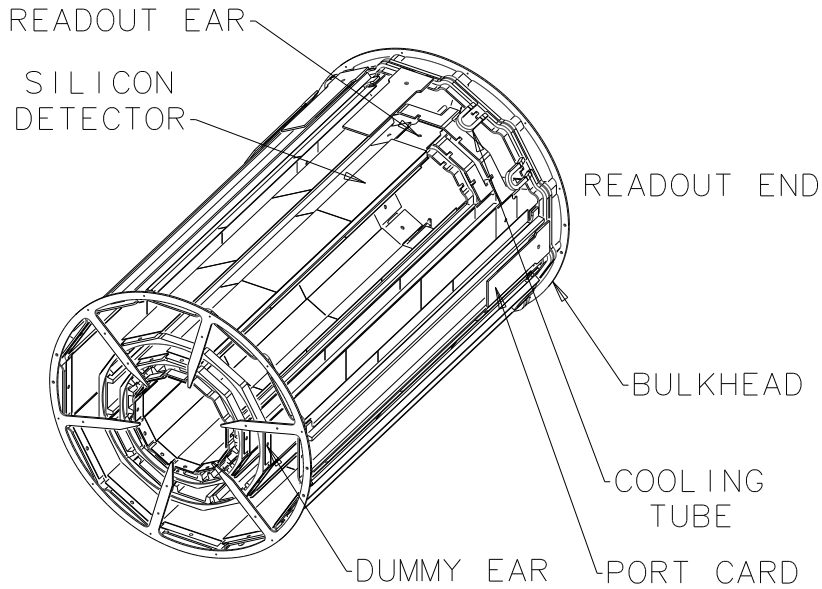

Figure 3.3: Schematic view of an SVX barrel.

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 tracking detector is designed to detect secondary vertices from heavy flavor decay especially b hadrons, which have a macroscopic mean decay length[2]. It consists of two barrels aligned end-to-end on either side of the $z = 0$ point, and surrounding the 1.9 cm radius beam pipe. The total coverage of the SVX is 51 cm in the z direction which corresponds to an $|\eta| < 1.9$. As the spread of $p\bar{p}$ collisions is 60 cm along the z axis, the SVX track acceptance is roughly 60%. Figure 3.3 shows one

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

SVX barrel. In the azimuthal direction it features 12 wedges of 30° each, with the silicon strip detectors arranged in 4 concentric layers. The number of silicon strips changes with each layer. The strip lines run parallel to the beam line and provide hit information in the $r - \phi$ plane. The SVX single hit resolution is measured to be 13 μm in $r - \phi$ plane, and the resolution of the impact parameter relative to the primary vertex is 17 μm. The SVX system plays a major role in just about all areas of top quark physics.

The vertex drift chamber lies outside the SVX. The VTX consists of 28 modules of octagonal time projection chambers along the z axis. 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 drift regions. The ionization electrons drift to the two endcaps containing sense wires running perpendicular to the 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 function of the VTX is two-fold. The VTX is able to identify the z position of the primary vertex with a resolution of $\sim 1\ mm$. In the case of events with multiple interaction[3], the VTX is used to establish which interaction a primary vertex and associated tracks belong to.

The CTC is a 3.2 m long cylindrical drift chamber with a 1.3 m radius. The CTC covers a pseudorapidity range of $|\eta| < 1.0$. It consists of 84 layers of sense wires arranged into 9 superlayers. There are two types of superlayers alternating in radial direction: *axial* superlayers (0, 2, 4, 6, 8) are composed of cells with 12 sense wire layers and provide $r - \phi$ track information; *stereo* superlayers (1, 3, 5, 7) are

[3]There are roughly 10^{10} particles in the proton and antiproton bunches. On average we expect one interaction per beam crossing but in some cases there are more than one

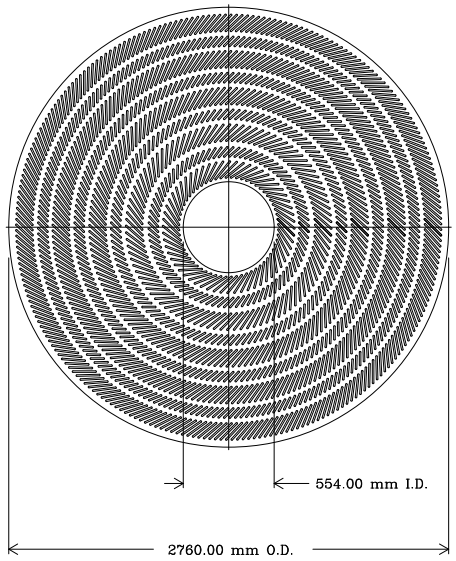

Figure 3.4: Transverse view the Central Tracking Chamber (CTC) which shows the alternating superlayers of axial and stereo cells.

composed of cells with 6 sense wire layers and provide tracking information in the $r-z$ plane. Figure 3.4 shows the transverse view at the endplate of the CTC. Each cell is tilted by 45° with respect to the radial direction to ensure that the electrons drift perpendicularly to the radial direction, which simplifies track reconstruction. There are over 6000 sense wires in the CTC, with each wire having a design resolution of $\approx 200\ \mu m$ in the $r-\phi$ plane, and $4\ mm$ in the $r-z$ plane. The momentum resolution of a charged particles is given by $\delta \mathrm{P_T}/\mathrm{P_T} = 0.001\ GeV^{-1}c \times \mathrm{P_T}$ when the information from the SVX and CTC are combined. The transverse momentum is given by $P_T = P sin\theta$, where P is the total momentum of the particle measured by the CTC. Further information about the SVX, VTX, and CTC can be located in references [13], [14], and [15] respectively.

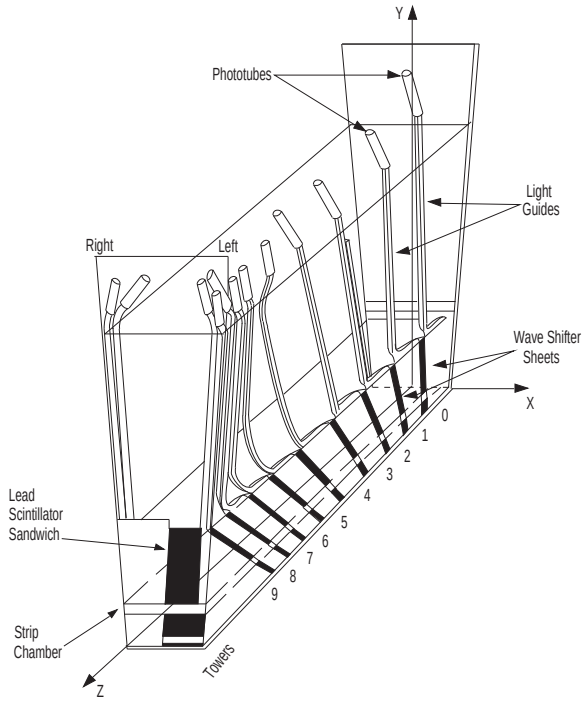

Figure 3.5: A central calorimeter wedge. There are 24 wedges covering $\Delta\phi = 15°$ on either side of the transverse plane $z = 0$.

3.2.2 Calorimeters

The CDF calorimeters, located outside the solenoid, are segmented into towers (in $\eta - \phi$ space) that point back to the nominal interaction point (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* 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 $\Delta\eta = 0.1$. Calorimeter energies are given by the transverse energy, $E_T = E sin\theta$, where E is the total measured energy. A closer look at each calorimetry system follows.

The central calorimeter system is composed of the central electromagnetic calorimeters (CEM), and central hadronic calorimeters (CHA). There are 48 individual wedges each containing 10 towers. Figure 3.5 shows 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 CEM is constructed by alternating layers of lead and scintillator. The central electromagnetic shower counter (CES) is located at the CEM shower maximum to provide shower position information in the $z-\phi$ view. The CES is a proportional strip and wire chamber. The central preradiator (CPR) composed of proportional tubes is placed between the solenoid and the CEM. The CPR samples the early development of electromagnetic showers and is used to differenciate between elctrons and hadrons. The CHA and WHA calorimeter consist of alternating layers of iron and scintillator. Further information regarding the physical characteristics of the CEM, CHA and WHA calorimeters can be found in Table 3.3. A detailed description of each is given in references [16], [17].

In order to determine the energies of the forward jets, the plug and forward calorimeters are required. This information is also used to extract the missing energy $\not\!\!E_T$ in an event (see Section 8.5). As in the central calorimetry design, both the plug and the forward calorimeters feature electromagnetic and hadronic sections: PEM and FEM relating to the former and PHA and FHA describing the latter. 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 for the FEM and steel for the FHA, as summarized in Table 3.4.

	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.

	PEM	PHA	FEM	FHA
Coverage ($\|\eta\|$)	$1.1-2.4$	$1.3-2.4$	$2.2-4.2$	$2.3-4.2$
Tower Size ($\Delta\eta \times \Delta\phi$)	$0.1 \times 5°$	$0.1 \times 5°$	$0.1 \times 5°$	$0.1 \times 5°$
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 last row, the symbol $\oplus$ indicates that the constant term is to be added in quadrature to the resolution.

3.2.3 Muon Systems

A majority 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 [18] offers a coverage of $|\eta| < 0.6$, and is located inside the central calorimeter wedges radially outside of the CHA (Figure 3.6). In each wedge the CMU is segmented into three towers, each consisting of 4 radial layers of 4 drift cells. The two outer layers have an offset of 2 mm to resolve the left-right ambiguity of track measurement in azimuth respect to the two inner layers. The central muon upgrade (CMP), which is located behind an additional 0.6 m of steel, reduces the background due to the hadrons which "punch-through" to the CMU. It consists of 4 layers of drift cells and in effect has the same η coverage as the CMU, but it covers most of the CMU gaps. With the addition of the CMX the muon coverage increase to a range of $0.6 < |\eta| < 1.0$. The CMX is structured as four conical arches of drift tubes.

3.2.4 Event Triggers

In Run IB, the $p\bar{p}$ collisions were occurring at a rate of 280 kHz, which considerably exceeds the rate of writing out the events to tape (~ 10 Hz). To solve this problem, CDF devised a three-level trigger system that rejects most of the events, while retaining only those events that are potentially useful for various physics analyses. The trigger requirements are applied serially. Events passing the Level 1 trigger criteria are passed to Level 2; the events passing Level 2 are allow to propagate to

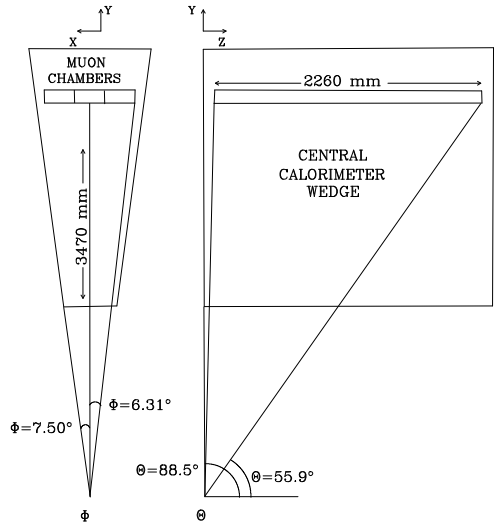

Figure 3.6: The location of the CMU system. Inside each calorimeter wedge there is a 2.4° ϕ interval not covered by the CMU.

Level 3 which imposes farther cuts before deciding whether an event should be written to tape or not. Typical acceptance rates for the three trigger levels are 1:300, 1:50, and 1:5 for Level 1, Level 2, and Level 3 respectively.

Level 1 selects events based on identification of energy clusters in the calorimeters or muon tracks in the muon chambers. The time required to make the decision is less than 3.5 μs, leaving little dead time in the system. The Level 1 requirements reduce the event rate down to 1 kHz from 280 kHz.

The Level 2 decision time is approximately 20 μs during which further beam crossings are ignored by the detector resulting in a dead time of a few percent. The decision to accept events is based on track information and clustered energies. The central fast tracker (CFT), a highly efficient hardware processor, reconstructs high

momentum tracks using the hit information from the CTC. Similarly, calorimeter clusters are formed by searching for a seed tower above a certain threhold and adding the neighboring towers. The Level 2 output rate is 20-30 Hz.

The Level 3 trigger is a software reconstruction trigger on a farm of Silicon Graphics processors. If an event passes the Level 2 requirements, the whole detector is read out, and the event is reconstructed using a simplified version of the $offline$ analysis code. This includes the reconstruction of the $t\bar{t}$ decay objects: lepton, jets, and $\not\!\!E_T$. The detector readout time is 3 ms which on average results in 10% dead time. Once an event is accepted by Level 3, it will be written to 8 mm tape as part of a particular data $stream$. The Level 3 output rate is 5-10 Hz.

CHAPTER 4

INTRODUCTION TO NEURAL NETWORKS

The mass of the top quark is one of the most important fundamental parameters in the Standard Model. Previous analyses of the top mass have resulted in the most accurate mass measurement among the quarks. However, there are ways the mass analysis can be improved by using Neural Networks (NN).

In this chapter, we give a brief historical background on Neural Networks. The first network models are introduced giving special attention to the *perceptron*. This computational model is the foundation of the multi-layer Neural Network. A description of the components of the multi-layer NN are given next. In the last few sections, we explain how these networks work and are able to perform a wide range of tasks.

4.1 Historical Background

Artificial Neural Networks are loosely based on how the human brain works. About 100 years ago, Santiago Ramon y Cajal, realized that the brain was made up of discrete units called *neurons*, the Greek word for nerves. He described neurons as polarized cells that receive signals via highly branched extensions, called *dendrites*, and send information along unbranched extensions, called *axons*. This information is transmitted in the form of electrical pulses called *action potentials*. These potentials

cause the neuron to release a chemical which excites or inhibits the neuron's membrane. The effect of all excitations or inhibitions may produce an action potential in the next neuron (i.e the neuron fires). An artist rendition of a biological neuron is shown in figure 4.1.

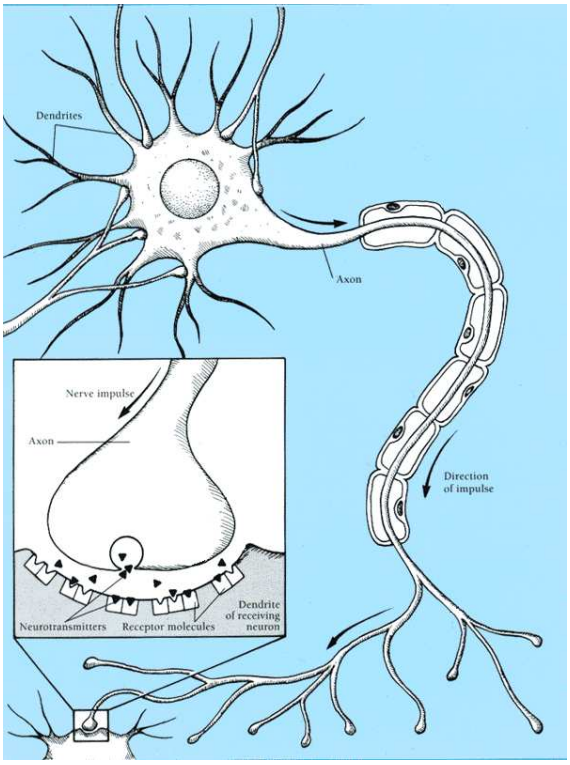

Figure 4.1: A biological neuron. The inset shows the synapse which is a specialized connection between neurons.

The first computational model of the neuron was introduced by McCulloch and Pitts in the 1940s. They proposed a binary unit for which the output, y, is equal to 1 when an action potential is generated, and 0 otherwise. A weight value w_i is associated to each of the ith connections of the neuron. These weights are *excitatory* if

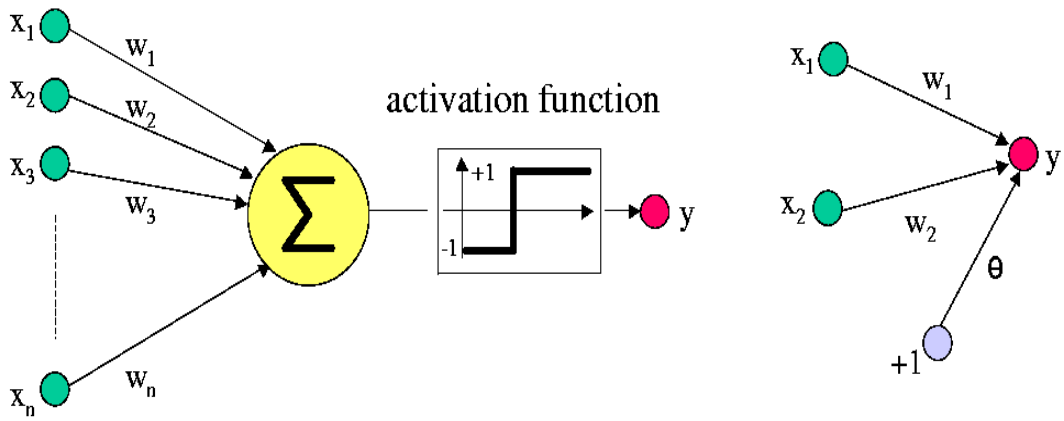

Figure 4.2: (Left) McCulloch-Pitt model for the computational neuron. The different weights are either excitatory $w_i = +1$ or inhibitory $w_i = -1$. (Right) Simple graphical represention of a perceptron with one output and two inputs.

$w_i = +1$, and inhibitory if $w_i = -1$. A neuron fires when the effect of the excitations and inhibitions is larger than a certain threshold (bias) θ. A graphical representation of this model is shown in figure 4.2.(L). Here, x_i are the input values fed into the network.

In 1958, Rosenblatt introduced a computational model of the neuron called the perceptron. Figure 4.2.(R) shows the graphical representation. The biggest improvement from the previous model is that it introduced numerical interconnection weights instead of the simple inhibitory/excitatory connections used before. The weights could now have any real value which are determined according to the task the perceptron has to perform. The mathematical model is the following:

$$y(t) = \begin{cases} +1 & \text{if} \quad \sum_{i=1}^{n} w_i(t)x_i > \theta(t) \\ -1 & \text{if} \quad \sum_{i=1}^{n} w_i(t)x_i < \theta(t), \end{cases} \tag{4.1}$$

where $y(t)$ is the output of the perceptron, $w_i(t)$ is the weight of the input x_i, and $\theta(t)$ is the bias. One can think of the bias (see Figure 4.2.(R)) as the weight of a unit

whose value is always set to one. Thus, the biases in the network are determined in a similar way as the weights.

In Equation 4.1, the output of the network is a function of time because it changes as the weights and biases are modified to perform a task. In Appendix A [20] we explain how the updating of the weights and biases is accomplished. Rosenblatt proved a remarkable theory about perceptron learning. However, the initial euphoria was soon replaced by disillusion after Minsky and Papert [21] found that the perceptron has severe restrictions on what it can represent as described in the next section.

4.2 The Multi-layer Perceptron

The single layer preceptron has severe restrictions as only linear classifiers can be constructed, or in the case of function approximation, only linear functions can be represented. Minsky and Papert showed that many of these restrictions can be overcome by introducing an extra layer of units, but they did not present a solution to the problem of how to update the weights from input to hidden units. The answer to this question was given by Rumelhart, and it is described in Section 4.3.3. The mathematical derivation is given in Appendix B [20].

A multi-layer perceptron, or feed-forward Neural Network, has a layered structure. An examples is shown in Figure 4.3. Each layer consists of units which receive their input from units from a layer directly to the left and send their output to units in a layer directly to the right of the unit. There are no connection within a layer. The N_i *input* units take their input value from outside the NN, and they push this information to the layer of $N_{h,1}$ *hidden* units. No processing takes place in the input units. The activation of the hidden units is a function $\mathcal{F}_i$ of the weighted inputs, plus

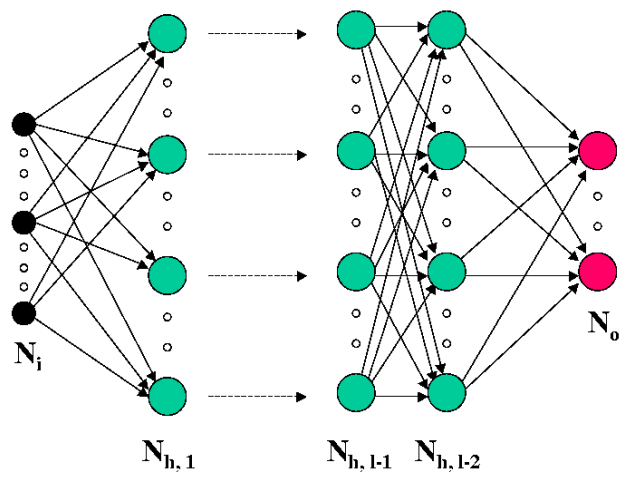

Figure 4.3: A multi-layer network with l layers of units. The information goes from the input units (left) to the output units (right).

a bias as given by

$$y_k(t+1) = \mathcal{F}_k(s_k(t)) = \mathcal{F}(\sum_j w_{jk}(t)y_j(t) + \theta_k(t)), \tag{4.2}$$

where y corresponds to the output value of the units, w_{jk} are the weights, and θ_k are the biases. The output of the hidden units is distributed to the next layer of N_{h_2} hidden units, until the last layer of hidden units whose outputs are fed to a layer of *output* units. The output units may or may not have the same activation function as the hidden units. The purpose of the output units is to send the data out of the Neural Network. Figure 4.3 shows a graphical representation of a multi-layer perceptron.

In general the activation function, $\mathcal{F}$(y) for the hidden and output units is a threshold function such as the sigmoid function. The only requirement is that the function must be differentiable with respect to the weights and biases in the network.

When dealing with muti-layer networks, it is common to refer to the different number of input, hidden, and output units as the *architecture* of the Neural Network. For instance, a NN with five inputs, ten units in the hidden layer, and one output corresponds to an architecture of 5-10-1.

4.3 Understanding Artificial Neural Network

As we have seen Neural Networks are loosely based on the human brain. They consist of a series of units arranged in layers that are connected to each other through weighted connections. In this section we explain how Neural Networks work. First we state the perceptron *learning rule* by which a network tries to learn a specific task. For this learning rule there exists a *convergence theorem* that says that starting from random weights, a network will converge to some solution in a finite number of steps. In order for the network to reach a solution, there must be a mechanism that updates the different weights at each step. We finish the section by describing this mechanism, called *Back-propagation*.

4.3.1 The Perceptron Learning Rule

A Neural Network has to be configured such that the application of a set of input values produces the desired set of output values. There are various methods we can use to set the weights. One way is to set the weight explicitly, using *a priori* knowledge. A more common way called *supervised learning* 'trains' the network by feeding it teaching patterns and letting it change the weights according to the learning rule. These teaching patterns, $(\vec{x}, d(\vec{x}))$, are chosen by the user and they must characterize the task that the Neural Network must perform. Much study goes into the selection of these learning samples.

Suppose we have a set of learning samples consisting of an input vector $\vec{x}$ and a desired output $d(\vec{x})$. The perceptron learning rule can be stated as follows:

1. Start with random weights for the connections.
2. Select a input vector $\vec{x}$ from the set of learning samples.
3. Modify all the weights so that the perceptron's response is as close to $d(\vec{x})$ as possible.
4. Return to step 2.

In the next sections we give the mathematical foundation which allows the Neural Network to resolve the challenges at each of the steps presented above.

4.3.2 The Convergence Theorem

The convergence theorem states that if there exists a set of weights $\vec{w}^*$ which allows the perceptron to perform a specific task (pattern classification, function appoximation, etc), the perceptron learning rule will converge to some solution (which may or may not be the same as $\vec{w}^*$) in a finite number of steps for any initial set of weights.

We define

$$cos(\vec{w}) = \frac{\vec{w} \cdot \vec{w}^*}{||\vec{w}||} \tag{4.3}$$

where cos$\vec{w}$ is a measure of how much the network's current weights, $\vec{w}$, differ from the correct set of weights $\vec{w}^*$. Because of the nature of the activation function, the length (norm) of the correct set of weights does not play a role in this proof. Thus, we let $||\vec{w}^*|| = 1$.

After n iterations it can be shown that the dot product between the correct set of weights and the n-th set of weights must be $\vec{w}(n) \cdot \vec{w}^* \geq n\Delta$, where Δ is the amount the weights change from one iteration to the next. Similarly, it can be shown that the length of the n-th set of weights is $\|\vec{w}(n)\| \leq n$. We can combine these results to rewrite equation 4.3 as:

$$cos(\vec{w}(n)) = \frac{\vec{w}(n) \cdot \vec{w}^*}{\|\vec{w}(n)\|} > \frac{n\Delta}{\sqrt{n}}. \tag{4.4}$$

In the limit that the number of iterations necessary to update the weights, goes to infinity we have that $lim_{n\to\infty}\ cos(\vec{w}(n)) = \infty$ but by definition we know that $\cos(\vec{w}(n)) \leq 1$. Therefore, there must exist an upper limit n_{max} for the number of iterations that the NN updates the weights. After n_{max} modifications of the weights the network is "correctly"[4] performing the task. For a more detailed description of this derivation see [22].

4.3.3 The Back-propagation Algorithm

The *Back-propagation* method is a generalization of the Delta rule derived in appendix A. For a mathematical derivation of the Back-propagation algorithm we refer the reader to appendix B. It is conveninent to become familiar with the Delta rule and then proceed with the Back-propagation method. In this section we explain the most important points of how the weights are updated without having to recite all the equations.

When the values of a learning pattern, $\vec{x}$ are fed to the Neural Network, these initial values are propagated to the output units, and the actual network output is

[4]Correctly does not mean that the network is right 100% of the time. This is not possible in most real cases. In section 4.5 we describe a series of factors that influence the NN performance

compared to the desired output values, $d(\vec{x})$. From this comparison we obtain an error for each of the output units. Let's defined e_o as the error corresponding to output unit o. This error is a measure of the difference between the network output and the desired output. The goal is to bring this error to a minimum.

The simplest way to do this is the greedy method: we change the connection, weights, in the NN so that the error e_o will be zero for this particular pattern. From equation B.7, we know that in order to reduce the error corresponding to the output units, we have to adapt the weights according to

$$\Delta w_{ho} = (d_o - y_o)y_h = \delta_o y_h, \tag{4.5}$$

where Δw_{ho} is the change in the weights connecting the output unit(s) to the hidden units, d_o is the desired output, y_o is the output from the output unit(s), and y_h is the output from the hidden units. However, applying the above equation alone does not complete the job since the weights from input to hidden units are not updated. If we stopped here we would not be using the full representational power of the Neural Network. To adapt the weights from input to hidden units we need to apply the Delta rule once more. The problem is that we do not have the value for δ, the difference between the Neural Network output and the desired output, for the hidden units. Not to worry, since δ for a given hidden unit is equal to the sum of the δ's from each output it is connected to multiplied by the weight of that connection ($\delta_h = \sum_o \delta_o w_{ho}$). The amount the hidden units change is proportional to the amount the output units change times the strength of their connections. Now, we just have to apply the activation function, and the Back-propagation method can continue with the next pattern.

4.4 The Universal Approximation Theorem

A multilayer perceptron trained with the back-propagation algorithm may be viewed as a tool to perform nonlinear input to output mapping. The input-output relationship of the network defines a mapping from the Euclidean input space to the Euclidean output space. The most important aspect of this input to output mapping becomes the number of hidden layers necessary to accomplish it.

The *universal approximation theorem* [23] states that only one hidden layer suffices to approximate any function with finite discontinuities to arbitrary precision, provided that the activation functions are non-linear.

Let $\mathcal{F}(\cdot)$ be a nonconstant, bounded, and monotonically-increasing continuous function. Let I_N denote the N-dimensional unit hypercube $[0,1]^N$, and C(I_N) be the space of continuous functions on I_N. Then, given any function $f \ni C(I_N)$ and $\epsilon > 0$, there exist a set of real constants α_i, θ_i, and w_{ij} where $i = 1, \ldots, L$ and $j = 1. \ldots, N$ such that we may define

$$F(x_1, \ldots, x_N) = \sum_{i=1}^{L} \alpha_i \mathcal{F}(\sum_{j=1}^{N} w_{ij} x_j + \theta_i) \tag{4.6}$$

as an appropiate realization of the function $f(\cdot)$; that is

$$|F(x_1, \ldots, x_N) - f(x_1, \ldots, x_N)| < \epsilon \tag{4.7}$$

for all $x_1, x_2, \ldots, x_N$ that lie in the input space. The universal approximation theorem says that a single layer of hidden units is sufficient to approximate any continuous function to within a certain error ϵ, However, it is not implied that a single layer is optimum in the sense of learning time, ease of implementation, or more importantly generalization. Thus, the architecture of the NN must be determined by the nature of the problem at hand.

4.4.1 Neural Networks and Fourier Series

The reader should notice that the universal approximation theorem is basically a generalization of approximations by Fourier series. From Fourier analysis it is known that any continuous function can be written as an infinite sum of sine and cosine terms:

$$f(x) = \sum_{n=0}^{\infty}(a_n \; cos \; nx + b_n \; sin \; nx) = a_0 + \sum_{n=1}^{\infty} c_n \; sin(nx + \theta_n), \tag{4.8}$$

where $c_n = \sqrt{(}a_n^2 + b_n^2)$ and $\theta_n = \arctan(\mathrm{b/a})$. The above equation has the same form as equation 4.2, so we can interpret it as Neural Network with a single activation unit x, a single output unit corresponding to $f(x)$, and hidden units with an activation function $\mathcal{F} = sin(s)$. The term a_0 corresponds to the bias of the output unit, the factors c_n correspond to the weights from the hidden units to the output unit, the phase factor θ_n corresponds to the bias term of the hidden units, and the factor n corresponds to the weights between the units of the input and the hidden layer.

The basic difference between the Fourier approach and the Neural Network method is that in the Fourier approach the n factors (weights) are integer numbers which are analitically determined. In contrast, in the NN method, these weights are real numbers and are determined through learning.

Many people are reluctant to use Neural Networks because of their fancy name. If we look beyond this fact we would be surprised to see that sometimes Neural Networks are nothing more than a Fourier expansion!

4.5 NN performance

Neural Networks have been around for several decades. There have been many improvements over the first initial models of the computational neuron which have led

to the use of Neural Networks in many different fields. Although, Neural Networks have been very successful in performing different tasks, this success depends on a series of different aspects that impact the network's performance. We describe the most important aspects and how they impact the analyses presented here:

- **Information:** a neural network will not be able to perform any task if it is not given the correct information. If one wants to fit a distribution you must use all the *relevant* information which describes that distribution. Similarly, for event classification we want to use variables that have the most distingushing power between the different classes of events.

- **Architecture:** the number of input and output units is determined by the function one wants to fit. The only thing one must worry about is the number of hidden units. For function approximation one wants to have enough hidden units so that the distribution represented by the NN is close to the original. Two few hidden units will result in a poor fit. If one uses too many, the NN will "fit the noise" (statistical fluctuations) instead of making a smooth approximation. In the case of event classification, the number of inputs is determined by the number of variables with distingushing power. The number of outputs corresponds to the different types of events we want to classify. To figure out how many hidden nodes are needed, it is useful to try many different architectures and then choose the one that performs best.

- **Number of learning samples:** in function approximation, we want to have enough patterns (points) so that the shape of the distribution is well represented. For instance, a straight line can be represented by two points, but a sine function

clearly requires more points. For event classification, we want to use at least ten times the number of learning samples as hidden plus output units are in the NN. In any case it is better to have more learning samples than less.

- **Number of iterations or** *epochs*[5]**:** in the case of function approximation we define a χ^2 (Equation 7.1) function which compares the shape of the distribution we want to fit to the NN output. We update the weights for as many iterations as needed so that $\chi^2 \sim 1$. At this point, the weights are frozen and this defines the network which gives a functional form that is able to approximate the function. In an event classification task, we stop updating the weights when the performance of the testing sample begins to increase (see Figure 8.7. At this point, the NN is begining to learn the specific features of the learning sample and therefore losing its generality. Function approximation and pattern classification are explained in detail in Sections 7.1.1 and 8.2.2 respectively.

[5]An epoch has passed after all the lerning samples have been presented to the network once

CHAPTER 5

EVENT SAMPLE

The Run I data collected by the CDF detector from 1992 to 1995 corresponds to an integrated luminosity of 106.0 $\pm$ 4.1 pb^{-1}. During that time, roughly 50 million events were written to tape. This chapter describes how events that "look" like $t\bar{t}$ are selected for the mass analysis. First, we describe how the different objects in the $t\bar{t}$ decay are identified. We then give the selection criteria used to improve the signal to background ratio for both the Run I data sample and the Monte Carlo samples used in this study. In the final section, we describe the Monte Carlo simulation and generation for both $t\bar{t}$ signal as well as background.

5.1 High P_T Leptons

From the signature of the lepton+jets decay channel, we are only interested in leptons coming from the decay of a W boson. In general, these leptons are isolated (there is low jet activity around them) and they are more energetic than the leptons coming from the semileptonic decay of the b and c quarks.

From the Run I data we select a sample of events which contain a high-E_T electron, $E_T > 20$ GeV/c, located in the central region of the detector, $|\eta| < 1$. Backgrounds from photon conversions and charged hadrons are removed by cutting on several

variables. For a detailed description of all the different cuts we refer the reader to Ref. [24]. Here we describe those cuts which provide the largest discrimination between signal and background. We require the electrons to have a CTC track pointing towards the electron shower in the CEM. The ratio of the energy in the hadronic calorimeter and the energy in the electromagnetic calorimeter, HAD/EM, must be less that 5%. We also require the shower energy divided by the momentum of the associated track to be less than 1.8. Each of the electrons must be matched to a VTX track, and the invariant mass of this track associated with any other CTC track must be greater than 0.5 GeV/c^2. Using this cut we can remove electrons from photon conversions with an efficiency of 88% [24]. The energy of the electrons is measured by the calorimeter energy of the tower to which the CTC track points to plus the adjacent towers [27]. The energy resolution for high-$\mathrm{E_T}$ electrons is given by $\sigma(\mathrm{E_T})/\mathrm{E_T} = 13.5\%/\sqrt{(}\mathrm{E_T}) \oplus 2\%$.

The high-$\mathrm{P_T}$ muon sample is created by selecting central muons, $|\eta| < 1$, with a $\mathrm{P_T} > 20$ GeV/c. Muon candidates are identified by a matching a track segment in the CMU, CMP, or CMX to one in the CTC. The momentum of the high-$\mathrm{P_T}$ muon is measured by constraining the CTC track to the average beam position. The muon transverse momentum resolutions is given by $\sigma(\mathrm{P_T})/\mathrm{P_T} = 0.11\%\mathrm{P_T}$. The primary background in the muon sample is due to charged hadrons which "punch through" the calorimeter and produce tracks in the muon chambers, and cosmic rays. We remove the background from the charged hadrons by requiring that their energy deposition in the calorimeter be characteristic of a minimum ionizing particle. The cosmic ray background is removed by requiring that the track extrapolates back (in $r - \phi$) to within 3 mm of the beamline, and that it is within 5 cm (at r = 0) of the primary

vertex in the $r - z$ plane. The other muon selection requirements are given in Ref. [24].

From these high-P_T lepton samples, we only select those events in which the lepton is isolated from jet activity [28]. In the lepton+jets channel we require only one lepton which corresponds to the $W \to \ell\nu$ decay. We refer to this lepton as the primary lepton of the event.

5.2 Missing Transverse Energy, $\not\!\!E_T$

The neutrino produced in the leptonic decay of the W boson does not interact with the detector. Thus, its existence is inferred indirectly by an imbalance in the total transverse energy measured in the calorimeter. In $p\bar{p}$ collisions, the proton and anti-proton's initial momentum only have a nonzero z-component. Thus, from energy conservation[6] we expected the total energy in the transverse plane to be equal to zero. The $\not\!\!E_T$ is then given by the following

$$-\vec{\not\!\!E}_T = \vec{E}_T(lepton) + \sum_{i=1}^{4} \vec{E}_T(jet) + \vec{X}_T \tag{5.1}$$

such that the $\not\!\!E_T$ plus the total calorimeter energy cancel each other. The definition of the quantity $\vec{X}_T$ is given in Section 6.2. The above expression is highly correlated with the jet energies. Thus, the $\not\!\!E_T$ is not an independent parameter in the mass fit.

5.3 Jet Reconstruction

Jets will deposit their energy in a localized cluster of calorimeter towers. The cluster identification begins by determining the tower with the highest transverse energy. This seed tower is required to have an $E_T > 3.0$ GeV. An energy weighted

[6]In the high energy limit, energy and momentum are equivalent up to the constant c factor

centroid is calculated for the cluster, and only the energy from towers within a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.4$ are included to determine the energy of the jet. This raw jet energy, E_T^{raw}, must be corrected for the various energy loses (see Section 6.1) to obtain the true energy of the original parton. A discussion of the jet reconstruction algorithm can be found in Ref. [29, 37].

5.3.1 Secondary Vertex Tagging of B Jets

The Silicon Vertex Tagging (SVX) algorithm [24, 25] searches for displaced vertices due to B hadron decays. B hadrons have a life time of about 1.5 picoseconds and they are expected to be very energetic in top decays. Thus, b quarks can travel up to 1 or 2 cm before they decay. The location of the B hadron decay, *secondary* vertex, can then be distinguished from the $p\bar{p}$ interaction point, *primary* vertex. This is shown in Figure 5.1.

In order to be considered for tagging, SVX tracks must be associated with a jet that has a raw $E_T > 15$ GeV and $|\eta| < 2.0$. These tracks must also be within a cone of $\Delta R < 0.4$ and they must have hits in the silicon vertex detector. The algorithm employs a two step process to find a secondary vertex. Initially, it attempts to reconstruct displaced vertices with three or more tracks with $P_T > 0.5$ GeV/c, for which at least one of them must have $P_T > 1.0$ GeV/c. If it fails, it searches for a displaced vertex with only two tracks with tighter quality cuts. This algorithm has an average efficiency for Run 1b of 48% for tagging at least one of the b jets in $t\bar{t}$.

5.3.2 Soft Lepton Tagging of B Jets

The Soft Lepton Tagging (SLT) algorithm [24, 30] searches for additional leptons resulting from the semileptonic B hadron decay. These occur via $b \to \ell\nu_\ell X$, or

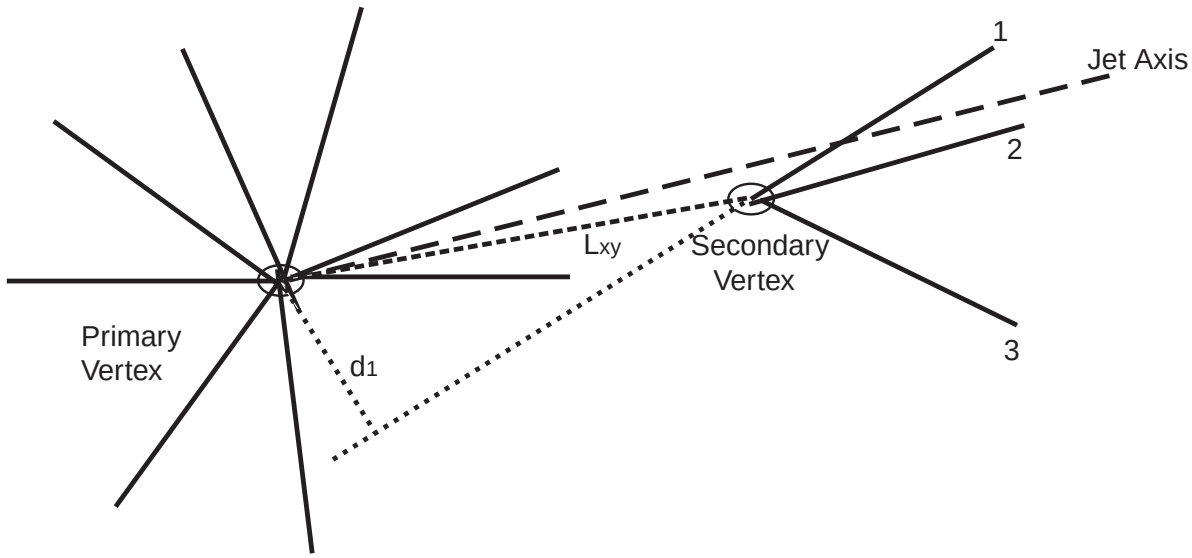

Figure 5.1: Schematic view of an event containing a secondary vertex shown in the transverse ($r - \phi$) view. The primary vertex is the point where the $p\bar{p}$ interaction occurs. The lines denoted 1, 2, and 3 are charged-particle tracks reconstructed in the SVX which yield the secondary vertex. 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 [24]).

$b \to c \to \ell\nu_\ell X$. The lepton is required to have a $\mathrm{P_T} > 2$ GeV/c and to be within a $\Delta R < 0.4$ of a jet with raw energy, $\mathrm{E_T} > 8$ GeV. The efficiency for tagging at least one of the the b jets in $t\bar{t}$ is roughly 15%.

The SVX tagging algorithm obtains a higher purity and efficiency than the SLT method. However, the SLT algorithm is employed because it uses information which is independent from the SVX method and adds significantly to the acceptance. Finally, we note that not all the jets tagged by either of the algorithms come from the decay of the B hadron. Displaced vertices are also produced by other long-lived particles such as c quark, τ, K_S^0, Λ, etc.

5.4 Top Mass Candidate Sample

In Section 5.5 we describe the largest background processes to $t\bar{t}$ production. These background events do not have any top mass information. Thus, we define a series of cuts that are intended to keep most of the $t\bar{t}$ events while removing most of the background events. We choose events that have four jets, an energetic lepton, and $\not\!\!E_T$ from the leptonic decay of the W boson. All the events in the mass sample must pass the following cuts:

1. The event must have an electron (muon) with $\mathrm{E_T} > 20$ GeV ($\mathrm{P_T} > 20$ GeV/c) and $|\eta| < 1$.

2. The raw missing energy, $\not\!\!E_T$, must be greater than 20 GeV. If the primary lepton is a muon, the $\not\!\!E_T$ includes a correction for the muon momentum.

3. One isolated lepton is required (see Section 5.1)

4. Dilepton events ($t\bar{t} \rightarrow \ell^+\bar{\nu}\ell^-\nu b\bar{b}X$), defined according to the selection criteria described in Ref. [26] are removed.

5. We remove events consistent with Z boson decays $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$. These events are characterized by two oppositely charged, same flavor high-$\mathrm{P_T}$ leptons ($\mathrm{P_T} > 20$ GeV/c) that have an invariant mass between 75 and 105 GeV/c^2. Events with a high-$\mathrm{P_T}$ photon and whose $\ell\ell\gamma$ invariant mass falls within the Z mass window are also removed.

6. The primary vertex of the event must be within 60 cm of $z = 0.0$.

7. At least three jets with $\mathrm{E_T} > 15$ GeV and $|\eta| < 2.0$.

8. An additional jet with $E_T > 15$ GeV and $|\eta| < 2.0$ or $E_T > 8$ GeV and $|\eta| < 2.4$.

9. After the the mass reconstruction is performed, events are required to pass a goodness-of-fit, $\chi^2 < 10.0$, where the variable χ^2 is defined in Section 6.3.

In the Run I data, a sample of 324 events pass cuts one through seven, and are the same events used in the CDF measurement of the $t\bar{t}$ cross section [25]. Imposing cut number eight reduces the sample to 163 events. The final selection requirement removes 12 events, from which we obtain an inclusive sample of 151 W+multi-jet events. There are 34 of these events that have jets which are tagged by the SVX or SLT algorithm. To improve the signal-to-background (S/B) ratio for the untagged events we require them to have four jets with $E_T > 15$ GeV and $|\eta| < 2.0$. There are 75 untagged events which do not meet this requirement. Below we discuss the different mass subsamples used in this analysis.

5.4.1 Mass Subsamples

There are a total of 76 events in the Run I mass sample. These events are arranged into four non-overlapping subsamples according to their tagging information. Monte Carlo simulations show that the statistical uncertainty on the mass measurement decreases by 17% when the results from the fits from the different subsamples are combined together. The four non-overlapping subsamples are described as follows:

- SVX Double: Events with two SVX tags;
- SVX Single: Events with one and only one SVX tag;
- SLT: Events with one or two SLT tags, but no SVX tags;

- No Tags: ≥ 4 jets with $E_T > 15$ GeV and $|\eta| < 2.0$.

The number of Run I data events for each subsample is given in Table 5.1. In the top quark mass analysis we assume that the four highest E_T have come from the decay of the $t\bar{t}$ pair. Thus, we only consider the tags on these four jets when we distribute the events into the subsamples. The table also shows the S/B ratio for the different subsamples using the background estimates presented in Section 7.2.

Data Sample	Number of Events	Expected S/B
SVX Double	5	24
SVX Single	15	5.3
SLT	14	0.8
No Tags	42	0.4

Table 5.1: Subsamples used in the lepton+jets mass analysis and the expected signal to background ratio (S/B) for each. See Section 7.2 for background estimates for these subsamples.

5.5 Backgrounds

Background events can mimic the same final state as top events, but they are produced by a different physical process. The sources of background in the lepton+jets channel come from W+multijet events with heavy flavor, non-W events, mistags, single top events, diboson events (WW, ZZ), $Z \to \tau^+\tau^-$, and Drell Yan. Next, we describe the main sources of background for both the SVX and SLT tagging algorithms.

The largest source of background in the SVX tagged sample is inclusive W production in association with jets containing a b or a c quark, for example $p\bar{p} \to Wg(g \to$

$b\bar{b}$). The efficiency for tagging each background source is calculated from Monte Carlo samples as a function of jet multiplicity. These effiencies along with the background fractions are compared to the number of W+jets events observed in the data to estimate the contribution from each source. The second largest background source comes from mistags. These are jets which do not contain a true displaced vertex. From Monte Carlo studies, the transverse decay length, L_{xy}, distribution is symmetric about zero. The negative side of the L_{xy} distribution comes primarily from reconstruction errors in light quarks. To calculate the background from mistags, we parametrize the negative L_{xy} measured in generic jet data as a function of E_T, η, and the number of SVX tracks in the jet. This parametrization is applied to the W+jets data to predict the number of mistag events.

The dominant source of background in the SLT tags comes from "fake" soft lepton tags. These are tags due to particles which did not originate from heavy flavor decay. Fake tags include non-leptons that pass the lepton requirements such as pions faking an electron or muon, electrons from photon conversions, pions or kaons decaying on flight, etc. The fake background is calculated by measuring the fraction of tags per track in a generic jet sample as function of track P_T.

The remaining background sources are estimated similarly for both taggers. The non-W background, expected to be mostly $b\bar{b}$ events, is calculated from the data by measuring the number of tags as function of lepton isolation and $\not\!\!E_T$. The single top background in determined from Monte Carlo studies by measuring the W^* and W-gluon production, and normalizing them to the respective theoretical cross sections. The rest of the background sources are relatively small for both taggers and are evaluated from Monte Carlo predictions.

5.6 Monte Carlo Simulation

We would like to note that Monte Carlo studies are one of the most important parts of particle physics. Using Monte Carlo samples we can calculate detector efficiencies, understand the detector response to a particular physics process, develop new analyses techniques, or obtain background calculations.

Monte Carlo numerical methods are described as statistical simulation techniques used to simulate a physical process, which is described by probability density functions (pdf's), by random sampling from these pdf's. In our case, the physical process that we want to simulate is the collection of $t\bar{t}$ events by the CDF detector.

Monte Carlo event generators are used to simulate the physics of $p\bar{p}$ collisions. The event generators output a list of four-vectors of the final state particles which are used as the input to the CDF detector simulation. Jet and lepton identification are accomplished using the same algorithms that are applied to the real data. This allows the same selection criteria to be applied to the Monte Carlo and the Run I data.

5.6.1 Top Samples

The Monte Carlo program used to generate $t\bar{t}$ events is PYTHIA 6.129a [32]. Additional checks are provided by Monte Carlo $t\bar{t}$ events generated with HERWIG [31]. PYTHIA is based on the leading order QCD matrix elements for hard scattering processes. It uses the parton-shower approach for initial and final state radiation, and partons are fragmented using the Lund string model [33]. The default set of parton distributions functions is GRV 94L [56]. B hadron decays are modeled with the CLEOMC package [34].

5.6.2 Background Sample

The Monte Carlo program used to study the kinematic characteristics of the background is VECBOS [35]. This program is based on tree-level matrix element calculations for W+jets production. The simulated events produced by VECBOS contain a W boson and up to four additional final state partons. These partons are input into the same parton shower evolution and cluster hadronization used in the HERWIG program. The VECBOS events generated for this analysis use the W+3 parton matrix elements, with a required additional jet produced during parton showering. The VECBOS Monte Carlo generator has been shown to reproduce distributions of a wide range of kinematic variables in a large sample of Run I W+jets events [36].

CHAPTER 6

MASS RECONSTRUCTION

This chapter describes the corrections applied to all the $t\bar{t}$ decay products to better approximate the energy of the original partons. The raw energies are corrected for non-instrumented regions, non-linear response of the calorimeter, multiple interactions at high luminosity, and other effects. These corrected energies serve as inputs to a χ^2 from which the corrected four momenta of all the particles in the lepton+jets decay chain, $p\bar{p} \rightarrow t\bar{t} \rightarrow \ell\nu q\bar{q}' b\bar{b}X$, can be reconstructed by applying a series of energy-momentum kinematic constraints. Therefore, the reconstructed top quark mass, M_{rec} , can be calculated in an event-by-event basis.

6.1 Jet Energy Corrections

Before the jet energies are used to reconstruct the mass of the top quark, they are corrected to better estimate the original momenta of the daughter partons in $t\bar{t}$ decay. The corrections applied to the jets are done in two different step. First, a set of generic corrections are applied to all jets with a raw $\mathrm{E_T} > 8$ GeV and $|\eta| < 2.4$. These adjustments are intended to provide an estimate of the true jet energy from the observed jet energy. Then, the second set of corrections is applied to the leading four jets of the event which are assumed to be the $t\bar{t}$ decay partons. These corrections

are applied after the generic ones, and map the jet momenta to the momenta of the partons in top events.

The generic corrections account for the mismeasurement of the jet energies due to detector effects, energy falling outside the clustering cone, and contributions from the underlying event and multiple interactions. After accounting for all these factors, the corrected energy, $\mathrm{P_T^{cor}}$ can be written in terms of the raw energy, $\mathrm{P_T^{raw}}$, as follows:

$$\mathrm{P_T^{cor}}(R) = (\mathrm{P_T^{raw}}(R) \times f_{rel} - UEM(R)) \times f_{abs}(R) - UE(R) + OC(R). \qquad (6.1)$$

The parameter R is the cone radius and for this analysis it is set to 0.4. The different corrections are given below:

- f_{rel} is the relative energy scale, and it corrects for the relative response of the different calorimeter sections to that of the central region of the calorimeter, $0.2 < |\eta| < 0.7$.

- There are two kinds of underlying event corrections: $UEM(R)$ and $UE(R)$. The first correction takes into account the extra energy of the event due to multiple interactions. The $UE(R)$ correction takes into account the extra energy from the primary $p\bar{p}$ interaction due to fragmentation of particles not associated with the hard scattering.

- f_{abs} is the absolute energy scale, and it attempts to map the raw jet energy observed in a cone of radius R into the average true jet energy. This correction accounts for the detector response and any effects due to the fact that particles interact in the calorimeter and are experimentally observed as jets.

- $OC(R)$ corrects for the energy expected to be outside of the jet cone radius of 0.4.

The four leading jets in the $t\bar{t}$ decay undergo an additional energy correction which depends on the jet type. In the lepton+jets channel the jets originate from four different sources: a light quark, a hadronically decaying b quark, a b quark decaying semileptonically into the electron channel, and a b quark decaying semileptonically into the muon channel. To derive the $t\bar{t}$ specific jet momentum corrections we generate a Monte Carlo sample that is processed using the CDF simulation and reconstructed in the same way as the data. First, we match the Monte Carlo partons to the reconstructed jets in $\eta - \phi$ space. Then, an average correction factor is obtained by comparing the $\mathrm{P_T}$ of the Monte Carlo partons to the $\mathrm{P_T}$ of the reconstructed jets. The correction factor is a function of the jet's $\mathrm{P_T}$ and it is given by the median of the distribution of $\Delta_{\mathrm{P_T}} = (\mathrm{P_T}(parton) - \mathrm{P_T}(jet))/\mathrm{P_T}(jet)$.

Figure 6.1 shows the size of the $t\bar{t}$ specific corrections for the four types of jets: (A) jets from hadronic W decays. (B) jets from b quarks, (C) b jets containing an electron, (D) b jets containing a muon. As we can see the larger corrections correspond to b jets containing a soft lepton in their decay. This is due to the undetected neutrino in the semileptonic decay, and for the muon channel, the correction also accounts for the fact that muons deposit very little energy in the calorimeter.

6.2 The Unclustered Energy

In general, $t\bar{t}$ events are very energetic, and they show high jet activity. Thus, we write the $t\bar{t}$ production mechanism as $p\bar{p} \rightarrow t\bar{t} + X$ where X is made up of all the unspecified particles which recoil against the $t\bar{t}$ system. The quantity $\mathrm{X_T}$ is given by all the energy in the event not counting the primary lepton and the four leading jets.

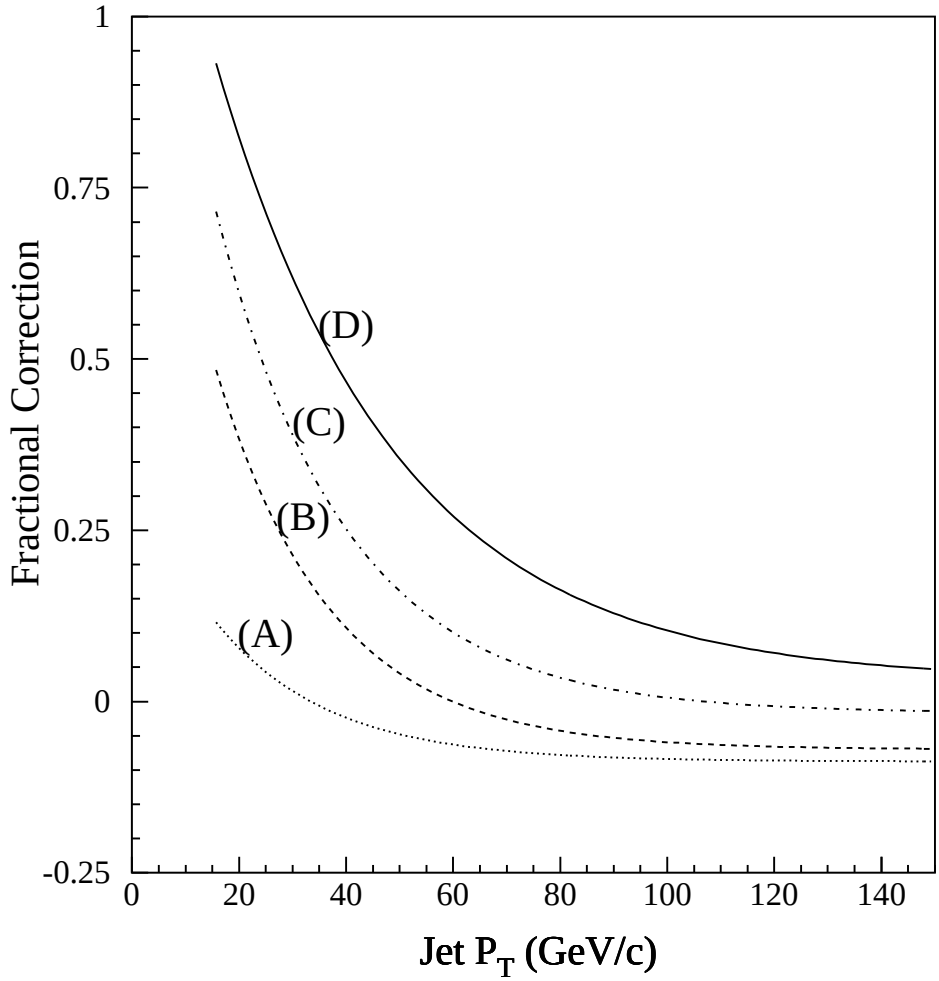

Figure 6.1: The $t\bar{t}$-specific corrections applied to jets according to available b-jet information. The curves show the fractional change to the corrected jet P_T after all the generic jet corrections have been applied. The curves are for: (A) jets from the decay of W bosons, (B) jets from all b quarks, (C) jets from b quarks containing an electron, and (D) jets from b quarks containing a muon.

The equation can be written as:

$$\vec{X}_T = \vec{U}_T + \sum_{i=5}^{N_{jets}} \vec{E}_T(jet) \tag{6.2}$$

where the transverse unclustered energy, $\mathrm{U_T}$, is given by all the energy in the calorimeter which is not clustered into jets. Each component of the unclustered energy is corrected with a single factor $f_{u.e.} = 1.6$ based on studies of Z boson events with no extra jets, where the Z boson is well measured by the two leptons it decays into. The precision of the unclustered energy is not very well understood so the uncertainty of each component of $\vec{U}_\mathrm{T}$ is taken to be 100%.

6.3 Event Reconstruction

In Chapter 5 we describe how the lepton and jets are identified in the CDF detector. From the tagging information it is possible to determine whether a jet has come from a b quark or not. We have also shown how the existence of the neutrino is inferred by an imbalance in the total transverse energy measured in the calorimeter. Then, by correcting the raw energies of these top decay products the original energy of the t and $\bar{t}$ daughter partons can be better estimated. The challenge now is how to combine the decay products to reconstruct the mass of the top quark.

A fitting algorithm is used to calculate the four momenta of the t and $\bar{t}$ for a given event by reconstructing the four momenta of the six particles in the decay: ℓ, ν, b, $\bar{b}$, q, and $\bar{q}$[7]. From the corrected observed quantities and their uncertainties along with a series of kinematic constraints we are able to construct a χ^2, which can be minimized to yield the best estimates for the four momenta of the particles in the $t\bar{t}$

[7]It should be noted that although we sometimes refer to the decay products at the parton level, it is the mass fitting code which determines their final assignment

decay. For the minimization we assume the masses of the partons to be 0.5 GeV/c^2, except for the b, and $\bar{b}$ which are assigned a mass of 5.0 GeV/c^2.

The hypothesis of Standard Model $t\bar{t}$ implies the production process

$$p\bar{p} \rightarrow t + \bar{t} + X, \tag{6.3}$$

followed by the decays

$$t \rightarrow W^+ + b, \tag{6.4}$$

$$\bar{t} \rightarrow W^- + \bar{b}, \tag{6.5}$$

$$W^\pm \rightarrow \ell^\pm + \nu, \tag{6.6}$$

$$W^\mp \rightarrow q + \overline{q}'. \tag{6.7}$$

where the quantity X is described in Section 6.2.

The χ^2, which is minimized with the program MINUIT [38], is given below. This χ^2 formulation includes the energy and kinematic constraints imposed by the $t\bar{t}$ decay hypothesis. The χ^2 is expressed as:

$$\begin{aligned}\chi^2 &= \sum_{\ell,jets} \frac{(\hat{P}_T - P_T)^2}{\sigma^2_{P_T}} + \sum_{i=x,y} \frac{(\hat{U}'_i - U'_i)^2}{\sigma^2_{U'_i}} + \frac{(\hat{M}_{\ell\nu} - M_W)^2}{\sigma^2_{M_W}} + \\ &\quad \frac{(\hat{M}_{jj} - M_W)^2}{\sigma^2_{M_W}} + \frac{(\hat{M}_{\ell\nu j} - M_{rec})^2}{\sigma^2_{M_{rec}}} + \frac{(\hat{M}_{jjj} - M_{rec})^2}{\sigma^2_{M_{rec}}}.\end{aligned} \tag{6.8}$$

The notation is as follows: ℓ represents the primary lepton, ν refers to the neutrino, and j refers to one of the four leading jets in the event. The first sum is over the lepton and all the jets with $E_T > 8.0$ GeV and $|\eta| < 2.4$. The second sum is over the transverse components of the unclustered energy. The hatted symbols in the sums represent quantities that are modified by the fit procedure, whereas unhatted symbols represent the input values. M_W is the mass of the W boson and it is 80.4 GeV/c^2

with an uncertainty, σ_{M_W}, set to its theoretical width of 2.1 GeV/c^2. The uncertainty in the top mass is set to its theoretical width of 2.5 GeV/c^2.

In summary, there are five effective constraints in our χ^2: the transverse momentum of the components of the $t\bar{t}$ system must be zero, the invariant mass of the $\ell\nu$ system must be equal to the W mass, the invariant mass of the $q\bar{q}'$ system must also be equal to the mass of the W boson, the two three-body invariant masses, M_{jjj} and $M_{\ell\nu j}$, must each be equal to the reconstructed top mass, M_{rec}. After minimizing the χ^2 with respect to all of the available information, we obtain an event-by-event determination of the top quark mass, M_{rec} .

6.3.1 Combinatorics

Depending on the tagging information of the event there are many different ways we can combine the top decay partons to reconstruct the top mass. If none of the jets are tagged by the SVX or the SLT algorithms, there are 12 different ways we can assign the four leading jets to the partons in the fit. Combined with the $\mathrm{P_Z}$ of the neutrino $\mathrm{P_Z^\nu}$ ambiguity[8], there are 24 different configurations. If one of the jets is tagged as a b candidate, the number of possible configuration reduces to six, 12 if we include the $\mathrm{P_Z^\nu}$ ambiguity. We note that if a jet is tagged by either tagging algorithm, it is automatically assigned to one of the b quarks. If two of the jets have been tagged there are two possible jet assignments, the two b jets can be interchanged, and therefore four possible configurations.

When we reconstruct the mass of the top quark, all the possible configurations are ran through the fitting code and the combination with the lowest χ^2 is chosen as the

[8]The neutrino $\mathrm{P_Z}$ is determined from the W mass constraint. Thus solving $M_W = \sqrt{E_W^2 - P_W^2}$ leads to two possible solutions.

best solution. We require this solution to have a $\chi^2 < 10.0$ as stated in Section 5.4. The efficiency for this cut for each of the mass subsamples is given in the table below 6.1. The table shows the result for $t\bar{t}$ signal as well as the VECBOS background. The efficiency of the χ^2 cut in the SVX Double background subsample is not available because there are only a handful of events.

Mass subsample	χ^2 efficiency ($t\bar{t}$)	χ^2 efficiency (VECBOS)
SVX Double	$73.4 \pm 0.6\%$	N/A
SVX Single	$90.3 \pm 0.4\%$	$81.9 \pm 0.8\%$
SLT	$89.7 \pm 0.6\%$	$80.7 \pm 1.3\%$
No Tags	$96.5 \pm 0.2\%$	$95.4 \pm 0.4\%$

Table 6.1: Efficiency of the χ^2 cut for each of the mass subsamples. The probability of finding a solution with $\chi^2 < 10.0$ increases as the number of possible configurations increases.

6.3.2 Gluon radiation

From the Monte Carlo plus detector simulation we find that roughly 50% of the $M_{top} = 175$ GeV/c^2 $t\bar{t}$ events contain at least one jet that cannot be matched to any of the top decay partons. These extra jets are emitted during the production of the $t\bar{t}$ pair, initial state radiation (ISR), or during the $t\bar{t}$ decay, final state radiation (FSR) [39]. From a theoretical perspective, the jets that are produced after the creation of the $t\bar{t}$ pair should be included into the mass reconstruction only if the top quark is on shell. These jets carry a fraction of the energy of the top quark and therefore their inclusion would translate in a more accurate top mass estimate. However, it is not possible to distinguish jets from ISR to those from FSR. Furthermore, production and decay stage radiation cannot be differentiated from one another in the detector.

Thus, in a large fraction of the Monte Carlo events the four highest-E_T jets do not correspond to the top decay partons.

A possible solution is to add one more jet to the mass fitting code, but this would increase the number of combinations by a factor of three, four, and five for the SVX Double, SVX Single or SLT, and No Tags events respectively. The increase in the number of possible configurations decreases the probability of choosing the correct solution. This is why only the four highest-E_T jets are considered for the mass analysis.

A more elegant solution to this problem would be to determine the amount of gluon radiation in the data. Chapter 10 describes how this can be accomplished by using Neural Networks.

CHAPTER 7

TEMPLATE-BASED MASS ANALYSIS

The best estimate of the top mass is obtained by comparing the shape of the M_{rec} distribution to Monte Carlo expectations. To improve the mass measurement these Monte Carlo distributions are fitted with a single functional form. Using a Neural Network method we can parametrized any distribution with mass information. A maximum likelihood method is used to extract M_{top} from the parametrized forms for $t\bar{t}$ signal and background, along with the constraints of the background fractions.

7.1 Neural Network Template Parametrization

There is only a finite number of Monte Carlo statistics we can generate. In $t\bar{t}$ Monte Carlo samples, it is observed that the shape of the M_{rec} distributions, depend on the generated top quark mass. For different top masses, the shape of the reconstructed mass distributions are very similar. They tend to peak around their generated top mass and they have asymmetric tails. This is quite reasonable, since we only change the mass of the top quark but the physics processes responsible for the distributions remain the same. This is the reason why it is reasonable to model the parent distribution of the templates by a single functional form dependent only upon the mass of the top quark.

In previous analyses, a combination of a Gaussian plus a Gamma function (GG) has been used to fit the mass templates as a function of M_{top} [40]. The use of these distributions is motivated by the shape of the M_{rec} distributions, but it should be noted any function that could fit the templates could be used. Thus, we propose using a Neural Network to perform this fits. This technique has been employed by the ALEPH experiment to fit templates of different Higgs masses [41].

We use the PYTHIA Monte Carlo plus the detector simulation to model the shape of the $t\bar{t}$ events. For this study we have generated samples raging from a top mass of 120 GeV/c^2 to 230 GeV/c^2. To model the shape of the background we use the VECBOS Monte Carlo program. All the MC generated events are put through the detector simulation.

7.1.1 Fitting the M_{rec} Distributions

Figure 7.1 shows a 2-D histogram of the reconstructed mass (M_{rec}) versus the generated top mass (M_{top}) for the Single SVX subsample. We want to use a Neural Network to model the $t\bar{t}$ reconstructed mass distribution for any given value of the input top quark mass. For this purpose we have chosen the Neural Network package MLPfit [42]. MLPfit uses the sigmoid function, $\mathcal{F}(y) = \frac{1}{(1+e^y)}$ for the hidden units, and the identity function $\mathcal{F}(y) = y$, for the output units.

As explained in chapter 4, the nature of the problem dictates how many input and output units are needed to solve the problem. In this case the NN has two input units: the first input is associated with the reconstructed mass, the x-axis in Figure 7.1, while the second input corresponds to M_{top} , the y-axis in Figure 7.1. For a more detailed description of the NN fitting procedure we refer the reader to Appendix C.

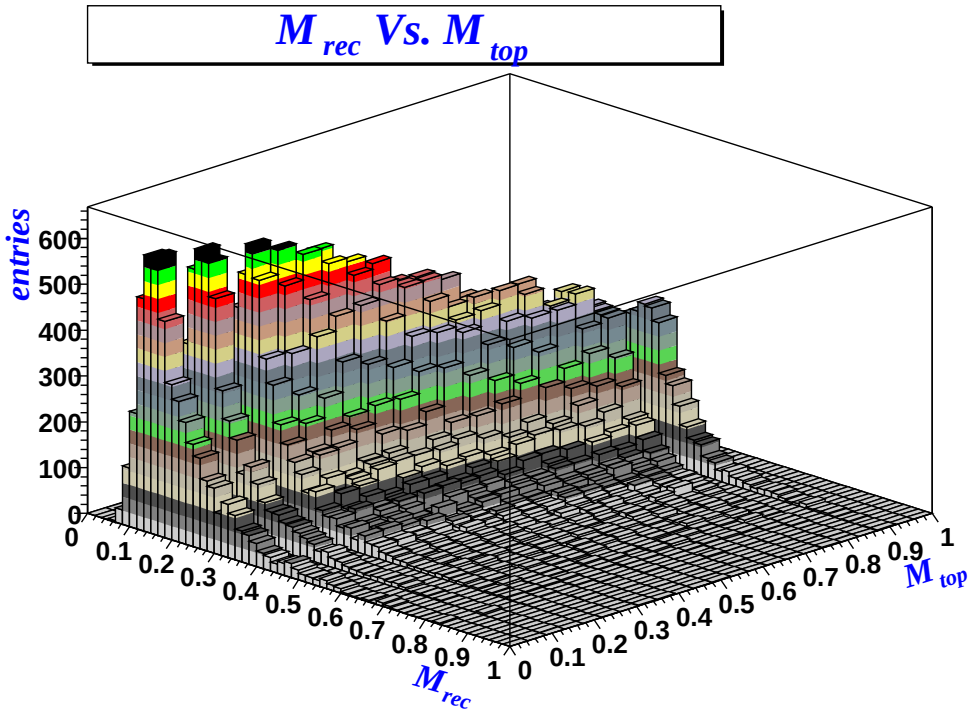

Figure 7.1: We show a 2-D histogram of all the Single SVX mass templates as a function of top mass. The M_{rec} axis is normalized to be between 0 (M_{rec} = 80 GeV/c^2) and 1 (M_{rec} = 380 GeV/c^2). The M_{top} axis is also normalized to be between 0 (M_{top} = 120 GeV/c^2) and 1 (M_{top} = 230 GeV/c^2).

The architecture of the networks for the $t\bar{t}$ signal is 2-4-4-1. This corresponds to 37 possible weights and biases (there are 12 free parameters in the Gaussian plus Gamma fit). We would like to emphasize that this is not the only solution to the problem. It is possible to obtain similar results using a single hidden layer. However, in this case we have found that with two hidden layers the network converges to a solution more rapidly and has better generalization.

To determine whether or not the networks are able to fit the distributions correctly we define a χ^2 function as follows:

$$\chi^2 = \sum_{M_{top}=120...230} \sum_{\text{bin i}}^{60} \frac{(b(i, M_{top}) - \mathcal{O}(M_{rec}, M_{top}))^2}{\sigma^2_{b(i,M_{top})^2}} \tag{7.1}$$

where $b(i,\ M_{top}\ )$ is the content of th i-th bin of the histogram obtained with a top quark mass M_{top} , $\sigma^2_{b(i,M_{top})^2}$ is the corresponding bin error, and $\mathcal{O}(M_{rec}\ ,\ M_{top}\ )$ is the output given by the NN. The fit χ^2 per degree of freedom (dof) are shown in Table 7.1. We also show the fit χ^2 per dof from fitting the same distributions using the previous Gaussian plus Gamma parametrization method.

Top Subsample	NN templates	GG templates
SVX Single	0.98 for 919 dof	1.29 for 944 dof
SVX Double	1.27 for 778 dof	1.30 for 812 dof
SLT	1.08 for 831 dof	1.24 for 856 dof
No Tags	1.37 for 920 dof	1.59 for 945 dof

Table 7.1: We show the χ^2 per degree of freedom for each of the four subsamples. The results from both fitting techniques are given.

The results from both fitting techniques are shown in Figures 7.2 through 7.5 where the blue curves correspond to the NN fits while the green curves show the results from the GG fits. Both sets of curves look fairly similar for low top masses. However, at high top masses the NN curves peak higher. Thus, the NN shows better separation between the different masses.

The background distributions do not have any mass dependence. Thus, the architecture of the networks is 1-5-1, which corresponds to 16 weights and biases (there are three free parameters in the Gaussian plus Gamma fit). The one input is associated with reconstructed mass, and the desired output value is set to the number of entries in each bin of the background subsample. Figure 7.10 shows the parametrization of the background distributions for the Single SVX, SLT, and No Tags subsamples. The results from the GG fits are given by the green curves, while the NN fit results

are given by the blue curves. Because of limited statistics and low probability for obtaining two SVX tagged jets in the VECBOS Monte Carlo simulation, we assume the same background distribution for the SVX-tagged subsamples.

This analysis uses the same VECBOS samples generated for past analyses [43]. The default background samples were generated with a scale $Q^2 = \langle P_T \rangle^2$. To study the effect of the background shape on the top measurement, another set of templates was generated with a scale $Q^2 = \langle M_W \rangle^2$. To increase the background statistics we have combined both sets of events. We have performed a series of test, which show this combination does not affect the mass of the top quark (Appendix D).

7.1.2 Fitting the $\mathrm{H_T}$ Distributions

One of the benefits of fitting a distribution with the NN method is that we do not need to make an *a priori* decision of what the underlying function describing the distribution is. Thus, a NN can approximate any distribution independently of its shape.

Previous analysis have shown that there are other kinematic variables besides M_{rec} that have mass information [47]. One of these variables is the total transverse energy of the event, $\mathrm{H_T}$. In Chapter 8, we study the possibility of using the mass information from several variables to calculate the top mass. Here, we are interested in showing how the NN fitting method is applicable to any distribution of a single variable. The fit χ^2 per dof for the Single SVX, Double SVX, SLT, and No Tags are: 936.5 for 914 dof, 766.3 for 835 dof, 783.1 for 866 dof, and 896.1 for 886 dof. The signal $\mathrm{H_T}$ fit results are shown in Figures 7.6 through 7.9. The background $\mathrm{H_T}$ fit results are shown in Figure 7.11. Figures only show the NN parametrization results.

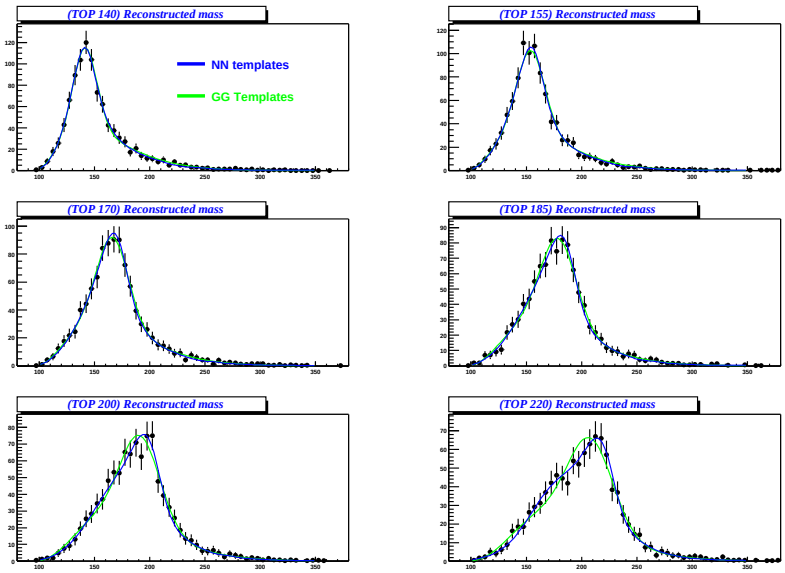

Figure 7.2: Reconstructed mass distributions for single SVX events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the GG templates (green) and the NN templates (blue) are also shown.

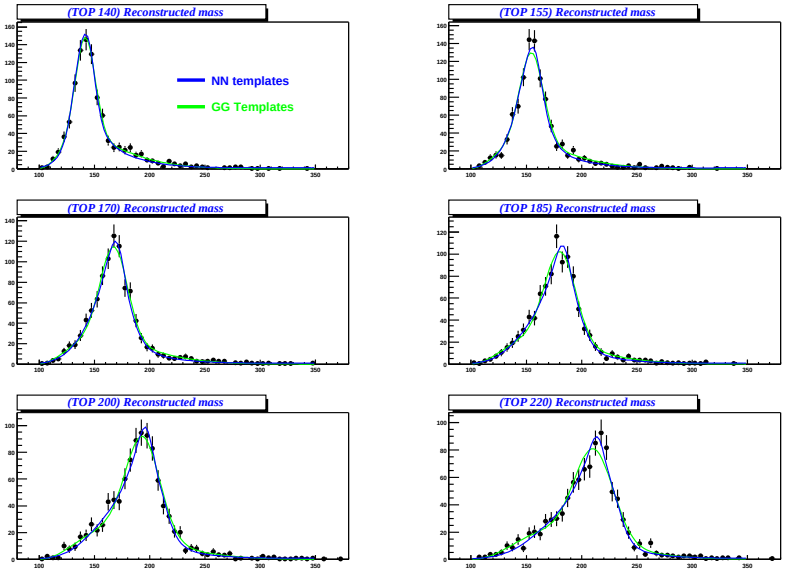

Figure 7.3: Reconstructed mass distributions for double SVX events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the GG templates (green) and the NN templates (blue) are also shown.

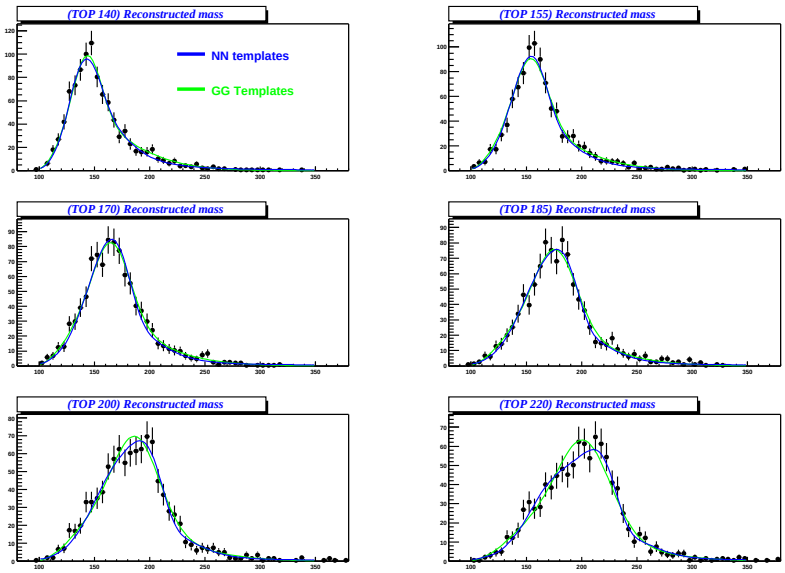

Figure 7.4: Reconstructed mass distributions for SLT events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the GG templates (green) and the NN templates (blue) are also shown.

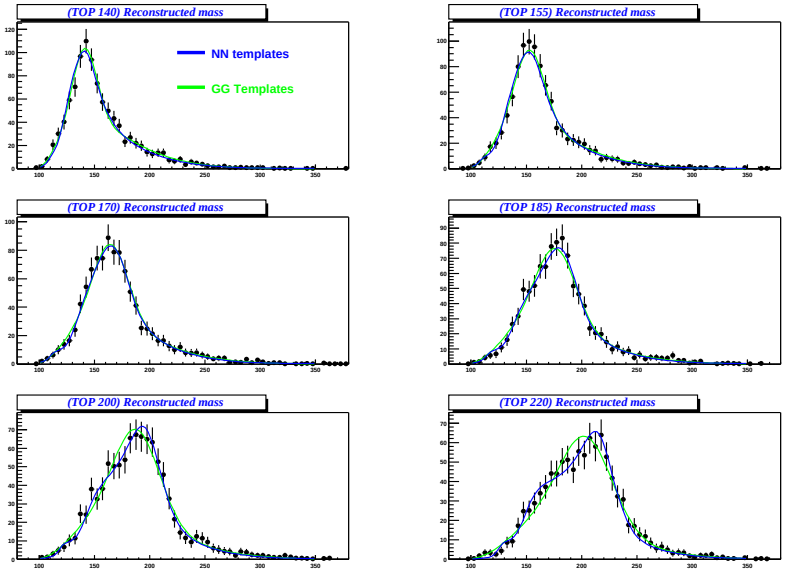

Figure 7.5: Reconstructed mass distributions for events with no SLT or SVX tags in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the GG templates (green) and the NN templates (blue) are also shown.

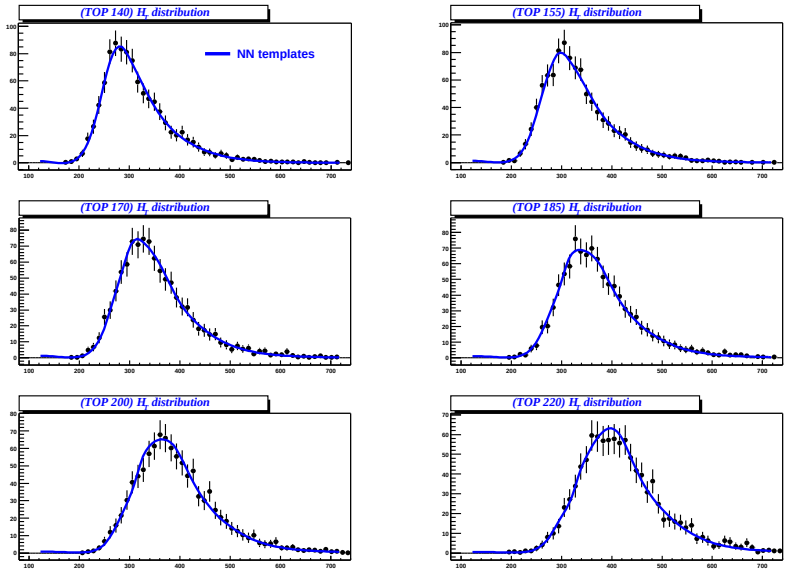

Figure 7.6: H_T distributions for single SVX events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the NN templates are also shown.

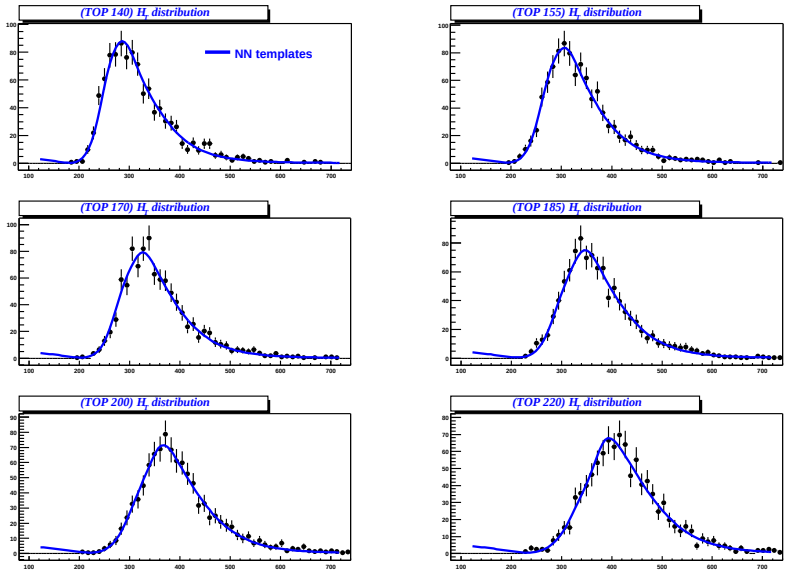

Figure 7.7: H_T distributions for double SVX events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the NN templates are also shown.

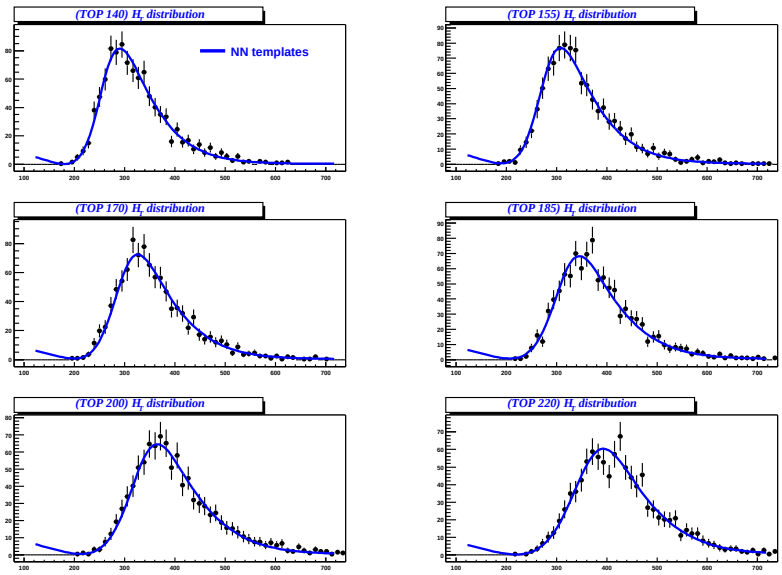

Figure 7.8: H_T distributions for SLT events in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the NN templates are also shown.

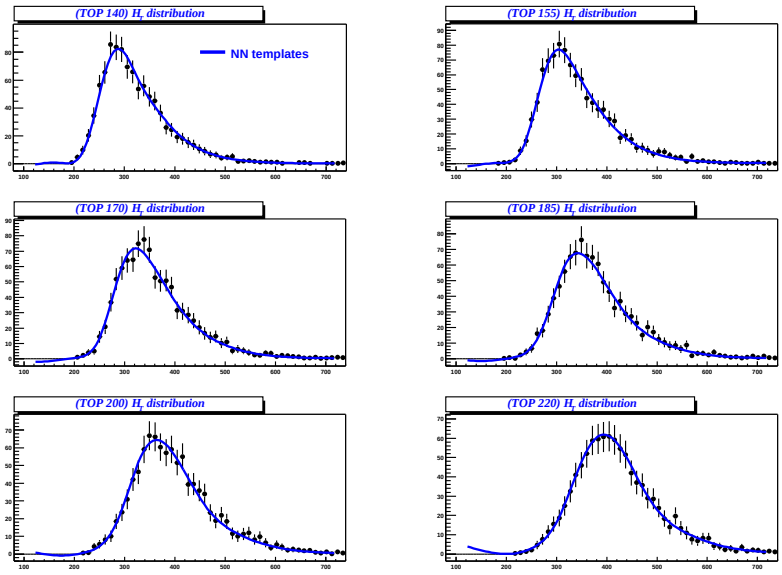

Figure 7.9: H_T distributions for events with no SLT or SVX tags in $t\bar{t}$ Monte Carlo generated with a top quark mass of (a) 140, (b) 155, (c) 170, (d) 185, (e) 200, and (f) 220 GeV/c^2. The fits from the NN templates are also shown.

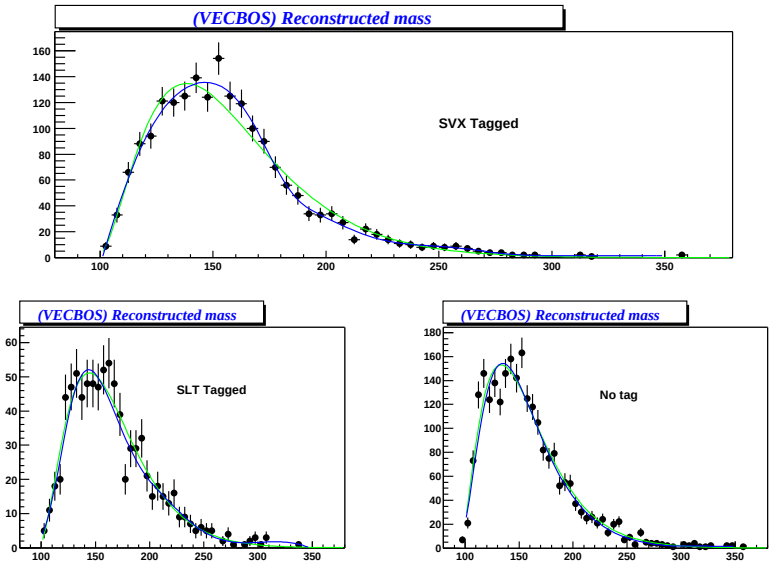

Figure 7.10: Reconstructed mass distributions for VECBOS background, which contains events generated with $Q^2 = \langle P_T \rangle^2$ and $Q^2 = \langle M_W \rangle^2$ (see Appendix D). The fits from the GG templates (green) and the NN templates (blue) are also shown.

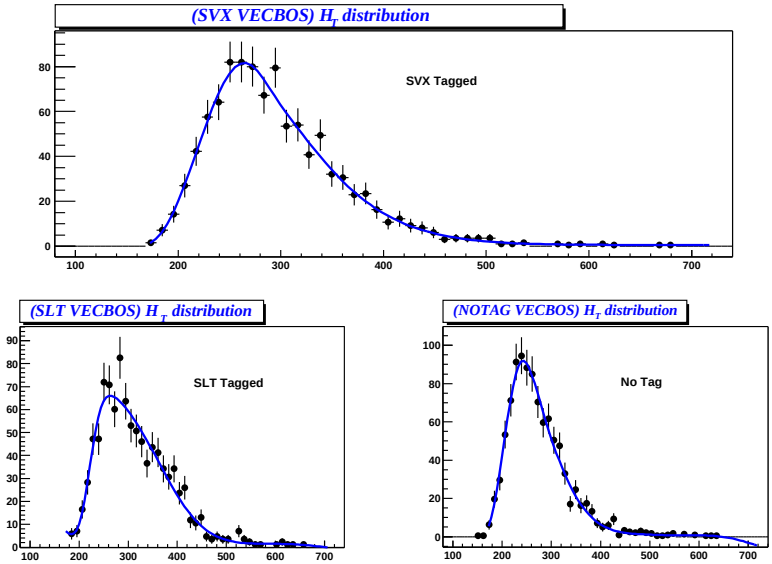

Figure 7.11: H_T distributions for VECBOS background, which contains events generated with $Q^2 = \langle P_T \rangle^2$ and $Q^2 = \langle M_W \rangle^2$ (see Appendix D). The fits from the NN templates are also shown.

7.2 Background Calculation

The computation of the background fraction in each mass subsample is first achieved by estimating the number of background events from the known processes described in Section 5.5 for:

- class I: events that have exactly three jets with $E_T > 15$ GeV and $|\eta| < 2$ and one or more jets with $E_T > 8$ GeV and $|\eta| < 2.4$.
- class II: events with four or more jets with $E_T > 15$ GeV and $|\eta| < 2$. This class of events has a larger S/B ratio than class I.

The expected number of $t\bar{t}$ and background fractions in the top mass sample are then estimated using a maximum likelihood fit that compares the observed rates of SVX and SLT tagged events with predicted rates. The $t\bar{t}$ fraction is a free parameter in the fit, and it is allowed to vary to optimize the agreement between the number of observed and predicted tagged events. The returned $t\bar{t}$ fraction value is combined with the SVX and SLT tagging probabilities to estimate the expected amount of $t\bar{t}$ signal and background in each of the mass subsamples. The detailed calculation of the background is given in Ref. [40]. Here, the results are presented in Table 7.2.

7.3 Likelihood Procedure

A maximum likelihood method is used to extract the value of M_{top} for each mass subsample. The likelihood is used to characterize the similarity between the reconstructed masses of the data events and the $t\bar{t}$ and background Monte Carlo samples. In the fit, the background fraction x_b is constrained by a Gaussian distribution. The only parameter that is completely unconstrained in the fit is M_{top}. Since the four

Item #	Process	SVX Single	SVX Double	SLT (no SVX)	No Tags	Total
1	non-W/Z	0.5	0.0	1.0	4.6	6.1
2	WW	0.1	0.0	0.1	0.6	0.8
3	WZ	0.0	0.0	0.0	0.1	0.1
4	ZZ	0.0	0.0	0.0	0.1	0.1
5	$Z \to \tau\tau$	0.1	0.0	0.2	0.5	0.8
6	Single Top	0.2	0.0	0.1	0.2	0.4
7	$Wc + Zc$	0.2	0.0	0.8	1.7	2.7
8	$Wb\bar{b} + Zb\bar{b}$	0.8	0.2	0.4	1.1	2.5
9	$Wc\bar{c} + Zc\bar{c}$	0.4	0.0	0.8	2.0	3.2
10	$W/Z + u, d, s$	0.2	0.0	4.1	19.6	23.9
	Background sum	$2.4^{+0.8}_{-0.7}$	0.2 ± 0.1	7.6 ± 1.3	$30.4^{+4.3}_{-4.7}$	40.7
11	$t\bar{t}$	12.6	4.8	6.4	11.6	35.3
	Observed events	15	5	14	42	76

Table 7.2: Number of background events for the four mass subsamples from various processes.

subsamples are completely independent, we can construct a joint probability for the mass sample by multiplying together the four individual likelihood functions. These four likelihoods have the same form:

$$\mathcal{L} = \mathcal{L}_{shape} \times \mathcal{L}_{background} \tag{7.2}$$

where:

$$\mathcal{L}_{shape} = \prod_{i=1}^{Nevents} [(1 - x_b) f_s(M_i, M_{top}) + x_b f_b(M_i)], \tag{7.3}$$

$$\mathcal{L}_{background} = P(x_b). \tag{7.4}$$

The term $\mathcal{L}_{shape}$ is the joint probability for a sample of N events to have come from a parent distribution with a signal fraction of $(1 - x_b)$ and a background fraction x_b. The probability distributions, f_s and f_b, are derived from either the GG templates

or the NN templates described in the previous sections. $\mathcal{L}_{background}$ constrains the fraction of background events to the value within its uncertainties.

In the past, the likelihood function included a $\mathcal{L}_{param}$ term to account for the limited number of statistics used to obtain the continuous functional forms of the templates [40]. This term depends on the parameters which determine the f_s and f_b distributions and the error matrices associated with them. It is possible to construct a $\mathcal{L}_{param}$ term for the Neural Network fitting method by using the Hessian matrix of the network. However, computing the Hessian can be extremely time consuming as the number of network weights increases [22]. Furthermore, Monte Carlo studies show that the inclusion of this term in the likelihood changes the returned mass and statistical error by an average < 0.01 GeV/c^2.

To calculate the top quark mass for each subsample, we minimize the -log $\mathcal{L}$ with respect to M_{top} and x_b. The statistical uncertainty in the top mass is given by the change in M_{top} which results in a 0.5 unit increase in the -log $\mathcal{L}$. The overall estimate of M_{top} and its uncertainty for the entire mass sample is obtained by multiplying the individual likelihoods together.

7.4 Results from Simulated Experiments

We ran 2000 pseudoexperiment consisting of four subsamples with the same number of events (N^i_{obs}, $i = 1, \ldots, 4$) as observed in the Run I data (Table 5.1). The number of background events in each of the simulated experiments is calculated in a two-step process:

1. Calculate random numbers from a Gaussian distribution: $N^*_{bkg} = \text{Gauss}(N^i_{exp}, \sigma^i_{exp})$

2. Calculate random numbers from a Poisson distributions: $N^i_{bkg} = \text{Poisson}(N^*_{bkg})$

where the index i runs over the four subsamples in the mass analysis N^i_{exp} and σ^i_{exp} correspond to the calculated background numbers shown in Table 7.2. The number of $t\bar{t}$ signal events is then given by $N^i_{sig} = N^i_{obs} - N^i_{bkg}$. The distinct values used for the pseudoexperiments are chosen randomly from the signal and background templates following the shape of the distributions. For each simulated experiment we obtain a top quark mass, a statistical uncertainty and a maximum likelihood value. The likelihood method is tested by comparing these results with expectations.

Table 7.3 shows the median mass and uncertainty resulting from the 2000 simulated experiments. As we can see the median statistical uncertainty when using the M_{rec} NN templates is roughly 12% better than the uncertainty obtained by using the M_{rec} GG templates. Although, we have not added any new information to the top mass analysis, we find that by fitting the distributions more precisely we obtain a smaller uncertainty in the mass measurement. The results obtained from using the $\mathrm{H_T}$ NN templates are also presented.

Fitting Technique	M_{top} from M_{rec}	M_{top} from H_T
GG templates	175.1 ± 7.3 GeV/c^2	N/A
NN templates	174.8 ± 6.5 GeV/c^2	174.6 ± 11.3 GeV/c^2

Table 7.3: These are the results obtained by using the NN fitting technique and the GG fitting method for the reconstructed mass. We only show the results from the H_T NN templates since we were unable to obtain continuous templates with the previous Gaussian plus Gamma method.

The top mass distribution and pull distribution from the M_{rec} NN templates are shown in Figure 7.12. Similar plots for the M_{rec} GG templates and $\mathrm{H_T}$ NN templates are shown in Figures 7.13 and 7.14. The pull is constructed following the recipe

detailed in [46], and it is given by:

$$\text{If } \mathrm{M_{fit}} \leq \mathrm{M_{true}} \quad : \quad g = \frac{M_{true} - M_{fit}}{\sigma_+},$$
$$\text{otherwise} \quad : \quad g = \frac{M_{fit} - M_{true}}{\sigma_-}. \tag{7.5}$$

M_{fit} is the value return by MINUIT, while M_{true} is 175 GeV/c^2. The positive and negative MINOS errors [53] are given by σ_+, and σ_- returned by the fit. This definition guarantees that the percentage of pull distributions between -1 and $+1$ equals the coverage of the error interval returned by MINOS. This error interval corresponds to a coverage of 68.27% if we require 1σ errors. As expected the pull distributions shown have a mean of zero and a width close to one. The pull distributions are not exactly one because of the limited statistics in the pseudoexperiments. When we increase the number of events in the simulated experiments, the width of the pull distributions are closer to one. Furthermore, this slightly wider pull distributions have also been seen in Ref. [40]

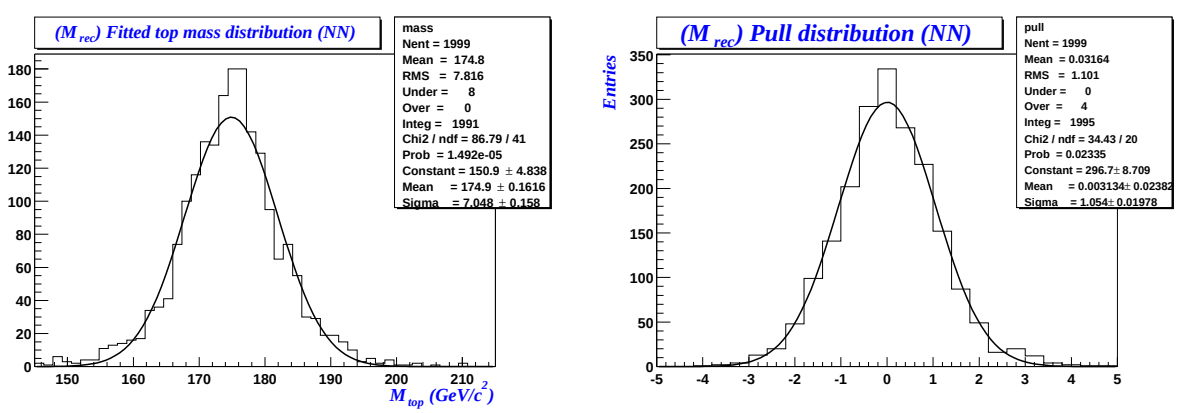

Figure 7.12: Results for 2000 pseudoexperiments using M_{rec} NN templates. (Left) Fitted top mass distribution. (Right) Pull distribution. Signal and background fractions as observed in the Run I data.

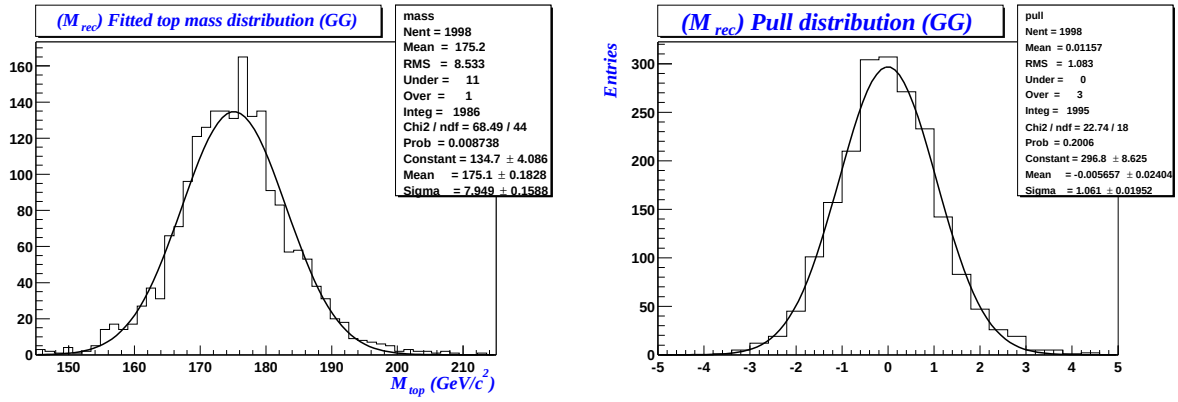

Figure 7.13: Results for 2000 pseudoexperiments using M_{rec} GG templates. (Left) Fitted top mass distribution. (Right) Pull distribution. Signal and background fractions as observed in the Run I data.

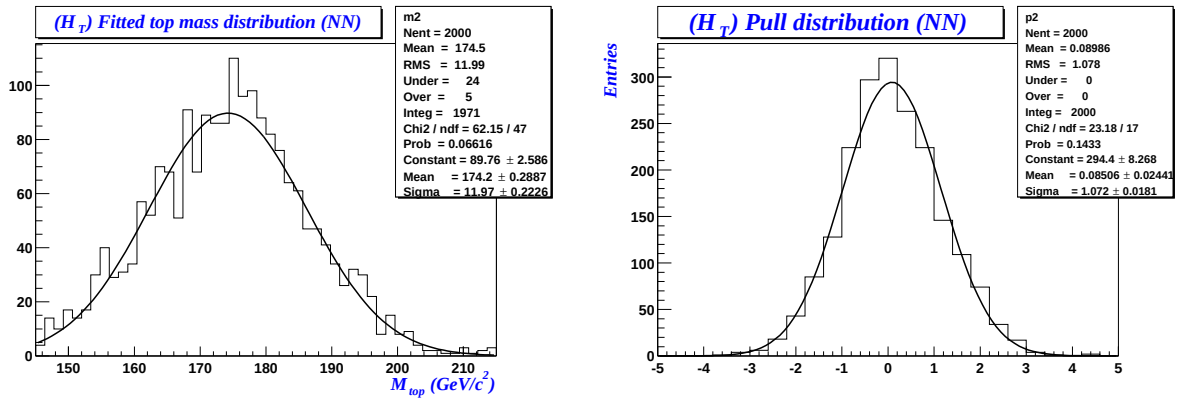

Figure 7.14: Results for 2000 pseudoexperiments using H_T NN templates. (Left) Fitted top mass distribution. (Right) Pull distribution. Signal and background fractions as observed in the Run I data.

7.5 Combining the Top Mass Results

We want to combine the results obtained using M_{rec} and $\mathrm{H_T}$. We ran 1000 pseudoexperiments in which we calculate the top mass using M_{rec} and $\mathrm{H_T}$ independently from one another. For each subsample, we construct a single top mass measurement as described in Appendix E. When we perform this combination, the correlation between $M_{top}^{M_{rec}}$ and $M_{top}^{\mathrm{H_T}}$ is set to the mean correlation value seen in the Monte Carlo. At this point we have combined the top mass measurement for each of the four subsamples. Since, these results are 100% independent from one another the final top measurement is given by the weighted average of the four results from the different subsamples. Figure 7.15 shows the results from the pseudoexperiments.

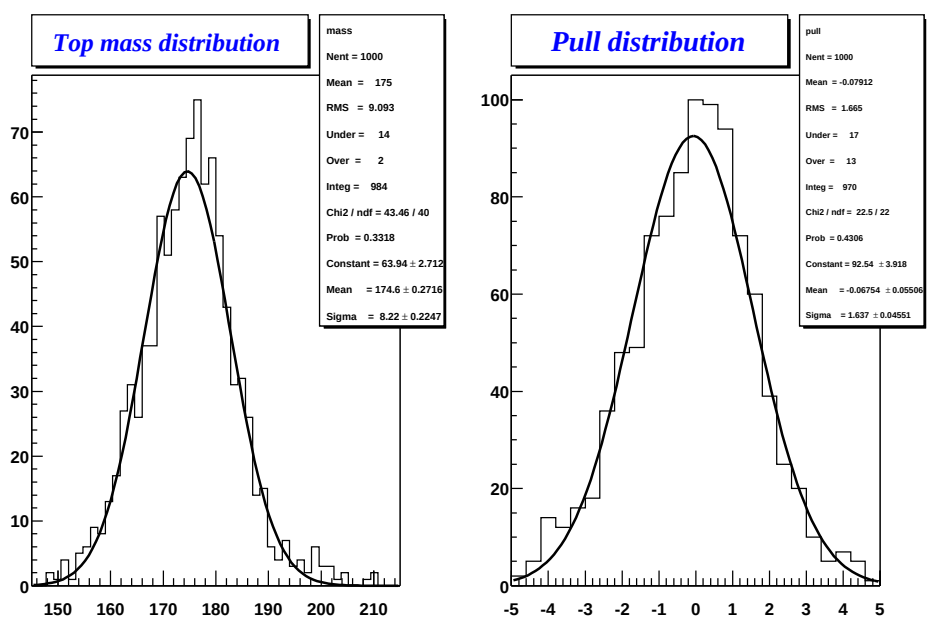

Figure 7.15: Results from 2000 pseudoexperiments in which we combine the top mass measurements obtained using the M_{rec} templates and the $\mathrm{H_T}$ templates. The left histogram shows the top mass distribution, and the right histogram shows the pull distribution.

As expected, the top mass distribution peaks around 175 GeV/c^2. However, the width of the distribution is wider than we would have expected. Recall that the M_{top} distribution obtained from only using the M_{rec} NN templates (Figure 7.12.(L)) has a width, which is 1 GeV/c^2 smaller that what we get when we combine the M_{rec} and $\mathrm{H_T}$ results. Furthermore, we see from the pull distribution that the errors returned by the fit are underestimated by 60%.

Combining the two correlated measurements becomes a very difficult task using traditional methods. Although, the pseudoexperiments return the correct mass, the error returned is incorrect. We have looked at two different methods based on Neural Network to combine the information from different variables. These techniques are described in the following chapters.

CHAPTER 8

STANDALONE NEURAL NETWORK MASS ANALYSIS

The goal of this analysis is to measure the top mass as accurately as possible. Recall that the previous best measurement of the mass only uses the reconstructed mass variable. However, other analyses have shown that other kinematic quantities have top mass information. The challenge is how to combine this information. A Neural Network provides a simple and elegant way of adding new information to the mass analysis. This chapter describes a standalone Neural Network method to extract the top quark mass.

8.1 Input Variables

We want to create a Neural Network with the capability to distinguish events that come from top quarks from different top masses. Thus, we need to find a set of mass dependent variables [47]. We have looked at different kinematic quantities and have found that the M_{rec} , the $\mathrm{H_T}$, the invariant mass of the $t\bar{t}$ system, $M_{t\bar{t}}$, and the sum of the $\mathrm{P_T}$'s of the two leading jets in the event, $\mathrm{P_T}(1) + \mathrm{P_T}(2)$ exhibit the greatest mass dependence. On the other hand, variables such as the $\not{E}_T$ or the $\mathrm{P_T}$ of the primary lepton have very little mass information. The first four quantities are described below in further detail:

- M_{rec} is described in Section 6.3.

- $\mathrm{H_T}$ is the sum of all the transverse energy in the event including the neutrino as well as the unclustered energy.

- $M_{t\bar{t}}$ is the invariant mass of the $t\bar{t}$ system. It only includes the four highest $\mathrm{E_T}$ jets are well as the lepton and the neutrino.

- $\mathrm{P_T}(1) + \mathrm{P_T}(2)$ is the sum of the two highest $\mathrm{E_T}$ jets. No tagging requirement is made.

All of these variables are constructed after applying the mass analysis corrections described in Section 6.1.

A good figure of merit to determine whether a variable has mass information or not is given by the quantity $\frac{RMS}{slope}$, where the RMS is the average RMS of the different distributions, and the *slope* is the slope given by the means of the distributions for the different top masses. A good variable will be narrow, with a small RMS, and it will show good separation between the peaks of the distributions, with a large slope. The results from our kinematic study are given in Table 8.1. Variables with a small $\frac{RMS}{slope}$ are more effective at measuring the top mass.

Variable	Avg. RMS	slope	$\frac{RMS}{slope}$
M_{rec}	27.9	0.62	45.1
H_T	72.3	1.35	53.6
$M_{t\bar{t}}$	83.2	1.58	52.6
$P_T(1) + P_T(2)$	43.7	0.77	56.8
Missing E_T	25.7	0.16	161.8
lepton P_T	27.1	0.15	181.1

Table 8.1: We show the *slope* and RMS values for the single SVX-tagged events.

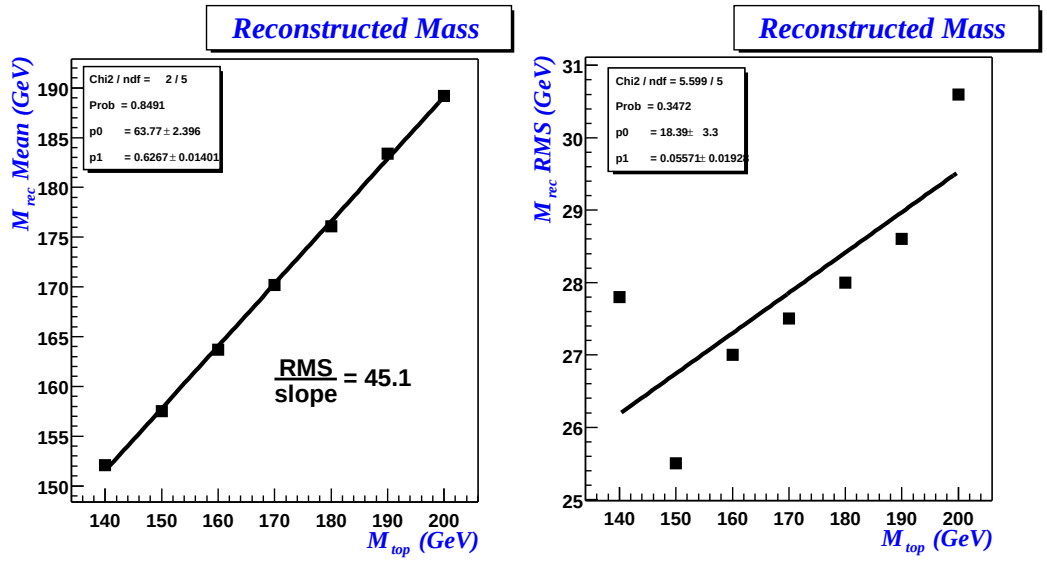

Figure 8.1: (Left) Mean and (Right) RMS, of the reconstructed mass distribution versus the generated top mass.

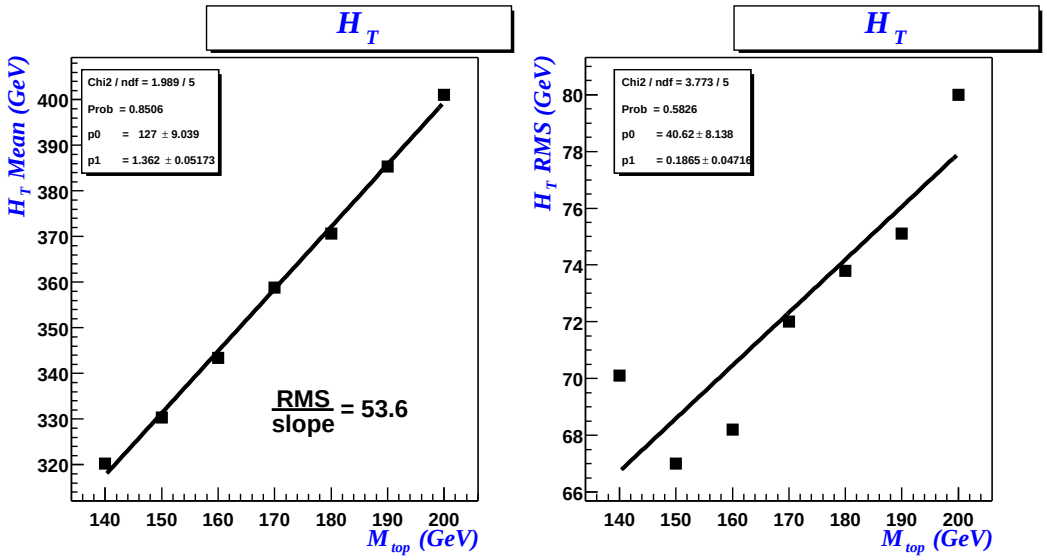

Figure 8.2: (Left) Mean and (Right) RMS, of the H_T distribution versus the generated top mass.

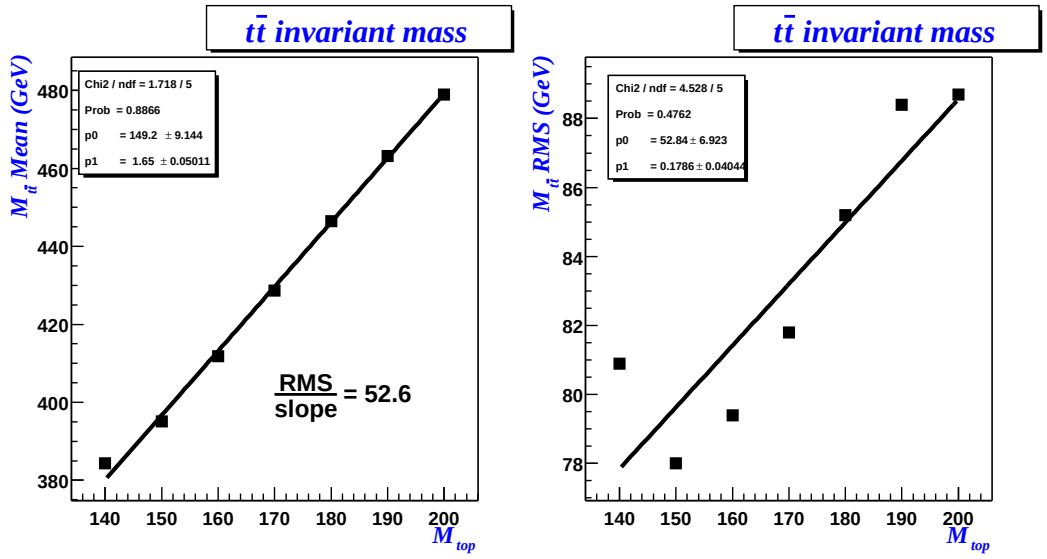

Figure 8.3: (Left) Mean and (Right) RMS, of the invariant $t\bar{t}$ mass distribution versus the generated top mass.

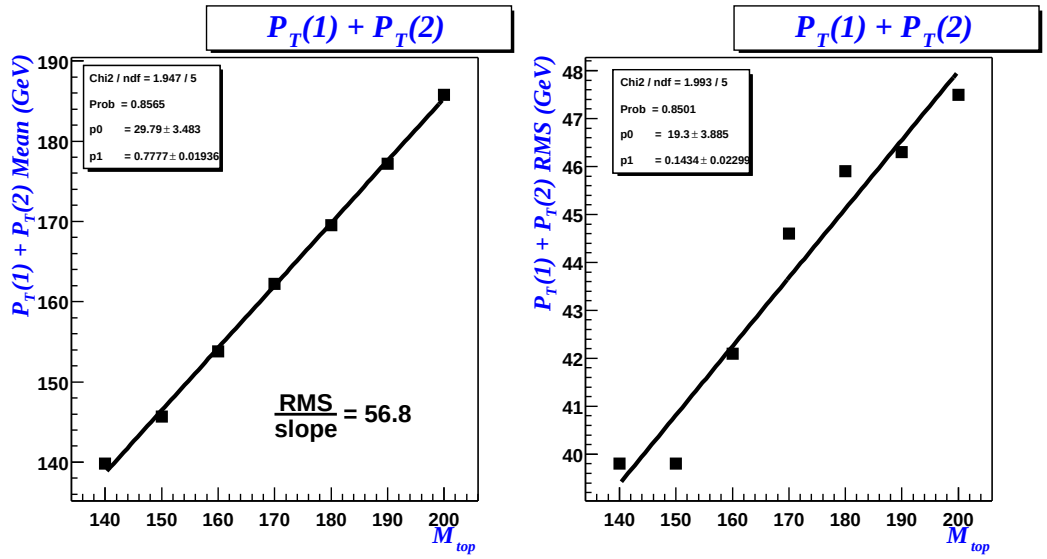

Figure 8.4: (Left) Mean and (Right) RMS, of the $P_T(1) + P_T(2)$ distribution versus the generated top mass.

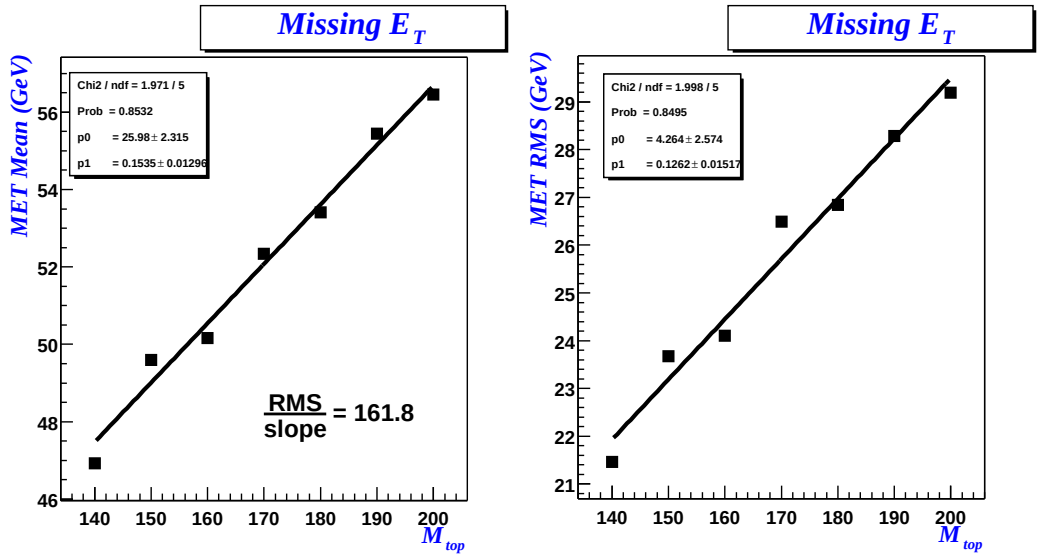

Figure 8.5: (Left) Mean and (Right) RMS, of the missing E_T distribution versus the generated top mass.

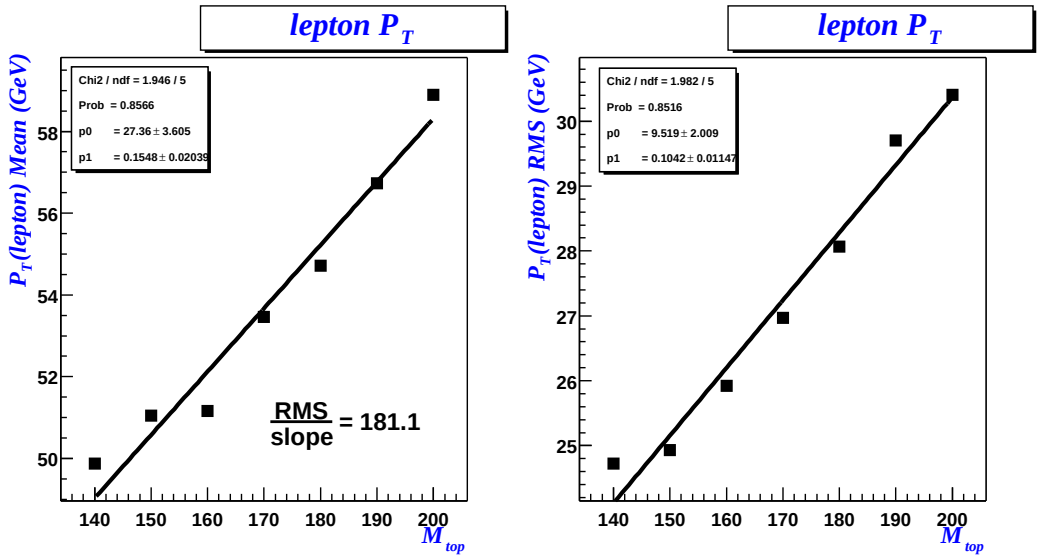

Figure 8.6: (Left) Mean and (Right) RMS, of the lepton P_T distribution versus the generated top mass.

Figures 8.1 through 8.6 show a graphical representation of the results presented in Table 8.1. As we can see, the change in the mean of the distributions is linear for all the kinematic quantities. However, this shift is minimal for variables with poor mass information, Figures 8.5.(L) and 8.6.(L) as can be seen by the value of the slope (fit parameter $p1$). Another interesting aspect is that variables such as the $\mathrm{H_T}$ or $M_{t\bar{t}}$ have a slope much larger than the slope for M_{rec} . However, Figures 8.2.(R) and 8.3.(R) show that these distributions are much broader than the M_{rec} distributions. Thus, when we construct the $\frac{RMS}{slope}$ we see that the M_{rec} variable has the most mass information. The distributions shown correspond to events with two SVX tagged jets.

Not all the variables are used for the different subsamples. The decision to include or exclude an input variable in the Neural Network is based upon the performance of the network. A description of the networks used in this study is given in Section 8.2.2.

8.2 Extracting M_{top} with a Neural Network

The Neural Network analysis also takes advantage of the different S/B ratios for the different mass subsamples. There are a total of four networks, which classify events into signal and background. For this analysis we have generated $t\bar{t}$ signal Monte Carlo files corresponding to a top mass of: 120, 130, 140, 145, 150, 155, 160, 165, 167.5, 170, 172.5, 175, 177.5, 180, 182.5, 185, 190, 195, 200, 205, 210, 220, and 230 GeV/c^2. Therefore, our networks have a total of 24 output units, 23 of them are associated with signal, and the last one is background. For this study we have used the MLPfit Neural Network package. All the networks were trained using the BFGS method [48].

8.2.1 Neural Network Architecture

In the previous section we found that there are four variables that show good mass information. Thus, the networks used in this analysis have a maximum of four input nodes, and 24 output nodes (23 associated with $t\bar{t}$ and one with background). To determine the final architecture of the Neural Networks we try multiple architectures and the ones that produce the best Neural Network performance are chosen. This performance is defined in the next section. The different architectures for each subsample are shown in Table 8.2. We note that for the SLT and the No Tags subsamples not all the input variables were used. This decision was solely based upon the performace studies. An explanation of how the different networks were created is given next.

Top Subsample	NN architecture	Input Variables	Training Evts.	Testing Evts.
SVX Single	4-60-24	M_{rec}, H_T, $M_{t\bar{t}}$, and $P_T(1) + P_T(2)$	38400	7200
SVX Double	4-50-24	M_{rec}, H_T, $M_{t\bar{t}}$, and $P_T(1) + P_T(2)$	24000	7200
SLT	2-50-24	M_{rec}, and H_T	19200	6958
No Tags	2-50-24	M_{rec}, and H_T	50400	7200

Table 8.2: Neural Network architecture. The choice of variables depends solely on the network's performance.

8.2.2 Training and Testing

Creating a Neural Network to perform pattern classification requires *training* and *testing* of the network at each epoch. During training and testing, all 23 different $t\bar{t}$ signal samples as well as background are used.

During training, we set the desired output target value of each class to 1. For instance, in the 24-dimensional output space $M_{top} = 175$ GeV/c^2 corresponds to (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0) and the background target value is given by (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1). The advantages of having a different output for each class is explained in [50]. With this convention, the output values can be interpreted as a-posteriori Bayes probabilities, and therefore they sum to one. This condition can also be used to determine whether the network was properly trained or not. The sum of the outputs for one of our networks is given in 8.7.(L).

Once the target values for each class are set, the network is presented a training sample that is used to modify the weights and biases so that the classification task can be performed. Once all patterns in the training sample have been presented the weights and biases are frozen. The training performace of the network is evaluated by

$$E^p = \frac{1}{2}\sum_{o=1}^{N_o}(d_o^p - y_o^p)^2. \tag{8.1}$$

This is the same equation shown in B.3 where the sum is over all the outputs of the network, d_o^p is the desired output for a given pattern p, and y_o^p is the network's output for that same pattern p. The total error is obtained by summing over all the learning patterns, $E = \sum_p E_p$. A small error E corresponds to a high performance and vice versa. As the number of epochs increases, the total error decreases because the network is able to learn the characteristics of the different classes. However, the continuing improvement in the network's performace is in part due to the fact that the network is learning the characteristics of the training samples. The testing stage is used to determine when to stop training the Neural Network.

During testing, the network obtained from the training session is presented with a testing sample. These testing patterns, which are different than those in the training sample, are used to calculate the testing performance of the network at each epoch. This is done in the same fashion as explained above. As the number of epochs increases, the testing performance decreases. However, as the network begins to learn the characteristics of the training sample it loses its generality. At this point, the performance obtained from the testing samples begins to increase. Figure 8.7.(R) shows the training performance, blue curve, and the testing performace, red curve. The inset shows the epoch at which the network begins to learn the characteristics of the training sample. At this point the weights and biases are frozen and this is the network used to extract the mass of the top quark.

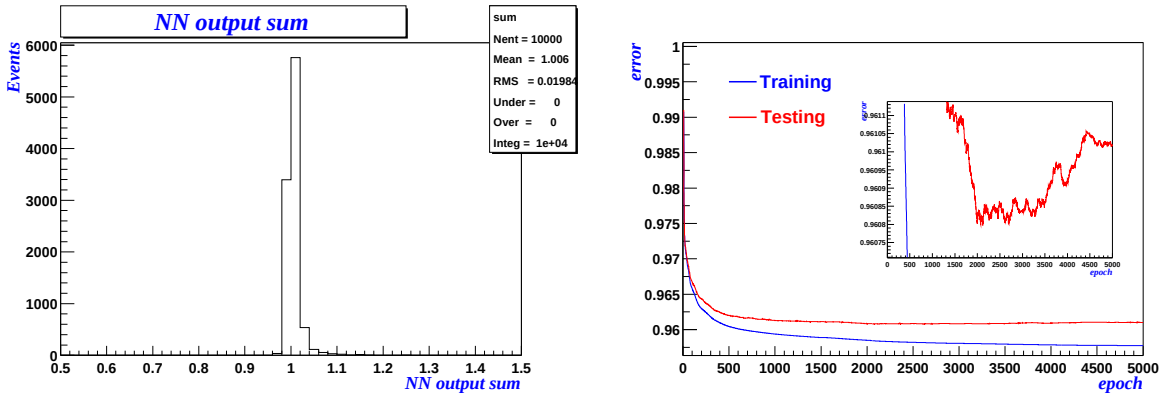

Figure 8.7: (Left) Sum over the output values of the Neural Network. (Right) Shows the training and testing error. The inset shows a close view of the testing error curve. When this error begins to increase the training is stopped.

8.2.3 Likelihood Procedure

For each event, the NN provide a set of probabilities that it has come from any of the top masses used in this analysis, as well as background. Recall that each of

the probabilities is associated with one of the 24 outputs in our network. From this information we extract M_{top} using a discrete maximum likelihood technique. As in the template based method we have two distinct parameters in the fit: the background fraction, x_b, which is allowed to vary between its uncertainties, and the top mass. The joint probability is constructed by multiplying together the four individual likelihood functions. These likelihoods are written as:

$$\mathcal{L}^{bin} = \mathcal{L}^{bin}_{shape} \times \mathcal{L}_{background} \tag{8.2}$$

$bin = 1, \ldots, 23$ corresponds to M_{top} = 120, ..., 230 GeV/c^2 and:

$$\mathcal{L}^{bin}_{shape} = \sum_{i=1}^{Nevents} [(1 - x_b)\mathcal{O}^{bin}_i + x_b\mathcal{O}^{24}_i], \tag{8.3}$$

$$\mathcal{L}_{background} = P(x_b). \tag{8.4}$$

The Neural Network provides the probability that a given event, i, has come from any of the generated top masses, $\mathcal{O}^{bin}_i$, as well as the probability that it has come from background, $\mathcal{O}^{24}_i$. Therefore, each of the bins in the likelihood histogram contains an admixture of signal and background. The $\mathcal{L}_{background}$ term uses a Gaussian to constraint the background fractions to be within the values shown in Table 7.2.

To calculate the top mass, we first minimize the -log $\mathcal{L}$ with respect to x_b. For each mass subsample, we use MINUIT to modify the values of x_b between their limits. We chose the values of x_b that minimize the log likelihood distribution. Once the values for x_b are obtained, the top mass and its uncertainties are calculated by fitting the -log $\mathcal{L}$ with 3rd degree polynomial. This is done to account for asymmetric errors. The polynomial is centered around the minimum of the -log $\mathcal{L}$ distribution, and its limits are placed at the points for which the -log $\mathcal{L}$ increases by at least five units. The errors in M_{top} are given by the 0.5 unit increase in the -log $\mathcal{L}$. Figure 8.8 shows

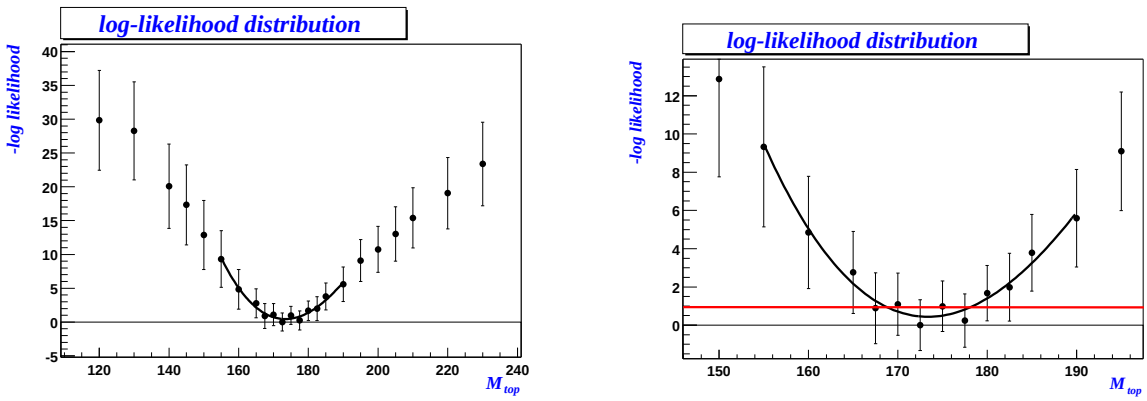

Figure 8.8: (Left) The -log $\mathcal{L}$ distribution for a given Run I standalone NN pseudoexperiment. (Right) Close-up of the left pseudoexperiment. The black curve is the 3rd degree polynomial used to calculate the mass and uncertainties. The red line shows the point at which -log $\mathcal{L}$ changes by 0.5 units.

the results from a pseudoexperiment. As a first approximation, the error in the value of the -log $\mathcal{L}$ for each top mass was set to the RMS of the -log $\mathcal{L}$ distribution for that given mass. These errors probably overestimate the actual errors in the -log likelihood.

8.2.4 Results from Pseudoexperiments

We ran 2000 pseudoexperiments with an admixture of signal and background as shown in Table 7.2. The top mass distribution is shown in Figure 8.9.(L). The width of the distribution is smaller than the widths of the top mass distributions shown in Figures 7.12.(L) and 7.13.(L). The median statistical error for this method is 5.6 GeV/c^2, which corresponds to a 30% statistical improvement over the template based method that uses the Gaussian plus Gamma parametrization technique. The statistical improvement with respect to the NN parametrization technique is roughly 16% (when using the M_{rec} variable). These improvements correspond to an increase

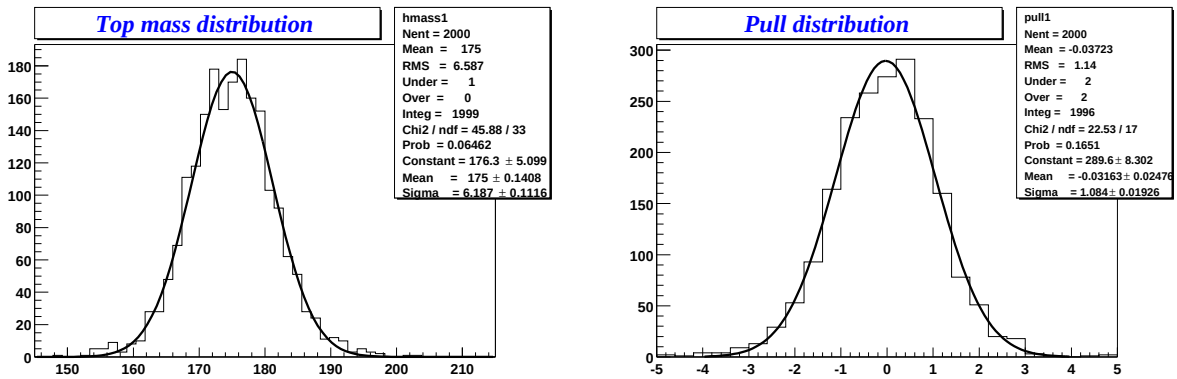

Figure 8.9: Results for 2000 pseudoexperiments using the standalone NN method. (Left) Fitted top mass distribution. (Right) Pull distribution. Signal and background fractions as observed in the Run I data.

in the effective luminosity of over 70% for the GG template based analysis, and over 34% for the NN template based analysis. To make sure the errors are calculated correctly, we construct the pull as

$$g = \frac{(M_{fit} - M_{true})}{\sigma}, \tag{8.5}$$

where M_{fit} is the mass returned by the standalone NN method, M_{true} is the input top mass, and σ are the errors calculated as described above. The pull distribution is shown in Figure 8.9.(R).

CHAPTER 9

SYSTEMATIC UNCERTAINTIES

The systematic errors in the top mass calculation are due to any biases that may have occurred in the measurement procedure as well as uncertainties in the simulation used to model the underlying physics. This chapter describes how the different systematic uncertainties are determined. A summary of the systematic errors associated with each analysis is given in the last section.

9.1 Jet energy scale

The event reconstruction algorithm uses the measured energy of the lepton, jets, and the unclustered energy to estimate the mass of the top quark. The electron energy scale and muon momentum scale are known with a precision better than 0.2% [27]. From studies, it is observed that large variations in the unclustered energy have very little effect on the top mass. The $\not{E}_T$ is calculated from the lepton, jets, and the unclustered energy so it is not an independent measurement. Therefore, the energy scale uncertainty is dominated by the uncertainty in the jet momenta.

Section 6.1 describes the different corrections applied to the jets to better estimate the original parton momenta. However, potential systematic uncertainties arise from the difference in the jet E_T scales between the Monte Carlo and the data. This

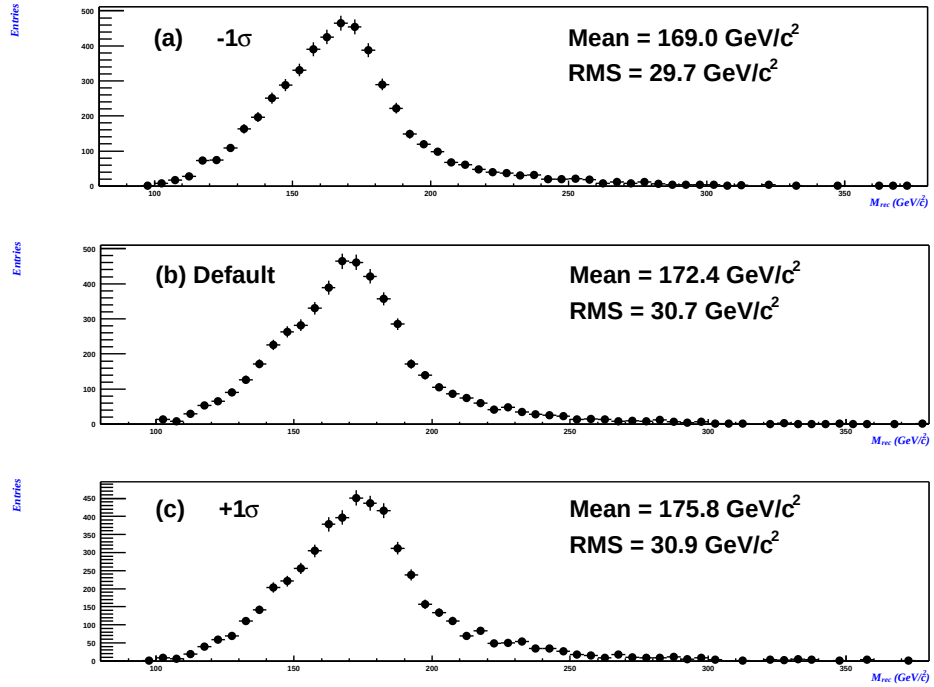

Figure 9.1: M_{rec} distribution for Single SVX tagged events from the PYTHIA $t\bar{t}$ Monte Carlo simulation using an input top mass of 175 GeV/c^2 for events with (a) a -1σ shift in the jet momenta, (b) no shift in the jet momenta, and (c) a $+1\sigma$ shift in the jet momenta.

uncertainty is due to two main sources: detector effects and soft gluon effects [55]. Figure 9.1 shows the M_{rec} distribution for (a) Monte Carlo events in which we have applied a -1σ shift to the jet momenta, (b) default Monte Carlo events, and (c) Monte Carlo events in which we have applied a $+1\sigma$ shift in the jet momenta.

To obtain the jet $\mathrm{E_T}$ uncertainty we apply a $+1\sigma$ and -1σ shift to the jet momenta in $t\bar{t}$ signal and background, and measure the effect on the top mass estimate. We ran two set of simulated experiments. In the first set, we draw events randomly from the sample in which the jet momenta is shifted by $+1\sigma$. In the second set we choose events from the sample in which the jet momenta is shifted by -1σ. Both sets of simulated

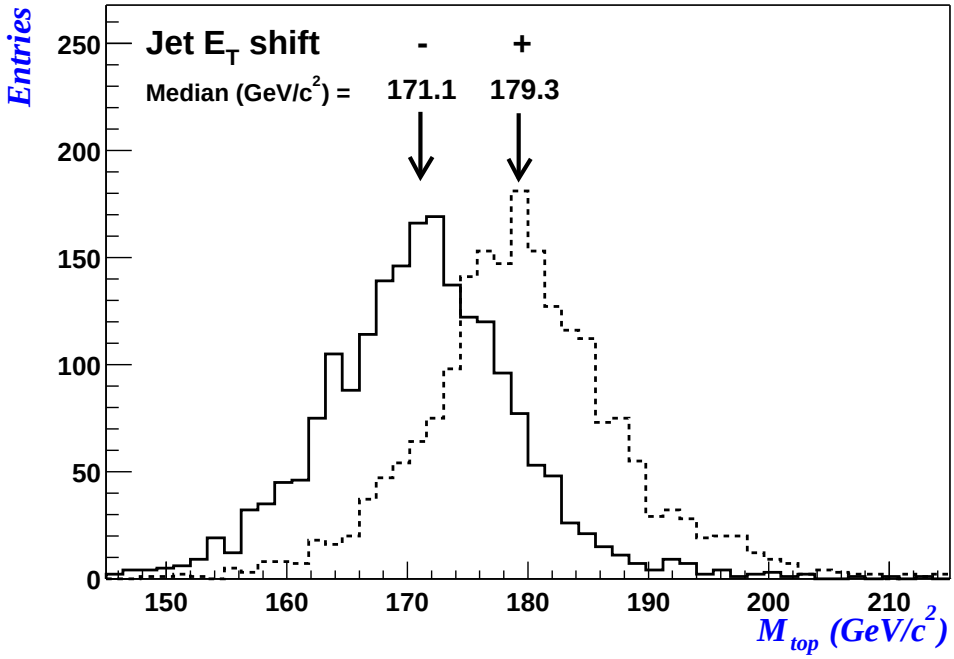

Figure 9.2: M_{top} distribution returned from the likelihood procedure from simulated experiments using the NN-fitted M_{rec} templates. The solid histogram shows the distribution when the jet E_T scale is shifted down by -1σ, and the dashed histogram shows the distribution when the jet E_T scale is shifted up $+1\sigma$. The systematic uncertainty is half the difference between the medians of the distributions.

experiments are fit using the standard templates and likelihood technique described in Chapter 7[9]. The error associated with the energy scale is half the difference between the medians of the -1σ and $+1\sigma$ pseudoexperiment results. Figure 9.2 shows the difference between the two distributions.

9.2 Initial and final state radiation

Radiated gluons play a very important roles in the measurement of the top quark mass. This is not only due to the fact that most $t\bar{t}$ events are accompanied by

[9]For the standalone NN method we use the same network and likelihood procedure described in Section 8.2.

additional gluon jets, but also to the fact that gluons are radiated during the production of the $t\bar{t}$ pair and the decay processes. This complicates the top momentum reconstruction as well as it makes it very difficult to correctly identify the top decay partons.

The uncertainty associated with initial state radiation (ISR)[10] is determined by comparing the median mass obtained using the standard Monte Carlo PYTHIA sample to the median mass obtained using a sample for which the ISR has been turned off. After a large number of simulated experiments, the uncertainty is given by half the difference in the median mass between the standard PYTHIA and the no-ISR PYTHIA samples.

Obtaining the error due to the final state radiation (FSR) is a little bit more complicated since PYTHIA describes the formation of a jet through a parton shower. From the no-ISR PYTHIA sample we select events that have exactly four jets which meet the selection criteria described in Section 5.4, and are uniquely matched to the $t\bar{t}$ decay partons. Using the procedure of simulated experiments, the uncertainty is given by half the difference in the median mass between the no-ISR PYTHIA sample with no restriction on the number of jets, and the no-ISR PYTHIA sample with only four jets that are uniquely matched to the top decay partons. The total systematic uncertainty is obtained by adding the ISR uncertainty and the FSR uncertainty in quadrature.

The ISR/FSR uncertainty is the second largest systematic error associated with the mass analysis. In Chapter 10 we explore the possibility of using a Neural Network to measure the gluon content in $t\bar{t}$ events.

[10]ISR is the radiation that occurs before the $t\bar{t}$ pair has been produced. FSR is the radiation which occurs after the top quarks has been produced.

9.3 b-Tagging

There is a systematic error associated with the uncertainty in the SVX and SLT tagging efficiencies. The uncertainty in the SVX tagging efficiency comes primarily from the possible E_T dependence. The SVX tagging efficiency is determined from Monte Carlo and then corrected by a scale factor. The systematic error is calculated by varying this scale factor. This results in a mass shift of only 0.1 GeV/c^2.

The SLT error is determined by using large data samples of $\psi \to \mu\mu$ and $\gamma \to ee$. The relevant uncertainty for SLT tags arises in the ratio of true to fake tags in $t\bar{t}$ events. This ratio has a 10-20% uncertainty. To study the effect of this ratio on the top mass we generate Monte Carlo $t\bar{t}$ events in which all SLT tags are either (a) true tags, or (b) fake tags. We ran a large number of pseudoexperiments in which all the events have come from set (a) or from set (b). These events are then compared to the standard Monte Carlo sample. The error associated with the SLT tagging efficiency is half the difference between the median mass of set (a) and set (b).

9.4 Parton distribution function

All the top Monte Carlo samples were generated using PYTHIA with the default parton distribution function (PDF) GRV 94L. Other PDF choices are available such as CTEQ 3L [57]. We take the shift in the median top mass samples generated with the two PDF's to be the systematic uncertainty.

9.5 Monte Carlo generators

The uncertainty associated with the Monte Carlo generators is obtained by comparing the results from PYTHIA to HERWIG. We run a large number of simulated

experiments, and the uncertainty is given by half the difference between the median mass of the PYTHIA and HERWIG simulations.

9.6 Summary of Systematic Errors

All the systematic uncertainties described in the previous sections are given in the Table 9.1. We provide the systematic errors for the NN-fitted M_{rec} templates, the NN-fitted H_T templates, and the standalone Neural Network analysis. The total systematic uncertainty is obtained by adding the individual errors in quadrature.

	Uncertainty (GeV/c^2)		
Source	M_{rec} NN templates	H_T NN templates	Standalone NN
Jet E_T scale	4.1	7.5	4.6
ISR/FSR	1.7 (0.5/1.6)	4.5 (2.1/4.1)	2.1 (0.9/1.9)
b-Tagging	0.3	1.5	0.4
PDF	1.1	2.0	0.9
MC generators	0.4	1.1	0.3
Total	4.6	9.2	5.2

Table 9.1: Summary of the systematic uncertainties associated with each of the different Neural Network analysis methods described in this thesis. In parenthesis we give the (ISR/FSR) individual systematic errors.

CHAPTER 10

MEASUREMENT OF THE GLUON CONTENT

In the top mass analysis, we assume that the four highest E_T jets from an event correspond to the top decay partons. However, our Monte Carlo predicts that in roughly 50% of the events, one or more of these jets cannot be matched within a cone of $\Delta R < 0.4$ to any of the decay partons of the $t\bar{t}$ pair. Jets which are not within a $\Delta R < 0.4$ cone of any of the Monte Carlo partons are considered to be gluon jets.

Gluon jets are produced during the production of the $t\bar{t}$ pair, initial state radiation (ISR), or during its decay, final state radiation (FSR). To obtain a more accurate measurement of the top mass we would only want to include the gluon jets which carry energy from the top quarks. This is an extremely difficult task. However, an improvement of the top mass can be obtained by determining the number of events with at least one gluon jet among the jets used in the top mass reconstruction. This information can be used to construct templates that model the data more closely. We describe a technique to measure the gluon content in the SVX tagged Run I data.

10.1 Gluon Effect on the Top Mass

A number of features of the mass analysis from Run I are consistent with the data having fewer gluon events than the Monte Carlo would predict. Examples include

the width of the observed mass distribution and the number of extra jets. We first address whether a more accurate knowledge of the gluon content in the top mass sample would lead to a better mass measurement.

10.1.1 Gluon Fraction in the Monte Carlo

By matching the simulated jets to the Monte Carlo top decay products we are able to determine which events have all the top decay products present and those which do not. We look at the top decay partons at the Monte Carlo level. Then, we take the four leading jets of an event after it has gone through the detector simulation and we match each of these jets to the partons from the Monte Carlo. If a jet is not within a cone of $\Delta R < 0.4$ from any parton we consider this jet to be a gluon jet. We note that this definition is somewhat arbitrary. The Monte Carlo samples are divided into two different sets: *gluon templates*, and *non-gluon templates*. Gluon templates contain only events with at least one gluon jet. Non-gluon templates only contain events in which all jets are uniquely matched to the top decay products. In the $t\bar{t}$ Monte Carlo we find:

- $52.7\% \pm 0.7\%$ of the single SVX-tagged events contain at least one gluon jet as one of their four highest E_T jets.
- $48.7\% \pm 1.1\%$ of the events in the double SVX-tagged sample contain at least a gluon jet as one of their four highest E_T jets.

Figures 10.1 and 10.2 show some examples for gluon and non-gluon templates. Gluon templates tend to peak at a lower mass than the non-gluon templates. Also, non-gluon templates show better separation between M_{top} masses and they are narrower than the gluon templates. We construct the $\frac{RMS}{slope}$ quantity to assign a figure of merit

to both set of templates (recall from Section 8.1 that this quantity is an indication of how much mass information a given quantity exhibits):

$$\text{Gluon Templates}: \frac{\text{RMS}}{\text{slope}} = 81.6 \text{ (Single SVX)}, \quad \frac{\text{RMS}}{\text{slope}} = 69.3 \text{ (Double SVX)},$$
$$\text{Non}-\text{gluon Templates}: \frac{\text{RMS}}{\text{slope}} = 43.3 \text{ (Single SVX)}, \quad \frac{\text{RMS}}{\text{slope}} = 33.9 \text{ (Double SVX)}.$$

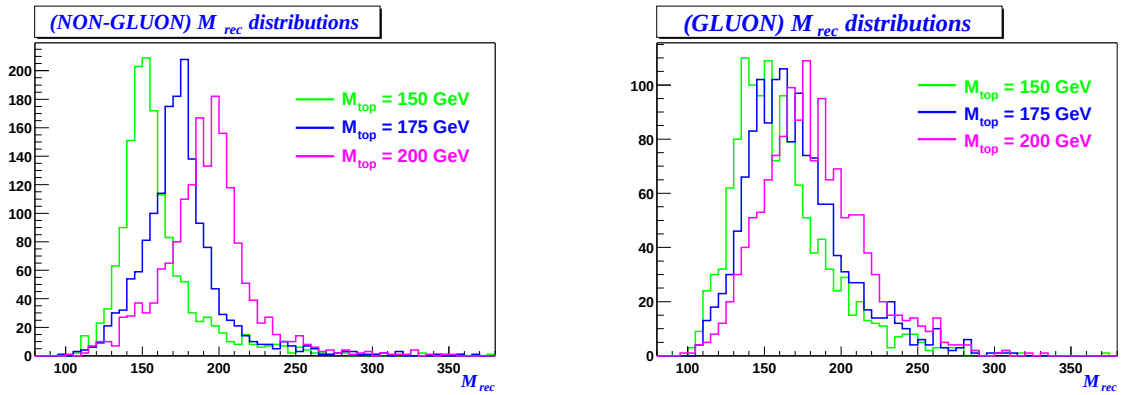

Figure 10.1: Reconstructed top mass distributions for the single SVX Monte Carlo sample for (Left) non-gluon events and (Right) gluon events.

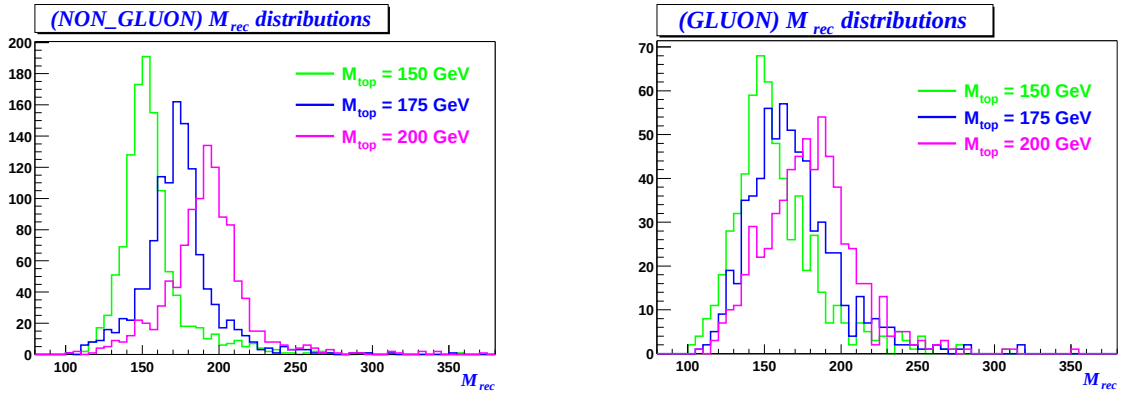

Figure 10.2: Reconstructed top mass distributions for the double SVX Monte Carlo sample for (Left) non-gluon events and (Right) gluon events.

10.1.2 Pseudoexperiments

The goal is to study what happens when we compare the data to the "wrong" set of templates, which do not have the same fraction of gluon events as the pseudoexperiment data. If the data does not have the same mix as the MC, this is what will result in the data measurement.

In the reconstruction of the top mass, we use a set of top mass templates ranging from M_{top} = 150 GeVc^2 to 200 GeVc^2. For each simulated experiment we perform a discrete likelihood fit [49] to determine which top mass is the most likely to have produced a M_{rec} distribution similar to what is seen in the pseudoexperiment data. To keep the study simple, we do not parametrize the templates and background is not used. The top mass and its uncertainty is calculated by integrating the -log $\mathcal{L}$.

We generate a set of templates for which the percentage of gluon events is varied. Thus, we have six sets of templates containing 0%, 20%, 40%, 60%, 80%, and 100% of gluon events. We ran 1000 pseudoexperiments where we draw the events from each of these templates and we compare them to the default Monte Carlo. Recall that the default Monte Carlo templates contain roughly 50% of gluon events. Tables 10.1 and 10.2 show the results from this study for the Single SVX and Double SVX subsamples respectively. These results can be summarized as:

- Increasing the gluon content in the templates results in a decrease of the median top mass returned by our fit, and an increase in the median uncertainty. Gluon events are mismeasured low and they widen the template's distribution.

- Comparing data events to the wrong templates results in a mismeasurement of the top mass that can be significant.

Gluon Content	Median M_{rec}	Median error	Pull mean	Pull width
0%	177.1	8.3	0.3	0.84
20%	176.0	8.5	0.2	0.91
40%	174.7	8.7	-0.02	0.93
60%	173.6	8.9	-0.2	1.0
80%	172.2	9.1	-0.3	1.1
100%	169.3	9.2	-0.6	1.3

Table 10.1: We show the results from the 1000 pseudoexperiments for the Single SVX subsample. In each of the pseudoexperiments, we draw 15 events from the templates, which we have constructed with the appropriate amount of gluon events, and we compare them to the default Monte Carlo templates. Background is not included.

Gluon Content	Median M_{rec}	Median error	Pull mean	Pull width
0%	176.5	9.7	0.2	0.85
20%	176.2	9.9	0.2	0.91
40%	175.2	10.0	0.04	0.93
60%	173.9	10.2	-0.09	0.95
80%	172.9	10.4	-0.2	1.1
100%	170.3	10.6	-0.4	1.3

Table 10.2: We show the results from the 1000 pseudoexperiments for the Double SVX subsample. In each of the pseudoexperiments, we draw 5 events from the templates, which we have constructed with the appropriate amount of gluon events, and we compare them to the default Monte Carlo templates. Background is not included.

- Comparing the data events to the wrong templates, results in a top mass error that is not accurate. When the gluon content in the templates is too low the errors are overestimated. On the other hand, if the gluon content is too high the errors are underestimated.

When we compare the data to the wrong templates, the top mass is biased and the statistical uncertainty reported is not correct. It is very important that we construct

templates that closely mimic the data. An accurate measurement of the gluon content in our top mass sample will lead to a better measurement of the top quark mass.

10.2 Measuring the Gluon Content

In the previous section, we showed that the measurement of the gluon content would lead to a better estimate of the top mass and the reported uncertainty. We have developed a NN technique that attempts to measure the amount of gluon events in the data. In order to reduce the amount of background in our sample, we are only using events which pass all the mass selection cuts with at least one SVX-tagged jet.

10.2.1 Neural Network Input Variables

We want to develop a NN that can distinguish between gluon events and non-gluon events. Three variables, shown in Figures 10.3 and 10.4, with good discriminating power were selected for this analysis:

- The di-jet invariant mass is constructed with the untagged jets in the event. In the Single SVX case there are three possible ways to form the di-jet invariant mass. Only two of them are used since the di-jet mass constructed with the two least energetic jets differs very little from gluon events to non-gluon events. Each di-jet mass is used as an input for the Single SVX network. For double SVX-tagged events there is only one way to form the di-jet mass, which is using the two untagged jets in the event. In this subsample, the di-jet distribution is clearly narrower for non-gluon events than for gluon events. Also, the di-jet mass distribution peaks around 81 GeV/c^2, the W mass, when there are no gluon jets in the event.

- The number of extra jets with $E_T > 8.0$ GeV and $|\eta| < 2.4$ besides the four jets used in the mass reconstruction. Gluon events have more jet activity than those without gluons.

- The χ^2 returned by the mass reconstruction code (see Section 6.3). Events that contain all the top decay partons tend to have a lower χ^2 than those with gluon jets.

All the above variables have good information regarding whether or not gluon jets are present in the events. A Neural Network is well suited for this measurement since it provides us with a simple way of combining all of the different variables. Including more information will decrease the error in the gluon content measurement.

10.2.2 Neural Network Training and Testing

For this analysis, we are using the JETNET [44] subroutines interfaced to ROOT via the Root_Jetnet [45] package. We use two different Neural Networks: one for events with a single SVX-tagged jet, and another for events with two SVX-tagged jets. They are both three-layered feed-forward perceptrons with the standard activation function (tanh scaled to (0, 1)). The Neural Network architectures are the following:

- **NN_1SVX** $\rightarrow$ 4-8-1. The single tagged Neural Network has four different input variables since two di-jet masses are used separately.

- **NN_2SVX** $\rightarrow$ 3-6-1.

We trained the Neural Networks using two different Monte Carlo samples: one containing gluon events, and the other one without any gluon events. During training, the desired Neural Network output for non-gluon events was set to 1, while the output

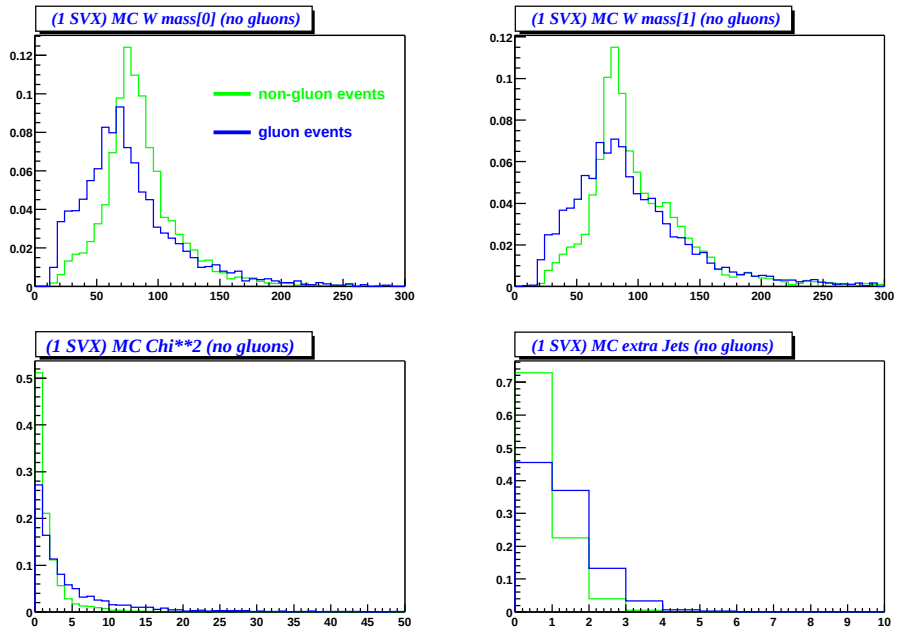

Figure 10.3: The four input variables for the SVX Single Neural Network. (a) Di-jet invariant mass between the two most energetic untagged jets. (b) Di-jet invariant mass between the most and least energetic untagged jets. (c) χ^2 returned by the top mass kinematic fit. (d) Number of extra jets with $E_T > 8.0$ GeV and $|\eta| < 2.4$. Background is not included.

corresponding to the gluon events was set to 0. The updating of the weights was done using the standard back-propagation algorithm with default training parameters. We ran a large number of training loops keeping track of the classification performance of the Neural Network during each of the loops. Recall, this performance is determined by the number of events in which the Neural Network is able to classify each class (gluon events, and non-gluon events) correctly. When the classification is best we freeze the weights. These final weights constitude our optimal Neural Network. Figures 10.5.(L) and 10.5.(R) show the output for both networks for gluon events, non-gluon events and background events, which look like gluon events to the Neural Networks.

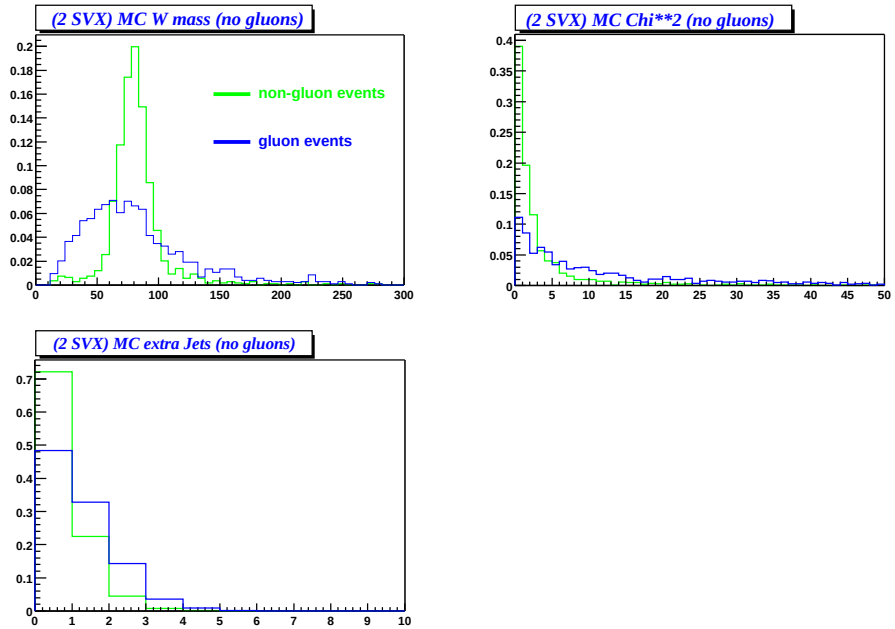

Figure 10.4: The three input variables for the SVX Double Neural Network. (a) Di-jet invariant mass between the two untagged jets. (b) χ^2 returned by the top mass kinematic fit. (c) Number of extra jets with $E_T > 8.0$ GeV and $|\eta| < 2.4$. Background is not included.

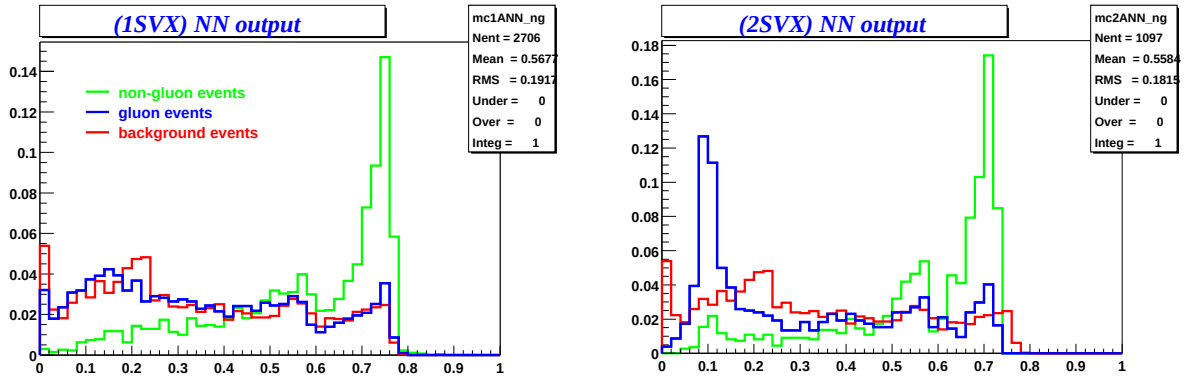

Figure 10.5: (Left) NN output distribution for the Single SVX subsample. (Right) NN output distributions for the Double SVX subsample. The single SVX-tagged background is used for both subsamples.

10.2.3 Definition of the Likelihood

Monte Carlo studies show that the amount of gluon events in the Single SVX and Double SVX subsamples is very similar. We construct individual binned likelihoods for the two subsamples. Since they are statistically independent from one another, we can construct a joint binned likelihood for the entire sample by multiplying together the two individual binned likelihood functions. The gluon fraction, x_g, and the non-gluon fraction, x_{ng}, are obtained from a maximum likelihood fit applied to the Neural Network output distributions. Both x_g and x_{ng} are unconstrained in our fit. The background fractions for each subsample, x_b, are allowed to vary about their calculated value and within their uncertainties (see Table 10.3). The individual binned likelihood has the form:

$$\mathcal{L} = \mathcal{L}_{shape} \times \mathcal{L}_{count} \times \mathcal{L}_{background} \tag{10.1}$$

where:

$$\mathcal{L}_{shape} = \sum_{i=1}^{\text{N bins}} \frac{N_g f_g(\mathcal{O}) + N_{ng} f_{ng}(\mathcal{O}) + N_b f_b(\mathcal{O})}{N_g + N_{ng} + N_b}, \tag{10.2}$$

$$\mathcal{L}_{count} = \frac{e^{-(N_g+N_{ng}+N_b)}(N_g + N_{ng} + N_b)^N}{N!}, \tag{10.3}$$

$$\mathcal{L}_{background} = P(N_b). \tag{10.4}$$

and N_g, N_{ng}, and N_b can be written in term of the fraction of events as:

$$N_g = x_g \cdot N, \quad N_{ng} = x_{ng} \cdot N, \quad N_b = x_b \cdot N. \tag{10.5}$$

The quantities N_g, N_{ng}, and N_b, represent the number of gluon events, non-gluon events, and background respectively, N is the number of observed events, and $P(x_b)$ constraints the backgrounds to be around their central value and within their uncertainties. The variable $f_g(\mathcal{O})$ is the Neural Network output distribution for gluon

events. The NN output distribution for non-gluon events is given by $f_{ng}(\mathcal{O})$, and the background NN output distributions are $f_b(\mathcal{O})$. To calculate x_g, and x_{ng} we minimize the -log $\mathcal{L}$, and the statistical uncertainties in our measurements are given by changing -the log $\mathcal{L}$ by half a unit respect to its minimum.

10.2.4 Pseudoexperiments

In the Run I Single SVX subsample there are a total of 17 events. Two of these events do not pass the χ^2 selection criteria and are not used in the mass analysis. In the Run I Double SVX subsamples there are a total of five events. The χ^2 cut is removed to increase the acceptance . The efficiencies of the χ^2 cut shown in Table 6.1 are used to calculate the new background fractions. Table 10.3 shows the results.

Mass Sample	Data Events Observed	Predicted Background
Single SVX	17 (15)	2.93 ± 0.92 (2.4 ± 0.75)
Double SVX	5 (5)	0.24 ± 0.12 (0.2 ± 0.1)

Table 10.3: We show the expected number of events without the χ^2 cut. In parenthesis we show the expected number of events after the χ^2 cut. These numbers are taken from Table 7.2.

We ran 5000 pseudoexperiments with the expected admixture of signal and background. For each simulated experiment we compare the NN output distribution of the data to a combination of the NN output distribution for gluon events, non-gluon events, and background. Figure 10.6 show the distribution for the gluon fraction and non-gluon fraction for the pseudoexperiments. The mean of these distributions is 43.5% in the gluon case and 41.4% for the non-gluon fraction as expected from the Monte Carlo. The remaining contribution is from the background. We expect the

pull distributions (Equation 7.5) to have a mean of zero and a width of one. However, we find that the pull distributions are too narrow suggesting that the errors returned by the fit are overestimated. This effect is due to the fact that our simulated experiments have very low statistics. Therefore, we scale the errors returned by the fit by a factor of 0.75 for the gluon events, and 0.74 for the non-gluon events [46]. After rescaling the errors the median statistical uncertainty for the gluon and non-gluon content is 18.9% and 17.4% respectively. Figure 10.7 shows the pull distributions.

During Run IIa, we expect to collect 2 fb^{-1} of data or roughly 20 times the amount of statistics collected during Run I. Assuming Run I conditions, we ran 5000 pseudoexperiments in order to estimate the expected uncertainty in the large data sample. The study shows we should be able to measure the gluon content in Run IIa with an accuracy of roughly 5.6% for gluon events, and 4.3% for non-gluon events. The pseudoexperiment results are summarized in Table 10.4.

10.3 Measurement of the Gluon Fraction in the Run I Data

We apply our analysis to the Run I data. The statistical errors returned are scaled by the same factor used in the Monte Carlo study. We fit the Run I data and we find:

- $23.7\% \pm 20.0\%$ of the events in our single plus double SVX-tagged mass sample contain a gluon jet among their 4 highest E_T jets.
- $62.2\% \pm 21.0\%$ of the events are non-gluon events.
- $17.0\% \pm 5.3\%$ of the 17 single tagged events are background events.
- $4.8\% \pm 2.3\%$ of the 5 double tagged events are background events.

Data Sample	Predicted uncertainty in x_g	Predicted uncertainty in x_{ng}
Run I (106 pb^{-1})	18.9%	17.4%
Run IIa (20× Run I)	5.6%	4.3%

Table 10.4: This is a summary of the Run I, and Run IIa pseudoexperiment results.

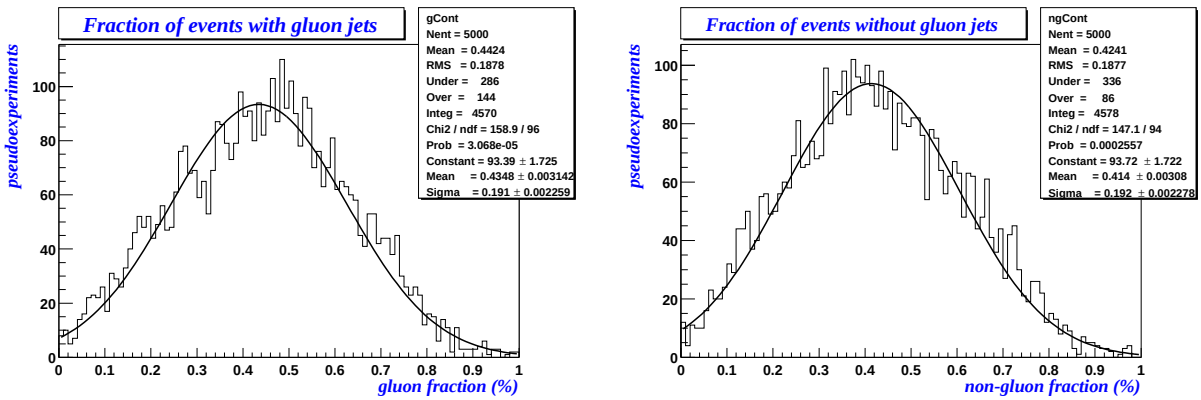

Figure 10.6: Results from 5000 pseudoexperiments for (Left) fraction of gluon events and (Right) fraction of non-gluon events.

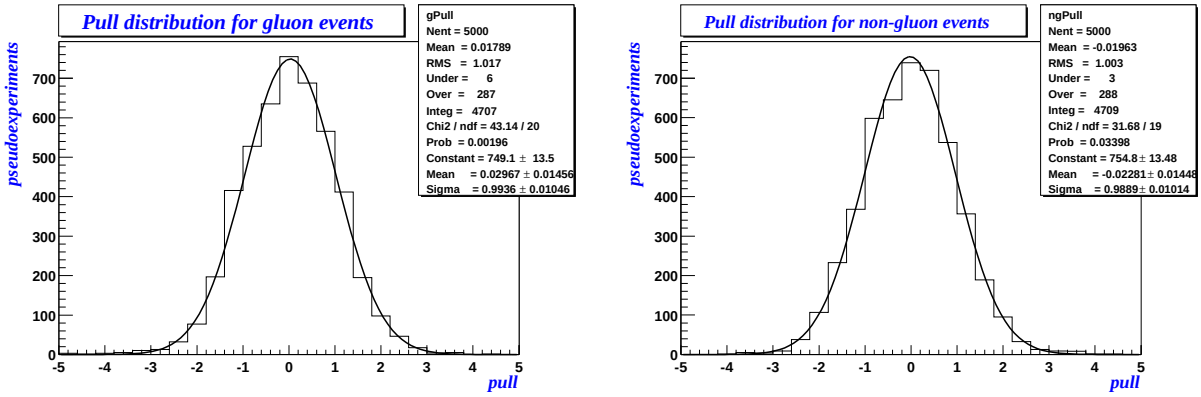

Figure 10.7: Pull distribution from 5000 pseudoexperiments for (Left) fraction of gluon events and (Right) fraction of non-gluon events.

In Figure 10.8.(L), we compare the M_{rec} distribution for the Single SVX Run I data and the Monte Carlo. In the Monte Carlo histograms we show the contributions from the background, yellow, background plus gluon events, blue, and background plus gluon plus non-gluon events, green. Each contribution is set according to our fit results. Figure 10.8.(R) shows the same plot for the Double SVX Run I data. A lower fraction of gluon events produces slightly narrower M_{rec} distributions.

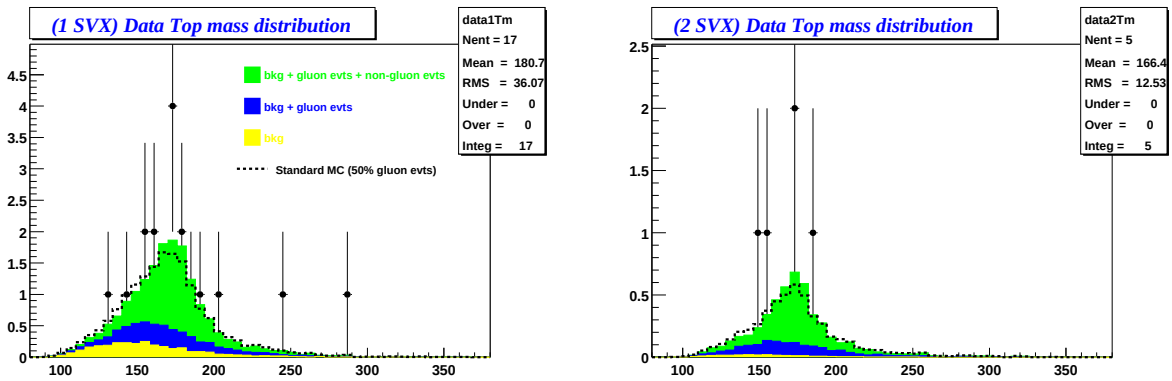

Figure 10.8: Reconstructed mass distribution for (Left) Single SVX events and (Right) Double SVX events. The gluon fraction in the Monte Carlo is set to that measured in the Run I data. The dashed line histogram has the standard MC gluon fraction ($\sim 50\%$). Signal and background fractions as observed in the Run I data.

10.4 Systematic Errors

For this analysis we have looked at two different sources of uncertainty in the gluon content measurement. First, we looked at the energy scale. We apply $+1\sigma$ and -1σ shifts to the jet momenta in $t\bar{t}$ signal and background events, and measure the effect on the gluon content estimate. The error is given by half of the difference between the

medians of the $+1\sigma$ and -1σ distributions. We find that the error associated with the energy scale is less than 0.1% in both the gluon fraction and non-gluon fraction.

In measuring the gluon content we have assumed a top mass of 175 GeV/c^2. We change the mass of the top quark to 170 GeV/c^2 and 180 GeV/c^2 to see whether or not this has any effect on our measurement. The error is given by the half difference between the medians of the 170 GeV/c^2 results and the 180 GeV/c^2 results. In this case we find that the error associated with the top mass is also less than 0.1%.

10.5 Application to the Mass Analysis

The likelihood procedure used to calculate the top mass is described in Section 7.3. We use the four mass subsamples for this analysis: Single SVX, Double SVX, SLT, No Tags. Using the events in these subsamples we construct two sets of templates:

- **Data-like templates:** these are templates in which the amount of gluon events and non-gluon events are set to the mean values returned by the fit in Section 10.3
- **MC-like templates:** these templates contain the default mixture of gluon events and non-gluon events.

We fit these templates using the NN method (Section 7.1.1 to obtain the top mass dependent functional form.

We run pseudoexperiments for which we draw the data events from the Data-like templates and we compared them to the Data-like templates and to the MC-like templates. Each simulated experiment contains an admixture of signal and background as described previously. The results from 2000 pseudoexperiments are summarized in Table 10.5.

Pseudo. samples	Compare to	Result (GeV/c^2)
Data-like samples	Data-like templates	174.5 ± 5.5
MC-like samples	MC-like templates	174.8 ± 6.5
Data-like samples	MC-like templates	175.8 ± 6.1

Table 10.5: Median top mass and median statistical MINOS error results for the 2000 pseudoexperiments using Run I statistics.

We measure the gluon content in the Run I data sample to be lower than expected from the default MC. We have shown that samples with lower gluon content have smaller top mass errors. The above MC studies imply that the top mass statistical error is about 18% smaller for samples with Data-like gluon content than for samples with the default Monte Carlo gluon content. Also, we have shown that the top mass measurement is biased and its reported uncertainty is inaccurate when the gluon content of the pseudoexporiment sample and the templates' gluon composition differ. Monte Carlo studies imply that the top mass is biased by + 1.3 GeV/c^2 and the error increases by about (6.1/5.5) = 11% when samples with Data-like gluon content are compared to templates constructed from the default Monte Carlo.

CHAPTER 11

RUN I RESULTS

In this chapter, we apply the different mass analysis techniques to the Run I data sample. First, we summarize the Monte Carlo results from the previous chapters. The Run I data results are given next. After presenting these results, the consistency between the data and the Monte Carlo expectations for $t\bar{t}$ plus background events are checked for the different analyses methods.

11.1 Summary of Monte Carlo Results

Table 11.1 presents the different Monte Carlo results. We include the results obtained from using the templates constructed with the gluon content measured in the Run I data. However, it should be noted that the error in the gluon content is rather large ($\sim 20\%$).

11.2 Template-Based Data Results

The template based likelihood procedure is applied to the events in the four mass subsamples. To investigate the effect of the background constraining term the mass fit is performed two different ways: we constrain the backgrounds to be within their expected values, and the background constraint is removed. The top mass measurements

Fitting Method	Median error (GeV/c^2)
NN M_{rec} templates	6.5
GG M_{rec} templates	7.3
NN H_T templates	11.3
Standalone NN	5.6
NN M_{rec} templates (gluon)	5.4
GG M_{rec} templates (gluon)	6.0
NN H_T templates (gluon)	10.6

Table 11.1: Summary of the Monte Carlo results. The (gluon) description means that the templates were constructed with the gluon content measured in the data.

for each subsample and the combined result when using the M_{rec} templates fitted with a combination of a Gaussian plus a Gamma function, and the M_{rec} templates fitted with a Neural Network, are presented in Table 11.2 and 11.3 respectively. Figure 11.1 shows the results from the background constrained combined fit to all four subsamples using the NN-fitted M_{rec} templates. The results when using the H_T templates are given in Table 11.4 and shown in Figure 11.2.

Several observations can be made when comparing the mass fit results when the background fractions are constrained to those when they are not. The most important result is that when all the subsamples are used, independently of the templates or fitting method used, the top mass measurement shows very little sensitivity to the background constraint. For the M_{rec} templates, the unconstrained mass fits of the tagged subsamples yield a background fraction of zero, although with large uncertainties. Also, the fit on the No Tag subsample yields a background fraction lower than expected. For the H_T templates, the unconstrained fit of the SVX tagged subsamples yield a background fraction of zero with large uncertainties. The unconstrained fit of the SLT sample yields a background fraction larger than the one expected. The fit on

the No Tag subsample yields a similar background fraction whether the background is constrained or unconstrained. In general, we would expect an increase in the statistical error if the background constrained is removed since not all the information is being used. However, most of the measurements seem to indicate the opposite. This is due to the fact that in most of the measurements the unconstrained background fraction fitted to a smaller value thus increasing the number of signal events.

When we fit the M_{rec} templates with the Neural Network method we obtain a statistical error which is 4.75 GeV/c^2. This corresponds to a (5.8/4.75) = 22% statistical improvement over the result obtained when we fit the templates with a combination of a Gaussian plus Gamma functions. Using the $\mathrm{H_T}$ templates fitted with the Neural Network method yields a statistical error of 9.25 GeV/c^2. That is a (12.0/9.25) = 30% statistical improvement over the previous top mass measurement that used the $\mathrm{H_T}$ [51].

Data sample	N_{obs}	x_b^{fit}		Top mass (GeV/c^2)	
		Constrained	Unconstrained	Constrained	Unconstrained
SVX Double	5	0.04 ± 0.02	$0.0^{+0.3}_{-0.0}$	$169.8^{+11.4}_{-10.3}$	$169.6^{+11.1}_{-10.3}$
SVX Single	15	0.14 ± 0.05	$0.0^{+0.09}_{-0.0}$	$178.4^{+9.5}_{-8.6}$	$177.9^{+8.8}_{-8.0}$
SLT	14	$0.53^{+0.09}_{-0.08}$	$0.0^{+0.8}_{-0.0}$	$149.3^{+37.1}_{-21.7}$	$152.2^{+27.3}_{-20.5}$
No Tags	42	0.65 ± 0.09	$0.41^{+0.18}_{-0.17}$	$184.8^{+14.1}_{-10.8}$	$181.7^{+10.9}_{-8.9}$
All Subsamples	76	0.46 ± 0.08	0.23 ± 0.19	$177.3^{+6.0}_{-5.7}$	$177.0^{+5.6}_{-5.2}$

Table 11.2: Results of applying the mass likelihood procedure using the GG parametrized M_{rec} templates to the four subsamples and for all subsamples combined. We show the results when the background fractions are constrained to their expected values (Table 7.2), and when they are not. The combined background fraction is the average of the x_b fit results weighted by the number of events in the subsamples.

Data sample	N_{obs}	x_b^{fit}		Top mass (GeV/c^2)	
		Constrained	Unconstrained	Constrained	Unconstrained
SVX Double	5	0.04 ± 0.02	$0.0^{+0.4}_{-0.0}$	$172.6^{+9.2}_{-9.1}$	172.3 ± 9.0
SVX Single	15	0.15 ± 0.05	$0.0^{+0.1}_{-0.0}$	$178.1^{+8.7}_{-7.5}$	$177.5^{+7.9}_{-7.0}$
SLT	14	$0.54^{+0.09}_{-0.1}$	$0.0^{+0.7}_{-0.0}$	$168.4^{+20.7}_{-42.6}$	$145.9^{+29.5}_{-13.2}$
No Tags	42	0.63 ± 0.09	$0.37^{+0.18}_{-0.16}$	$183.4^{+9.5}_{-8.1}$	$181.3^{+7.4}_{-6.9}$
All Subsamples	76	0.48 ± 0.08	0.2 ± 0.18	$177.9^{+4.8}_{-4.7}$	177.5 ± 4.4

Table 11.3: Results of applying the mass likelihood procedure using the NN parametrized M_{rec} templates to the four subsamples and for all subsamples combined. We show the results when the background fractions are constrained to their expected values (Table 7.2), and when they are not. The combined background fraction is the average of the x_b fit results weighted by the number of events in the subsamples.

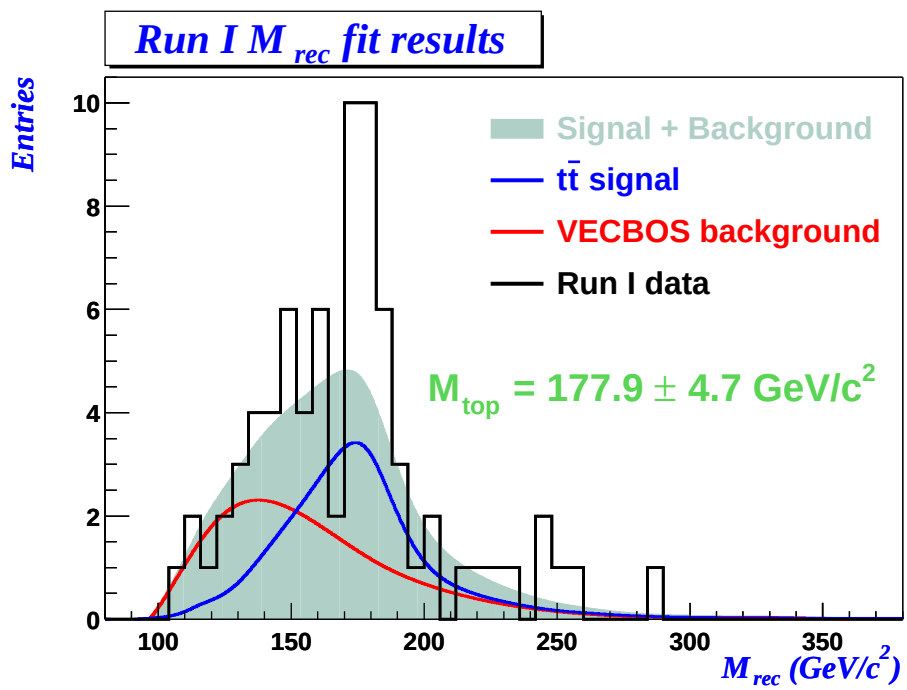

Figure 11.1: Results from applying the likelihood procedure to the combined four subsamples using the NN-fitted M_{rec} templates. The figure shows the Run I data (histogram), fitted background and signal (shaded region), fitted signal (blue curve), and fitted background (red curve).

Data sample	N_{obs}	x_b^{fit}		Top mass (GeV/c^2)	
		Constrained	Unconstrained	Constrained	Unconstrained
SVX Double	5	0.04 ± 0.02	$0.0^{+0.1}_{-0.0}$	$220.4^{+21.8}_{-19.4}$	$220.1^{+21.4}_{-19.1}$
SVX Single	15	0.14 ± 0.05	$0.0^{+0.07}_{-0.0}$	$196.5^{+14.1}_{-13.7}$	$194.4^{+12.9}_{-12.6}$
SLT	14	$0.56^{+0.09}_{-0.08}$	$0.77^{+0.2}_{-0.2}$	$199.4^{+38.4}_{-77.5}$	$216.7^{+51.3}_{-56.2}$
No Tags	42	0.65 ± 0.07	0.6 ± 0.1	$303.1^{+19.9}_{-134.7}$	$301.2^{+20.0}_{-180.2}$
All Subsamples	76	0.49 ± 0.07	0.47 ± 0.1	$204.4^{+9.3}_{-9.2}$	$199.9^{+9.1}_{-9.1}$

Table 11.4: Results of applying the mass likelihood procedure using the H_T templates to the four subsamples and for all subsamples combined. We show the results when the background fractions are constrained to their expected values (Table 7.2), and when they are not. The combined background fraction is the average of the x_b fit results weighted by the number of events in the subsamples.

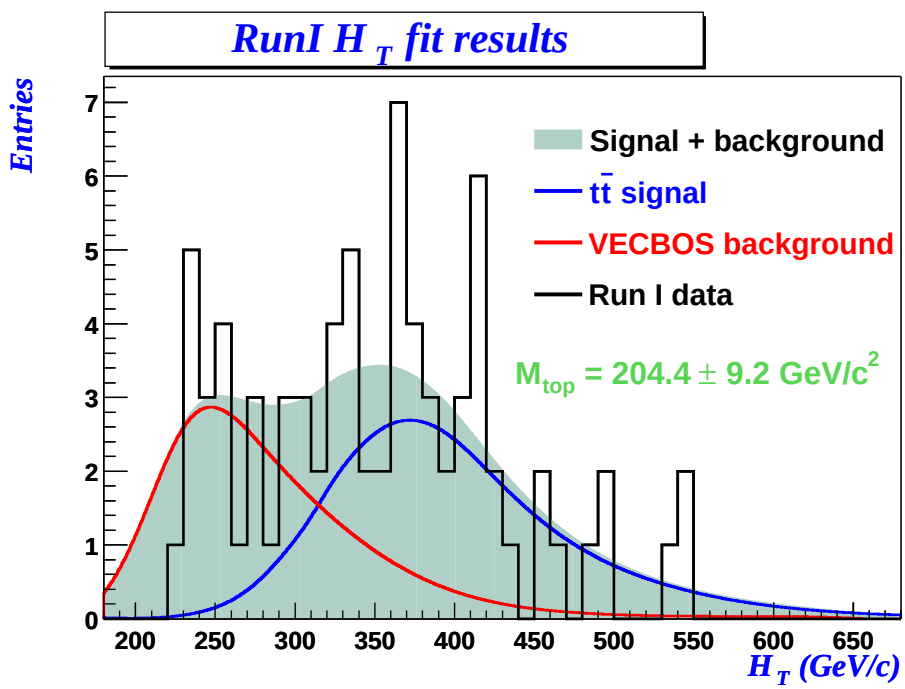

Figure 11.2: Results from applying the likelihood procedure to the combined four subsamples using the H_T templates. The figure shows the Run I data (histogram), fitted background and signal (shaded region), fitted signal (blue curve), and fitted background (red curve).

11.3 Comparison of Data to Expectations

Using the NN M_{rec} templates to estimate the mass of the top quark yields a result of 177.9 ± 4.8 GeV/c^2. The result obtained when using the $\mathrm{H_T}$ templates is 204.4 ± 9.3 GeV/c^2. Therefore, using different kinematic information from the same events produces results, which differ from one another by 26.5 GeV/c^2. We run 2000 pseudoexperiments for which we plot the difference between the top mass obtained using the NN M_{rec} templates ($M_{top}^{M_{rec}}$), and the top mass obtained from using the $\mathrm{H_T}$ templates ($M_{top}^{\mathrm{H_T}}$). Figure 11.3 shows the distribution of the difference $M_{top}^{M_{rec}}$ - $M_{top}^{\mathrm{H_T}}$. We find that roughly 5% of the simulated experiments yield a difference in the top mass that is equal or larger than what we see in the data sample. While this number is small, it is not statistically unreasonable.

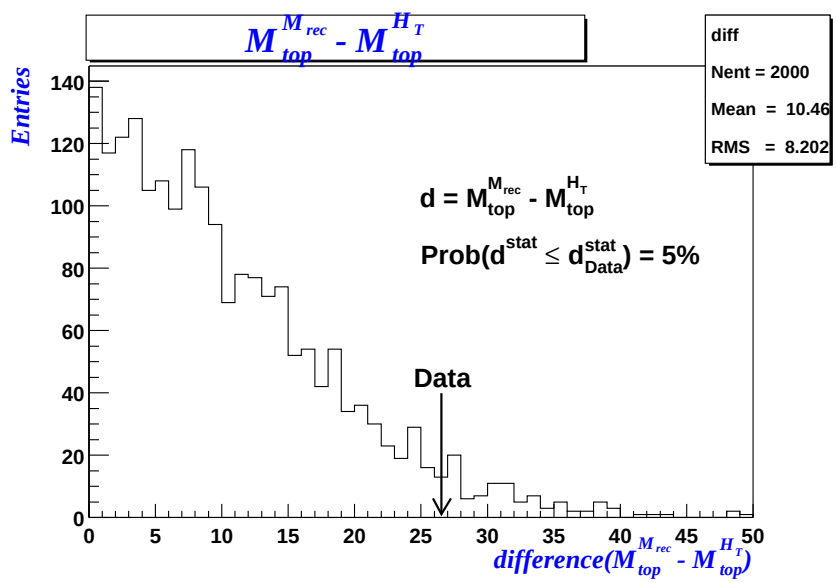

Figure 11.3: Distribution of the difference between the top mass obtained using the M_{rec} templates and the top mass obtained using the $\mathrm{H_T}$ templates from simulated experiments of $t\bar{t}$ ($M_{top} = 175$ GeV/c^2) plus background. Also shown is the difference obtained from the data sample.

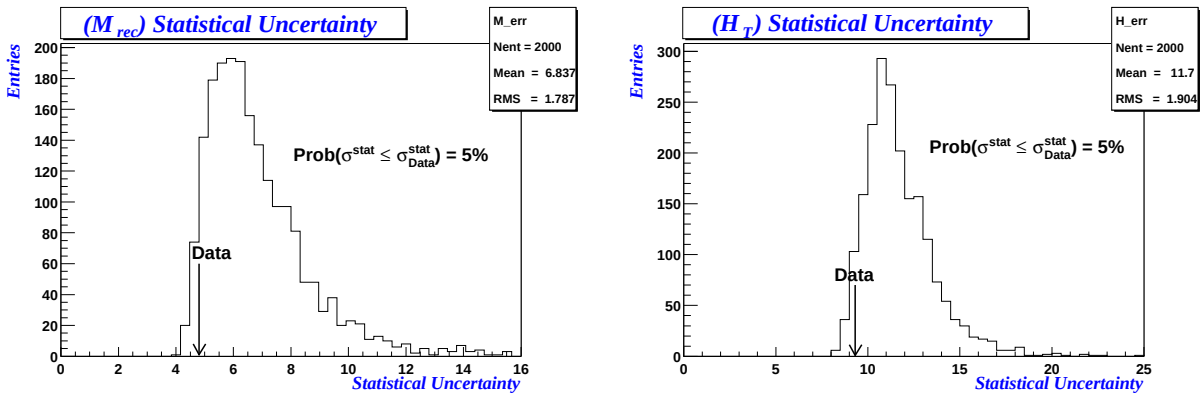

Figure 11.4: Statistical uncertainties from pseudoexperiments of $t\bar{t}$ (M_{top} = 175 GeV/c^2) plus background using (Left) the NN M_{rec} templates, and (Right) the H_T templates. Also shown in the statistical uncertainty obtained from the data sample.

The statistical uncertainty obtained from fitting the data with the NN M_{rec} templates is 4.8 GeV/c^2. Figure 11.4.(L) shows the distribution of the statistical uncertainties from the simulated experiments. We find that roughly 5% of the pseudoexperiments yield a statistical error which is equal or smaller to that obtained from the data. Figure 11.4.(R) shows the the distribution of the statistical uncertainty from the pseudoexperiments which use the H_T templates along with the data result (9.3 GeV/c^2). For this case, we find that 5% of the simulated experiments yield a statistical error which is equal or smaller to the one obtained from fitting the data with the H_T templates.

Finally, we check the returned log likelihood value from the pseudoexperiments to the one obtained from the data. Figure 11.5 shows the log likelihood distribution for pseudoexperiments (L) using the NN M_{rec} templates and (R) the H_T templates along with the log likelihood value from the data sample. We find that 50% of the pseudoexperiments that use the NN M_{rec} templates yield a log likelihood larger than that obtained from the data. Also, 87% of the simulated experiments that use the H_T templates yield a log $\mathcal{L}$ value larger than the one seen in the data.

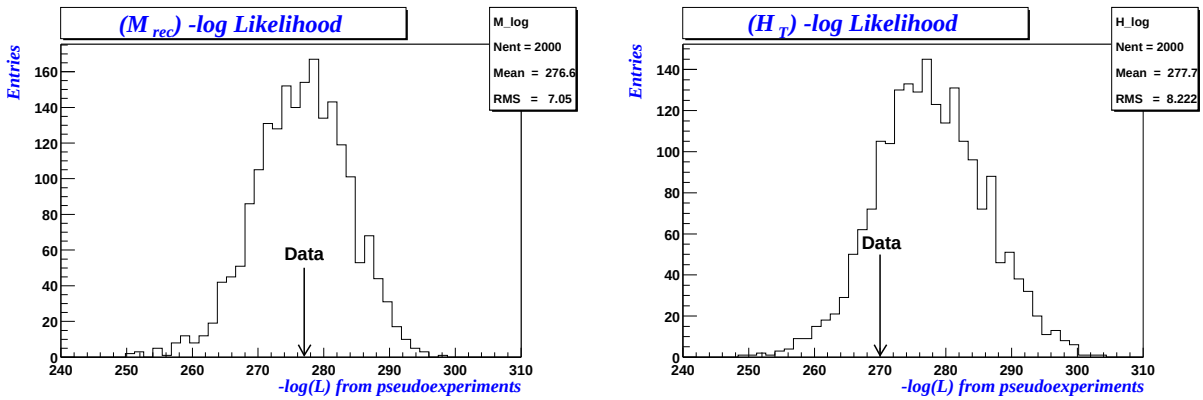

Figure 11.5: Distribution of the minimum value of -log $\mathcal{L}$ from pseudoexperiments of $t\bar{t}$ (M_{top} = 175 GeV/c^2) plus background using (Left) the M_{rec} templates, and (Right) the H_T templates. Also shown in the minimum -log $\mathcal{L}$ value obtained from the data sample.

11.4 Standalone NN Run I Data Results

The standalone NN method is applied to the events in the four mass subsamples. To investigate the effect of the background constraining term, the mass is performed two different ways: the backgrounds are constrained to be within their expected values, and the background constraint is removed. The results from applying the standalone NN method are given in 11.5.

As in the template-based method, the unconstrained mass fit of the SVX tagged subsamples yield a background fraction of zero, although with large uncertainties. The unconstrained fit to the No Tag samples yields a background fraction lower than expected. In general we would expect an increase in the statistical uncertainty if the background constraint is removed. This is not the case because the unconstrained fit favors fewer background events than what it is expected from Monte Carlo. Overall, the standalone NN method behaves in a similar fashion as the template-based techniques.

Data sample	N_{obs}	x_b^{fit}		Top mass (GeV/c^2)	
		Constrained	Unconstrained	Constrained	Unconstrained
SVX Double	5	0.04 ± 0.02	$0.0^{+0.2}_{-0.0}$	179.9 ± 10.6	180.0 ± 10.4
SVX Single	15	0.14 ± 0.05	$0.0^{+0.06}_{-0.0}$	172.7 ± 8.6	172.5 ± 8.4
SLT	14	$0.55^{+0.08}_{-0.08}$	$0.6^{+0.25}_{-0.37}$	167.4 ± 28.8	169.4 ± 30.0
No Tags	42	0.53 ± 0.08	$0.32^{+0.13}_{-0.1}$	189.1 ± 10.0	183.2 ± 9.8
All Subsamples	76	0.42 ± 0.07	0.29 ± 0.14	182.4 ± 5.1	182.0 ± 4.7

Table 11.5: Results of applying the standalone NN method to the four subsamples and for all subsamples combined. We show the results when the background fractions are constrained to their expected values (Table 7.2), and when they are not. The combined background fraction is the average of the x_b fit results weighted by the number of events in the subsamples.

It is surprising that the statistical error returned by the standalone NN method is higher (5.1 GeV/c^2) than the one returned when we only use the M_{rec} templates (4.8 GeV/c^2). Adding more information has the opposite effect that one would expect. This can be due to the fact that the M_{rec} and $\mathrm{H_T}$ variables favor a top mass which differs from one another by 26.5 GeV/c^2. Therefore, combining the information from these variables broadens the likelihood distribution, which results in a higher statistical uncertainty.

11.5 Comparison of Data to Expectations

We want to determine whether the correlation coefficients between the variables in the Monte Carlo pseudoexperiments agree with what we see in the Run I data. Figure 11.6 show the different correlation distributions for all the variables used in the standalone Neural Network method along with the correlation values seen in the data.

As we can see, the correlations between the different variables in the Monte Carlo agrees reasonably well with the data values. We note that the correlation between M_{rec} and $\mathrm{H_T}$ in the data falls in the middle of the Monte Carlo distribution. These two variables are the ones with the most mass information. We have also checked the uncertainty returned by the fit on the data. This error is reasonable.

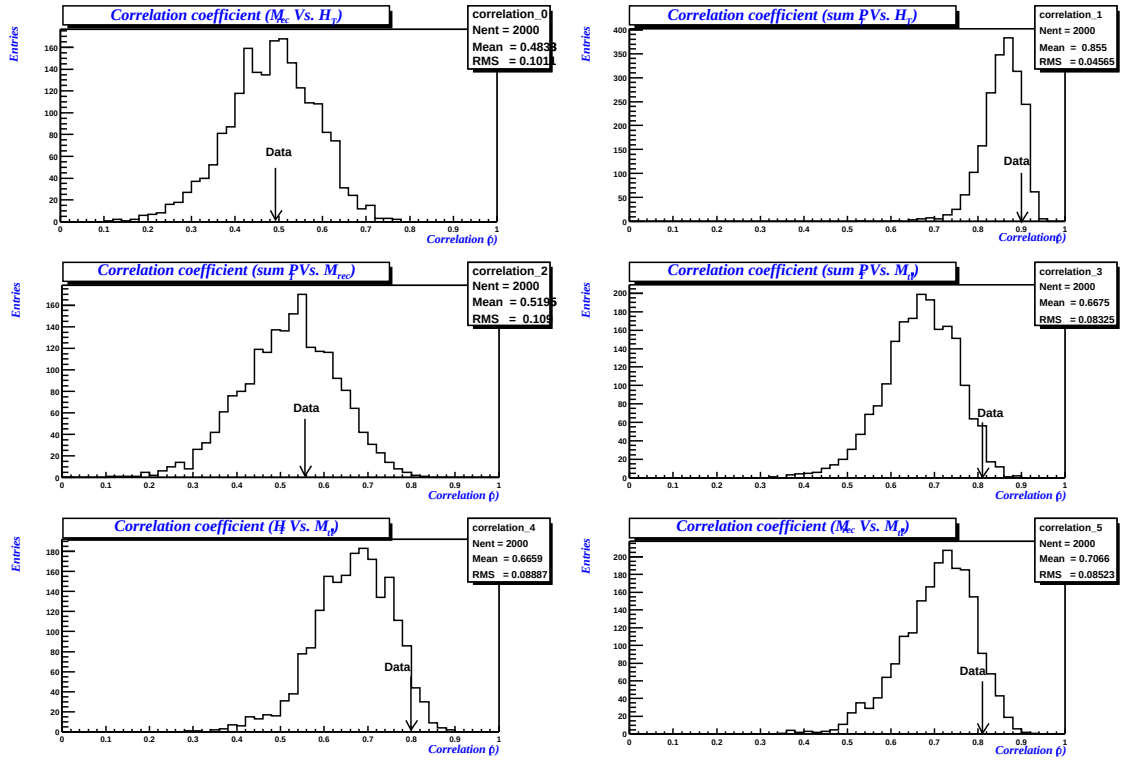

Figure 11.6: Distribution of the correlation coefficient between the different variables used in the standalone NN mass analysis along with the correlation value seen in the data sample. (a) Correlation between M_{rec} and $\mathrm{H_T}$, (b) correlation between $P_T(1) + P_T(2)$ and $\mathrm{H_T}$, (c) correlation between $P_T(1) + P_T(2)$ and M_{rec} , (d) correlation between $P_T(1) + P_T(2)$ and $M_{t\bar{t}}$, (e) correlation between $\mathrm{H_T}$ and $M_{t\bar{t}}$, and (f) correlation between M_{rec} and $M_{t\bar{t}}$.

CHAPTER 12

MEASURING THE TOP MASS IN RUN II

Run II started at the Tevatron in March 2001. In this chapter, we outline a method to measure the top mass during Run II. The technique uses the M_{rec} and $\mathrm{H_T}$ information simultaneously. Using this method we present a priori expectations on the top mass measurement.

12.1 Motivation

From Monte Carlo pseudoexperiments, we know that the best measurement of the top mass is obtained from using the standalone NN method. This technique combines the information from several kinematic variables to extract the top mass using a discrete likelihood function. However, this likelihood procedure may have some problems in Run II:

- For Run I, the discrete likelihood consisted of 23 different points, which were associated with a different top mass, spaced at intervals as close as 2.5 GeV/c^2. As the Run II statistics increase, we would have to add more points to the likelihood to take advantage of the expected smaller statistical uncertainty.

- Adding more points to the likelihood means that we would have to generate more $t\bar{t}$ Monte Carlo samples at more top masses closer together.

- Using all the Monte Carlo samples we would have to create new Neural Networks to perform the event classification. Because the networks are so large this would take large amounts of time.

In the next section we describe a new mass analysis method, which uses the two most powerful variables in the standalone NN along with a continuous likelihood procedure to extract the top mass. We note that this method was not applied to the Run I data sample because we did not have enough Monte Carlo statistics. However, we expect both methods to give comparable results.

12.2 M_{rec} Vs. $\mathrm{H_T}$ Templates

The biggest improvement in the standalone NN method comes from the combination of the M_{rec} and the $\mathrm{H_T}$ information. For the measurement of the top mass in Run II we propose the use of 2-D M_{rec} Vs. $\mathrm{H_T}$ templates. These templates would then be fitted with a NN to obtain a functional form dependent on the mass of the top quark.

One of the challenges that one faces when trying to fit 2-D histograms is that the amount of events needed to correctly populate the histograms is very large. Although we have large amounts of $t\bar{t}$ and background Monte Carlo samples we do not have enough to construct 2-D templates. However, outlining this procedure is still very important since it may represent an excellent technique to extract the mass of the top quark. Thus, Appendix F describes a method to generate "fake" Monte Carlo to increase the sizes of our samples. The resulting 2-D templates closely mimic the shape of the actual Monte Carlo results. Using these distributions, we investigate whether a Neural Network is able to fit them or not.

12.3 Fitting Procedure

Fitting a set of 2-D templates (M_{rec} Vs. H_T) is much like fitting a 1-D distributions. We just have to make sure that the network gets all the information, and that there are enough hidden units to correctly approximate the function. For the $t\bar{t}$ signal we have chosen a 3-8-8-1 architecture (113 weights and biases). Each of the inputs is associated with a different variable: M_{rec}, H_T, and M_{top}. During training, the desired NN output is set to the number of entries in each bin of the 2-D distributions. In our case, the 2-D distributions have a total of 1600 bins (there are 40 bins in the M_{rec} axis and 40 in the H_T axis), and since there are 19 different distributions our fits have over 30,000 degrees of freedom. Figure 12.1 shows a few distributions of our 2-D Neural Network fits for the SVX Single subsample. Although the network has to fit many different points, it performs the task very well as we can see from the χ^2 results shown in Table 12.1. The χ^2 function for the 2-D case has the same form as Equation 7.1 with the addition of an extra sum to take into account the two dimensions of our templates.

Top Subsample	χ^2 result
SVX Single	1.02 for 22408 *dof*
SVX Double	1.19 for 23433 *dof*
SLT	1.08 for 24045 *dof*
No Tags	1.06 for 23150 *dof*

Table 12.1: We show the χ^2 per degree of freedom for each of the four subsamples. The NN does an excellent job at approximating the 2-D surfaces.

For the background we remove the top mass dependency and the NN architecture becomes 2-5-5-1. This corresponds to a total of 41 weights and biases.

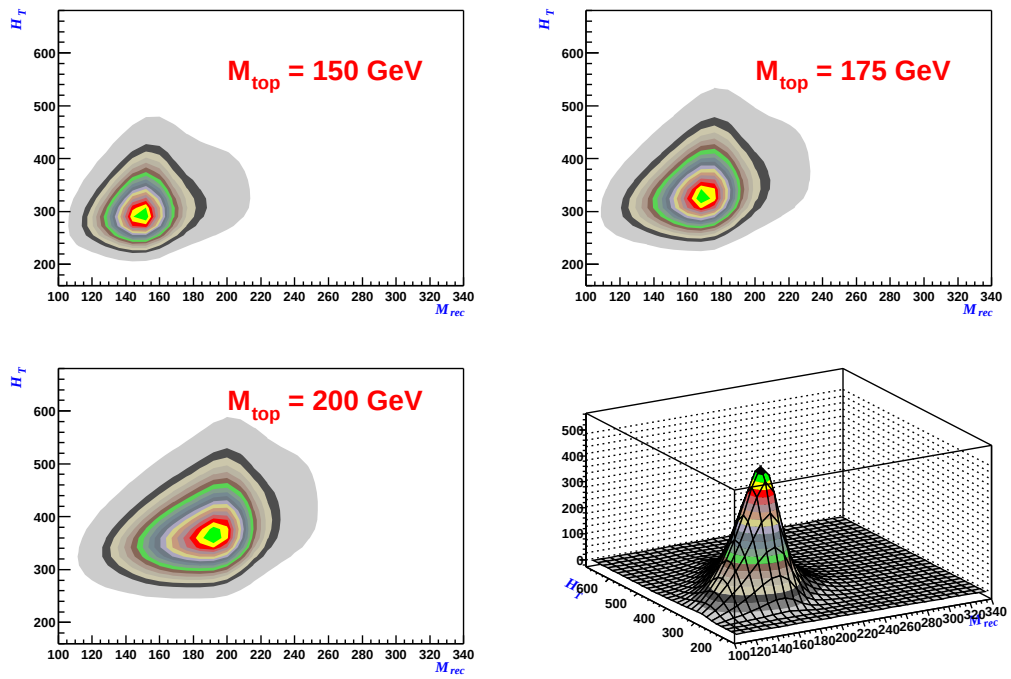

Figure 12.1: M_{rec} Vs. H_T surfaces generated from the functional form obtained by fitting the "fake" Monte Carlo Single SVX subsample. (a) Contour plot for M_{top} = 150 GeV/c^2. (b) Contour plot for M_{top} = 175 GeV/c^2. (c) Contour plot for M_{top} = 200 GeV/c^2. (d) Surface plot for M_{top} = 175 GeV/c^2.

12.4 Run II Expectations

To estimate the mass of the top quark we use the continuous likelihood method described in Section 7.3. In this case, the functional forms for the signal and background use both the M_{rec} and $\mathrm{H_T}$ information from each event. In the likelihood fit, the mass of the top is the only free parameter and the background fractions are constrained to be within their expected values. We run 2000 simulated experiments with the admixture of signal and background as expected in Run I. During Run II, we expect to collect over 1 fb^{-1} of data. We run a set of pseudoexperiments for which the

statistics are increased 2, 4, 6, and 10 times respectively. For each pseudoexperiment, we calculate the top mass using: the H_T NN-fitted templates, the M_{rec} GG-fitted templates, the M_{rec} NN-fitted templates, and the M_{rec} Vs. H_T NN-fitted templates. The results are shown in Figure 12.2.

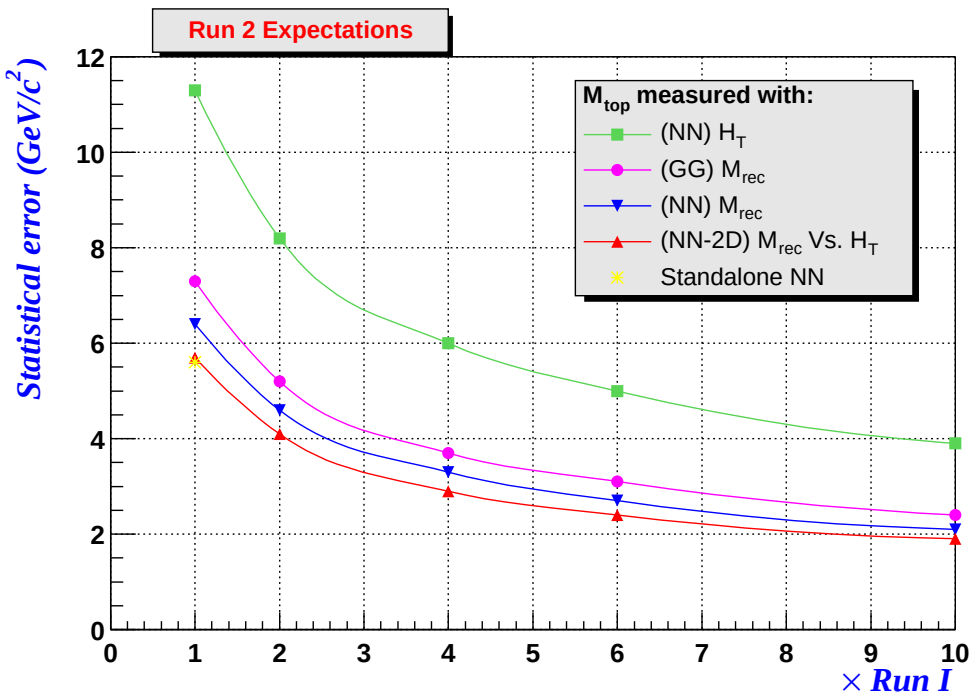

Figure 12.2: We apply the different template-based analyses methods to 2000 simulated experiments. The graph shows the median error obtained from the pseudoexperiments containing 1, 2, 4, 6, and 10 times the Run I statistics. For the standalone NN method we only show the results from the Run I simulated experiments.

Using the Run I statistics, the M_{rec} Vs. H_T template method returns a median error of 5.7 GeV/c^2. This error is slightly higher than that obtained by the standalone NN method technique (5.6 GeV/c^2), and (7.3/5.7) = 28% better than the M_{rec} template fitted with the Gaussian plus Gamma functions. Recall, that this

method gave the best Run I top mass measurement. This improvement remains the same when we use simulated experiments, which contain 10 times the amount of Run I statistics (this corresponds to roughly 1 fb^{-1}).

The goal for Run II is to measure the top mass with an error less than 2 GeV/c^2. From a statistical perspective, this goal can be reached with roughly 1 fb^{-1} of data using the M_{rec} Vs. H_T template method. At this time, the error in the top mass will be dominated by the systematic uncertainty.

APPENDIX A

THE DELTA RULE

For a single layer network with an output unit with a *linear* activation function the output is simply given by

$$y(t+1) = \sum_j w_j(t)x_j + \theta(t). \tag{A.1}$$

Such a simple Neural Network is able to represent a linear relationship between the value of the output unit and the values of the input units. We use this simple network to fit to a set of training samples consisting of input values x^p and desired output values d^p. For every given input sample, the output of the Neural Network differs from the desired output value by $(d^p - y^p)$, where y^p is defined above. The Delta Rule uses an error function based on these differences to adjust the weights.

The error function is just the *least mean square* (LMS) and the total error is defined to be

$$E = \sum_p E^p = \frac{1}{2}\sum_p (d^p - y^p)^2, \tag{A.2}$$

where the index p ranges over the set of input patterns, and E^p represents the error on pattern p. The LMS method uses the *gradient descent* method to find the weights that minimize the error function. The idea is to make the change in the weight, w_j,

proportional to the negative of the derivative of the error as measured on the current pattern, j, with respect to each weight, w_j:

$$\Delta_p w_j = -\gamma \frac{\partial E^p}{\partial w_j} \tag{A.3}$$

where γ is a proportionality constant. The derivative can then be written as

$$\frac{\partial E^p}{\partial w_j} = \frac{\partial E^p}{\partial y^p} \frac{\partial y^p}{\partial w^j}. \tag{A.4}$$

Because we are dealing with linear units (Eq. A.1),

$$\frac{\partial y^p}{\partial w^j} = x_j \quad \text{and} \quad \frac{\partial E^p}{\partial y^p} = -(d^p - y^p) \tag{A.5}$$

such that

$$\Delta_p w_j = \gamma \delta^p x_j \tag{A.6}$$

where $\delta^p = d^p - y^p$ is the difference between the NN output and the desired output pattern. The Delta rule modifies the weights according to the desired output pattern, the input pattern, and the actual NN output. These characteristics have opened up a wealth of new applications.

APPENDIX B

BACK-PROPAGATION

We must extend the Delta rule so that it can be applied to networks with many layers and with non-linear activation functions. This generalisation is called the *Back-propagation* method and the central idea is that errors of the weights associated with units in the hidden layers are determined by back-propagating the error of the weights associated with the output units.

We generalize the Delta rule to the set of non-linear activation functions. The activation, is a differentiable function of the total input given by

$$y_k^p = \mathcal{F}(s_k^p), \quad \text{in which} \quad s_k^p = \sum_j w_{jk} y_j^p + \theta_k, \tag{B.1}$$

where the index p corresponds to the different patterns, the index k is the number of units per layer, and the index j keeps track of the different weights connecting to each unit. To get the correct generalization of the Delta rule presented in Appendix A, we must set

$$\Delta_p w_{jk} = -\gamma \frac{\partial E^p}{\partial w_{jk}}. \tag{B.2}$$

The error measure E^p is defined in the same manner as before, and it corresponds to the quadratic error for pattern p at the output units:

$$E^p = \frac{1}{2} \sum_{o=1}^{N_o} (d_o^p - y_o^p)^2, \tag{B.3}$$

where d_o^p is the desired output for unit o when the pattern p is used. We further set the total error to $E = \sum_p E^p$. We can write

$$\frac{\partial E^p}{\partial w_{jk}} = \frac{\partial E^p}{\partial s_k^p} \frac{\partial s_k^p}{\partial w_{jk}}. \tag{B.4}$$

By equation B.1 we see that the second term is

$$\frac{\partial s_k^p}{\partial w_{jk}} = y_j^p. \tag{B.5}$$

When we define

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p}, \tag{B.6}$$

we have an update rule which is similar to the Delta rule derived in appendix A. In a multi-layer network with non-linear activation function, the weights change according to:

$$\Delta_p w_{jk} = \gamma \delta_k^p y_j^p. \tag{B.7}$$

The problem now is how to figure out what δ_k^p is for every unit k in the network. The interesting results, as we will see, is that there is a simple recursive computation of these δ's which can be implemented by propagating the errors backward through the network.

We rewrite equation B.7 using the chain rule as a product of two terms, one term reflecting the change in the error as a function of the output of the unit and another term reflecting the change as a function of the changes in the input of the unit. Therefore, we have

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p} = -\frac{\partial E^p}{\partial y_k^p} \frac{\partial y_k^p}{\partial s_k^p}. \tag{B.8}$$

From equation B.1, the second term is

$$\frac{\partial y_k^p}{\partial s_k^p} = \mathcal{F}'(s_k^p), \tag{B.9}$$

which is simply the derivative of the non-linear [11] function $\mathcal{F}$ for the k-th unit, evaluated at the net input s_k^p to that unit. To compute the first term in equation B.7, we consider two cases. First, assume that the unit k is an output unit $k = o$ of the network. In this case, it follows from the definition of E^p that

$$\frac{\partial E^p}{\partial s_k^p} = -(d_o^p - y_o^p), \tag{B.10}$$

which is the same result as obtained in appendix A. Substituting this result and equation B.10 into equation B.8, we get

$$\delta_o^p = (d_o^p - y_o^p)\mathcal{F}'(s_o^p) \tag{B.11}$$

for any output unit o. Secondly, if k does not correspond to an output unit but a hidden one $k = h$, we do not readily know the contribution of the unit to the total output error of the network. However, we can write the error as a function of the net inputs from hidden to output layer. That is, $E^p = E^p(s_1^p, s_2^p, \ldots, s_j^p, \ldots)$ and we use the chain rule to write

$$\frac{\partial E^p}{y_h^p} = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p}\frac{\partial s_o^p}{\partial y_h^p} = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p}\frac{\partial}{\partial y_h^p}\sum_{j=1}^{N_h} w_{jo}y_j^p = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} w_{ho} = -\sum_{o=1}^{N_o} \delta_o^p w_{ho}. \tag{B.12}$$

Substituting this into equation B.9 yield

$$\delta_h^p = \mathcal{F}'(s_h^p)\sum_{o=1}^{N_o} \delta_o^p w_{ho}. \tag{B.13}$$

Equations B.11 and B.13 give a recursive procedure for computing all the different δ's in the network. Equation B.7 is then used to modify the weights. This procedure constitudes the generalized Delta rule for feed-forward networks with non-linear activation functions.

[11] these functions are also called *squashing* functions such that $\lim_{\lambda\to\infty}\mathcal{F}(\lambda) = 1$, and $\lim_{\lambda\to-\infty}\mathcal{F}(\lambda) = 0$

APPENDIX C

FITTING WITH A NEURAL NETWORK

We want to use a Neural Network to fit a series of reconstructed mass distributions. An example showing the distributions for a top mass of 120 GeV/c^2, 175 GeV/c^2, and 230 GeV/c^2 is given in Figure C.1. As expected, the distributions peak around the true top mass used to generate each Monte Carlo sample. The goal of this project is to create a single Neural Network that is able to duplicate the shape of a distribution as a function of the true top mass.

We have to put all the information shown in the histograms of Figure C.1 into a format that the Neural Network can understand. First, all the histograms must be scaled so that they have the same number of events. Then, the Neural Network must have the information that for M_{top} = 120 GeV/c^2 the 10th bin has 122 entries, for M_{top} = 175 GeV/c^2 it has only 19 entries, and for M_{top} = 230 GeV/c^2 the 10th bin has four entries. There is only two ways we can introduce any information into a Neural Network: the input nodes and the output node.

We chose one of the input units to be the true top mass, M_{top}. The Neural Network must have the information that if the top mass changes the shape of the distribution changes. We chose the second input unit to be the bin location along the M_{rec} axis. Each of the reconstructed mass histograms range from 80 to 380

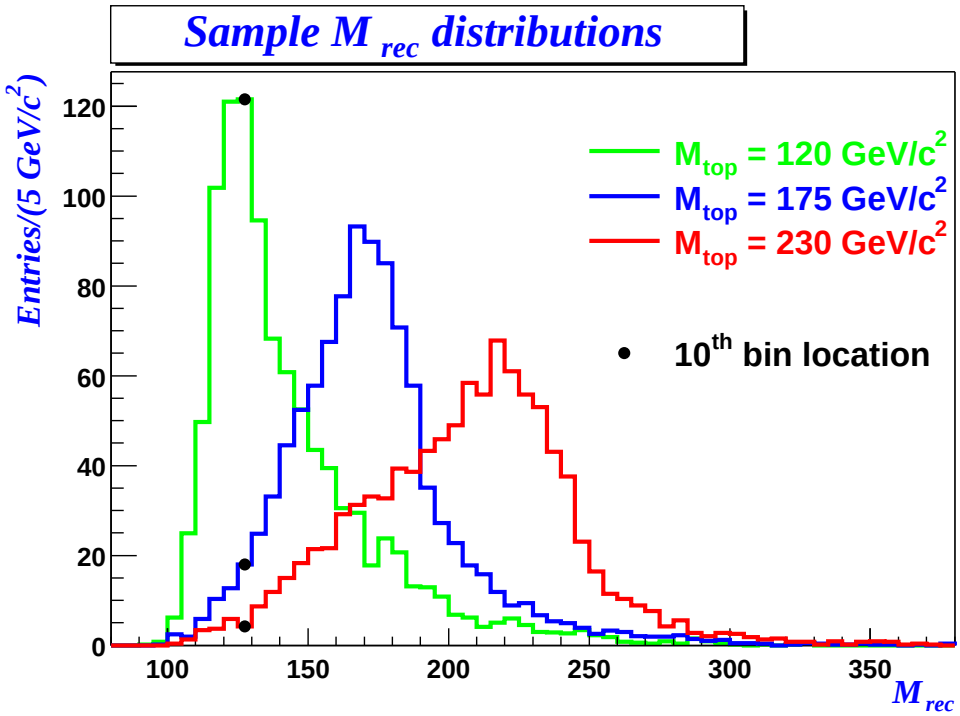

Figure C.1: We show the M_{rec} distributions for three different top masses. All of the histograms have been normalized to the same number of events.

GeV/c^2 and are divided into 5 GeV/c^2 bins In order for MLPfit to work properly we have to normalize the input variables so that they are between 0 and 1. Thus, the normalization equations are the following:

$$\text{input}_1 = \frac{(\text{Bin location} - 80.0)}{300.0} \tag{C.1}$$

$$\text{input}_2 = \frac{(M_{top} - 120.0)}{110.0} \tag{C.2}$$

The last piece of information comes in through the output node. In classification problems we are used to setting the output node to a single value such as 1 for signal and 0 for background. That is not the case when we deal with fitting. Now, we set the target value of the output node to the bin content associated with both the bin

location and the true top mass. An excerpt of a training file is given below. The first column corresponds to the bin location, and the second column is M_{top} . The numbers on the right have been normalized to unity according to Equation C.1 and C.2. The last two columns is what is ultimately fed into the Neural Network.

$$
\begin{array}{cccccc}
127.5 & 120.0 & \Longrightarrow & 0.16 & 0.0 \\
122.0 & & \Longrightarrow & 122.0 & \\
\vdots & \vdots & & \vdots & \vdots \\
127.5 & 175.0 & \Longrightarrow & 0.16 & 0.5 \\
19.0 & & \Longrightarrow & 19.0 & \\
\vdots & \vdots & & \vdots & \vdots \\
127.5 & 230.0 & \Longrightarrow & 0.16 & 1.0 \\
4.0 & & \Longrightarrow & 4.0 &
\end{array}
$$

When this information is fed into the Neural Network, it tells the network that for $M_{top} = 120$ GeV/c^2 there are 122 entries in the 10th bin. Similarly, for $M_{top} = 175$ GeV/c^2 there are 19 entries in the 10th bin, and for $M_{top} = 230$ GeV there are four entries. The above is just a simple example with only three M_{rec} distributions. The actual file used to train the Neural Network is made up of 1140 entries, which correspond to 19 different distributions with 60 bins each. Training is stopped when the χ^2 function defined in Section 7.1.1 is close to one.

Once an optimal Neural Network is obtained, we use it much like a regular function. The network takes two input parameters, the M_{rec} and M_{top}, and the output of the Neural Network is interpreted as a probability density function. This output is used in the mass analysis likelihood function described in Section 7.3.

APPENDIX D

VECBOS BACKGROUND

To increase the number of background events used in this analysis we have combined all the VECBOS Monte Carlo generated into a single sample. Thus, our background sample combines events generated with a scale $Q^2 = \langle P_T \rangle^2$ and $Q^2 = \langle M_W \rangle^2$. In this Appendix we present the studies that support that combining both kinds of background events does not have an effect on the top mass measurement.

In Section 8.1 we introduced the kinematic variables that contain the most mass information. We plot these distributions for both kinds of backgrounds to see whether or not they differ from one another. Figure D.1 shows the M_{rec} , $\mathrm{H_T}$, $M_{t\bar{t}}$, and $P_T(1)+P_T(2)$ distributions for the SVX Single VECBOS sample. The only distributions that seem to be somewhat different are the M_{rec} distributions. It is important to point out that the SVX subsamples suffer from the least amount of background. Although not shown, the distributions for the SLT and No Tags subsamples show very little difference. We perform a series of Kolmogorov tests between the distributions in each subsample to obtain a measure of how different they are. Table D.1 shows the results from our K-S tests.

VECBOS Subsample	M_{rec}	H_T	$M_{t\bar{t}}$	$P_T(1) + P_T(2)$
SVX	2.72%	43.4%	77.9%	27.0%
SLT	44.9%	34.5%	23.7%	14.0%
No Tags	96.5%	97.3%	95.4%	100%

Table D.1: We give K-S tests results for each of the VECBOS subsamples. We note that the SVX Single background is used for both SVX subsamples.

The K-S tests are an indication that both sets of background events are fairly similar. Next, we investigate whether or not using the entire VECBOS background has an effect on the different mass analyses described in this thesis.

D.1 Background Shape Effect on NN Template Method

We separate the background events into the two different classes. Using the default $Q^2 = \langle P_T \rangle^2$ background events, we construct the M_{rec} background template which is then fitted with the NN method. We generate pseudoexperiments as explained in Section 7.3 and we calculate the top mass. Next, we generate pseudoexperiments for which we draw the background events from the $Q^2 = \langle M_W \rangle^2$ background, and we compare them to the default background distributions. When we compare the results from the previous two methods, the median top mass was found to differ by 0.3 ± 0.2 GeV/c^2. Thus, combining both classes of events into a large background does not change the top mass value.

D.2 Background Shape Effect on Standalone NN Analysis

The main reason why we want to combine all of the VECBOS events into a single background sample is to have enough statistics to properly train the networks used in the Standalone NN analysis.

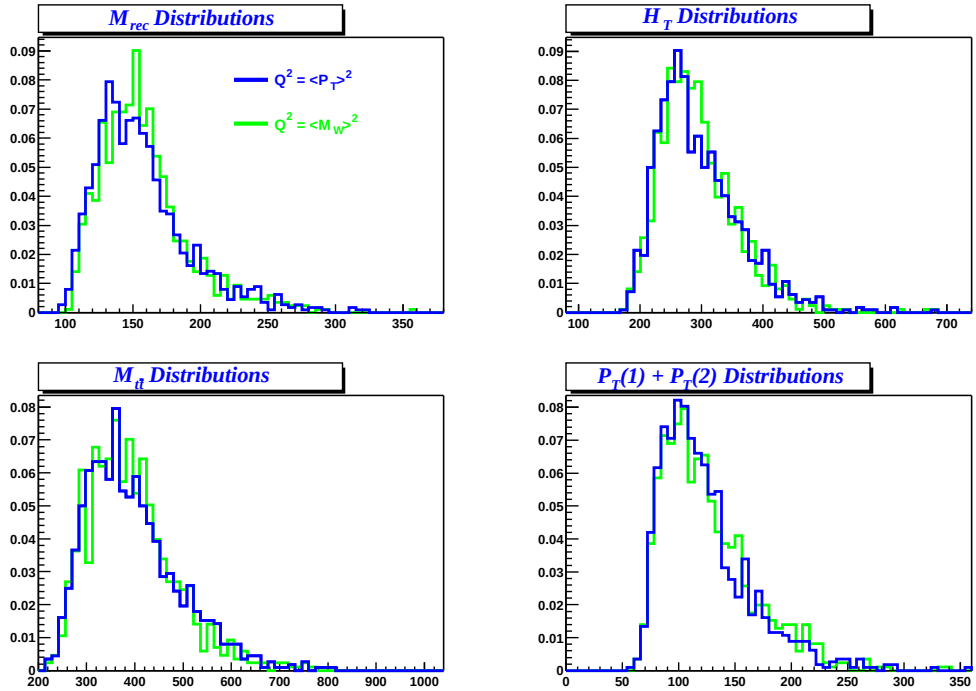

Figure D.1: Comparisons of the VECBOS background generated with two different values of Q^2 for SVX-tagged events for (a) reconstructed top mass, M_{rec} , (b) $\mathrm{H_T}$, (c) $t\bar{t}$ invariant mass, $M_{t\bar{t}}$, and (d) $P_T(1) + P_T(2)$.

We have generated two new networks: one for the Single SVX subsample, and another for the No Tags subsample. These new networks have the same architecture as the ones shown in Table 8.2. However, they are trained using only default background events. We have not generated a network for the SLT subsample because there are not enough events in the default background sample to train the network correctly.

We ran pseudoexperiments for the Single SVX subsample in which we draw the background events from the default background sample. Next, we ran pseudoexperiemnts for which we draw the background events from the $Q^2 = \langle M_W \rangle^2$ background sample. We find that the difference in the top mass median is 0.2 ± 0.2

GeV/c^2. We perform a similar study for the No Tags subsample and we find that the difference in the mass median is 0.0 ± 0.2 GeV/c^2.

Based on the results from the K-S tests, and the mass analyses tests we think that combining the $Q^2 = \langle P_T \rangle^2$ and $Q^2 = \langle M_W \rangle^2$ backgrounds has no effect on the top mass calculation.

APPENDIX E

COMBINING MEASUREMENTS

Suppose one is trying to determine the value of quantity s and there are n independent measurements s_i with their corresponding error σ_i. The best estimate for s is then given by

$$\overline{s} = \frac{\sum_i x_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}. \tag{E.1}$$

Now, suppose that we have n measurements of a quantity s but that each measurement is related to the other $n-1$ measurements. In this case, we say that the measurements are correlated. To find the best estimate of the quantity s we need more information than the individual errors. Thus, we define the *covariance matrix* V_s whose terms are given by

$$(V_s)_{ij} = \rho_{ij}\sigma_i\sigma_j, \tag{E.2}$$

and

$$\rho_{ij} = \frac{(V_s)_{ij}}{\sigma_i\sigma_j} \tag{E.3}$$

where the *correlation coefficient*, ρ_{ij}, is a measure of the relation between two variables, and it can be proved that it must lie between -1 and +1.

We want to construct the best linear estimator of the quantity s which accounts for all the measurements, including their errors and correlations. We define the estimator

to be

$$\overline{s} = \sum_i c_i s_i, \tag{E.4}$$

with an error given by

$$\sigma_{\overline{s}}^2 = \sum_{ij} c_i c_j (V_s)_{ij}. \tag{E.5}$$

In the previous two equations, c_i's are the coefficients that must be found. From any statistics book [52], we find that these coefficients are given by

$$c_i = \frac{\sum_k (V^{-1})_{ik}}{\sum_{jk} (V^{-1})_{jk}}. \tag{E.6}$$

Note that this expression for c_i is identical to equation E.1 when the errors are independent from one another.

E.1 Information

Imagine that we want to combine two measurements $s_1 \pm \sigma_1$ and $s_2 \pm \sigma_2$ into a single measurement $\overline{s}$ with error $\sigma_{\overline{s}}^2$. We plot the combined error $\sigma_{\overline{s}}^2$ for two different errors σ_1 and σ_2 as a function of the correlation, ρ, between the measurements s_1 and s_2. Figure E.1.(a) shows the results for the case when $\sigma_1 = 5.0$ and $\sigma_2 = 10.0$. As we can see from the graph, $\sigma_{\overline{s}}^2(\rho)$ is not linear. As a matter of fact, the behavior of the function is quite surprising. The plot implies that when the measurements are 100% correlated ($\rho = +1$ or -1) , we will be able to know the quantity perfectly.

One expects the combined error to be at a minimum when the measurements are 100% uncorrelated and then to grow as the correlation coefficient increases. The naive interpretation is that as the measurements become more correlated the information one is adding decreases. To determine whether this is true or not we use the Fisher information variable [53]. We assume the two measurements have binormal

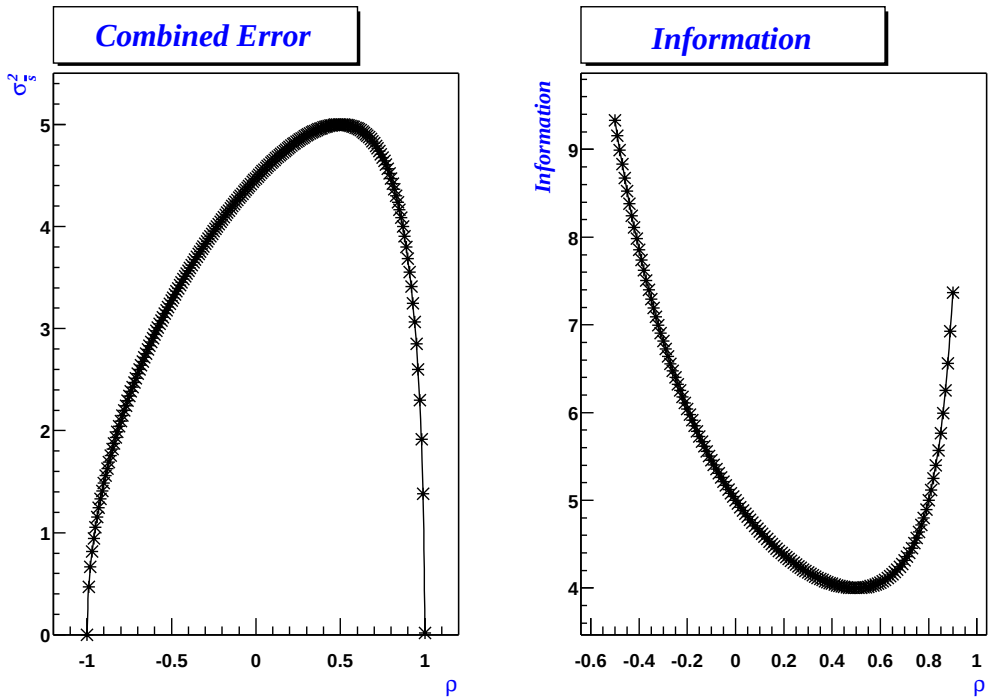

Figure E.1: (Left) Shows $\sigma^2_{\bar{s}}(\rho)$. The function is not linear and it is equal to 0 when the $\rho = +1$ or -1. (Right) The Fisher information is shown as a function of ρ. The information is maximum when the measurements are 100% correlated (values not shown because they approach infinity).

distributions and we derive:

$$I = \frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{\sigma_1^2 \cdot \sigma_2^2 \cdot (1 - \rho^2)} \tag{E.7}$$

Figure E.1.(R) shows that information decrease from $\rho = 0$ to $\rho = \sigma_1/\sigma_2$, after which it begins to increase. The explanation to this effect is somewhat simple. As the correlation increases beyond a certain point, you are limiting the freedom of the measurements to differ from the true value s_{true}. Therefore the variance of the combined measurement starts going down. This is most easily understood for the case where $\rho = 1$.

A correlation of 100% means that if s_1 shifts by $f \times \sigma_1$, then s_2 has to shift by exactly $f \times \sigma_2$. It is straightforward to check that under such a combined check the value of $\overline{s}$ does not change. In fact, since the expectation values of s_1 and s_2 are both s_{true}, and since they are 100% correlated, there must exist some real number r such that $s_1 = s_{true} + r \times \sigma_1$ and $s_2 = s_{true} + r \times \sigma_2$ [54]. If we substitute this expression into equation E.4 we find that $\overline{s} = s_{true}$. Therefore the result $\sigma^2_{\overline{s}}(\sigma_1, \sigma_2) = 0$ makes perfect sense.

APPENDIX F

GENERATION OF "FAKE" MONTE CARLO

The generation of "fake" Monte Carlo is only to see whether or not a NN can fit a set of 2-D distributions. It is not intended to replace the actual Monte Carlo in any way or form.

In Chapter 7 we describe how to obtain the M_{top} -dependent functional forms for M_{rec} and $\mathrm{H_T}$. These functional forms provide an infinite amount of statistics. However, these functions are 100% independent and from Monte Carlo studies we know that M_{rec} and $\mathrm{H_T}$ are somewhat correlated. The results from this correlation study are given in tables F.1. Looking at all distributions from all the generated top mass we find that the average correlation between these variables is roughly 38% for the $t\bar{t}$ signal and about 50% for the VECBOS background. Using this information we generate "fake" Monte Carlo using the following recipe:

- We draw 50 events[12] randomly from the M_{rec} and $\mathrm{H_T}$ functional forms. On average the correlation between the 50 events in these subsamples will be ~ 0. However, in some cases the correlation will be equal or greater to that observed in the Monte Carlo.

[12] A larger number of events will reduce the probability of getting a correlation close to what the Monte Carlo predicts.

Top Mass	SVX Single	SVX Double	SLT	No Tags
120 GeV	0.42	0.42	0.47	0.44
130 GeV	0.41	0.32	0.42	0.41
140 GeV	0.37	0.38	0.38	0.46
145 GeV	0.36	0.36	0.36	0.37
150 GeV	0.37	0.37	0.40	0.39
155 GeV	0.37	0.31	0.40	0.42
160 GeV	0.36	0.38	0.37	0.40
165 GeV	0.37	0.36	0.36	0.39
170 GeV	0.39	0.37	0.38	0.37
175 GeV	0.36	0.37	0.37	0.39
180 GeV	0.38	0.39	0.35	0.39
185 GeV	0.38	0.38	0.36	0.38
190 GeV	0.37	0.36	0.37	0.40
195 GeV	0.38	0.33	0.38	0.36
200 GeV	0.37	0.37	0.38	0.36
205 GeV	0.35	0.34	0.35	0.37
210 GeV	0.38	0.40	0.36	0.40
220 GeV	0.39	0.38	0.41	0.37
230 GeV	0.38	0.41	0.40	0.37
VECBOS	0.53	N/A	0.54	0.51

Table F.1: We show the correlation factor between M_{rec} and H_T for all the different Monte Carlo $t\bar{t}$ samples.

- We only keep the subsamples for which $\rho_j \geq \rho_{MC}$ (this is true for the $t\bar{t}$ signal as well as the background). If a subsample meets the ρ requirements we add it to a larger sample for which

$$\rho_{total} = \frac{\sum_{j=1}^{N} \rho_j}{N}. \tag{F.1}$$

  By only selecting subsamples with a specific correlation we are able to create a sample with a correlation coefficient similar to the one observed in the real MC.

- The process is repeated until we have 50,000 events for each of the signal and background mass subsamples. We point out that we do not really need that

many events to populate the 2-D histograms. 10,000 or fewer events would be sufficient when dealing with the real Monte Carlo.

The "fake" Monte Carlo mimics the shape of the real Monte Carlo quite well as we can see by overlaying the projections from our "fake" 2-D templates to the actual results from the 1-D functional forms. Figure F.1 shows an example from the Single SVX subsample. Therefore, we think that fitting these templates with a Neural Network will tell us whether or not this method can be used on the real Monte Carlo.

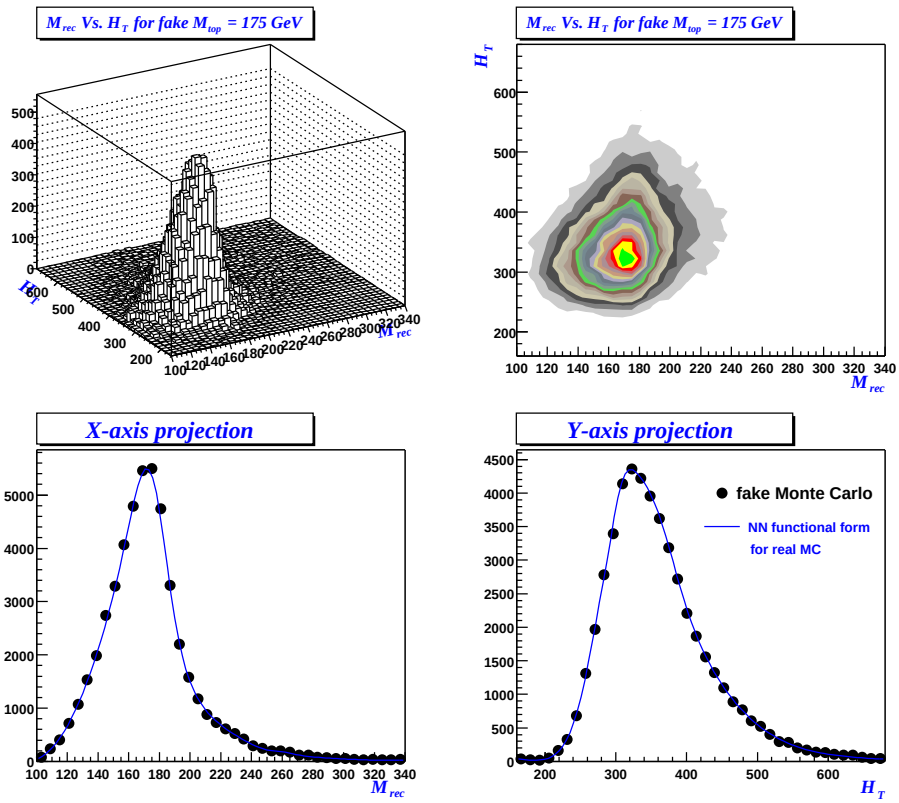

Figure F.1: (a) Lego histogram of our "fake" MC distribution for $M_{top} = 175$ GeV/c^2. (b) Contour plot of the same distribution. (c) Projection along the M_{rec} axis of our 2-D plot. The blue curve corresponds to the 1-D NN fit of the true MC for a top mass of 175 GeV/c^2. (d) Projection of the 2-D plot along the H_T axis. The blue curve corresponds to the 1-D NN fit of the true MC for a top mass of 175 GeV/c^2.

BIBLIOGRAPHY

[1] D. Griffiths, *Introduction to Elementary Particles*, Harper and Row Publishers, (1987); F. Halzen and A. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, Jhon Wiley and Sons, (1984); V. Barger and R. Phillips, *Collider Physics*, Addison Wesley Publishing Company, (1987).

[2] M. Cacciari, *Symposium of Twenty Beautiful Years of Bottom Physics-b20*, AIP, Woodfury, NY (1998); S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi, *Heavy Flavors II*, Word Scientific, River Edge, NJ (1997).

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

[4] R. Bonciani, S. Catani, M. L. Mangano, and P. Nason, Report No. CERN-TH/98-31 (1998).

[5] A. D. Martin, R. G. Roberts, and W. J. Stirling, Phys. Lett. B **387** 419, (1996).

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

[7] A. D. Martin, W. J. Stirling, and R. G. Roberts, Phys. Lett. B **306**, 145 (1993).

[8] E. L. Berger and H. Contopanagos, Phys. Rev. D **57**, 253 (1998).

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

[10] M. Kobayashi and T. Meskawa, Prog. Theor. Phys. **49**, 652 (1973).

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

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

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

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

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

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

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

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

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

[20] B. Kröse, P. van der Smagt *An Introduction to Neural Networks*, University of Amsterdam (1996).

[21] M. Minsky and S. Papert *Perceptrons: An Introduction to Computational Geometry*, MIT Press (1969).

[22] S. Haykin *Neural Networks: A Comprehensive Foundation*, 2nd edition, Prentice Hall (1999).

[23] H. White, A. R. Gallant, K. Hornik, M. Stinchcombe and J. Wooldrige *Artificial Neural Networks: Approximation and Learning Theory*, Blackwell (1992).

[24] F. Abe *et al.,* Phys. Rev. D **50**, 2966 (1994); Phys. Rev. Lett. **73**, 225 (1994).

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

[26] F. Abe *et al.,* Phys. Rev. Lett. **80**, 2779 (1998).

[27] F. Abe *et al.,* Phys. Rev. D **52**, 4784 (1995); F. Abe *et al.,* Phys. Rev. Lett. **75**, 11 (1995).

[28] The isolation, I, is defined as the E_T (P_T) within a cone of $\Delta R < 0.4$ centered on the lepton, but excluding the lepton energy, divided by the E_T (P_T) of the lepton.

[29] F. Abe *et al.,* Phys. Rev. D **45**, 1448 (1992).

[30] D. Kestenbaum, Ph.D. Thesis (unpublished), Harvard University (1996)

[31] G. Marchesini and B.R. Webber, Nucl. Phys. **B310**, 461 (1988); G. Marchesini *et al.,* Comput. Phys. Comm. **67**, 465 (1992).

[32] T. Sjöstrand, Computer Physics Commun. **39**, 347 (1986) ; T. Sjöstrand and M. Bengtsson, Computer Physics Commun. **43**, 367 (1987); T. Sjöstrand and M. Bengtsson, Computer Physics Commun. **46**, 43 (1987).

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

[34] P. Avery, K. Read, G. Trahern, Cornell Internal Note CSN-212, March 25, 1985 (unpublished). We use Version 9_1 with a b lifetime of 1.45 ps.

[35] F.A. Berends, W.T. Giele, H. Kuijf and B. Tausk, Nucl. Phys. **B357**, 32 (1991).

[36] F. Abe *et al.,* Phys. Rev. Lett. **79**, 4760 (1997).

[37] F. Abe *et al.,* Phys. Rev. D **47**, 4857 (1993).

[38] MINUIT Reference Manual, CERN Program Library Entry **D505** (1994).

[39] L. Orr, T. Stelzer, and W. J. Stirling, Phys. Rev. D **52**, 124 (1995); Phys Lett. B **354**, 442 (1995).

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

[41] B. Tuchming, ALEPH Note 99-073: *Two dimensional fits of Higgs mass shapes as a function of Higgs mass with Neural Networks*, (1999).

[42] J. Schwindling and B Mansoulie, MLPfit Neural Network, see:
`http:/schwind.home.cern.ch/schwind/MLPfit.html`

[43] A clear description of the VECBOS files generated for the Run I mass analysis can be found at:
`http://www-cdf.fnal.gov/internal/physics/top/w_mc_samples.txt`

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

[45] C. Ciobanu, R. Hughes, P. Koehn, C. Neu, and B. Winer, *A ROOT script interface to JETNET*, CDF Note 5434 (2000). The Root_Jetnet package can be downloaded from:
`http://cpdfpc2.mps.ohio-state.edu/root_to_jetnet/rtj.html`

[46] L. Demortier, L. Lyons, *Everything you always wanted to know about pulls*, CDF Note 5776 (2002).

[47] A. Beretvas, M.Binkley, *Overview of the kinematics of $t\bar{t}$ events*, CDF Note 4232 (1996).

[48] R. Fletcher, *Practical methods of optimization*, second edition, Wiley (1987).

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

[50] M. D. Richard, R. P. Lippmann, Neural Computation 3 (1991).

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

[52] R. J. Barlow *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, Wiley (1995).

[53] W. Eadie, D. Drijard, F. James, M. Roos, B. Sadoulet *Statistical Methods in Experimental Science*, North-Holland, Amsterdam (1971).

[54] Luc Dermortier, private communication.

[55] N. Eddy, Ph. D. Thesis (unpublished), University of Michigan (1998).

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

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