

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Search for t -Channel Single Top Quark Production in $p\bar{p}$ Collisions
at 1.96 TeV

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Philip Michael Perea

June 2006

Dissertation Committee:

Dr. Ann Heinson, Co-Chairperson

Dr. John Ellison, Co-Chairperson

Dr. Robert Clare

Copyright by
Philip Michael Perea
2006

The Dissertation of Philip Michael Perea is approved:

Committee Co-Chairperson

Committee Co-Chairperson

University of California, Riverside

Acknowledgements

I've waited so long for this moment... finally getting to write the acknowledgements. The first thing to be read, and the last thing to be written (after 6 years!). Well let's just get to it then. This thesis goes out to the many people along the way, who helped me, guided me, and were there to just help me blow off some steam when my brain was about to explode. This is to late nights in the DØ trailer and smashing big purple bouncy balls to relieve the stress. It's to making amazing friends from around the world, and learning about their cultures as much as cultivating our knowledge of physics. To all of those that helped me stay awake on the sometimes grueling owl shifts. To ending an eve shift and finding a party on the grass in the village - I shouldn't be surprised though because I know these people too well.

This thesis was a life-changing experience. I fulfilled a childhood dream of doing “big physics” and got to work side-by-side with many great minds. Thanks to the professors who gave me my foundation during my first two years at UCR. Thanks to John Ellison for taking me under his wing for a period and introducing me to the plethora of cryptic HEP-speak used in so many papers. Thanks to Ia Iashvili for mentoring me as I moved to Fermilab and began my first studies on the Jet Energy Scale. For her endless patience in explaining the physics and tools necessary to do the job. Thanks to Eric Kajfasz, Petros Rapidis, and Leonard Christofek for teaching me how the SMT worked, and for the interesting nights working inside of the detector. Marumi, I still owe you one for helping out a struggling and nervous graduate student give his first big talk at the Beaune conference. And the day to day could have been much more plain if I wasn't surrounded by such a great bunch of characters. Finding the track ball in my mouse missing, CAUTION tape engulfing my desk, or having to be on the lookout for unidentified flying objects coming over the cubicle and wacking me on the head. Thanks Chris, Tim, Phil, Kyle, Rick, and Raymond for keeping the office interesting. And to

the entire English Posse, thanks for inducting me into the fine tradition of tea time - pulling me away from my keyboard just long enough to chill the mind, and bringing a little social atmosphere to the trailer.

The Single Top Group. We put the group together, we poured our heart and souls into the group and have come amazingly far. The day to day sometimes felt a little overwhelming, but looking back on the effort, we have accomplished a lot. Thank you everyone for helping me with the fundamentals of an analysis, helping me to understand the single top theory, helping me to become a better speaker and to give better presentations. Reinhard, you had a big hand in the success of my analysis. Thank you for the constant guidance, patience, explanations, and always challenging me along the way. I'll admit sometimes you drove me crazy, but I wouldn't have changed a thing. Shabnam and Thomas, it was fun working with you guys in the DØ code mines and building up the analysis from scratch. I really enjoyed working with you both. Len, thanks for helping me wrap my head around those neural networks - they were one of the coolest tools I got to use in this analysis and you helped me to rip open the "black box" and understand the inner workings. Nice.

This thesis was accomplished by hard-work as much as it was by the necessary breaks inbetween. Much love to the Wicker Park posse for the many nights out (some of which are still a little fuzzy...). Mac's, Cans, 1056, Rainbo - you know what I'm talking about. To the Presidential Towers posse, ALL my fermi village people, and everyone who came out to visit me in Chi - especially my wonderful sister Stephanie. And I can't forget about all my friends on the other side of the ring at CDF. To all the good times!

And now that I've kept you in suspense Ann, I have a few words to say. I've heard all kinds of nightmare stories about advisors, and have to say I was lucky enough to avoid them all. You were great to me. Thanks for the fruitful singletop discussions before I moved to Fermilab, and thanks for setting me up with wonderful people to help me along the way. Thanks for all the guidance and quick responses despite being 2,000

miles away. You gave me the freedom to explore my ideas in the analysis, and still kept me on track. Thanks for all the useful comments on the many talks I have given, and helping me to become a more proficient presenter. Writing up has been stressful at times, but you worked with me through the entire process and I appreciate it. Thanks for everything Ann.

Mom, thanks for talking me through the crazy times, when I was so fed up that I thought I would quit. Mom and Dad, thanks for everything. And I mean everything. Especially relevant to this thesis may be you guys taking that crazy little boy of yours to some funny museum at “Los Alamos” because he read about it in a book. Now you see the result.

Finally, thanks to a friend that has always made me smile and enjoy my time at Fermilab even more. A kooky someone who livened up the office whether she was giggling from her cubicle or just rushing down the hall like a crazed lunatic with the force of a 1,000 things that all had to be done that second. I’m so glad you made that final trip back to the lab. Thanks Amber, for listening, letting me vent, making me laugh, making me smile, giving me strength, and just keeping me sane in the final stretch of this Ph.D.

*Dedicated to Mom and Dad
for years of love, support, and always
encouraging their little mijo and his
curiosity of the world around him*

ABSTRACT OF THE DISSERTATION

Search for t -Channel Single Top Quark Production in $p\bar{p}$ Collisions at 1.96 TeV

by

Philip Michael Perea

Doctor of Philosophy, Graduate Program in Physics

University of California, Riverside, June 2006

Dr. Ann Heinson, Co-Chairperson

Dr. John Ellison, Co-Chairperson

I have performed a search for t -channel single top quark production in $p\bar{p}$ collisions at 1.96 TeV on a 366 pb^{-1} dataset collected with the DØ detector from 2002–2005. The analysis is restricted to the leptonic decay of the W boson from the top quark to an electron or muon, $tq\bar{b} \rightarrow l\nu_l b q\bar{b}$ ($l = e, \mu$). A powerful b -quark tagging algorithm derived from neural networks is used to identify b jets and significantly reduce background. I further use neural networks to discriminate signal from background, and apply a binned likelihood calculation to the neural network output distributions to derive the final limits. No direct observation of single top quark production has been made, and I report expected/measured 95% confidence level limits of 3.5/8.0 pb.

Contents

1	What is Single Top and Why is it Interesting?	1
1.1	Introduction	1
1.2	Standard Model	3
1.2.1	The Particle Zoo	3
1.2.2	The Forces	4
1.2.3	Gauge Theories	5
1.3	The Top Quark	7
1.3.1	Discovery	7
1.3.2	Properties	7
1.3.3	Pair Production	8
1.4	Single Top	10
1.4.1	Introduction	10
1.4.2	What Can Single Top Teach Us?	11
1.4.2.1	Standard Model Electroweak Coupling	11
1.4.2.2	Non-Standard Model Electroweak Coupling	14
1.4.2.3	New Fields, Mediators and Particles	16
1.4.3	Single Top Quark Modeling and Calculation	18
1.4.4	Decay Modes	19
1.4.5	Current State of Affairs	23
1.5	Analysis Overview	24
2	Creating Single Top Quarks	26
2.1	Introduction	26
2.2	Theory	26
2.3	Acceleration Chain	28
2.3.1	Proton Production	28
2.3.2	Antiproton Production	30
2.4	Luminosity	31
2.5	The DØ Detector	33
2.5.1	Coordinates	33
2.5.2	Central Tracking Systems	34
2.5.3	Silicon Microstrip Tracker	34
2.5.4	Central Fiber Tracker	35
2.5.5	Preshower Detector	37

2.5.6	Calorimeter	38
2.5.7	InterCryostat Detector and Massless Gaps	41
2.5.8	Muon System	41
2.6	Data Acquisition	44
3	Finding Single Top Quarks	47
3.1	Introduction	47
3.2	Data Samples	48
3.2.1	$e+jets$	48
3.2.2	$\mu+jets$	49
3.2.3	$e + e$ (DIEM)	49
3.2.4	$\mu + \mu$ (DIMU)	51
3.3	Object Reconstruction	54
3.3.1	Particle Tracks	54
3.3.2	Primary Vertices	56
3.3.3	Muons	57
3.3.4	Electromagnetic Clusters	58
3.3.5	Jets	58
3.3.6	b -Quark Jets	59
3.3.7	Neutrinos	61
3.4	Object Identification	61
3.4.1	Primary Vertices	62
3.4.2	Muons	62
3.4.3	Electromagnetic Clusters	63
3.4.4	Jets	66
3.4.5	b -Quark Jets	70
3.4.6	Neutrinos	73
3.5	Event Selection	74
4	Modeling Single Top Quark Backgrounds	79
4.1	Introduction	79
4.2	Masters of Disguise	80
4.3	Monte Carlo Simulation	81
4.3.1	Correction Factors	83
4.3.1.1	Primary Vertex Correction Factors	83
4.3.1.2	Electron Correction Factor	83
4.3.1.3	Muon Correction Factor	84
4.3.1.4	Jet Corrections	84
4.3.1.5	Trigger Efficiency	85
4.3.1.6	Tag-Rate Functions	88
4.3.2	Normalization	93
4.4	Monte Carlo Samples	94
4.4.1	Single Top Signal Samples	94
4.4.2	Top Pair Production Samples	94

4.4.3	W +jets Samples	95
4.4.4	Diboson Samples	95
4.5	Data Samples	95
4.6	Background Fraction Estimation – The Matrix Method	96
4.6.1	Electron Calculation	98
4.6.2	Muon Calculation	101
4.7	Certifying the Model	103
4.7.1	Acceptances	103
4.7.2	Yields	105
4.7.3	Distributions	107
4.7.4	Systematic Uncertainties	107
4.7.4.1	Integrated Luminosity	107
4.7.4.2	Cross Sections	110
4.7.4.3	Branching Fractions	110
4.7.4.4	Triggers	111
4.7.4.5	Primary Vertex Identification	111
4.7.4.6	Electron Identification	111
4.7.4.7	Muon Identification	112
4.7.4.8	Jet Fragmentation	112
4.7.4.9	Jet Reconstruction and Identification	112
4.7.4.10	Jet Energy Resolution	112
4.7.4.11	Jet Energy Scale	114
4.7.4.12	Flavor-Dependant Tag Rate Functions	114
4.7.4.13	Matrix Method	114
4.7.4.14	Final Uncertainties	117
5	Isolating Single Top Quarks From Background	120
5.1	Introduction	120
5.2	Building A Brain	121
5.3	Learning (In Theory)	124
5.4	Learning (In Practice)	125
5.5	Our Neural Network Approach	126
5.5.1	Variable Optimization	129
5.6	Discriminating Variables	131
5.6.1	Final State Object Reconstruction	132
5.6.2	b -Tagged Jet Variables	132
5.6.3	Variables	133
5.7	Optimized Neural Networks	142
5.8	Optimized Super Neural Network	159
6	Setting Limits on Single Top Quark Production	163
6.1	Introduction	163
6.2	Bayesian Approach to Limits	164
6.3	Cross Section Limits	167

7	Summary	174
8	What's to Come	175
8.1	Observation	175
8.2	A New Window — The Large Hadron Collider	176
A	Physics Object Distributions	187
B	Calculation of Reconstructed Physics Objects	204
C	Decision Trees	206

List of Tables

1.1	Fundamental Particles	4
1.2	Fundamental Forces	5
1.3	Top Quark Pair Production Decay Modes	9
1.4	Top Quark Mass and Cross Section Measurements	9
1.5	Single Top Quark Production Cross Sections	19
1.6	Single Top Quark Production Published Limits	24
2.1	Silicon Microstrip Tracker Module Specifications	36
2.2	Calorimeter Radiative and Absorptive Lengths For Particle Traversal . .	40
3.1	Data Samples and Triggerlist Versions	52
3.2	L1, L2, and L3 Trigger Conditions	53
3.3	Primary Vertex Track Requirements	62
3.4	Secondary Vertex Requirements	72
3.5	Neural Network Tagger Operating Points	73
3.6	Data Sample Sizes	77
4.1	Average Values for Trigger Weights	88
4.2	Neural Network Tagger Mistagging Efficiencies	93
4.3	Monte Carlo Sample Details	96
4.4	Signal Acceptances	105
4.5	Sample Yields for the Electron Channel	106
4.6	Sample Yields for the Muon Channel	106
4.7	Sample Yields for the Electron Channel	107
4.8	Sample Yields for the Muon Channel	110
4.9	Systematic Uncertainties for the Trigger Efficiencies	111
4.10	Systematic Uncertainties for the Jet Identification	113
4.11	Systematic Uncertainties for the Jet Energy Resolution	113
4.12	Systematic Uncertainties for the Jet Energy Scale	114
4.13	Systematic Uncertainties for the Tag Rate Functions	115
4.14	Systematic Uncertainties for the Electron Single Tag Channel	118
4.15	Systematic Uncertainties for the Electron Double Tag Channel	118
4.16	Systematic Uncertainties for the Muon Single Tag Channel	119
4.17	Systematic Uncertainties for the Muon Double Tag Channel	119

5.1	Individual Background Model Fractions	127
5.2	YIELD and TRAIN Sample Sizes	128
5.3	Decision Tree Ranking for Wbb Network Variables	145
5.4	Decision Tree Ranking for Wjj Network Variables	146
5.5	Decision Tree Ranking for $t\bar{t} \rightarrow \ell + jets$ Network Variables	147
5.6	Decision Tree Ranking for $t\bar{t} \rightarrow \ell\ell$ Network Variables	148
5.7	Decision Tree Ranking for <i>multijet</i> Network Variables	149
5.8	Final Neural Network Input Variables	152
5.9	Final Single Tag Network Variables	153
5.10	Final Double Tag Network Variables	154
5.11	Neural Network Optimized Training Parameters (Electron Channel) . . .	155
5.12	Neural Network Optimized Training Parameters (Muon Channel) . . .	155
5.13	Super Neural Network Optimized Parameters	160
5.14	Super Neural Network Input Variables	160
6.1	Count-Based Cross Section Limits	171
6.2	Expected Individual Background Neural Network Limits	171
6.3	Final Shape-Based Cross Section Limits	171
6.4	Super Neural Network Input Optimization Trials	173

List of Figures

1.1	Top Quark Pair Production Feynman Diagrams	9
1.2	Single Top Quark Production LO Feynman Diagrams	11
1.3	Top Quark Decay and Polarization Scenarios	14
1.4	Higgs Production Feynman Diagrams	17
1.5	Flavor Changing Neutral Current Feynman Diagrams	17
1.6	t -channel Feynman Diagrams	20
1.7	s -channel Feynman Diagrams	20
1.8	W -gluon Fusion Interference Diagrams	21
1.9	SingleTop Monte Carlo b Quark Distributions	22
1.10	SingleTop Monte Carlo Parton Level Distributions	23
2.1	CTEQ6M and MRST2001 Parton Distribution Functions	27
2.2	Fermilab Acceleration Chain	28
2.3	Cross Section of the DØ Detector	33
2.4	Cross Section of the DØ Tracking Systems	35
2.5	Silicon Microstrip Tracker (SMT)	36
2.6	DØ Calorimeter	39
2.7	A Quarter Cross Section of the DØ Calorimeter	39
2.8	Muon Scintillation System	43
2.9	Muon Drift Chamber System	44
2.10	Trigger System	45
3.1	Impact Parameter Diagram	61
3.2	Evolution of a Jet	68
3.3	Jet Energy Scale Correction Factors	69
3.4	Neural Network Tagger Output	73
3.5	“Triangle Cuts” ($\Delta R(\text{object}, \cancel{E}_T)$ vs \cancel{E}_T)	78
4.1	Single Top Background Feynman Diagrams	81
4.2	Monte Carlo Simulation Chain	82
4.3	Level 3 Trigger Turn-On Curves	85
4.4	Trigger Weight Distributions	88
4.5	Jet Taggability Functions	90
4.6	b -quark Tag Rate Functions	91
4.7	c -quark Tag Rate Functions	92

4.8	lightjet Mis-Tag Rate Functions	92
4.9	Real and Fake Electron Components of the W Trasnverse Mass	100
4.10	Real and Fake Muon Components of the W Trasnverse Mass	102
4.11	Data and Background Model Comparison Before Tagging	104
4.12	Data and Background Model Comparison in the Single Tag Channels . .	108
4.13	Data and Background Model Comparison in the Double Tag Channels . .	109
5.1	A Biological and an Artifical Neuron	122
5.2	A Simple Feed-Forward Neural Network	123
5.3	Illustration of Sigmoid Properties	124
5.4	Neural Network Testing and Training Curves	126
5.5	Sample Splitting Diagram	128
5.6	Neural Network Figure of Merit	129
5.7	Neural Network Single Tag Input Variable Distributions - I	137
5.8	Neural Network Single Tag Input Variable Distributions - II	138
5.9	Neural Network Single Tag Input Variable Distributions - III	139
5.10	Neural Network Double Tag Input Variable Distributions - I	140
5.11	Neural Network Double Tag Input Variable Distributions - II	141
5.12	Neural Network Double Tag Input Variable Distributions - III	142
5.13	Figure of Merit Optimization Plot	150
5.14	Final Figure of Merit Plots	151
5.15	Neural Network Output Signal versus Background Acceptance Plots . .	156
5.16	Neural Network Output Plots	157
5.17	Neural Network Output - Background Model and Data	158
5.18	Super Neural Network Output	161
5.19	Super Neural Network - Background Model and Data	162
6.1	Cross Section Posterior Probability Density Distributions	172
6.2	Expected Limit versus Super Neural Network Cut Point	173
8.1	Cross Section Significance versus Integrated Luminosity	177
A.1	Electron Channel Object Distributions - I	188
A.2	Electron Channel Object Distributions - II	189
A.3	Electron Channel Object Distributions - III	190
A.4	Electron Channel Object Distributions - IV	191
A.5	Electron Channel Cross Check Sample Distributions - I	192
A.6	Electron Channel Cross Check Sample Distributions - II	193
A.7	Electron Channel Cross Check Sample Distributions - III	194
A.8	Electron Channel Cross Check Sample Distributions - IV	195
A.9	Muon Channel Object Distributions - I	196
A.10	Muon Channel Object Distributions - II	197
A.11	Muon Channel Object Distributions - III	198
A.12	Muon Channel Object Distributions - IV	199
A.13	Muon Channel Cross Check Sample Distributions - I	200

A.14 Muon Channel Cross Check Sample Distributions - II	201
A.15 Muon Channel Cross Check Sample Distributions - III	202
A.16 Muon Channel Cross Check Sample Distributions - IV	203
C.1 Decision Tree Diagram	207

Chapter 1

What is Single Top and Why is it Interesting?

1.1 Introduction

A lone neon sign, glowing and flickering, beckons all with its incentive of “Ice Cold Beer.” And while considering the tempting offer, I ponder how odd a place for the remnants of man’s first particle accelerator to reside. These tubes, once called Geissler tubes after their inventor, Heinrich Geissler, consisted of a partially evacuated glass tube, neon gas, and an electrode at either end. With the application of a voltage across the electrodes, a current of electrons was accelerated from one end of the tube to the other, and in the process radiated that soft glow that calls us to our neighborhood bar. From these humble beginnings more rooted in entertainment and dazzling illumination, accelerators were designed (this time intentionally) to probe the very fabric of the universe. While man has always pondered the fundamental nature of matter, only in the last 150 years has technology evolved to the challenge - bringing forth the birth of particle physics.

Even with those same early tubes that Geissler had come up with, the first fundamental unit of matter had been discovered - the electron [1]. Today, such tubes are

but a speck in comparison to modern day accelerators and particle detectors. The field has grown immensely, and with each new energy regime provided by an accelerator, a new batch of particles has been discovered. Still, we toy with the same questions that have always intrigued us — what, how and why? What is the universe really made of, how do its fundamental units interact, and how does this influence the evolution of the universe? (The “Why’s” are usually more philosophical, and tend to evade scientific evaluation, while still providing engaging conversation topics). Ceaselessly seeking these answers, a large variety of experiments have sprung up around the world, each working day and night to collect enough data to prove or disprove our most imaginative theories. They are being run by international collaborations, and are working with one another to develop and test the current theory of particle physics - the Standard Model. The latter half of the 20th century witnessed the creation of these super experiments, and notable ones that made significant discoveries include the Stanford Linear Accelerator (SLAC) [2, 3], the Brookhaven Alternating Gradient Synchrotron (AGS) [4, 5, 6], and the European Center for Nuclear Research’s Super Proton Synchrotron (SPS) [7, 8], and later its Large Electron-Positron collider (LEP) [9, 10, 11]. Each of these experiments has yielded a new piece of the puzzle, and the current piece is being forged by the world’s leading high energy physics collider: the Tevatron at Fermi National Accelerator Laboratory (Fermilab).

In this dissertation, a description will be given of an analysis at the DØ experiment at Fermilab. A full description of a search for single top quarks has been completed with a dataset spanning 2002–2005. We begin with the fundamental theory that has led us to take on this endeavor, how and why a search for single top quarks should be pursued, and then give a detailed description of the analysis in the following chapters.

1.2 Standard Model

The Standard Model of particle physics describes our current understanding of matter and provides a mathematical framework for particle interactions, creation, decay, and annihilation. It is the work of many scientists spanning approximately the last 100 years. From J.J. Thompson’s discovery of the electron (1897), to the accelerator “particle explosion” (1950’s), and evidence for proton structure (1960’s), experimentalists have uncovered the fundamental building blocks of the universe. Working in parallel, many theorists have guided this quest for understanding, accurately describing equations for antiparticles, creating gauge theories such as the beautifully accurate theory of charged interactions (Quantum Electrodynamics), and predicting the massive mediators of the weak force. From the work of many, a brief review of the Standard Model is given here.

1.2.1 The Particle Zoo

All matter consists of three groups of particles: quarks, leptons, and force carriers, and their antiparticles. There are six quarks, named up, down, charm, strange, top, and bottom, arranged in three generations of weak isospin doublets, (u,d) , (c,s) , (t,b) . They are spin 1/2 fermions and the up-type quarks carry a charge of $2/3e$ and the down-type quarks a charge of $-1/3e$. The mass varies greatly from about 4 MeV for an up quark to 175 GeV for a top quark (Table 1.1). The leptons — electrons, muons, taus, and their neutrinos — are also arranged in three generations of doublets, (e, ν_e) , (μ, ν_μ) , (τ, ν_τ) . They are spin 1/2 fermions and each electron-type lepton carries a charge of $1e$ and is partnered with a chargeless neutrino of corresponding flavor. Neutrinos are nearly massless¹ and the other leptons range from 0.5 MeV to 1.8 GeV.

¹Recent evidence from neutrino oscillation measurements show that at least two of the neutrinos must have mass to account for observed neutrino mixing [12, 13]

Generation	Flavor	Charge	Spin	Colors	Mass
1	u	$+2/3$	$1/2$	Yes(3)	few MeV
	d	$-1/3$	$1/2$	Yes(3)	few MeV $m_d > m_u$
	e^-	-1	$1/2$	No	0.511 MeV
	ν_e	0	$1/2$	No	nearly massless
2	c	$+2/3$	$1/2$	Yes(3)	~ 1 GeV
	s	$-1/3$	$1/2$	Yes(3)	~ 0.1 GeV
	μ^-	-1	$1/2$	No	0.1057 GeV
	ν_μ	0	$1/2$	No	nearly massless
3	t	$+2/3$	$1/2$	Yes(3)	~ 175 GeV
	b	$-1/3$	$1/2$	Yes(3)	~ 4 GeV
	τ^-	-1	$1/2$	No	1.776 GeV
	ν_τ	0	$1/2$	No	nearly massless

Table 1.1: The fundamental particles and their basic properties (This table is a replication from [14], with updated Particle Data Group values [48]).

1.2.2 The Forces

The Standard Model further describes four forces which allow our zoo of particles to interact and evolve into the complex universe that we observe today (Table 1.2). The carriers of these forces are spin 1 bosons. The first force, gravity, is the familiar attractive force between all particles with mass or energy. It is mediated by the graviton, and while dominating at large distances, it is dwarfed by the other forces at short or nuclear ranges. The electromagnetic force occurs between all particles with electric charge, and is mediated by the massless photon. The strong force is mediated by gluons and allows color-charged particles to interact with one another. At short range it is the dominant force and is responsible for holding together atomic nuclei. Finally, the weak force is mediated by the charged W^\pm bosons and the chargeless Z^0 boson. The weak force allows interactions amongst all the particles except gluons. It is short ranged, but allows for many otherwise forbidden decays to occur.

Force	Mediator	Interact By	Charge	Spin	Mass
gravity	graviton	mass	0	1	Massless
electromagnetic	γ (photon)	charge	0	1	Massless
weak	W^\pm	—	± 1	1	~ 80 GeV
weak	Z^0	—	0	1	~ 91 GeV
strong	g (gluon)	color	0	1	Massless

Table 1.2: The force mediators and their properties.

1.2.3 Gauge Theories

Gauge theories describe physical systems based on the idea that there is a symmetry governing their actions, and that transformations by this symmetry can be performed both locally and globally. These symmetries, such as charge, parity, or time, in combination with quantization of a theory, have led to the development of quantum field theories.

In the late 1940's the first field theory, Quantum Electrodynamics (QED), was being developed by Richard P. Feynman, Julian S. Schwinger, and Tomonaga Shin'ichiro. One of the most elegant and precise theories to date (correctly predicting the anomalous magnetic moment of the electron to more than 10 decimal places [15]), it describes the interactions of all particles with electric charge. With this triumph, the theory was further extended by Sheldon Glashow, Abdus Salam, and Steven Weinberg in the 1960's and the electromagnetic and weak forces were united into one force, the electroweak force.

The electroweak gauge theory is described by a $SU(2)_L \otimes U(1)_Y$ group²; it predicted four force mediators, two charged, and two neutral. With the weak force being experimentally short ranged, it became necessary for some of these mediators to gain mass. This process, known as “spontaneous symmetry breaking,” correctly predicted the massive W^\pm and Z^0 bosons, kept the massless photon, and retained normalizability of the

² $SU(2)_L$ refers to left-handed doublets, while $U(1)_Y$ refers to the weak hypercharge, $Y^{weak} = 2(Q - I_3^{weak})$, where Q is the electric charge, and I_3^{weak} the weak isospin

theory. With such success, the model for strong interactions was based on the QED and electroweak theories, and the theory of Quantum Chromodynamics (QCD) was created.

QCD aims to describe the interactions of quarks and to explain the observed structure of all the mesons and baryons. As the electric charge is the fundamental parameter of QED, a “color” charge was created for QCD. This charge comes in three flavors, red (r), green (g), and blue (b), and their anticolors. The theory requires all unbound strongly interacting particles to exist in colorless states, which is possible only by combining all colors, rbg , all anticolors, \overline{rbg} , or a color and its anti-color, *e.g.* $g\bar{g}$. Quarks possess a single color, and the theory requires eight gluons, each possessing a color and a differing anticolor. These guidelines explain why all mesons are made of two quarks, like the $\pi^+(u_b\bar{d}_{\bar{b}})$, and all baryons are made up of three quarks, like the proton $(u_ru_gd_b)^3$.

QCD also consists of two peculiar properties: asymptotic freedom and confinement. Both deal with the nature of the strong force at varying distances. At short, or nuclear distances, the interaction between quarks and gluons diminishes and they behave as free, noninteracting particles. This feature has allowed high energy physics to flourish, yielding accurate models of hadrons and thus, accurate predictions of scattering cross sections. While quarks are free at short distances, the strength of the strong force increases with distance indefinitely. Only an infinite amount of energy would allow quarks to roam free, so they remain bound. If there is enough potential energy in the field between two quarks, another pair of quarks may appear to relieve the tension, creating two separate, colorless hadrons.

The group that describes QCD is $SU(3)_C$ and together with the electroweak theory, the Standard Model is described by the group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. A final note should be made concerning the origin of mass. The same mechanism that gave the electroweak mediators their mass, spontaneous symmetry breaking, requires a new field

³evidence for “pentaquarks” has been recently submitted, a bound state of five quarks, for example $u_ru_gd_b\bar{d}_{\bar{b}}\bar{s}_{\bar{b}}$, that could still feasibly be colorless [16, 17]

to account for the process, and this field is the Higgs field. The boson associated with this field not only gives mass to the mediators, but to all fermions. It is an essential piece of the Standard Model, and is being actively sought today.

1.3 The Top Quark

1.3.1 Discovery

The first evidence for a third generation of quarks came with the discovery⁴ of the b quark in 1977. In order to fit into the current theory of particle physics, the down-type “bottom” or “beauty” quark was expected to have an up-type isospin partner, and the “top” or “truth” quark was postulated to exist. Nearly 20 years later, a particle accelerator finally generated enough center of mass energy to produce the top quark, and it was observed at Fermilab by the DØ and CDF collaborations in 1995 [19, 20].

1.3.2 Properties

The top quark exhibits all the natural properties of an up-type quark, being a spin 1/2 fermion, having a charge of $+2/3e$, and interacting via the strong and weak forces. It also exhibits several unusual properties, the most notable being its large mass: 178.0 ± 4.3 GeV/ c^2 [25]. The top quark is much more massive than its isospin partner or any other mediator or quark discovered thus far. This mass is close to the electroweak symmetry breaking scale (vacuum expectation value = 246 GeV), and makes the top quark a good place to look for deviations from the Standard Model [80]. This large mass further leads to an expected decay width $\Gamma_t/|V_{tb}|^2 = 1.39$ GeV, or a lifetime $\sim 0.4 \times 10^{-24}$ s [23, 24]. Considering strong interactions take place on the order of $\sim 3.0 \times 10^{-24}$ s ($1/\Lambda_{QCD}$), the top quark decays before hadronization, and its polarization is cleanly transmitted

⁴via the $\Upsilon(b\bar{b})$ decay seen by E288 at Fermilab [18]

to its decay products. As data samples grow, top quark polarization can be tested for anomalous couplings. The top quark couples to both the strong and weak fields and allows study via two production modes: pair production (Section 1.3.3) and single top quark production (Section 1.4).

1.3.3 Pair Production

The top quark was first observed at Fermilab through pair production via the strong force [19, 20]. At a $p\bar{p}$ collider, both the proton valence quarks, and the gluons and quarks from the parton sea can contribute to the production. At a center-of-mass energy of 1.96 TeV, about 85% of production comes from the quark process $q\bar{q} \rightarrow t\bar{t}$, and 15% from the gluon process $gg \rightarrow t\bar{t}$ (Figure 1.1) [21, 22]. Using the Standard Model with a top quark mass equal to 175 GeV, the NLO (*Next-to-Leading Logarithmic Order*) $t\bar{t}$ cross section has been calculated as 6.77 ± 0.42 pb [22]. With a branching fraction of nearly 100% for $t \rightarrow Wb$, the pair production decay channels are defined by the decay of the W bosons. These channels include $t\bar{t} \rightarrow$ alljets ($\mathcal{B}=36/81$), lepton+jets ($\mathcal{B}=36/81$), and dileptons ($\mathcal{B}=4/81$), where the branching fractions come from all available fermion final states (Table 1.3). Since the heavier τ further decays, it is not considered part of the dilepton channel, and makes up the remaining 5/81 of the branching fraction. The dilepton channel consists of ee +jets, $e\mu$ +jets, or $\mu\mu$ +jets, and suffers from a problem in reconstructing the W bosons. There are two neutrinos in the event, but they can only be reconstructed as one object in the detector, missing transverse energy, thus making reconstruction of both W bosons and a measurement in this channel difficult. The alljets channel consists of two b -quark jets from the top quark decays, and four or more light-quark jets from the decay of the W bosons. This channel is hindered by the combinatorics of correctly identifying the jets from each decayed top quark. The lepton+jets channel is the “golden channel” for $t\bar{t}$ measurements, yielding one clean W

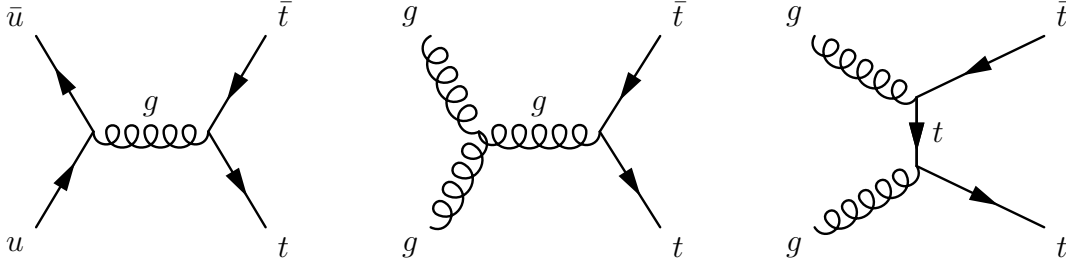


Figure 1.1: The production modes for $t\bar{t}$ pair production. The left figure is mediated by two incoming quarks, while the center and right figures show initial state gluons producing the reaction. The right figure has an additional diagram with the incoming gluons crossed, but it is not shown.

boson from the lepton and \cancel{E}_T , and greatly reducing the combinatorics from the other jets. Cross section measurements and mass measurements have been made in all of these channels, and results are shown in Table 1.4.

Top Pair Production Decay					
Modes of the W boson					
	e^+	μ^+	τ^+	u/d	c/s
e^-	1	1	1	3	3
μ^-	1	1	1	3	3
τ^-	1	1	1	3	3
\bar{u}/\bar{c}	3	3	3	9	9
\bar{d}/\bar{s}	3	3	3	9	9

Table 1.3: The available decay modes of each W boson from the decays of the top quarks in pair production. The number of final states from the W^+ (top row) and the W^- (left most column) decays are given. Factors of three arise from the color charge of the quarks.

Top Quark Precision Measurements	
Top Quark Mass	$M_t = 172.5 \pm 1.3(\text{stat}) \pm 1.9(\text{sys}) \pm 2.3(\text{tot}) \text{ GeV}/c^2$
$t\bar{t}$ Cross Section (DØ Best)	$\sigma(t\bar{t}) = 8.1^{+1.3}_{-1.2}(\text{stat+sys}) \pm 0.6 \text{ (lumi) pb}$
$t\bar{t}$ Cross Section (CDF Best)	$\sigma(t\bar{t}) = 8.2 \pm 0.6(\text{stat}) \pm 1.0(\text{sys}) \text{ pb}$

Table 1.4: A large number of precision top quark measurements are being pursued by the DØ and CDF experiments. The latest world average for the top quark mass is given [25], as well as the most precise $t\bar{t}$ production cross section measurement to date for both DØ [26] and CDF [27]. Both cross section measurements take place in the lepton + jets decay channel.

1.4 Single Top

1.4.1 Introduction

Single top quark production occurs via the weak interaction and provides an exciting new perspective with which to study the top quark. Single top processes have been the subject of many previous studies [34]–[47]. At hadron colliders, there are three production modes that contribute and each is defined by the virtuality (Q_W^2), or four-momentum squared, of the W boson mediating the production. Each process is described below for top quark production, and all conjugate (\bar{t}) processes can be easily deduced. The leading order Feynman diagrams are illustrated in Figure 1.2 (higher order diagrams are shown in Figures 1.6 and 1.7).

- **t - or u -channel ($Q_W^2 < 0$)** The t -channel is mediated by a spacelike W boson and is the dominant production mode at the Tevatron. There are two constructively interfering diagrams that contribute, a $2 \rightarrow 2$ and a $2 \rightarrow 3$ process. The $2 \rightarrow 2$ process involves a u valence quark from the proton and a b quark from the antiproton sea. The $2 \rightarrow 3$ process differs by requiring a gluon from the antiproton sea to pair produce instead of an incoming b quark. Both the $2 \rightarrow 2$ and $2 \rightarrow 3$ processes produce a high momentum light quark and a top quark, but the $2 \rightarrow 3$ process produces an additional soft \bar{b} antiquark. The $2 \rightarrow 3$ process is also responsible for the t -channel’s alias “ W -gluon fusion”, and may also be referred to as tqb .
- **s -channel ($Q_W^2 > 0$)** The s -channel is mediated by a timelike W boson and only includes a $2 \rightarrow 2$ process at leading order. The process involves two incoming quark isospin partners, generally a valence u quark from the proton and a valence \bar{d} quark from the antiproton. The process produces a top quark and a hard \bar{b} antiquark, and may also be referred to as the W^* or tb mode.

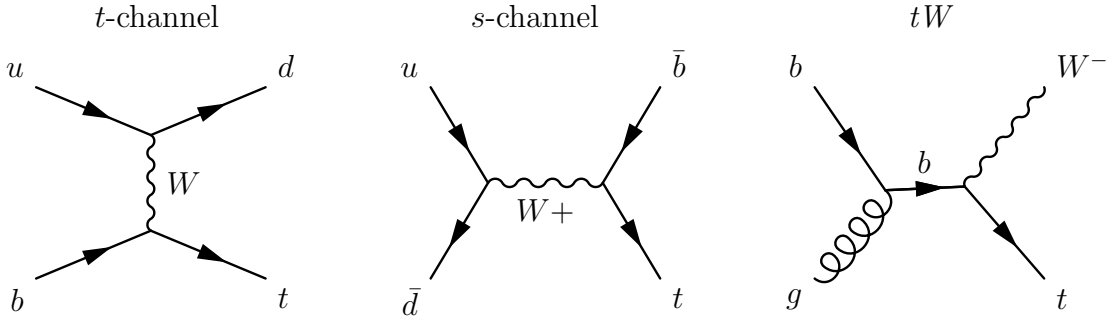


Figure 1.2: Leading order single top quark production modes.

- **tW ($Q_W^2 = 0$)** The tW process involves an on-shell W boson and is the smallest process at our collider. This process involves an incoming b quark from either the proton or antiproton sea, and a gluon from the other sea. The products are a real W boson and a top quark. While this is a unique signature to search for in the detector, the low cross section leaves the tW channel out of reach at this time, and so it is ignored for the duration of this thesis. Further discussion can be found at [64].

1.4.2 What Can Single Top Teach Us?

The wide scope of single top processes allows for verification of the standard model and a chance to possibly observe new physics. The weak coupling of the top quark can be measured, while looking for new fields that couple to the top quark, and studying the top quark polarization. Each of these inquiries can yield interesting results, and each is discussed below in detail.

1.4.2.1 Standard Model Electroweak Coupling

The electroweak coupling is mediated by the W and Z bosons; it allows for weak isospin partners to interact and for quarks and leptons to interact. The coupling is

$V-A$ (vector – axial vector), or left-handed in nature, and a charged current interaction is described by

$$-\frac{g}{2\sqrt{2}}[W_{\mu}^{-}\bar{\Psi}\gamma^{\mu}(1-\gamma^5)\Phi + W_{\mu}^{+}\bar{\Phi}\gamma^{\mu}(1-\gamma^5)\Psi] \quad (1.1)$$

where g is the weak coupling constant ($g = e/\sin\theta_W$), W^{\pm} represent the W boson field, Ψ and Φ the wave functions of the interacting particles, and $\gamma^{\mu}(1-\gamma^5)$ the $V-A$ coupling (there will be an additional constant for quark interactions to account for QCD corrections) [48]. In addition to describing the interactions of weak isospin partners, the electroweak theory describes mixing between up- and down-type quarks via the Cabbibo-Kobayashi-Maskawa (CKM) matrix. In the standard model with three quark generations, the CKM matrix takes on the form

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (1.2)$$

with each up(down)-type quark actually being a linear combination of all three up(down)-type quarks. The $V_{qq'}$ terms represent the strength of mixing from any up-type quark to any down-type quark. If there are only three quark generations, then unitarity of the CKM matrix and current measurements of several elements in the matrix tightly restrict the range of values for the remaining elements, including V_{tb} , as shown in Equation 1.3 [48].

$$\begin{pmatrix} 0.9739 \text{ to } 0.9751 & 0.221 \text{ to } 0.227 & 0.0029 \text{ to } 0.0045 \\ 0.221 \text{ to } 0.227 & 0.9730 \text{ to } 0.9744 & 0.039 \text{ to } 0.044 \\ 0.0048 \text{ to } 0.014 & 0.037 \text{ to } 0.043 & 0.9990 \text{ to } 0.9992 \end{pmatrix} \quad (1.3)$$

If we relax the assumption of three generations, then the values of the remaining ma-

trix elements become virtually unconstrained. In Equation 1.3, the values for the three generation CKM matrix are shown and the value of V_{tb} is relatively fixed, ranging from $0.9990 - 0.9992$. Comparing this range with that allowed by an unconstrained matrix (Equation 1.4), one sees that a measurement of the V_{tb} element could shed light on the total number of quark generations. This is accomplished through a single top cross section measurement, given that the cross section is directly related to the matrix element, which is related to the V_{tb} element: $\sigma \propto |\mathcal{M}|^2 \propto |V_{tb}|^2$.

$$\begin{pmatrix} 0.9730 \text{ to } 0.9746 & 0.2174 \text{ to } 0.2241 & 0.0030 \text{ to } 0.0044 & \dots \\ 0.213 \text{ to } 0.226 & 0.968 \text{ to } 0.975 & 0.039 \text{ to } 0.044 & \dots \\ 0 \text{ to } 0.08 & 0 \text{ to } 0.11 & 0.07 \text{ to } 0.9993 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (1.4)$$

A useful example of possible observable deviations calculated by Tait and Yuan [58] considers changes in the V_{ts} and V_{tb} elements. In the three generation model, the maximum value of V_{ts} (0.043) would contribute $< 1\%$ of the total t -channel single top quark cross section. If we allow this V_{ts} element to be maximized in the range of an unconstrained CKM matrix to 0.11⁵, and the V_{tb} element to saturate the unitarity requirement for any number of quark generations, the standard model t -channel cross section nearly doubles. This dramatic increase is due to the larger parton density of the s quark compared to the b quark. Such an increase would be easily detected in a cross section measurement and possibly point to an extra generation of quarks. In the same circumstances, the s -channel would drop to 70% of its standard model value, and the tW mode would also nearly double.

⁵Tait and Yuan used the 2000 Particle Data Group upper value for the unconstrained CKM matrix, $V_{ts} = 0.55$

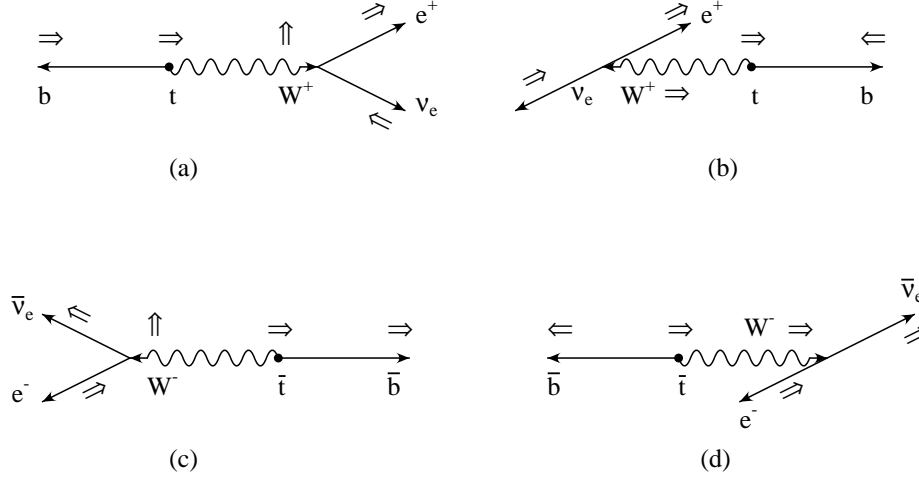


Figure 1.3: The polarization of the top quark and its decay products for various scenarios. The larger arrow above each particle represents the preferred direction of the polarization, and the skinnier arrows represent the preferred direction of the momentum in the top quark rest frame. For the t ((a) and (b)) and \bar{t} ((c) and (d)) quark decay, the case of a longitudinally ((a) and (c)) and a left-handed ((b) and (d)) W boson are pictured. For t (\bar{t}) decays, the e^+ (e^-) prefers to travel along (against) the polarization of the t (\bar{t}) quark. In all cases, the lepton polarization tends to align with the top quark polarization. (Diagram courtesy of [58])

1.4.2.2 Non-Standard Model Electroweak Coupling

While a measurement of single top quark production could allow us to confirm the $V-A$ nature of the electroweak force, studies of single top quark kinematics could allow us to probe a $V+A$ or other nonstandard coupling. This new coupling would most easily be recognized through deviations in the expected top quark polarization and can be easily studied in single top quark production via the final decay products. As discussed, the top quark decays before hadronization and cleanly transmits its polarization to the resulting W boson and b quark. This polarization is again translated in the left-handed decay of a W boson to its decay products, leaving all the necessary information to reconstruct the original polarization. Before discussing the observable effects of any additional couplings, it is important to describe the top quark bases that could give us the power probe such effects: the “helicity basis” and the “optimal basis” [56, 57, 58].

The helicity basis measures the component of the top quark's spin along its axis of motion in the center-of-mass frame. The optimal basis takes advantage of standard model dynamics and finds an increased top quark polarization along the axis of the incoming protons or outgoing spectator quarks. To understand why these bases are advantageous, it helps to look at several cases of top quark decay and how the polarization is transmitted (Figure 1.3).

In the top quark rest frame, the left-handed nature of the weak interaction leaves a left-handed b quark and either a longitudinally polarized W boson, or a left-handed W boson. In the longitudinal case (Figure 1.3(a)), the W 's momentum is preferentially aligned with the top polarization. The leptonic decay products tend to align with the W polarization direction, but inheriting its momentum, they are also aligned with the top polarization. In the left-handed case (Figure 1.3(b)), the b quark is forced to travel along the direction of the top polarization, and the W thus recoils in the opposite direction. In the ensuing W leptonic decay, the charged lepton is forced to be right-handed and thus prefers to travel in a direction opposite to the W parent. The cases for an antitop quark follow similar arguments and are illustrated in Figures 1.3(c) and (d).

The four previously discussed decay scenarios can be studied by measuring the angle between the top quark polarization and the lepton momentum [58, 56, 57, 47]. This distribution is modeled by

$$\frac{1}{\Gamma} \frac{d\Gamma}{d\cos \theta} = \frac{1}{2} (1 + \cos \theta) \quad (1.5)$$

where θ is the angle between the top polarization and the lepton momentum in the top quark rest frame, and Γ is the partial width for a semileptonic top quark decay in the Standard Model [59]. In practice, one could measure the top polarization along a given axis, and fit a distribution of the form

$$F(\cos \theta) = \frac{A}{2} (1 + \cos \theta) + \frac{1 - A}{2} (1 - \cos \theta) \quad (1.6)$$

where A is the degree of polarization along the chosen axis [58].

For t -channel single top, the basis of choice is the “optimal” basis. While the helicity basis would prove strong for a $2 \rightarrow 2$ process, yielding 100% left-handed top quarks, the $2 \rightarrow 3$ process muddies this polarization by a varying amount depending on the model and parameters used in the calculation. The optimal basis takes advantage of the fact that the top quark polarization is always along the axis of the incoming antiquark spectator in the reaction. The result of this basis is a top quark that is about 96% polarized along the direction of the spectator quark [58].

Similar studies have been made for the s -channel, and the helicity basis results in about 75% of the top quarks being left-handed, while the optimized basis results in about 97% of the top quarks being aligned with the incoming antiquark momentum.

Overall, measurements of the top quark polarization can help to validate the Standard Model, or could allow us to observe new terms in the top quark couplings. By studying top polarization in both bases, one could pin down the nature of any new couplings and define their strengths.

1.4.2.3 New Fields, Mediators and Particles

A final avenue that can be explored through single top quark production is the possibility of new fields and particles that only begin to noticeably couple to heavier fermions such as the top quark. As previously discussed, extra generations of standard model quarks could alter the measured single top quark production cross sections, but new extensions to the Standard Model involving extra gauge bosons, or new models such as supersymmetry (SUSY), could have noticable effects also. Each of these possibilities will be discussed.

The much anticipated Higgs boson has yet to be found and both the Standard Model’s neutral Higgs and the minimal SUSY charged Higgs could play a role in altering single top quark production rates [58, 45, 62]. The standard model neutral Higgs could produce

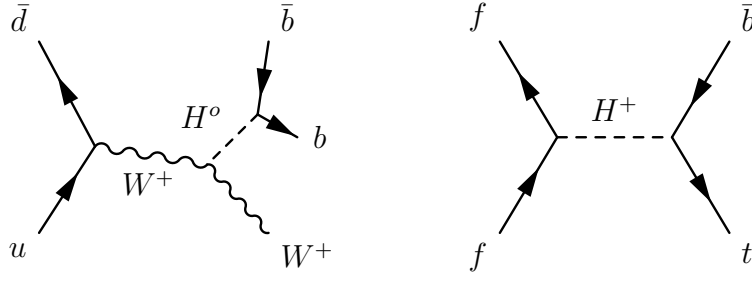


Figure 1.4: Higgs boson processes that mimic the single top quark signal. A standard model neutral higgs boson is shown (left) and a nonstandard model charged higgs boson (right).

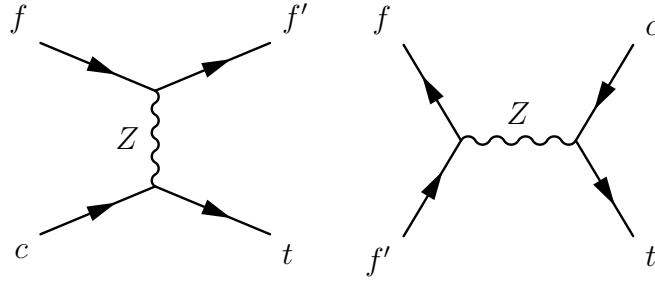


Figure 1.5: Possible flavor-changing neutral current processes involving Zqq' vertices. Both a t -channel (left) and s -channel (right) process are shown.

the same decay signature as s -channel single top, Wbb , while a SUSY charged Higgs could produce the exact same final state, $t\bar{b}$ (Figure 1.4). The t -channel could also be mimicked by this process if a b -quark is mis-reconstructed as a light quark. In parallel with other observations, an excess in single top quark production due to the indistinguishable Higgs processes could finally prove that the Higgs exists.

Flavor-changing neutral currents (FCNC) would also have a noticeable effect through single top quark production [58]. An additional term in the Standard Model interactions, or a completely new field could create neutral currents such as those illustrated in Figure 1.5. Such an interaction would have varying effects on the different single top processes, and by measuring each of these individual cross sections, one could compare the relative changes to provide evidence for this new physics. A scenario involving a Ztc vertex is used to illustrate how this study could be done in practice.

A FCNC could have dramatic effects on the t -channel process. Since the single top

process and the FCNC process have a completely different set of incoming quarks, there would be no interference between the two reactions, and a cross section measurement would simply involve the addition of the two individual processes. Although one may expect the strength of the FCNC to be small, the incoming vertex does not necessitate a b quark like the single top vertex, and therefore has a much larger incoming parton density. This would translate to a larger production cross section, and thus a much more visible effect on t -channel production.

The s -channel process follows a similar train of reasoning to the t -channel, but owing to its already small cross section, and the notorious difficulty in isolating this channel, it is nearly impossible to measure a FCNC effect. The channel would be swamped by $t\bar{t}$ and t -channel single top backgrounds, and a useful measurement would require correctly identifying the b quark from the Wtb production vertex and the additional b quark from the top quark decay vertex.

The tW channel acts as a sort of control channel, being unaffected by the addition of a FCNC. Since the tW channel produces a final state top and W boson, it can only show effects relating to the Wtb vertex. The FCNC process additionally has a different final state and would have no consequence for tW . By combining the information gained from a t and tW channel single top cross section measurement, one could find evidence for FCNCs.

1.4.3 Single Top Quark Modeling and Calculation

The next-to-leading order (NLO) cross sections for $p\bar{p}$ collisions at 1.96 TeV are shown in Table 1.5 [63, 64]. Many considerations have been taken into account with these calculations and are explained in the literature, but the most notable are correctly summing all NLO diagrams (illustrated in Figures 1.6 and 1.7), and dealing with the interference terms and double counting in the $2 \rightarrow 3$ t -channel process. In W -gluon

fusion, the incoming gluon pair-produces before the weak interaction vertex occurs. This gluon may produce a $t\bar{t}$ or $b\bar{b}$ pair and thus the final outgoing top quark is strongly produced in one process, and weakly produced in the other. The destructive interference between these two terms and its effect on the cross section are shown as a function of the top quark mass in Figure 1.8. Double counting becomes an issue in the calculation of t -channel production when considering the origin of the incoming b quark. In the $2 \rightarrow 2$ process, the b quark comes from the parton sea, which really involves a $g \rightarrow b\bar{b}$ splitting that is accounted for in the proton's parton distribution function. The $2 \rightarrow 3$ process explicitly includes this $g \rightarrow b\bar{b}$ splitting in the initial state and so one must be careful not to effectively calculate the same diagram twice. For the numerical calculation of these cross sections, values for the factorization (μ_F) and renormalization (μ_R) scales must be selected. For the s -channel, a good choice is $\mu^2 = M_{t\bar{b}}^2$ (the invariant mass of the $t\bar{b}$ system) and for the t -channel, $\mu^2 = Q_W^2$ for the Wud vertex, and $Q_W^2 + M_t^2$ for the Wtb vertex. A top quark mass of 175 GeV is used for the values given in Table 1.5.

Name	Cross Section (pb)
t -channel	1.98 ± 0.08
s -channel	0.88 ± 0.05
tW	0.09 ± 0.024

Table 1.5: The NLO single top quark cross sections at 1.96 TeV in a $p\bar{p}$ collider, assuming a top quark mass $M_t = 175$ GeV. Errors reflect the uncertainty in the choice of scales (μ_F, μ_R).

1.4.4 Decay Modes

There are several final decay modes for single top quark events and each will be discussed. In the t -channel, there is a final state top quark and a light quark. In the s -channel we find a top quark and a \bar{b} quark. The W boson from the top quark decay defines how we describe the final state and what we can expect to find in the detector. The branching fractions for W boson decay are 1/9 per lepton, and 2/3 for hadronic decays [48]. If

t-channel Feynman Diagrams

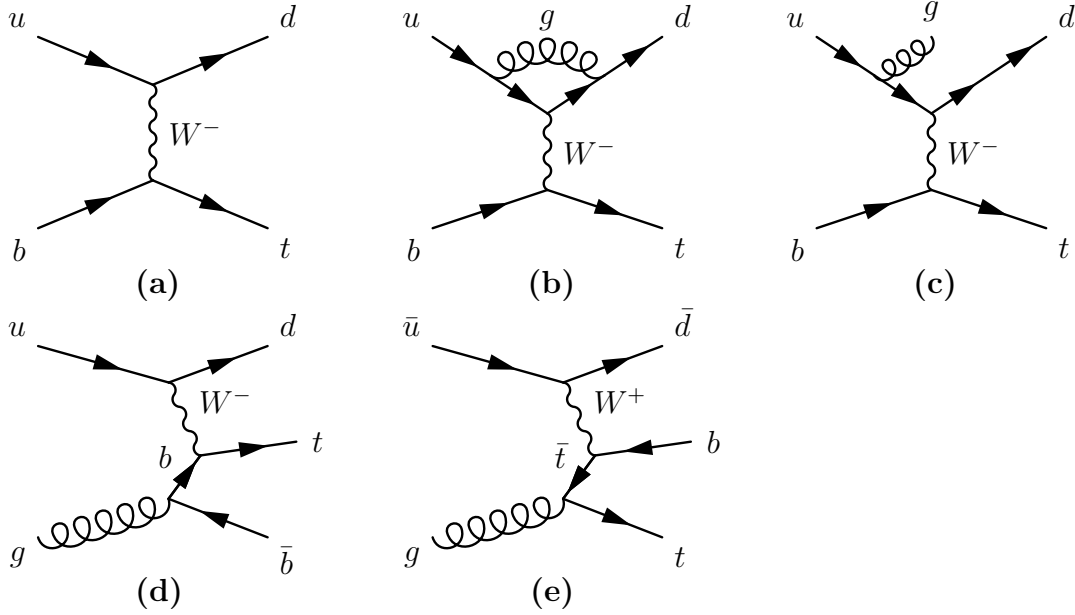


Figure 1.6: Relevant Feynman diagrams for t -channel single top quark production. The leading order (a), example next-to-leading order gluon contributions (b,c), W -gluon fusion (d) and its interference diagram from top quark pair production (e) are shown.

s-channel Feynman Diagrams

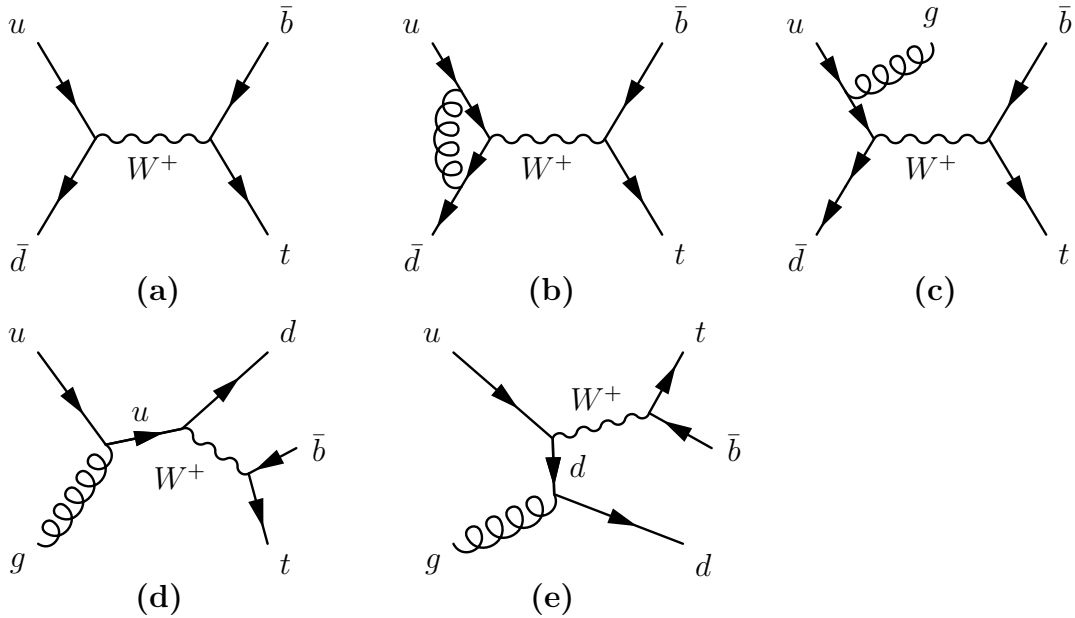


Figure 1.7: Relevant Feynman diagrams for s -channel single top quark production. The leading order (a), example next-to-leading order gluon contributions (b,c) and initial state gluon diagrams (d,e) are shown.

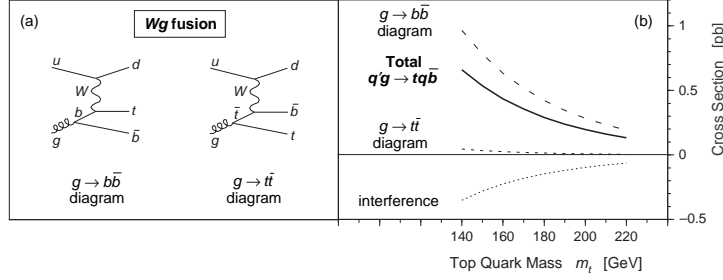


Figure 1.8: (a) Feynman diagrams for the weak and strong production of single top quarks in the $2 \rightarrow 3$ process are shown. (b) Cross sections for both $2 \rightarrow 3$ processes are shown as a function of the top quark mass, as well as the large destructive interference and their final sum [44].

the decay yields a τ lepton, its branching fractions are $\sim 17\%$ for e and μ each, and $\sim 66\%$ for hadronic decay. Single top production via a hadronic W decay is challenging to find in an enormous sea of backgrounds, and so we search for it in the electron and muon modes. Our final decay products in the t -channel are thus a light spectator quark, and the decay products of the top quark, a b quark, lepton, and neutrino. The $2 \rightarrow 3$ process can yield an extra b quark in the event. The s -channel leaves a \bar{b} quark from the Wtb interaction vertex, and a lepton, neutrino, and another b quark from the top quark decay. For both channels, initial and final state radiation (ISR/FSR) can increase the number of hadronic objects found in the final state.

We model these single top quark events by the monte carlo generator SingleTop[50], which is based on the CompHEP generator [49]. SingleTop applies the correct matrix element calculation for accurate NLO distributions and correctly accounts for the polarization of the top quark in its production and decay. In the t -channel, special care is taken to attain the correct NLO simulation. The issue involves the additional b quark from ISR in the $2 \rightarrow 2$ process and the explicit b quark in the $2 \rightarrow 3$ process. The PYTHIA generator creates ISR b quarks by backtracking through the event after it has been generated and creating a sensible initial state that avoids collinear singularities which could appear through the assumption of a massless b quark. The consequences of

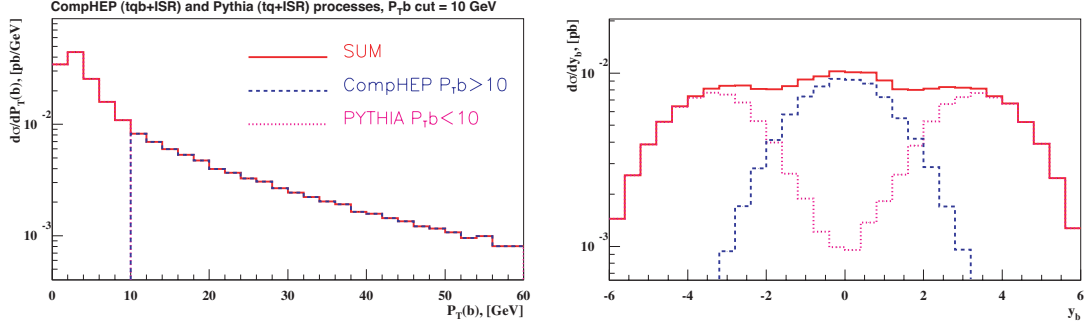


Figure 1.9: The t -channel monte carlo simulation requires the correct mixture of the $2 \rightarrow 2$ and $2 \rightarrow 3$ production modes. The matching of the $p_T(b)$ distributions for these modes at 10 GeV are shown (left), as well as the corresponding $\eta(b)$ distributions normalized to unity (right).

this process become evident when looking at the propagator of the b quark in the matrix element:

$$\frac{1}{p_b^2} \approx -\frac{1}{EE^*(1 - \cos(\theta))} \quad (1.7)$$

where E is the energy of the gluon from the proton or anti-proton, E^* is the energy of the b quark, and θ the angle between the two. As $\theta \rightarrow 0$, one can see that the matrix element blows up. Therefore, most of the events generated by PYTHIA have soft, forward b quarks. This is very different to the NLO calculation, which tends to have harder, more central b quarks, and so a more accurate mixture of the two is provided by the SingleTop generator. This mixture was selected by making the p_T distribution of the extra b quark continuous and resulted in using the $2 \rightarrow 2$ process for events with $p_T(b) < 10$ GeV, and the $2 \rightarrow 3$ process for events with $p_T(b) > 10$ GeV (Figure 1.9). An overall normalization was selected to retain the theoretically calculated t -channel production cross section. The s -channel simulation only applies a LO matrix element calculation, but this has been shown to accurately reproduce the NLO distributions for all partons [51].

Several parton level distributions for our t - and s -channel monte carlo samples are

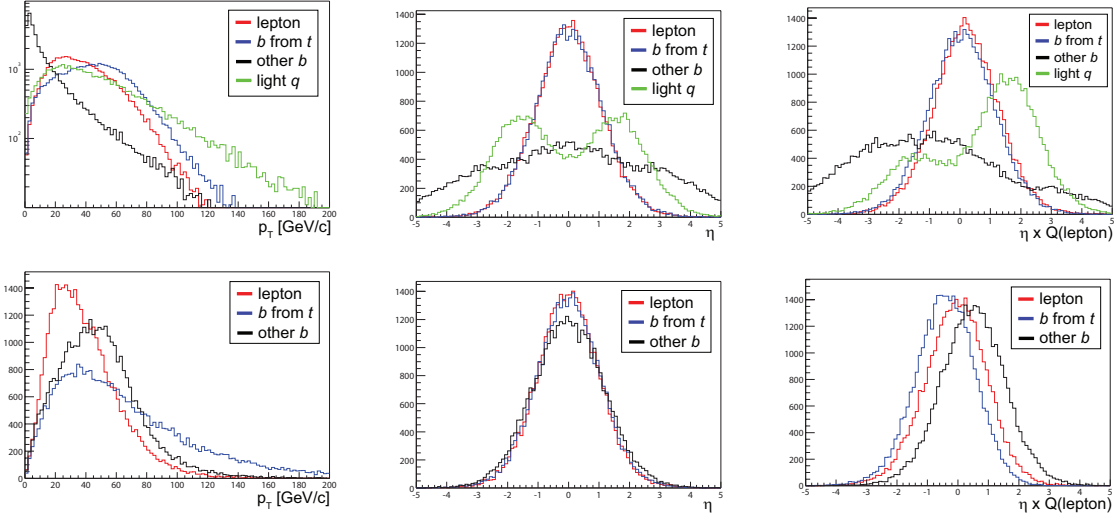


Figure 1.10: Parton level distributions for the t - (above) and s -channel (below) monte carlo simulation. The parton p_T (left), η_{parton} (middle), and $\eta_{\text{parton}} \times Q_{\text{lepton}}$ (right) distributions are shown.

shown in Figure 1.10. These are taken into account when considering our event selection criteria and the corresponding signal acceptance (Chapter 3).

1.4.5 Current State of Affairs

To date, several searches for single top quark production have been conducted without significant evidence being found, and cross section limits have been set [65]–[70]. In Table 1.6, the 95% confidence level measured and expected limits for each of these searches are given. Run I refers to the data recorded at the Tevatron between 1992 and 1996. Run II refers to the data being recorded from 2002 onwards. The limits are given for each of the detectors located at the Tevatron. The first measurement by DØ applied cuts on the scalar sum of various combinations of jet, lepton, and missing energies, and derived limits based on the final number of events. The latter two measurements by DØ used neural network techniques to arrive at a final distribution used for limit calculation. The first CDF measurement used a fit to a kinematic variable (H_T), and the second used an angular variable for the individual s and t channel measurements

$(Q_{\text{lepton}} \times \eta_{\text{untagged jet}})$, and the same H_T variable for the combined measurement. These variables are discussed in Section 5.6.

		95% CL Measured(Expected) Cross Section Limits [pb]				
		Published	\mathcal{L} (pb $^{-1}$)	t -channel	s -channel	Combined
Run I	DØ	2000	90	58	39	—
	DØ	2001	90	22	17	—
	CDF	2002	106	13	18	14
Run II	DØ	2005	230	5.0 (5.8)	6.4 (4.5)	—
	CDF	2005	162	10.1 (11.2)	13.6 (12.1)	17.8 (13.6)

Table 1.6: The published 95% confidence level measured (expected) upper limits for single top quark production. For each measurement, the data taking period, experiment, year of publication, integrated luminosity, and limits are given for each channel.

Referring to Table 1.6, one can see that our limits are quickly approaching the Standard Model predictions for single top quark production. Soon, we will make a first observation of single top quark production, or perhaps even more exciting, uncover an unexpected drop of new physics.

1.5 Analysis Overview

In the following chapters, the process of collecting data, analyzing it, and extracting a final limit will be discussed. Chapter 2 will describe the Fermilab facility, the chain of accelerators necessary to achieve final proton and anti-proton collision energies, and the DØ detector in which we collide these protons and record the aftermath. The triggering and data acquisition system are also discussed here. In Chapter 3, the reconstruction of real physics objects from the detector readout is explained. The cuts we apply to these physics objects to ensure good data quality and to begin isolation of our signal are also given. Chapter 4 discusses the process of producing accurate simulations of backgrounds in the search for single top. The important backgrounds are identified, the

features of their production via monte carlo discussed, and a test of the accuracy of our models is shown with a comparison to data. After the background model is shown to be consistent with data within the assigned uncertainties, we move on to the techniques used to isolate single top events in Chapter 5. An overview of the many variables we have looked at for background discrimination is given, and distributions comparing signal and background for these variables are shown. Our final discrimination technique involving neural networks is explained, as well as the multi-network approach we are taking to isolate our signal. Final input variables to the networks and their output distributions are shown. Once the neural networks have been optimized, the output distributions are used to derive a cross section limit. Chapter 6 introduces the fundamentals of deriving a limit, and explains our method of applying a binned likelihood calculation to the neural network distribution to attain a cross section limit. After final results are summarized in Chapter 7, we look to the future in Chapter 8, and discuss when an observation of single top quark production can be expected, and the precision of subsequent measurements at both the Tevatron, and future hadron colliders.

Chapter 2

Creating Single Top Quarks

2.1 Introduction

To create the smallest particles in the universe, it has become necessary to build the world's largest machines. Our hunt for single top quarks begins with protons and antiprotons, created and accelerated to ultra-relativistic velocities by the Tevatron at Fermi National Accelerator Laboratory in Batavia, Illinois. The Fermilab Tevatron consists of a chain of accelerators ranging from several meters in length to a circular ring over six kilometers in circumference. These accelerators each push protons and antiprotons towards their final collision energy of 0.98 TeV or 1.96 TeV in the center of mass frame. At their final energy, these protons are traveling at 99.99995% the speed of light and are ready to produce the wide range of physics processes available at the Tevatron.

2.2 Theory

In order to produce single top quarks in particle collisions, two conditions must be met: there must be sufficient energy available to produce the rest mass of the top quark, and there must be a pathway via the incoming particles that produces the correct final

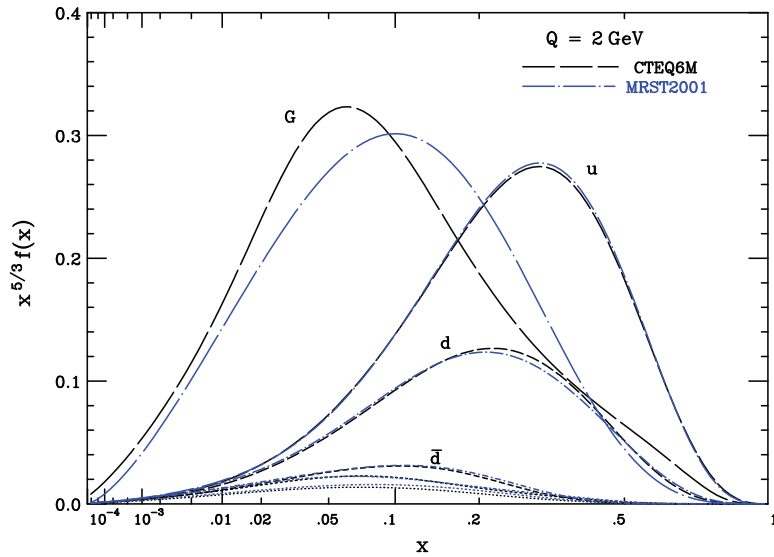


Figure 2.1: Parton distribution functions(PDFs) for protons at the tevatron. Both the CTEQ6M (used in this analysis) and the MRST2001 PDFs are shown. $f(x)$ represents the parton distribution function, and x the momentum fraction of the proton carried by that parton. The unlabeled curves are \bar{u} and $s = \bar{s}$ [71].

state. To accomodate both of these conditions, we collide extremely energetic protons and antiprotons. While the 0.98 TeV of energy contained by each proton is sufficient to produce a top quark, this energy is distributed throughout the components of the proton, amongst its three valence quarks (uud), and a bubbling sea of partons produced from their self-interactions. It is therefore necessary for the hard scattered partons from the proton and the antiproton to contain the energy for the entire reaction. The parton distribution functions (PDFs) give the probabilities for a specific parton with a specific momentum to be involved in the hard scatter (Figure 2.1). The PDFs are based on fits to many sets of data and they allow us to correctly model the $p\bar{p}$ collisions and predict scattering cross sections. Single top quarks are most often produced at the Tevatron by the up valence quark from the proton and the antidown valence quark from the antiproton (or the charge conjugate process for \bar{t}), but they can also be produced from the parton sea via any pair of weak isospin partners, or from an incoming b quark from the sea.

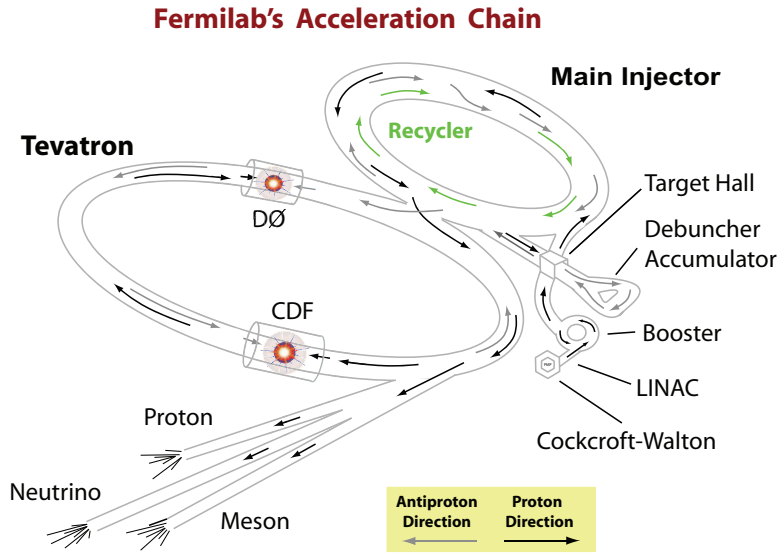


Figure 2.2: Fermilab's acceleration chain.

2.3 Acceleration Chain

2.3.1 Proton Production

The acceleration chain begins with the proton source in the preaccelerator [72, 73]. Here, a pressurized tank of hydrogen gas (H_2) is slowly sprayed across two high-voltage molybdenum electrodes, 1 mm apart. The diatomic molecules are ripped apart and bond to a cesium plate in the source chamber. The low work function of the cesium allows the hydrogen atoms to acquire an extra electron each to complete their s orbital shell and become negatively charged (H^-). These ions are then knocked off the cesium surface by the next wave of incoming hydrogen atoms and are directed by an electric field into the Cockcroft-Walton accelerator.

The Cockcroft-Walton accelerator gives the hydrogen ions their first push (Figure 2.2). It accelerates them through a large electric potential created through a voltage multiplier ladder (consisting of diodes and capacitors). Through this potential, a continuous beam of ions is ramped up to 750 keV and injected into the Linac [75]. The

Linac, or linear accelerator, consists of an alternating array of cylindrical drift tubes and empty gaps. A radiofrequency (RF) electric field is applied to these drift tubes such that the hydrogen ions are accelerated in the gaps, and shielded in the drift tubes while the field switches polarity. The oscillating RF field only gives positive acceleration in the gaps at certain moments of its cycle, and so the continuous beam from the Cockcroft-Walton accelerator condenses into ion bunches. The first portion of the 130 m Linac uses the original design from 1971 and brings the ions to 116 MeV. An upgrade was added in 1993 to bring the ions to their final energy of 400 MeV.

At this point the hydrogen ions are sent through a thin carbon foil and stripped of their electrons. The bare protons are then shot into the first synchrotron of the chain, the Booster. This circular accelerator uses an RF field to accelerate the protons, and 96 copper wire electromagnets to contain them in a circular orbit. Using the basic Lorentz force law, the strength of the magnetic field is increased from 0.0074 T to 0.7 T to maintain the radius of the proton orbit as it gains energy with each lap. The protons will make about 16,000 laps around the 475 m circumference of the Booster before reaching an energy of 8 GeV. These protons are then delivered to the Main Injector.

The Main Injector is a second, much larger synchrotron that was added to the acceleration chain in 2001. [72, 76]. Its 3,320 m circumference houses 344 electromagnetic bending magnets (dipole) and 208 beam focusing magnets (quadrupoles). It has several tasks: to accelerate protons to 150 GeV for injection into the Tevatron, to accelerate protons to 120 GeV for antiproton production, and to accelerate antiprotons to 150 GeV for Tevatron injection. The antiproton production mechanism is discussed in Section 2.3.2. Using one beampipe and one set of magnets, 8 GeV protons and antiprotons are accelerated in opposite directions around the Main Injector to an energy of 150 GeV. Following this, they are prepared for acceleration in the Tevatron.

The Tevatron is the final accelerator in the chain [72]. It is the largest synchrotron at the laboratory, with a radius of 1 km, and also the world's first synchrotron to use

superconducting magnets. It was built in 1983, and houses about 1000 superconducting magnets. Each dipole magnet is about 6.4 m long and made with approximately 68,400 km of niobium-titanium alloy wire (8 μm diameter). At maximum acceleration, they are each producing a magnetic field of 4.2 T. In the final stage of acceleration, 36 bunches of protons and of antiprotons, containing 10^{11} and 10^9 particles respectively, are injected into the Tevatron and ramped up to 0.98 TeV. These bunches are then held at constant energy for the duration of the store (10-20 hours), and repeatedly collided at the interaction points in the centers of the two detectors on the Tevatron ring, the DØ and CDF detectors.

2.3.2 Antiproton Production

Antiproton production begins with 120 GeV protons from the Main Injector [73]. This proton beam is collided with a 7 cm thick nickel target and results in a large spray of interaction products, including about 20 antiprotons for each million incoming protons. The wide spray of particles is collimated with a lithium lens, and sent through a bending magnet (essentially a mass spectrometer) to isolate the antiprotons. While being contained in a very tight bunch, the spread of energy is large around a mean value of 8 GeV. This spread is unacceptable for further acceleration, and so the antiproton energy spectrum is tightened in the Debuncher.

The Debuncher is a triangular shaped synchrotron, consisting of several bending and focusing magnets. As the antiproton bunch circles the Debuncher, more energetic particles travel at a larger radius than less energetic ones. This effect alters the arrival time at the RF cavity in such a way that all particles are accelerated towards the 8 GeV mean. After many revolutions about the accelerator, the energy distribution has been tightened, at the expense of elongating the bunch. The antiprotons are then sent to another synchrotron, the Accumulator, to be stored until their total number is large enough for

injection into the Main Injector, and then the Tevatron. Both the Debuncher and the Accumulator share the same tunnel, and both apply stochastic cooling techniques [79] to further focus the antiproton bunches and maintain the tight energy spectrum.

As the production of antiprotons is the limiting factor in preparing for collisions in the Tevatron, dumping antiprotons at the end of a store is a considerable waste. In an attempt to salvage these costly antiprotons and use them in the next store for collisions, the Recycler was added to the Fermilab accelerator chain in 2001. The Recycler shares the same tunnel as the Main Injector, but has its own beampipe and magnets. The Recycler uses permanent strontium ferrite magnets (the same compound in refrigerator magnets) to bend the beam. The use of permanent magnets removes the noise and problems that arise in mechanical elements and leads to a very stable and reliable storage system. With this new system, antiproton bunches can be saved after the termination of a store in the Tevatron and this significantly decreases the time needed to accumulate enough antiprotons for the next store. In addition, the Recycler can act as a post-Accumulator center and store antiprotons once the Accumulator has been filled to capacity. Together, these two functions of the Recycler have helped to raise the overall luminosity of the Tevatron.

2.4 Luminosity

The purpose of the Fermilab acceleration chain is not only to bring protons and antiprotons to an energy of 0.98 TeV for collisions, but to collide these particles at a rate high enough to study a vast array of rare processes, and increase the precision of already measured processes. The rate of collisions is determined by the energy and density of particles in both the proton and antiproton bunches. The unit of measurement used to describe the scattering cross section of these bunches is the barn (b)¹. The rate at which

¹1 b = 10⁻²⁴ cm²

the effective cross sectional area of the proton or antiproton bunch is being bombarded is the luminosity, given in units of particles per $\text{cm}^{-2} \text{s}^{-1}$.

To record the delivered luminosity at the $D\bar{O}$ interaction point, luminosity monitors [77] have been installed around the beam pipe at $z = \pm 140 \text{ cm}$. Both monitors are made up of two layers of 24 radiation-hard plastic scintillator wedges, with a photomultiplier tube mounted on each one. The wedges have a radius of 15 cm, and span the region $2.7 < |\eta| < 4.4$. By measuring the sum of single diffractive, double diffractive, and hard scattered $p\bar{p}$ collisions in each bunch crossing, the luminosity can be determined through the equation,

$$\mathcal{L} = \frac{f\overline{N}_{LM}}{\sigma_{LM}} \quad (2.1)$$

where \mathcal{L} is the luminosity, \overline{N}_{LM} is the average number of collisions, f the frequency of bunches, and σ_{LM} the effective cross section of the monitor, including acceptance and efficiency (LM = Luminosity Monitor). To accurately determine \overline{N}_{LM} , the number of hard scatters per bunch crossing is assumed to follow a Poisson distribution, and the distribution is then constrained to fit the number of bunch crossings with no interactions.

Increasing the luminosity at the Tevatron has been a ceaseless effort since its inception. Run I luminosities ranged from $10 - 20 \times 10^{30} \text{ cm}^{-2} \text{s}^{-1}$ and brought $D\bar{O}$ an integrated luminosity of $\sim 120 \text{ pb}^{-1}$ over four years. During this period, there were 6 proton and 6 anti-proton bunches circling the accelerator. In order to increase the Run II luminosity while keeping the number of hard-scattered interactions per bunch crossing at an acceptable level, the Tevatron switched to 36 proton and 36 anti-proton bunches. Luminosities began at $10 \times 10^{30} \text{ cm}^{-2} \text{s}^{-1}$ and have steadily increased to a current value of $120 \times 10^{30} \text{ cm}^{-2} \text{s}^{-1}$. The integrated luminosity delivered so far exceeds 1 fb^{-1} , and an order of magnitude increase is expected by the end of Run II in 2009.

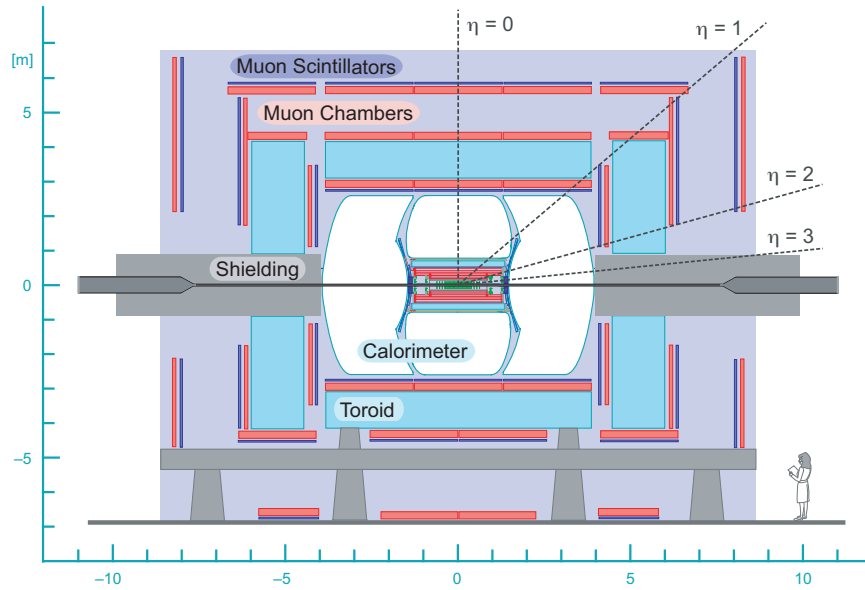


Figure 2.3: A cross section of the DØ detector. The marked scale is given in meters and the main components are labeled.

2.5 The DØ Detector

The DØ Detector is a formidable beast. Proposed in 1984 for Run I, it is 13 m high \times 11 m wide \times 17 m long, and weighs about 5500 tons. It can be divided into three main subdetectors — the central tracking region, the calorimeter, and the muon system (Figure 2.3). To accomodate the increased luminosity in Run II, the detector went through extensive upgrades. The central tracking and muon systems were completely redesigned, but the calorimeter has remained the same, with a replacement of all electronics. A description of the current design is given based largely on the technical report recently submitted [77].

2.5.1 Coordinates

The DØ detector is centered at the $p\bar{p}$ interaction point. From this point, several coordinate systems can be defined, and are in active use by the collaboration. A cartesian coordinate system is defined with \hat{z} in the proton direction, \hat{y} pointing up, and \hat{x} point-

ing radially inward, towards the center of the Tevatron ring. A cylindrical right-handed coordinate system is defined with the \hat{z} direction along the beam pipe in the proton direction, radial distance r , and azimuthal angle ϕ , where $\phi = 0$ coincides with the $+\hat{x}$ -axis. Finally there is an angle θ , defined as the angle from the $+\hat{z}$ -axis. To ease the transformation of coordinates from one relativistic reference frame along \hat{z} to another, θ is further converted into rapidity (y) and in the approximation that mass is small compared to energy, pseudorapidity (η).

$$y = \frac{1}{2} \ln \left[\frac{E + P_z c}{E - P_z c} \right] \quad \eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (2.2)$$

2.5.2 Central Tracking Systems

Starting from the interaction point, the first system a hard scattered particle encounters is the central tracking. The central tracking consists of several subsystems, the Silicon Microstrip Tracker (SMT), the Central Fiber Tracker (CFT), a 2T superconducting solenoid, and the preshower detectors. These systems are new for Run II and an overview of the more important characteristics is given.

2.5.3 Silicon Microstrip Tracker

The SMT is designed to give r - ϕ - z hits from particles as they traverse the detector. The fundamental detecting unit is a silicon wafer with many parallel microstrips across its surface. The strips are held at a bias voltage, and as charged particles pass through, they ionize the silicon and produce current flows measured by the wire strips known as “hits”. These silicon wafers are arranged in six barrels concentric with the beam pipe, 12 F-disks perpendicular to it, and 4 larger H-disks (Figure 2.5). The barrels are composed of rectangular wafers, called ladders, arranged in four concentric layers around the z axis. Most of the ladders are doubled sided, and layers 1 and 3 have microstrips oriented at 90°

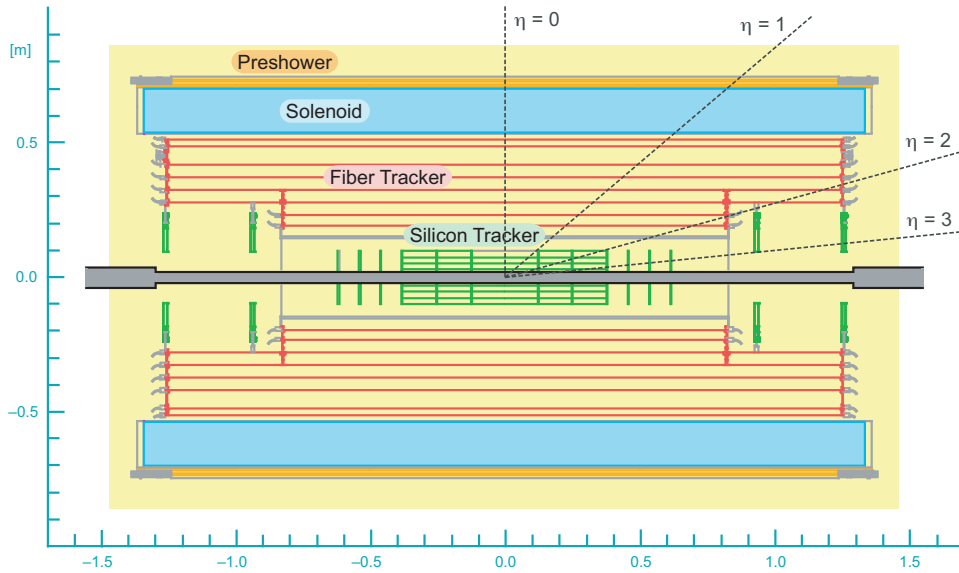


Figure 2.4: A schematic of the central tracking systems.

stereo, and layers 2 and 4 are 2° stereo. The distance between microstrips (pitch) varies from ladder to ladder, but is $\sim 50 \mu\text{m}$ (Table 2.1). The F-disks are interspersed between the barrels, and an additional three more are placed at each end of the detector. Each F-disk is made of 12 double sided trapezoidal shaped silicon wedges. The double sided silicon microstrips have $\pm 15^\circ$ stereo orientation. Finally, the four H-disks are located at the ends of the SMT, with the furthest one at $z = \pm 121 \text{ cm}$. Each is made of 24 wedges and each wedge consists of two single sided “half” wedges glued back-to-back, providing an effective 15° stereo coverage. All together, there are $\sim 793,000$ thousand channels available for readout from the SMT for each event, but many of these are suppressed on an event-by-event basis as they only contain noise and no “hits” from the $p\bar{p}$ interaction.

2.5.4 Central Fiber Tracker

Surrounding the SMT is the Central Fiber Tracker. This device uses scintillating fibers to detect the passage of charged particles. As particles travel through the fibers, a small fraction of their energy is deposited and converted to light. This light travels through

Module	Type	Layer	Pitch (μm)	Inner Radius (cm)	Outer Radius (cm)
F-disks	DS	-	50/62.5	2.57	9.96
H-disks	SS	-	40 80 readout	9.5	26
Central Barrels (4)	DSDM	1,3	50/153.5	2.72	7.58
	DS	2,4	50/62.5	4.55	10.51
Outer Barrels (2)	SS	1,3	50	2.72	7.58
	DS	2,4	50/62.5	4.55	10.51

Table 2.1: Design specifications of the SMT. Types are double-sided (DS), single-sided (SS), and double-sided double metal (DSDM)

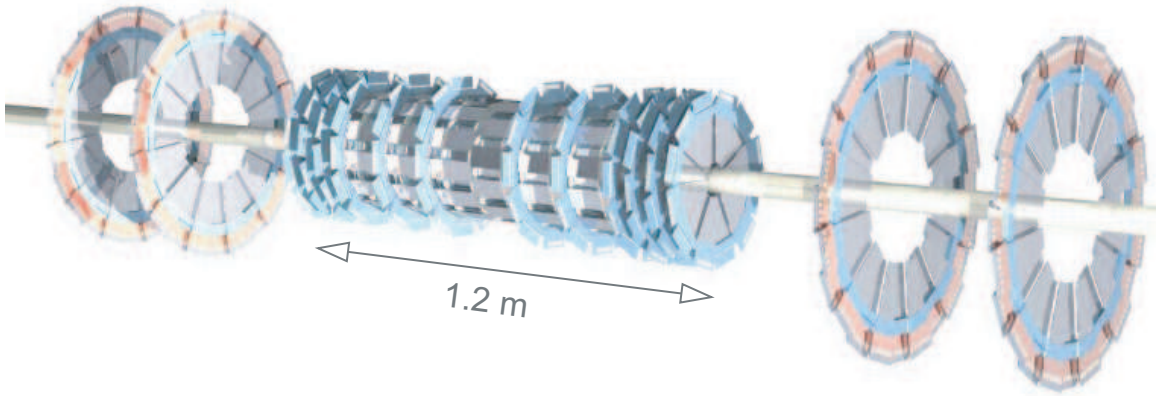


Figure 2.5: A schematic of the silicon microstrip tracker.

a wave guide and is collected by a “visible light photon counter” (VLPC) outside the detector, signaling a hit in the fiber tracker. The 835 μm fibers are arranged in eight concentric cylinders about the beam pipe, with the innermost at a radius of 20 cm and the outermost at 52 cm. Each cylinder is composed of a double layer of fibers, with one layer parallel to the z axis and the other layer $\pm 3^\circ$ (alternating between cylinders). The location of a particle in a layer is identified by the intersection of any two fibers that received a signal.

Encasing both of the tracking systems is a 2T superconducting solenoid. Its magnetic field aids in the measurement of momentum by curving the paths of charged particles. Working together, the SMT and CFT are able to resolve the primary interaction vertex to 35 μm along the beamline and help locate secondary vertices from b -quark decays with a resolution of 15 μm (at $\eta=0$ and $p_T > 10$ GeV).

2.5.5 Preshower Detector

To increase the resolution of the calorimeter after traversing the material in the solenoid, the remaining 5 cm gap between the solenoid and calorimeter is instrumented with another detector - the preshower. The preshower acts both as a final layer to the central tracking and as an extension to the calorimeter. It is used to help match tracks in the central tracking to energy clusters in the calorimeter, and acts as an extra calorimeter layer for the less penetrating electromagnetic showers. The central preshower (CPS) consists of three layers of scintillating fibers, triangular in cross section, arranged at parallel, $+23.77^\circ$, and -24.02° angles to the z axis. There is also a thin (7/32”) radiative lead plate encased by two stainless steel plates (1/32” each) between the solenoid and the CPS. These plates add about two radiation lengths of material for particles normal to the z axis, and up to four at larger η .

The forward preshower (FPS) rests in the thin cavity between the central and end

calorimeters. It consists of two layers of double layered scintillators, separated by an 11 mm lead-steel-lead plate. The first layer is known as the minimum ionizing particle or MIP layer, and the second layer following the lead plate is known as the absorber layer. All charged particles will leave a hit in the MIP layer, and electrons and photons will begin to shower in the absorber layer, producing a cluster of hits. Each layer is made of 48 wedges, and each wedge is double layered. Scintillating fibers in the layers are placed at 22.5° with respect to one another. The layers are slightly offset such that particles do not cross more than one set of uninstrumented cracks between the wedges.

2.5.6 Calorimeter

The calorimeter is the main instrument used to measure electron, photon, and hadron jet energies. It completely encompasses the central tracking and is separated into three units, one central calorimeter (CC), and north and south end calorimeters (ECN, ECS) (Figure 2.6). The calorimeters are further divided into an electromagnetic region (EM), a fine hadronic region (FH), and a coarse hadronic (CH), as you move further from the interaction point. Each calorimeter is segmented into many cells containing an absorber medium and an active medium, and are arranged in projective towers radiating from the interaction point (Figure 2.7). Liquid argon is used as the active medium in all cells, which requires the calorimeters to each be housed in a cryostat, held at approximately 80 K. The cell's absorbing medium varies with the region: depleted uranium (EM), uranium-niobium (2% alloy) (FH), copper (CC CH), stainless steel (EC CH). A typical calorimeter cell consists of an electrically grounded absorber layer, a 2.3 mm gap of liquid argon, and a G-10 coated copper signal board held at a positive voltage. Particles are slowed down by the absorber layer, and ionize the liquid argon as they shower through it, leaving a measured current proportional to the energy deposited in that cell.

The CC extends out to an $|\eta| \sim 1.0$. It is composed of three concentric cylinders,

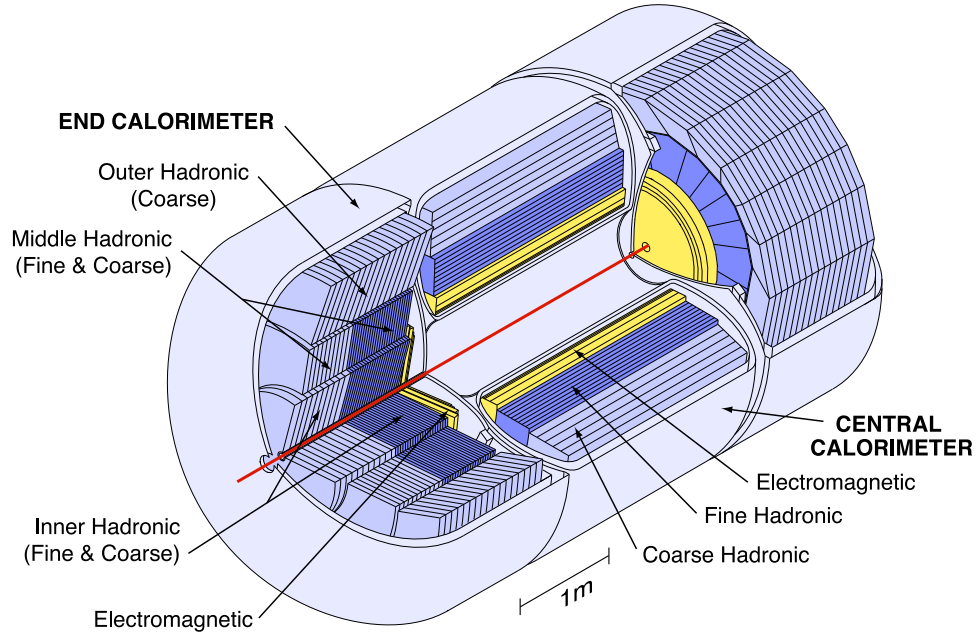
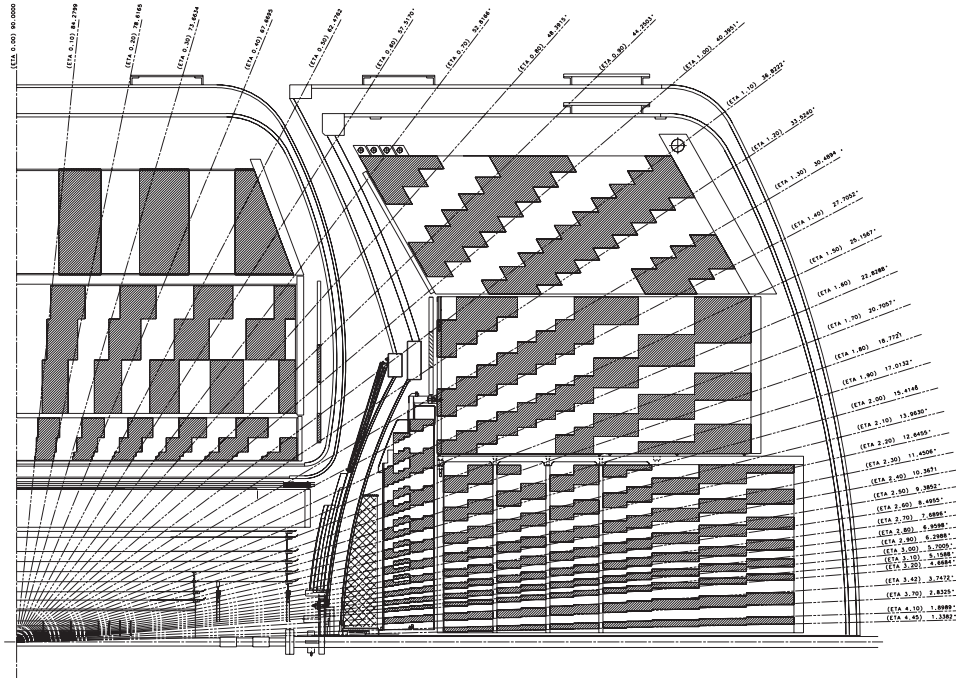


Figure 2.6: A diagram of the DØ calorimeter. The central and end cryostats are shown and the segmentation of the various regions is labeled.



making up the EM, FH, and CH regions. The transverse size of cells is typically $\Delta\eta = 0.1$ and $\Delta\phi = 0.1$. In ϕ , the EM region is split into 32 cells, and the FH and CH are split into 16 cells. The layers are offset in ϕ to avoid cracks. The EM region comprises the first four layers of the calorimeter. These thinner layers are enough to contain most of the energy in an electromagnetic shower, and can further help to distinguish differences in shower shape with the extra resolution of the third layer, $\Delta\eta = 0.05$ and $\Delta\phi = 0.05$. The FH region consists of three layers and aids in measurement of the further penetrating hadronic showers. The CH region consists of a single thick layer used to effectively contain the remaining energy in the particle shower. Absorptive and radiative lengths for particle traversal through the calorimeter are shown in Table 2.2.

Layer	Absorption (λ_A)		Radiation (X_o)	
	CC	EC (IH/MH/OH)	CC	EC
EM 1	0.76 (EM 1-4)		2.0	0.3
EM 2			2.0	2.6
EM 3			6.8	7.9
EM 4			9.8	9.3
FH 1	1.3	17.6/0.9/—		
FH 2	1.0	17.6/0.9/—		
FH 3	0.76	17.6/0.9/—		
FH 4		17.6/0.9/—		
CH 1	3.2	4.1/4.4/6.0		
Total				

Table 2.2: The absorptive and radiation lengths of the material composing each of the layers in the calorimeter.

The ECN and ECS are identical and have the same regions as the CC, but are segmented differently to deal with axial showers instead of radial ones. An EC consists of three concentric cylinders. The innermost contains four EM layers, and several FH and CH layers that make up the inner hadronic region. The next cylinder is called the middle hadronic and contains a FH and CH region. The final cylinder is called the outer

hadronic and only contains a CH region. Absorptive and radiative lengths can be found in Table 2.2.

2.5.7 InterCryostat Detector and Massless Gaps

To deal with the uninstrumented regions between the central and end calorimeters from $|\eta| = 0.8 - 1.4$, two systems have been added: the massless gaps and the intercryostat detectors. The massless gaps are essentially calorimeter cells without the absorptive layer. They are located inside of the cryostats on the walls facing the gap between the CC and ECs (Figure 2.7). The intercryostat detectors are located on the outer wall of the ECs and cover a region from $|\eta| = 1.1 - 1.4$. The system is made of a layer of scintillating tiles, each spanning an area $\Delta\eta \times \Delta\phi = 0.3 \times 0.3$.

2.5.8 Muon System

The muon system is the outermost layer of the DØ detector (Figure 2.3). It is designed to locate and aid in the momentum measurement of the deeply penetrating muons from the hard scatter and secondary decays. As a heavier version of the electron, muons have relatively fewer electromagnetic interactions as they traverse the detector, allowing them to sustain their energy as they reach the muon system and even travel beyond its boundaries. Two types of detectors are used to measure these muons, scintillators and drift chambers. Scintillators are used for triggering on events involving a muon by detecting the light given off as these charged particles traverse them. Drift chambers are used to measure the position of the muon as well as its momentum. Each chamber consists of two walls with cathode pads held at a negative voltage, and a wire running along the center held at a positive voltage. The chamber is filled with a gas that becomes ionized as muons shower through the detector, and the position is identified by two values: the difference in time between a chamber and its nearest neighbor detecting a particle

from the signals on the sense wires, and the integrated charge on the two pads along the walls. To aid in momentum measurement, a 109 cm thick, 1.9 T toroidal magnet is situated between the layers of the muon system, and the magnetic deflection of the track is used to assess momentum. The system has been superseded by the improved central tracking for momentum measurement, but has a comparable resolution at high energies.

The central muon system ($|\eta| < 1.0$) consists of two layers of scintillation tiles, an A- ϕ layer between the calorimeter and the toroid, and a cosmic layer after the toroid (Figure 2.8). Each layer has four walls surrounding the detector, parallel to the beamline. Both layers produce a fast readout for triggering and associating muons with the correct bunch crossing, while the cosmic cap is also used to reject out-of-time or cosmic-ray muons. The central muon system further includes three layers of proportional drift chambers (PDTs). The A layer is sandwiched between the calorimeter and the toroid, while the B and C layers are outside the toroid (Figure 2.9). Each of these layers is further made up of three or four decks (sublayers) of drift chambers that can be used to construct a particle track. The PDTs use a fast operating gas consisting of argon (84%), methane (8%), and CF_4 (8%), and the wires and pads are held at 4.7 and 2.3 kV respectively. The PDTs have a maximum electron drift time of ~ 500 ns, to be compared with the bunch spacing of 396 ns. About 55% of the central system is covered by all three layers of PDTs and about 90% with two layers.

The forward muon system ($1.0 < |\eta| < 2.0$) includes three layers of scintillation counters, three layers of mini drift-tubes (MDTs) and shielding around the beam pipe to reduce background in the detector. Each scintillation layer is perpendicular to the beam pipe and each octant includes about 96 individual detecting panels. The A layer is before the toroid and the B and C layers are found after it (Figure 2.8). The three layers of drift tubes are paired with the scintillation layers and are arranged parallel to them (Figure 2.9). While the PDTs in the central system were reused from Run

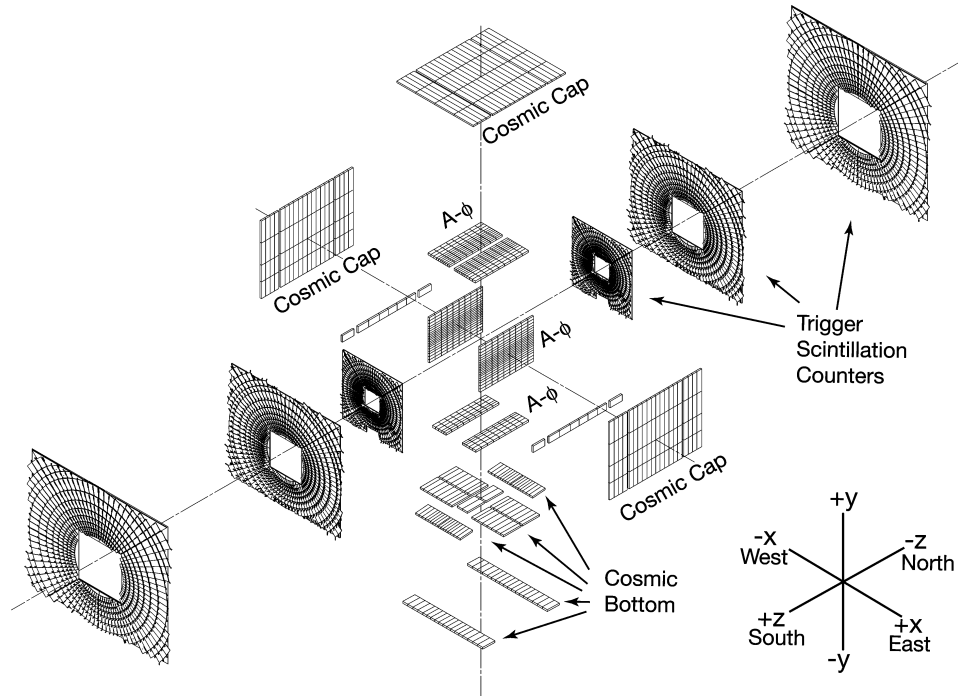


Figure 2.8: An exploded view of the scintillation layers in the muon system. From the interaction point, central muons will cross the A- ϕ Layer, the toroid, and the cosmic layer. Forward muons will cross the A layer, the toroid, and then the B and C scintillation layers.

I, the forward system is completely new, and the MDTs exhibit greater segmentation, are more radiation hard, and have a faster response time. They use a nonflammable, 90%-10% mixture of $\text{CF}_4\text{-CH}_4$ and have a maximum drift time of 60 ns. The A(B/C) layer consists of 4(3) decks of these MDTs and all MDTs are arranged along the field lines of the toroid. Again the A-layer falls before the toroid and the B/C-layer after.

Each forward system also includes a hundred ton block of shielding, surrounding the beam pipe. The innermost layer is made of a 51 cm thick slab of iron, followed by 15 cm of polyethylene, and 5 cm of lead [78]. Iron is used for its relatively short interaction (16.8 cm) and radiation (1.76 cm) lengths, and stops most electromagnetic and hadronic showers. The iron is nearly transparent to low energy neutrons and so polyethylene is added for its great absorptive capacity. The neutron capture process produces gamma rays, which are in turn absorbed by the final layer of lead shielding.

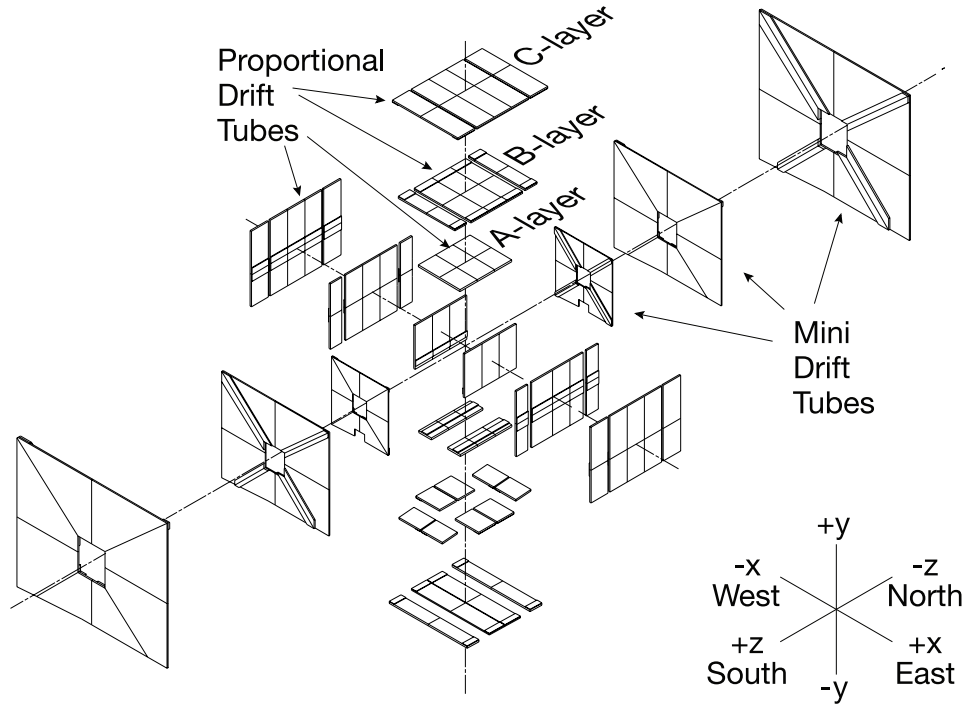


Figure 2.9: An exploded view of the wire chambers in the muon system. From the interaction point, muons traverse the A, B, and C layers.

2.6 Data Acquisition

During live collisions, the Tevatron ring is colliding $p\bar{p}$ bunches at ~ 1.7 MHz. Each bunch crossing yields a single hard scatter on average, but could also be absent of any collisions, or have several. Due to the constraints of cost for data storage and computing power, only select events will make it to permanent data storage to be used in an analysis. This selection process consists of three levels of triggers, with each requiring cuts on progressively better defined physics objects. The Level 1 or L1 trigger system is based on the fast readout from individual subdetectors, and using hardware triggers, can pass an event that fulfills a single detector requirement. Events that pass one of these requirements are then processed by the Level 2 or L2 trigger system. Here, rudimentary physics objects are constructed and the first global decision are made based on information from several subdetectors. L2 triggered events are sent to the Level 3 (L3)

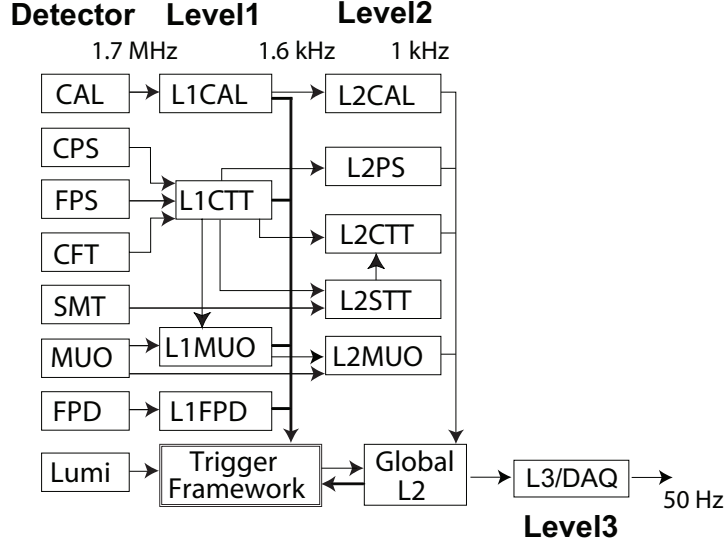


Figure 2.10: A schematic of the subdetector data flow for the DØ trigger system. From Global L2, data is passed on to the Global L3 processor, and then to tape (not pictured).

trigger system, where sophisticated software algorithms can be applied on the individual and combined subdetector readout. Physics objects are constructed, and basic cuts on energy, momentum, location, and object quantities can be applied. A diagram of the data flow for triggering is shown in Figure 2.10.

At Level 1, every single event is looked at in real-time, so the trigger requirements must be minimal enough such that all events can be processed in about 100 nanoseconds with the available computing power. Each subdetector uses on-board hardware cards and processors to determine if an event passes a basic trigger requirement. The Level 1 calorimeter (L1Cal) trigger looks for energy deposition in a 0.2×0.2 EM or hadronic tower block exceeding a given threshold. All 1,280 EM and 1,280 hadronic towers ($40 \eta \times 32 \phi$) are involved in this search. The Level 1 central track trigger (L1CTT) and Level 1 Muon (L1Muon) triggers look for a charged particle track with momentum exceeding a given threshold. The L1CTT uses the CFT and central preshower readout to locate tracks with a momentum exceeding 1.5 GeV. The 1.7 MHz bunch crossing rate is filtered down to about 1.6 kHz, and accepted events are sent on to the L2 trigger system.

The L2 system further reduces the 1.6 kHz rate to about 1 kHz. This is done by coordinating subdetector information and taking advantage of the extra processing time to construct simple physics objects. The Level 2 calorimeter (L2Cal) trigger processor creates electrons, photons, and jets, and calculates the missing transverse energy (\cancel{E}_T) in the event. Electron and photon decisions are based on a seed EM tower with $E_T > 1$ GeV and the energy of the next-highest neighbor, as well as the content of a 0.3×0.3 tower block centered on the seed tower. Jet requirements are based on an EM+hadronic seed tower with $E_T > 2$ GeV and a 0.5×0.5 tower block centered on the seed tower. The \cancel{E}_T is measured by using the primary vertex from the tracking systems, and then calculating a vector sum of the transverse energy of all calorimeter cells. The Level 2 muon (L2Muon) system applies calibrations and takes advantage of timing to improve muon candidates and cut on more precise p_T , position, and timing values. The Level 2 central track trigger (L2CTT) and Level 2 silicon track trigger (L2STT) create p_T -sorted lists of tracks to trigger on, and the improved reconstruction of the L2STT allows a measurement of the impact parameter (Section 3.3.1) for B -meson related triggers. The Level 2 global trigger combines all the information and decisions made by the individual L2 systems to pass or fail an event for L3 processing.

At Level 3, the trigger system has much more time to evaluate an event and reduces the 1.0 kHz rate to the final 50 Hz which is written to tape for permanent storage. An expandable farm currently using ~ 200 dual processor AMD 2000 or Pentium III 1 GHz machines handles the job. Each machine collects the fully digitized output from all subdetectors for a single event and provides a final trigger decision. Physics object algorithms very similar to the offline reconstruction algorithms are used to create and cut on the objects anticipated by each trigger. Descriptions of these objects and sample triggers are given later in Chapter 3. After a L3 accept, the fully digitized event (~ 250 kB) is sent to a robotic tape system about 3 km away from the detector.

Chapter 3

Finding Single Top Quarks

3.1 Introduction

The nature of particle physics is not to run an experiment once or twice and then draw a conclusion, but to run an experiment countless times (in our case about 30 trillion times a year), gather enormous amounts of data, and draw conclusions on a statistical basis. The fluctuations of intensity in the colliding proton beams, the complexity of the detector, the inherent noise in the detecting elements, and the enormous size of background processes being produced for any single signal process, require us to sift through large volumes of data and derive a statistical interpretation of any result. Experiments at the Tevatron accomplish such a feat by running and manning the experiment twenty-four hours a day, seven days a week, all year long. This chapter describes the data collected over the past several years. We discuss the process of reconstructing high quality physics objects in this data, as well as the event selection requirements used as a first attempt to narrow our search for single top.

3.2 Data Samples

The data samples used in this analysis were collected from August 2002 to January 2005. Several trigger lists were used during this period to collect the necessary samples for this analysis. As the instantaneous luminosity of the Tevatron grew, tighter definitions for triggers were required to keep the overall trigger rates and data flow within the limits of the data acquisition system. This evolution of triggers will be apparent. This section describes what we use each data sample for, which triggers were used to collect them, and the corresponding integrated luminosity for each.

3.2.1 $e+jets$

The electron+jets or “ $e+jets$ ” data sample is the signal sample expected to contain single top events with a leptonic decay from the W boson to an electron and a neutrino. The sample contains events with at least one electromagnetic object and at least two hadronic jets (of which the EM object counts for one). The total represented integrated luminosity is $366 \pm 24 \text{ pb}^{-1}$. These triggers are summarized in Tables 3.1 and 3.2.

- **EM15_2JT15** At Level 1 (L1), this trigger requires an EM calorimeter tower with $E_T > 10 \text{ GeV}$ and two jet towers with $E_T > 5 \text{ GeV}$. Level 2 (L2) requires at least a 10 GeV EM object with an electromagnetic fraction (Section 3.4.3) > 0.85 and two jets with $E_T > 10 \text{ GeV}$. The final Level 3 (L3) trigger requires $E_T > 15 \text{ GeV}$ with consistent calorimeter shower shape and at least two 15 GeV jets.
- **E1_SHT15_2J20** This trigger requires a L1 EM object with $E_T > 11 \text{ GeV}$. There is no L2 requirement, and at L3, an $E_T > 15 \text{ GeV}$ EM with consistent shower shape and two $E_T > 20 \text{ GeV}$ jets are needed.
- **E1_SHT15_2J_J25** This trigger also requires an $E_T > 11 \text{ GeV}$ EM object at L1. L2 further requires the EM to have $E_T > 15 \text{ GeV}$, and L3 requires at least

a 15 GeV EM with good shower shape, and two $E_T > 20$ GeV jets, with one of these further required to be above 25 GeV.

3.2.2 $\mu + jets$

The muon+jets or “ $\mu + jets$ ” data sample is the signal sample for single top events where the W boson decays leptonically to a muon and a neutrino. Triggers are listed in Tables 3.1 and 3.2 and the total integrated luminosity is $363 \pm 24 \text{ pb}^{-1}$.

- **MU_JT20_L2M0** The L1 trigger requires a muon scintillator and wire hit, and a calorimeter tower above 5 GeV. At L2, a medium muon with quality = 2 (Section 3.4.2) is sought and at L3, a jet with $E_T > 20$ GeV is required.
- **MU_JT25_L2M0** This trigger is similar to the previous one, with the exceptions of a L1 EM tower > 3 GeV and at L3 a more energetic jet with $E_T > 25$ GeV.
- **MUJ2_JT25** At L1, this trigger requires a muon scintillator and wire hit, and a calorimeter tower above 5 GeV. At L2, a medium muon with quality = 2 is required as well as a jet with $E_T > 8$ GeV, and at L3 the jet is required to exceed 25 GeV and have a detector $|\eta| < 3.6$.
- **MUJ2_JT25_LM3** This trigger is identical to the previous one, but further requires an L3 muon with $p_T > 3$ GeV.

3.2.3 $e + e$ (DIEM)

This sample consists of events where two electromagnetic objects have been produced such as electrons, photons, or pions. It is used to measure efficiencies for electrons to pass cuts that are applied in the main analysis. These efficiencies are then used to derive the Monte Carlo Correction Factors necessary to directly compare Monte Carlo samples with

data. The DIEM events we use are expected to have come from the leptonic decay of a Z boson, $Z \rightarrow ee$, and a cut is made above and below the reconstructed Z mass to ensure the quality of the correction factors. (Section 4.3.1.2). The DIEM triggers are listed in Tables 3.1 and 3.2 and the total integrated luminosity represented is $341 \pm 22 \text{ pb}^{-1}$.

- **2EM_HI** This trigger requires two EM trigger towers exceeding 10 GeV at L1. There is no L2 requirement, and at L3, one of the EM objects must have $E_T > 20 \text{ GeV}$ and have a detector $|\eta| < 3.0$.
- **E1_2L20** Requires an L1 EM object with $E_T > 11 \text{ GeV}$. There is no L2 requirement, and two EM objects with $E_T > 20 \text{ GeV}$ at L3.
- **E2_2L20** Requires two L1 EM objects with $E_T > 6 \text{ GeV}$. There is no L2 requirement, and two EM objects with $E_T > 20 \text{ GeV}$ at L3.
- **E3_2L20** Requires two L1 EM objects with $E_T > 3 \text{ GeV}$, one of which is also $> 9 \text{ GeV}$. There is no L2 requirement, and two EM objects with $E_T > 20 \text{ GeV}$ at L3.
- **E1_2L15_SH15** Requires an L1 EM object with $E_T > 11 \text{ GeV}$. There is no L2 requirement, and at L3, two $E_T > 15 \text{ GeV}$ EM objects, one of which also passes electron shape requirements.
- **E2_2L15_SH15** Requires two L1 EM objects with $E_T > 6 \text{ GeV}$. There is no L2 requirement, and at L3, two $E_T > 15 \text{ GeV}$ EM objects, one of which also passes electron shape requirements.
- **E2_2L15_SH15** Requires two L1 EM objects with $E_T > 3 \text{ GeV}$, one of which is also $> 9 \text{ GeV}$. There is no L2 requirement, and at L3, two $E_T > 15 \text{ GeV}$ EM objects, one of which also passes electron shape requirements.

3.2.4 $\mu + \mu$ (DIMU)

This sample consists of events where two muons have been produced. It is used in the analysis to measure efficiencies for muons to pass cuts that are applied in the main analysis. These efficiencies are then used to derive the Monte Carlo Correction Factors necessary to directly compare Monte Carlo samples with data (4.3.1.3). The DIMU events are expected to have originated from the leptonic decay of a Z boson, $Z \rightarrow \mu\mu$, and a cut is made above and below the reconstructed Z mass to ensure the quality of muon efficiencies. The DIMU triggers are listed in Tables 3.1 and 3.2 and the total integrated luminosity represented is $318 \pm 21 \text{ pb}^{-1}$.

- **2MU_A_L2M0** This trigger requires two muons with scintillator hits at L1. At least one muon must be of MEDIUM quality at L2, and there is no L3 requirement.
- **2MU_A_L2M0_TRK10** This trigger requires two muons with scintillator hits at L1. At least one muon must be of MEDIUM quality at L2, and at L3 a global track with $p_T > 10 \text{ GeV}$ is required.
- **2MU_A_L2M0_L3L15** This trigger requires two muons with scintillator hits at L1. At least one muon must be of MEDIUM quality at L2, and at L3, a muon must have $p_T > 15 \text{ GeV}$.
- **2MU_A_L2M0_TRK5** This trigger requires two muons with scintillator hits at L1. At least one muon must be of MEDIUM quality at L2, and at L3 a global track with $p_T > 5 \text{ GeV}$ is required.
- **2MU_A_L2M0_L3L6** This trigger requires two muons with scintillator hits at L1. At least one muon must be of MEDIUM quality at L2, and at L3, a muon must have $p_T > 6 \text{ GeV}$.

Triggers used to collect the Single Top Data Samples			
Final State	Trigger List	Trigger Name	Luminosity pb^{-1}
e +jets	v8.2-v11.0	EM15_2JT15	110
	v12.0-v13.0	E1_SHT15_2J20	210
	v13.0-v13.3	E1_SHT15_2J_J25	46
μ +jets	v8.2-v11.0	MU_JT20_L2M0	114
	v12.0-v13.0	MU_JT25_L2M0	210
	v13.0-v13.2	MUJ2_JT25	26
	v13.2-v13.3	MUJ2_JT25_LM3	13
DIEM	v8.0-v11.0 v12.0	2EM_HI	124
		E1_2L20	217
		E2_2L20	217
		E3_2L20	217
		E1_2L15_SH15	217
		E2_2L15_SH15	217
		E3_2L15_SH15	217
DIMU	v8.0 - v10.0 v11.0	2MU_A_L2M0	51
		2MU_A_L2M0_TRK10	57
		2MU_A_L2M0_L3L15	57
	v12.0	2MU_A_L2M0_TRK5	210
		2MU_A_L2M0_L3L6	210

Table 3.1: The triggers used for each portion of the data set and the associated integrated luminosity. As the instantaneous luminosity of the Tevatron increased over time, new versions of triggers had to be written with tighter requirements to keep the overall bandwidth within the limits of the data acquisition system.

Final State	Trigger List	Trigger Name	L1	L2	L3
e+jets	$\geq v8.2$ - 11.0 $v12.0$ - 13.0 $v13.0$ - 13.3	EM15_2JT15 E1_SHT15_2J20 E1_SHT15_2J_25	CEM(1,10)_CJT(2,5) CEM(1,11) CEM(1,11)	EM(.85,10)_JET(2,10) — L2CALEM(15,x)	ELE_SH_T(1,15)_JET(2,15) ELE_SHT(1,15)_JET(2,20) ELE_SHT(1,15)_JET(2,20)_JET(1,25)
μ +jets	$\geq v8.2$ - 11.0 $v12.0$ - 13.0 $v13.0$ - 13.2 $v13.2$ - 13.3	MU_JT20_L2M0 MU_JT25_L2M0 MUJ2_JT25 MUJ2_JT25_LM3	mulptxatxx_CJT(1,5) mulptxatxx_CJT(1,5) mulptxatlxCJT(1,5) mulptxatlxCJT(1,5)	MUON(1,med) MUON(1,med) MUON(1,med)JET(1,8) MUON(1,med)JET(1,8)	JET(1,20) JET(1,25) JET(1,25,3.6) MUON(1,LOOSE)JET(1,25,3.6)
DIEM	$v8.0$ $v9.0$ - 11.0 $v12.0$ $v12.0$ $v12.0$ $v12.0$ $v12.0$ $v12.0$ $v12.0$	2EM_HI 2EM_HI E1_2L20 E2_2L20 E3_2L20 E1_2L15_SH15 E2_2L15_SH15 E3_2L15_SH15	CEM(2,10) CEM(2,10) CEM(1,11) CEM(2,6) CEM(2,3)CEM(1,9) CEM(1,11) CEM(2,6) CEM(2,3)	— — — — — — — —	ELE_LOOSE(1,10) ELE(1,20) ELE(2,20) ELE(2,20) ELE(2,20) ELE(2,15)SH(1,15) ELE(2,15)SH(1,15) ELE(2,15)SH(1,15)
DIMU	$v8.0$ - 10.0 $v11.0$ $v12.0$	2MU_A_L2M0 2MU_A_L2M0_TRK10 2MU_A_L2M0_TRK5	mu2ptxatxx mu2ptxatxx mu2ptxatxx	MUON(1,med) MUON(1,med) MUON(1,med)	— TRK(1,10) TRK(1,5)

Table 3.2: The Level 1, 2, and 3 requirements for each of the triggers used to collect the data samples. The conditions are explained in the associated data sample section.

3.3 Object Reconstruction

A triggered event is really just 250 kB of 1's and 0's representing the various electrical and light signals measured in the detector. These bits must be ordered by complex algorithms to produce the more familiar physics objects we seek to study. The trigger system produces progressively more detailed shadows of such objects, and at L3, the triggered objects are fairly close to those used in the analysis. Still, the unrestricted time of the offline reconstruction code allows for more precise and clearly defined objects. This section discusses the offline algorithms used to reconstruct the physics objects that are evaluated in an analysis.

3.3.1 Particle Tracks

As charged particles traverse the tracking system, their paths are curved by the magnetic field of the solenoid. Along these paths, they deposit a small amount of energy in each layer of material in the tracker, otherwise known as a “hit.” The collection of hits from a single particle form a particle track. It is the duty of the DØ tracking algorithms to sift through the thousands of hits in any given event, and reconstruct the original particle tracks. The following algorithms have been developed to this end:

Histogramming Method (HTF) [82] The HTF method may begin with either SMT hits or CFT hits, and uses a transformation from x - y coordinates to ρ - ϕ (where ρ is the curvature of a track, and ϕ the azimuthal angle). The coordinate transformation facilitates track finding by locating a peak in the ρ - ϕ plane as the most likely track. The algorithm consists of:

- All possible pairs of hits in the event are converted to a point in $\rho - \phi$ space, and added to a histogram. Removing points below a minimum curvature, ρ_o , one simply finds a peak in the distribution, and has thus located a track and all the pairs of hits that constitute that track.

- The number of tracks is then reduced by a 2D Kalman filter [83] and another transformation. A Kalman filter is mathematically identical to a least-squares-fit, but is written to be computationally efficient for real-time processing. The second transformation adds in the z coordinate information, transforming from r - z space to z_o - C space, where z_o is the z position of the track at the origin, and C is the track inclination ($C = dz/dr$). Again, a peak is found in the z_o - C space, and is identified as the most likely set of points forming a track.
- Finally, a 3D Kalman filter is used to fit the remaining hits in a track, and further extrapolate the track to the other tracker, either the CFT or SMT.

Alternative Algorithm (AA) [81] The AA algorithm mainly uses SMT hits to seed the track finding, but CFT hits can be used where the SMT coverage is limited. The following steps are taken:

- Build a seed track in $r - \phi$ space with three SMT hits. First locate the beam spot and any single hit. Next, select a second hit within $\phi = \pm 0.08$. A third hit must then be found which creates a track with a minimum curvature ($r = 30$ cm), and a maximum impact parameter (Figure 3.1) with respect to the beam spot of 2.5 cm. The three-hit track requires a chi-squared fit $\chi_{trk}^2 < 16$.
- Further hits are added to the track from different layers of the SMT and CFT if they improve the overall χ_{trk}^2 . Hits from the same layer can be accepted and define multiple tracks if each case improves the χ_{trk}^2 .
- Tracks are then ordered by the number of hits they are composed of, and filtered or removed if there are too many missing hits along the track or at either end of it. Tracks sharing a portion of their hits may be kept if they pass other quality cuts. It should be noted that no z information is taken into account with this algorithm.

Most unphysical tracks are naturally removed by the algorithm, but the remaining ones are dealt with at the primary vertex reconstruction stage.

Global Track Reconstruction (GTR) [84] The GTR algorithm is applied last, using the reconstructed tracks from both the AA and HTF algorithms as inputs. A standard Kalman fit is applied to these tracks, and the final set of tracks in the event is defined.

3.3.2 Primary Vertices

The primary vertex is the three dimensional position of the hard scattered event from the proton and the antiproton. Its location is of paramount importance as many other variables, including energy, momentum and the position of secondary vertices, are calculated in terms of the primary vertex. Two algorithms have been developed to locate the primary vertex, and both show comparable results [85].

The **DØreco** algorithm consists of a two-pass procedure, and begins by locating the beam spot, or the x - y location of the $p\bar{p}$ interaction. This is done by fitting all tracks with an impact parameter (IP) significance, $S_{IP} = IP/\sigma_{IP} < 100$ (with respect to the $r - \phi$ origin), to a common vertex through minimization of the impact parameter of all tracks (Figure 3.1). If the worst fit track has a $\chi^2_{trk} > 10$, it is removed from the vertex, and the procedure is repeated. Once all tracks have $\chi^2_{trk} < 10$, a first pass primary vertex has been found. Additional primary vertices are then found with the remaining tracks, repeating the steps as outlined above, until all primary vertices are found. The beam spot is then identified through the position of all located vertices. Pass two is identical to pass one, but places a tighter S_{DCA} cut on all tracks, this time with respect to the beam spot. Once all pass two primary vertices have been constructed, they are each given a probability for coming from a minimum-bias event, or of not being the hard-scatter vertex. This vertex probability is defined as the product of the probability

for each track to be a minimum-bias track, weighted by the number of tracks involved in the vertex. The hard-scatter vertex has the smallest minimum-bias probability of all reconstructed vertices.

The **DØroot** algorithm shares the same pass-two procedure as the DØ reco algorithm, but differs in pass-one. First, tracks are clustered together in z to separate multiple vertices. Next, each cluster is fit with the Kalman filter technique. Once the pass-one list of primary vertices is constructed, the same pass-two procedure is performed.

The average number of tracks attached to a given vertex is 20 for a basic multijet data sample, with a reconstruction efficiency of $\sim 98\%$ (100% in the central region and tailing off outside of the SMT fiducial volume). The primary vertex resolution depends on the number of attached tracks, but beyond 15 tracks it is dominated by the resolution of the beam spot. The average resolution in the transverse plane is $\sim 35 \mu\text{m}$ [86].

3.3.3 Muons

Muons are reconstructed and categorized based on the pattern of hits in the muon system as well as by having a spatially matched track in the central tracking system [87]. Muons found only in the muon system are labeled “local”, and muons with an associated track in the central tracking are called “global”. Tight, medium, and loose muons are defined by the total number of hits found in the wire chambers and the scintillators. The quality of a track fit between these points is also taken into consideration. Muons are further subdivided by the number of segments (Nseg) or layers that a hit was found in. Nseg = 1 requires at least one A layer hit, Nseg = 2 requires at least one B or C layer hit, and Nseg = 3 includes at least one hit in the A layer and one in the B or C layer. From these categories, many types and qualities of muons can be defined, and the accepted ones are determined by the needs of a given analysis.

3.3.4 Electromagnetic Clusters

Electromagnetic (EM) clusters are localized deposits of energy in the calorimeter consistent with a purely electromagnetic interaction. These generally include showers from electrons, photons, and π_0 's and are completely contained in the first few layers of the calorimeter. For the central calorimeter, the algorithm looks for an EM seed tower above a given E_T threshold, and includes all surrounding EM towers in a radius $R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.2$. The end calorimeters require a similar seed tower, and then include all EM towers in a transverse radius of 10 cm [86]. In addition, the so-called t42 algorithm (Section 3.3.5) is applied to remove noisy cells.

3.3.5 Jets

In many processes, the $p\bar{p}$ interaction will lead to a scattering of the proton valence quarks or the production of new quarks. These quarks will instantly hadronize, and the resultant particles will radiate in a projective η - ϕ region through the detector in the direction of the original quark's momentum. As these hadrons traverse the material of the calorimeter, there will be both electromagnetic and strong interactions, and a shower of particles will form, depositing energy throughout the calorimeter and leaving the signature of a jet. Volumes have been written addressing the concept of a "jet" and how it truly relates to the hard scattered process, and while one can be referred to these works [88, 89] and references therein, only the DØ experiment's simple cone algorithm used in this analysis will be discussed here. The purest form of the algorithm is seedless, and simply requires testing every calorimeter tower as a "seed" and including all surrounding towers in a radius $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ as the contents of the jet. All test jets could be computed, and the best quality jets selected as real jets. With the constraint of computing power, this algorithm is modified such that only small clusters of towers exceeding a given energy threshold are used as seeds. In addition, there are issues with

jet cones overlapping and the algorithm studies midpoints between such jets in making a decision to reconstruct two separate jets, or one jet involving a fraction of the energy from each of the overlapping jets. Due to small levels of inherent noise in the calorimeter, an additional cell-killing algorithm is applied to improve the energy measurement. This algorithm, known as “t42” [90, 91], only allows calorimeter cells to be included in a jet or electromagnetic object if (a) the cell has a neighbor whose energy is 4σ above the noise threshold and (b) the cell itself is greater than 2σ above the noise threshold.

3.3.6 *b*-Quark Jets

Several algorithms for identifying *b*-quark mesons from the hard-scattered event have been developed at this experiment. All take advantage of the decay of the *b* meson: some look for a muon from *b* decay inside of a jet, while others seek the *b* decay vertex itself. Typically, a *b* meson will travel several millimeters before decaying into lighter particles. On average, simulations have shown [92] that the *b* meson leaves five charged decay products with $p_T > 0.5$ GeV/c, and that 99% of the time they are contained in a cone of $R = 0.5$. Additionally, the standard model tells us that the *b* meson will decay to a muon about 10% of the time. Four independent *b*-tagging algorithms have been constructed to take advantage of these observations, and each has been used successfully in various analyses. This analysis however, applies a newly developed neural network *b*-tagging algorithm, which combines the power of three of the previously independent *b* taggers. We will explain each of these individual *b*-tagging algorithms as well as the neural network *b*-tagger. A more thorough explanation of the multivariate neural network technique is given in Chapter 5.

Counting Signed Impact Parameter (CSIP) Algorithm [93] The CSIP tagger looks for tracks in a jet with a large perpendicular distance, known as the impact parameter (IP), with respect to the primary vertex (PV) (Figure 3.1). These tracks are

assumed to originate from the secondary decay vertex of a b hadron in the jet. The impact parameter significance, S_{IP} , is defined as the ratio of the impact parameter to its resolution. S_{IP} is positive (negative) if the track intersects the jet axis in the positive (negative) momentum direction. Once a minimum number of tracks have been located exceeding a given S_{IP} threshold, the jet is considered to be tagged (Section 3.4.5).

Jet Lifetime Probability Tagger (JLIP) Algorithm [94, 95] The JLIP tagger also uses the track impact parameter significance to locate b jets, but instead of requiring a given number of tracks passing a S_{IP} threshold, the significances of all relevant tracks are combined into a single probability for the jet to be a b jet. (Section 3.4.5).

Secondary Vertex Tagger (SVT) [92, 96] The SVT algorithm aims to locate the set of tracks from the b -hadron decay, and to reconstruct a secondary vertex. The first step involves reconstructing all $R = 0.5$ track-based jets. The tracks in these jets are first clustered by their distribution in z along the beamline, and are required to pass quality cuts (Table 3.4). Identified track-jets then go through a fitting process, building a vertex with all combinations of two-track seed vertices and adding a track at a time until the highest quality vertex is identified. This list of secondary vertices is then passed on for filtering at the identification stage (Section 3.4.5).

Neural Network Tagger (NN Tagger) [97, 98] The neural network tagger uses variables from the three taggers listed above to derive a function that effectively separates light-quark and b -quark jets. The function guides light-quark jets towards 0 and b quark jets towards 1. Several cut points on this neural network output define the various operating points of the tagger (Section 3.4.5), designed to maximize the tagging efficiency or to minimize the false-tag probability. By using all the available information, the NN Tagger is able to achieve higher b -tagging efficiency for lower false-tag efficiency than any of the three component tagging algorithms on their own.

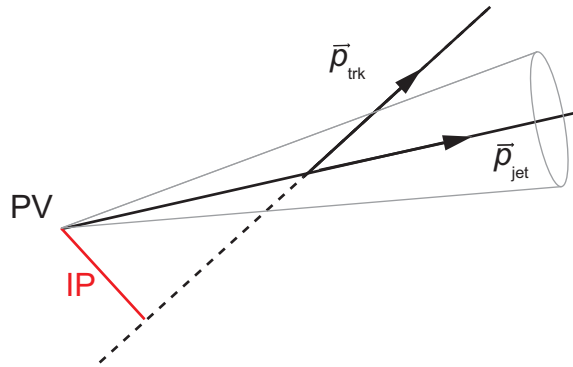


Figure 3.1: Illustrated above is the impact parameter (IP) of a track (\vec{p}_{trk}) with respect to a jet momentum axis (\vec{p}_{jet}). The IP is used both in vertex reconstruction and b -tagging algorithms. In the literature, the IP is also referred to as the distance of closest approach (DCA).

3.3.7 Neutrinos

Neutrinos are identified by the momentum and energy imbalance they leave in the detector. Passing through all detector material without interaction, they are revealed by the excess in the sum of all transverse energy (\cancel{E}_T) in the event, and their momentum vector is opposite to that. The z component of their momentum is left ambiguous as the initial z momentum of each of the hard-scattered partons from the proton and the antiproton is undetermined. The \cancel{E}_T algorithm sums the transverse energy of all calorimeter cells with respect to the primary vertex, except for those in the coarse hadronic region. These noisier cells are only included if they are part of an identified jet. The $t42$ algorithm described in Section 3.3.5 is also applied to remove noisy cells. Corrections for muon energy and calorimeter energy scale calibrations are also made (Section 3.4.6).

3.4 Object Identification

After each of the physics objects in an event has been reconstructed, quality cuts are made on these objects for acceptance into an analysis. These cuts are based on com-

parisons of data and Monte Carlo simulations, on how well we understand the relevant subdetector contributing to the object’s information, and the specific needs of an analysis. Below, the quality variables applied to each object and the cuts on these variables are given.

3.4.1 Primary Vertices

After first-pass primary vertex lists have been assembled (Section 3.3.2), the requirements listed in Table 3.3 are set for all tracks making up the final list of primary vertices. The hard-scatter primary vertex is selected as the vertex with the smallest minimum-bias event probability.

Track Variable	DØreco Vertex	DØroot Vertex
p_T	$\geq 0.5 \text{ GeV}/c$	$\geq 0.5 \text{ GeV}/c$
SMT hits	$\geq 2 \text{ (data)}, \geq 0 \text{ (MC)}$	≥ 2
IP significance	≤ 5.0	≤ 3.0

Table 3.3: The requirements for all tracks involved in the final reconstruction of primary vertices for both the DØreco and DØroot algorithms [85].

3.4.2 Muons

The muon identification used in this analysis [87] requires a minimum of the following:

A so-called “medium, Nseg = 3” muon, defined as

- \geq two A-layer wire hits
- \geq one A-layer scintillator hit
- \geq two wire hits in the B- and C-layers combined
- \geq one scintillator hit in the B- and C-layers combined
- a matched track to the central tracking systems, the silicon microstrip tracker (SMT) and central fiber tracker (CFT)

To reject cosmic-ray muons, we require

- A, B, and C-layer scintillation times < 10 ns from the bunch crossing time
- Central track $\chi^2 < 4$
- $|S_{IP}| < 3\sigma$
- $|\Delta z(\text{central track, primary vertex})| < 1$ cm

To ensure the muon has come from W boson decay, and not the decay of a b quark, we require the muon to be isolated from a jet

- $\Delta R(\mu, \text{jet}) > 0.5$
- Track halo isolation = $|\sum_{\text{tracks}} p_T/p_T(\mu)|$ for all tracks within a radius $R < 0.5$ of the muon track
- Calorimeter halo isolation = $|\sum_{\text{cells}} E_T/p_T(\mu)|$ for all cells within a radius $0.1 < R < 0.4$ about the muon's calorimeter track

These cuts have been optimized using both data and Monte Carlo simulations. In addition, the invariant mass of the Z boson in $Z \rightarrow \mu\mu$ events has been found to have better resolution and be shifted in Monte Carlo events compared to data. To correct this, a scale factor is applied to Monte Carlo muons to shift and smear their energy distribution. This scaling and smearing takes on the form

$$\frac{1}{p'_T} = \frac{1}{\alpha p_T} + [\text{Gaussian}(0, \sigma = S)] \quad (3.1)$$

where p'_T and p_T are the corrected and uncorrected muon momenta, and α and S are the energy scale and smearing parameters respectively. The fit values for α and S are:

	$\eta_\mu < 1.6$	$\eta_\mu > 1.6$
α	0.991	0.999
S	0.0023 c/GeV	0.0047 c/GeV

3.4.3 Electromagnetic Clusters

Once EM clusters have been reconstructed, we must further identify whether they are electron, photon, or pion clusters. This analysis seeks electron clusters and uses the following variables to identify and assess the quality of each EM object:

- $f_{\text{EM}} = E_{\text{EM}}/E_{\text{total}}$ The electromagnetic fraction, or f_{EM} , is the fraction of the total calorimeter energy in an $R = 0.2$ cone contained in the EM calorimeter region. It is expected to be close to one for a real electromagnetic shower.
- χ_{EM}^2 A covariance matrix (M_{ij}) has been developed to test how closely the EM cluster shower shape resembles that of an electron cluster. The matrix consists of the following seven variables - the energy in each of the four EM layers, the total EM energy, the primary vertex z position, and the transverse shower width in ϕ . Monte carlo electrons, that have been accurately modeled by test beam electrons, are used to build the matrix. The matrix is defined as

$$M_{ij} = \frac{1}{N} \sum_{n=1}^N (x_i^n - \bar{x}_i)(x_j^n - \bar{x}_j) \quad (3.2)$$

where the sum is over electrons in the reference Monte Carlo sample and x_i and x_j are the variables making up the covariance matrix. The χ_{EM}^2 for the event is then calculated as

$$\chi_{\text{EM}}^2 = \sum_{i,j=1}^7 (x_i^n - \bar{x}_i) M_{ij} (x_j^n - \bar{x}_j) \quad (3.3)$$

- The cluster isolation, f_{iso} , defined below, measures how localized the energy distribution is in the cluster by comparing energy in an EM tower cone of $R = 0.2$ with the total surrounding energy in an $R = 0.4$ cone.

$$f_{\text{iso}} = \frac{E_{\text{total}}(R < 0.4) - E_{\text{EM}}(R < 0.2)}{E_{\text{EM}}(R < 0.2)} \quad (3.4)$$

In addition, a match between the EM cluster and a central track is required for an electron candidate. With respect to the energy centroid of the EM in the calorimeter, this is defined as $|\Delta\phi_{\text{EM,trk}}| < 0.05$ and $|\Delta\eta_{\text{EM,trk}}| < 0.05$.

To select good electrons, an electron likelihood, \mathcal{L} , has been developed [101]. It is based on the following eight variables: f_{EM} , χ_{EM}^2 , $E_T^{\text{Cal}}/p_T^{\text{trk}}$, $Prob(\chi_{\text{SpatialEM-trk}}^2)$, $IP_{(e\text{-trk,PV})}$, $N_{\text{trks}}(R < 0.05)$, $\sum_{\text{trks}} p_T(R < 0.4)$, and $\text{CPS}_{\text{stripmax}}$. The probability $P(\chi_{\text{SpatialEM-trk}}^2)$ tests the spatial match between the calorimeter cluster and the matched track in ϕ and z . Central preshower $\text{CPS}_{\text{stripmax}}$ is the total number of hits in the u , v , and x layers of the largest cluster in the preshower within a 20 cm radius of the EM axis. All variable distributions are found in electron and non-electron samples, and the probabilities to be signal (P_{signal}) and background (P_{bkgd}) are calculated. The likelihood is created such that signal events tend towards 1, and background events tend towards 0. It is defined as a function of all input variables \vec{x} as:

$$\mathcal{L}(\vec{x}) = \frac{P_{\text{sig}}(\vec{x})}{P_{\text{sig}}(\vec{x}) + P_{\text{bkg}}(\vec{x})} \quad (3.5)$$

In addition to these cuts, an energy scale factor and smearing are applied to EM objects reconstructed in Monte Carlo samples. When comparing to data, these factors account for the better resolution and slightly shifted dielectron invariant mass found in $Z \rightarrow ee$ Monte Carlo samples. The EM energy scale is parameterized as

$$E' = E \times [\alpha + \text{Gaussian}(0, \sigma = \alpha S)] \quad (3.6)$$

where E' and E are the corrected and uncorrected calorimeter energy, α the data/MC EM scale factor, and S the smearing parameter. The values for α and S are given below. Electron identification is thoroughly discussed in References [99] and [100].

	Inside Fiducial (CC)	Outside Fiducial (CC)	(EC)
α	1.003	0.95	0.996
S	0.044	0.112	0.032

3.4.4 Jets

Owing to the size of jets, and the inherent noise in the calorimeter cells, a lengthy list of variables has been constructed to ensure the quality of these objects. We reconstruct $R = 0.5$ jets, and require all jets to pass selection cuts, listed in Section 3.5, on the following variables for use in our analysis:

- $f_{\text{EM}} = E_{\text{EM}}/E_{\text{total}}$ The electromagnetic fraction, or f_{EM} , is the fraction of the total calorimeter energy in the jet cone contained in the EM region. We require $0.05 < f_{\text{EM}} < 0.95$.
- $f_{\text{CH}} = E_{\text{CH}}/E_{\text{total}}$ The coarse hadronic fraction, or f_{CH} , is the fraction of the total calorimeter energy contained in the coarse hadronic layers of the jet cone. We require $f_{\text{CH}} < 0.4$.
- $f_{\text{Hot}} = E_{\text{cell1}}/E_{\text{cell2}}$ The hot fraction is the ratio of the energy in the most energetic cell of a jet to that of the second most energetic cell. It is designed to remove jets which are really just a very noisy or “hot” cell. We require $f_{\text{Hot}} < 10$.
- $n90$ This is the number of calorimeter towers in a jet that contain 90% of the total energy of the jet. This is used to remove noisy jets with unnaturally collimated energy deposition due to coherent noise in a calorimeter tower. We require $n90 > 1$.
- L1Conf A Level 1 trigger confirmation is applied to remove noisy jets. This variable requires two independent data acquisition chains, the L1 calorimeter readout and the final calorimeter precision readout, to yield similar measurements. A value greater than an optimized threshold is required for jets to be accepted. The L1conf is defined as

$$\frac{\sum_{\text{towers}} E_T}{E_T \times (1 - f_{CH})} \quad (3.7)$$

where the E_T sum is over all L1 trigger towers in the jet cone, and the E_T and f_{CH} are derived from the precision readout. We require $L1Conf > 0.4$ (CC,EC) and > 0.2 (ICR).

Jet Energy Scale Calibration

More than just ensuring the quality of reconstructed jets, we must make sure these jets represent the physics objects we are after — the hard-scattered quarks. After scattering, these quarks hadronize and form a jet of particles, and only after showering through the detector do they become the calorimeter jet that we identify (Figure 3.2). To correct the energy of our calorimeter jet back to that of the particle jet, the jet energy scale (JES) calibrations are applied [102] as defined by:

$$E_{\text{particle jet}} = \frac{E_{\text{meas}} - O}{R \times S} \quad (3.8)$$

where $E_{\text{particle jet}}$ is the corrected particle jet energy, E_{meas} is the reconstructed calorimeter jet energy, O is an energy offset, R is the calorimeter response to a jet, and S is the jet cone showering correction. The offset energy comes from the underlying event (other partons not involved in the hard scatter, but fragmented none-the-less), energy pile-up in the detector electronics from a consistently active region, events with multiple hard scatters, typical electronic noise, and noise from the nuclear radioactive decay of the uranium used in the calorimeter. The offset (O) is found by averaging over minimum bias data samples and averaging over ϕ , and is left as a function of detector η and instantaneous luminosity. The calorimeter response (R) is a measure of how efficiently the detector measures the energy deposited in it. It is determined by balancing the transverse energy in back-to-back γ +jet events, where the photon energy has been

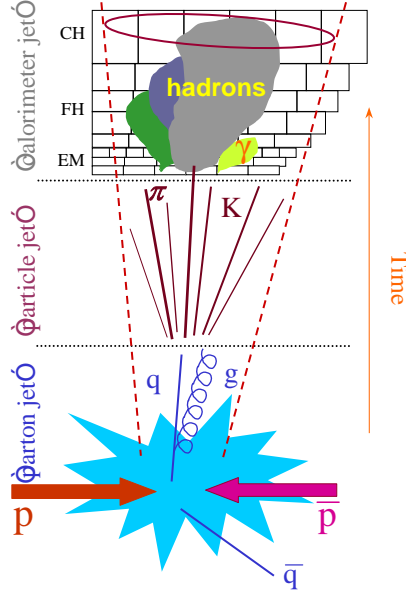


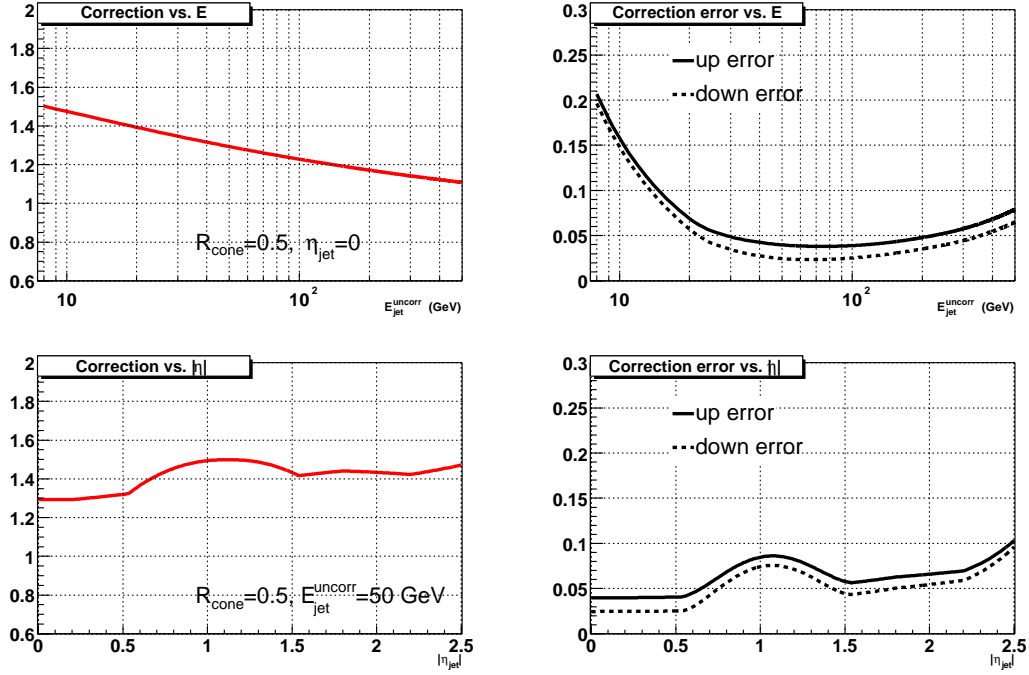
Figure 3.2: The evolution of a “jet”. After the $p\bar{p}$ interaction, a parton will hadronize and form a particle jet. This particle jet will then enter the detector and evolve into a particle shower as it interacts with the detector material. Electromagnetic jets will deposit most of their energy in the first layers of the calorimeter, and the shower from hadronic jets will penetrate deeper into the calorimeter.

measured very accurately, and the unbalanced energy (\cancel{E}_T) represents the inefficiency of the calorimeter. It is derived as a function of the jet energy and detector η . The showering fraction takes into account any energy that has leaked outside of the jet cone. The energy distribution of very isolated jets is studied (in back-to-back γ +jet and dijet events), and the average radius at which all cone energy has been contained is used to determine the showering fraction. The final jet energy scale corrections and their errors are shown in Figure 3.3.

Jet Energy Resolution

The jet energy resolution found in Monte Carlo is better than that found in data. To correct for these differences, the jet energy is smeared in Monte Carlo after the JES has been applied. The energy asymmetry in back-to-back γ +jet and dijet events are used to calculate the jet resolution, σ_{p_T} . A fit is performed in both data and Monte

Jet Energy Scale Correction Factor (DATA)



Jet Energy Scale Correction Factor (MC)

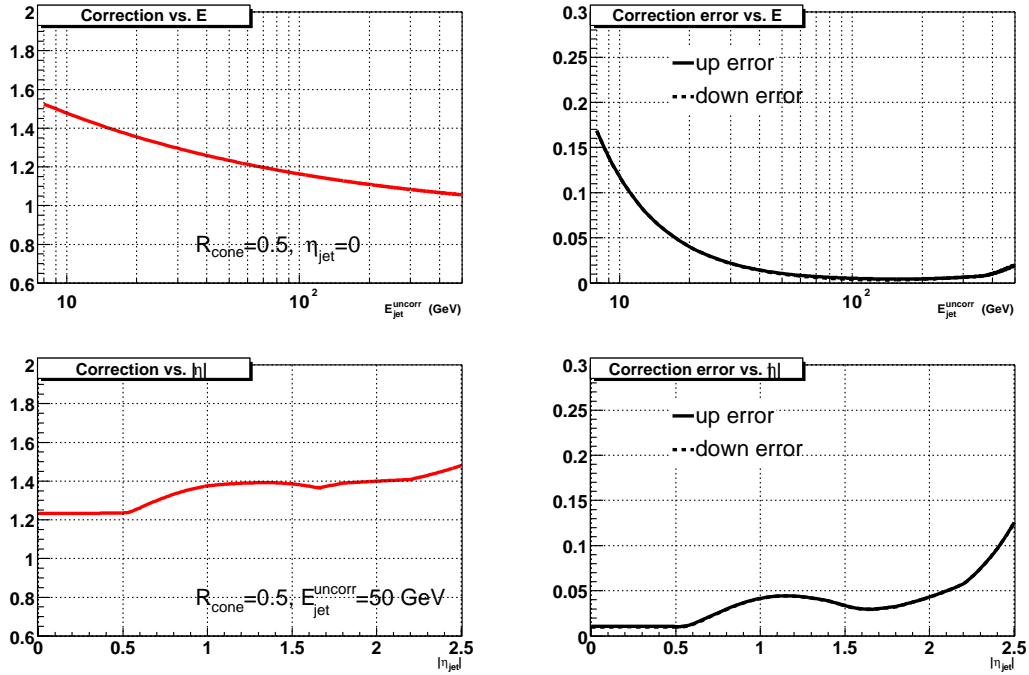


Figure 3.3: The jet energy scale correction factor (left) and its errors (right) are shown as a function of jet p_T (upper) and jet η (lower). The upper four plots are for data, the lower four plots for Monte Carlo.

Carlo (Eqn 3.9), and the appropriate scale factors are applied to Monte Carlo samples to correct the jet energy resolution.

$$\frac{\sigma_{p_T}}{p_T} = \sqrt{\frac{N^2}{p_T^2} + \frac{S^2}{p_T} + C^2} \quad (3.9)$$

The values for the constants N , S , and C are given below as a function of detector η for both monte carlo and data:

	<u>Monte Carlo</u>			<u>Data</u>		
	N	S	C	N	S	C
$0.0 < \eta < 0.5$	4.26	0.658	0.0436	5.05	0.753	0.0893
$0.5 < \eta < 1.0$	4.61	0.621	0.0578	0.0	1.20	0.0870
$1.0 < \eta < 1.5$	3.08	0.816	0.0729	2.24	0.924	0.135
$1.5 < \eta < 2.5$	4.83	0.0	0.0735	6.42	0.0	0.0974

3.4.5 b -Quark Jets

The neural network b -tagging algorithm [98] combines seven variables described here. Results from the tagger are discussed, as well as the loose and tight operating points we apply in this analysis.

Counting Signed Impact Parameter (CSIP) Input Variable

- CSIP Combined = $(6 \times 3s) + (4 \times 2s) + (3 \times 3w) + (2 \times 2w)$

Impact Parameter Significance $S_{IP} = \frac{IP}{\sigma_{IP}}$

CSIP 3s = the number of tracks with $S_{IP} > 3$

CSIP 2s = the number of tracks with $S_{IP} > 2$

CSIP 3w = the number of tracks with $S_{IP} < -3$ and $\Delta\phi(trk, jet) < 1.15$

CSIP 2w = the number of tracks with $S_{IP} < -2$ and $\Delta\phi(trk, jet) < 1.15$

CSIP Combined is the only CSIP tagging variable used in the neural network tagger. It linearly combines the impact parameter significance of several types of tracks into a single discriminating variable.

Jet Lifetime Probability (JLIP) Input Variables

- The positive and negative JLIP probabilities for a jet (Equation 3.10) are calculated from the product of the track probabilities for all tracks inside the jet cone. Only tracks with $IP > 0$ are used in the positive JLIP probability, and tracks with $IP < 0$ are used in the negative JLIP probability.

$$\mathcal{P}_{\text{jet}}^{\pm} = \Pi^{\pm} \times \sum_{j=0}^{N_{\text{trk}}^{\pm}-1} \frac{(-\log \Pi^{\pm})^j}{j!} \quad \text{where} \quad \Pi^{\pm} = \prod_{i=1}^{N_{\text{trk}}^{\pm}} \mathcal{P}_{\text{trk}}(S_{\text{IP}}^{\text{IP}>0}) \quad (3.10)$$

The Track Probability is defined as $\mathcal{P}_{\text{trk}}(S_{\text{IP}}) = \frac{\int_{-50}^{-|S_{\text{IP}}|} \mathcal{R}(s) ds.}{\int_{-50}^0 \mathcal{R}(s) ds.}$, where $\mathcal{R}(s)$ is a fit to the negative half of the signed impact parameter distribution for multijet data or Monte Carlo samples with little or no b -quark jet content.

Secondary Vertex Tagger (SVT) Input Variables

- The decay length significance, $S_{L_{xy}}$, is the variable used by this algorithm to identify a b -quark jet. Once a secondary vertex has been identified, and its corresponding track-jet has been matched to a calorimeter jet, a significance exceeding a threshold declares the jet b tagged. The decay length significance (Eqn. 3.11) is defined as the ratio of the transverse decay length from the primary vertex to the secondary vertex, and the resolution of this distance. The sign comes from the collinearity of the decay length and the momentum of the secondary vertex, i.e. if a secondary vertex and calorimeter jet are both on the same side of the primary vertex, L_{xy} is positive, and if they are found on opposite sides of the primary vertex, L_{xy} is negative. To increase the range of the distributions of the other SVT variables (for the neural network training), a very loose threshold for the significance has been set to identify a tagged jet. Table 3.4 shows the requirements for tracks and secondary vertices used in the SVT algorithm for the typical LOOSE

and TIGHT operating points, and the SUPER LOOSE point used in this neural network tagger.

$$S_{Lxy} = \frac{L_{xy}}{\sigma_{Lxy}} = \frac{|\vec{r}_{SV} - \vec{r}_{PV}|}{\sigma_{Lxy}} \quad (3.11)$$

- $\chi^2_{\text{dof}} = \chi^2$ per degree of freedom of the secondary vertex
- N_{tracks} = the number of tracks used to reconstruct the secondary vertex
- M_{SV} = the mass of the secondary vertex
- N_{SV} = the number of secondary vertices found in the jet

All five SVT variables listed above are used as inputs to the neural network tagger.

Variable	SUPER LOOSE	LOOSE	TIGHT
Secondary Vertex Track Criteria			
$p_T(\text{GeV}/c) >$	0.5	1.0	1.0
Number of SMT Hits $>$	2	2	2
r - ϕ DCA of tracks (cm) \geq	0.15	0.15	0.15
z DCA of tracks (cm) \geq	0.4	0.4	0.4
tracks $\chi^2 <$	15	10	3
IP significance of tracks \geq	0.0	3.0	3.5
Secondary Vertex Criteria			
track-to-vertex $\chi^2 \leq$	15	15	15
vertex $\chi^2 <$	100	100	100
vertex collinearity \geq	0.9	0.9	0.9
vertex decay length (cm) \leq	2.6	2.6	2.6
minimum track multiplicity \geq	2	2	2
b -quark Jet Criteria			
signed decay length significance \geq	2.5	5.0	7.0

Table 3.4: The requirements for secondary vertex tracks, track-jets, and reconstruction for the SUPER LOOSE, LOOSE, and TIGHT versions of the secondary vertex tagger [96, 98].

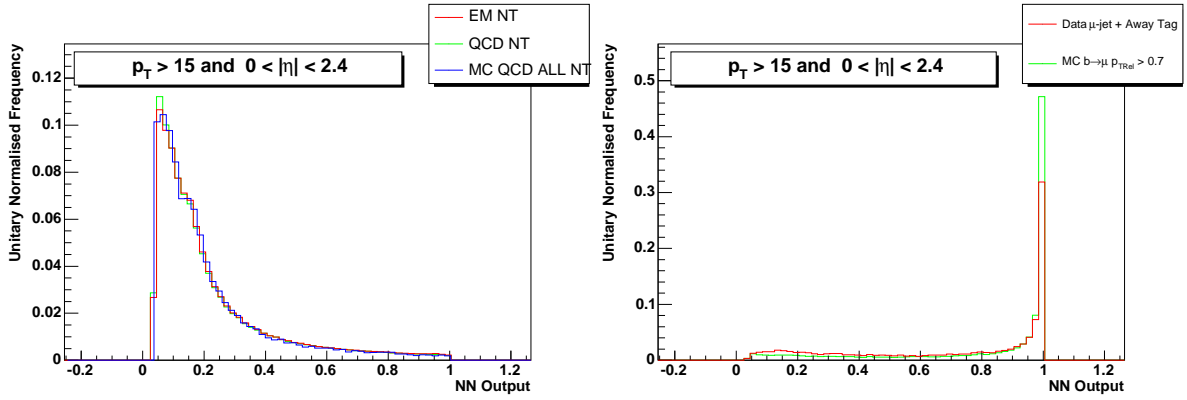


Figure 3.4: The neural network output for light-quark (left) and b -quark (right) jets for both Monte Carlo and data. (NT in the caption refers to a “negative tag”)

Neural Network b -Tagger

Using the seven variables described above, a neural network was trained with light-quark multijet events and b -jet enriched events for both data and Monte Carlo samples. LOOSE, MEDIUM, and TIGHT cut points (Table 3.5) are placed on the neural network output (Figure 3.4) to define b -tagged jets. In this analysis, we apply both the LOOSE and TIGHT versions of the neural network tagger.

	TIGHT	MEDIUM	LOOSE
NN Cut	> 0.775	> 0.65	> 0.5
Fake Rate	0.3	0.5	1.0

Table 3.5: The operating points for the neural network tagger. For each point, the neural network cut value, and the corresponding fake-rate are given [98].

3.4.6 Neutrinos

The neutrino momentum (\cancel{E}_T) must be corrected for all electrons, jets, and muons found in the event. Electron energies have been adjusted by the EM scale, and this adjustment is reflected in the missing transverse energy by adding it vectorially with the \cancel{E}_T . All jet energies have been adjusted by the jet energy scale, but only the response R (Section

3.4.4) of this scaling is compensated in the missing transverse energy. In addition, each muon in the event has left a minimum ionizing energy trail as it traversed the calorimeter (~ 2 GeV), but escaped with most of its energy. To correct the calorimeter \cancel{E}_T , the minimum ionizing energy is removed, and the muon p_T measured by the tracking system is used in the calculation. The final missing transverse energy after EM, JES, and muon corrections is denoted $\cancel{E}_T^{\text{Corr}}$.

3.5 Event Selection

With countless protons smashed to bits, armies of events fighting their way down a triggering gauntlet, and millions finally making it to the safe haven of permanent storage, it is time to select the chosen few. Our search for single top quark production with the leptonic decay of a W boson has led to the set of cuts listed below. This is the only round of selection in the analysis, and all events passing this selection will be used in the more sophisticated discrimination techniques applied later.

Single Top Guest List Requirements

Primary vertex

$$\begin{array}{ll} N_{trks} \geq 3 & \Delta z(\text{DØreco}, \text{DØroot}) < 5 \text{ cm} \\ |z_{\text{vtx}}| < 60 \text{ cm} & \end{array}$$

≥ 2 jets

$$\begin{array}{ll} \text{Leading Jet } E_T > 25 \text{ GeV}, \eta_{det} < 2.5 & f_{Hot} < 10 \\ \text{Second Jet } E_T > 20 \text{ GeV}, \eta_{det} < 3.4 & n_{90} > 1 \\ \text{Other Jets } E_T > 15 \text{ GeV}, \eta_{det} < 3.4 & f_{CH} < 0.4 \\ L1Conf > 0.4 \text{ (CC, EC)}, > 0.2 \text{ (ICR)} & 0.05 < f_{EM} < 0.95 \end{array}$$

≥ 1 b -quark jet

Single Tag: = 1 TIGHT tagged jet (which is also LOOSE by definition)
Double Tag: = 2 LOOSE tagged jets (at least one also being TIGHT)

≤ 2 bad hadronic jets

The number of jets failing the good jet cuts is restricted

Missing transverse energy (\cancel{E}_T)

15 GeV $< \cancel{E}_T < 200$ GeV
15 GeV $< \cancel{E}_T^{\text{Corr}} < 200$ GeV

Triangle cuts ($\Delta\phi(\text{object}, \cancel{E}_T)$ vs. \cancel{E}_T plane)

Leading Jet	Accept if below line	$\Delta\phi(jet, \cancel{E}_T) = \frac{\pi-1.5}{35} \cancel{E}_T + 1.5$
CC Electron	Accept if above line	$\Delta\phi(e, \cancel{E}_T) = -\frac{1.5}{35} \cancel{E}_T + 1.5$
Muon	Accept if above line	$\Delta\phi(e, \cancel{E}_T) = -\frac{1.5}{35} \cancel{E}_T + 1.5$

For the $e + jets$ analysis, = 1 electron, = 0 MEDIUM muons

EM cluster requirements

$ \eta_{\text{det}} < 1.1$ (central calorimeter only)	
$f_{\text{EM}} > 0.90$	$f_{\text{iso}} < 0.15$
$\chi_{\text{EM}}^2 < 50$	$E_T > 15$ GeV

EM tracking requirements

$ \phi_{\text{EM}} - \phi_{\text{trk}} < 0.05$	$ \eta_{\text{EM}} - \eta_{\text{trk}} < 0.05$
track $p_T > 10$ GeV/c	$\Delta z(e_{\text{trk}}, D\phi_{\text{root}}) < 1.0$ cm

EM likelihood

$\mathcal{L} > 0.4$	if $E_T \geq 30$ GeV
$\mathcal{L} > 0.85$	if $E_T < 30$ GeV

For the $\mu + jets$ analysis, = 1 MEDIUM muon, = 0 EM objects

Muon track quality

$$p_T > 15 \text{ GeV}/c$$

$$N_{seg} = 3$$

$$\chi_{trk}^2 < 4.0$$

$$\text{IP}(\mu\text{-trk,PV}) < 0.2 \text{ cm (No SMT hits)}$$

$$\text{IP}(\mu\text{-trk,PV}) < 0.02 \text{ cm (With SMT hits)}$$

Muon cosmic veto

$$\Delta t_{\text{A-scint}} < 10 \text{ ns}$$

$$S_{\text{IP}} < 3$$

$$|z_\mu - z_{\text{PV}}| < 1.0 \text{ cm}$$

$$\Delta t_{\text{BC-scint}} < 10 \text{ ns}$$

$$\chi_{\text{trk}}^2 < 4.0$$

Muon isolation

$$\Delta R(\mu, jet) > 0.5$$

$$\text{Track halo isolation} < 0.15, \quad \text{Cal halo isolation} < 0.15 \quad \text{if } p_T \geq 30 \text{ GeV}/c$$

$$\text{Track halo isolation} < 0.06, \quad \text{Cal halo isolation} < 0.08 \quad \text{if } p_T < 30 \text{ GeV}/c$$

Using the decay processes described in Section 1.4.4 as a guide, the selection criteria have been created. We require a clean, high quality primary vertex. From this primary vertex, all physics object four-vectors are calculated. Mainly two jet events are expected from s -channel decay, and two or three jets from t -channel decay, but we allow for additional jets from initial- and final-state radiation in the event, allowing up to four jets. To remove the large *multijet* background, a harder (higher p_T) and more central (smaller η) leading jet is required, as well as a harder second jet. In addition, a large fraction of the background is removed by requiring at least one b -quark jet tagged by the TIGHT version of the neural network tagger. To aid the multivariate discrimination techniques later applied, the sample is split into single- and double- tagged channels. To increase the signal acceptance in the double tagged channel, the second tagged jet is only required to have a LOOSE tag. To keep both tagged channels orthogonal, the single tagged channel does not allow events with more than one LOOSE tagged jet. A cut is made on the number of bad jets (jets failing the good-jet requirements) in the event to remove periods with excess noise in the calorimeter. A minimum \cancel{E}_T is required to assure a high- p_T neutrino is in the event. A maximum \cancel{E}_T is set to remove events

with coherent noise in the calorimeter, falsely magnifying the neutrino momentum. In the $e + jets$ analysis, we allow exactly one well constructed electromagnetic object, and require a matched track to identify this as an electron. We further veto on events with any MEDIUM quality, isolated muons. The opposite is required in the $\mu + jets$ analysis, requiring exactly one, isolated, MEDIUM quality muon, and no electromagnetic objects. Triangular shaped cuts in the $\Delta\phi(\text{object}, \cancel{E}_T) - \cancel{E}_T$ plane have also been implemented to reduce the large *multijet* background (Fig 3.5). The leading-jet triangle cut removes low \cancel{E}_T events where the \cancel{E}_T and the jet are back-to-back. Physically, these jets have most likely been incorrectly calibrated, and reconstructed with more energy than they actually possess. The electron triangle cut removes low- \cancel{E}_T events where the \cancel{E}_T and the electron are aligned. These cases involve a hadronic jet being misreconstructed as an EM cluster, therefore under-measuring the jet energy and increasing the aligned \cancel{E}_T . Muons may have their track p_T over or under measured, and the muon triangle cuts are used to remove dense regions of this nature.

In Table 3.6, the total number of events in each data sample are given, as well as the numbers before and after tagging. Final samples are split into single tagged events and double tagged events for neural network training and the final calculation of limits.

Data Sample	Total Events	Before Tagging	Selected Events	
			=1 Tag	=2 Tag
$e+jets$	13,411,834	5,351	270	53
$\mu+jets$	6,574,561	4,747	225	32

Table 3.6: The number of events in data samples passing various stages of the analysis. Shown are the total sample size, the numbers of events passing all selection criteria before applying the b -tagging requirement, and those passing with one or two b tags.

TRIANGLE PLOTS - ELECTRON CHANNEL

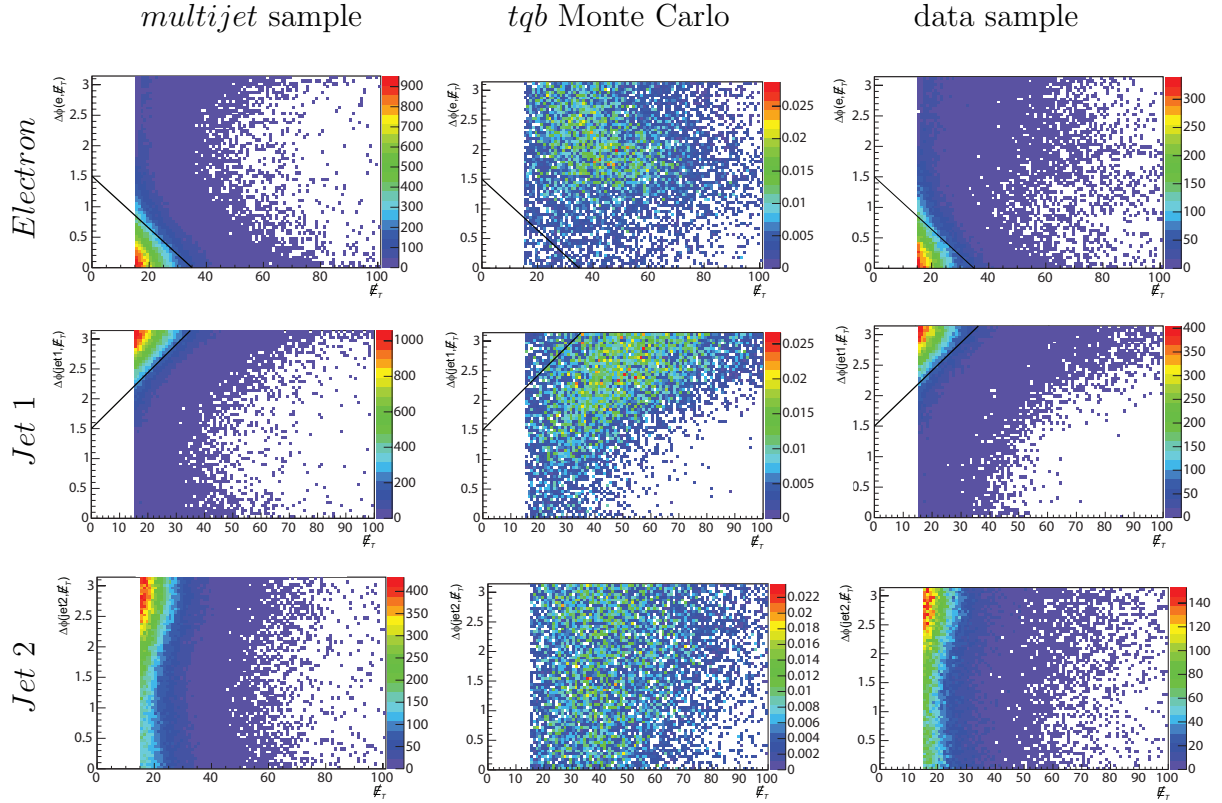


Figure 3.5: Triangle plots ($\Delta R(\text{object}, \cancel{E}_T)$ vs \cancel{E}_T) for the *multijet*, *t*-channel single top, and data samples. Shown are the electron triangle (first row), leading jet triangle (second row), and the second leading jet triangle (third row). The cut region is marked by the black line. These cuts were introduced to remove the abundant multijet background. After application of the lepton and leading jet triangle cuts, a second leading jet cut is unnecessary.

Chapter 4

Modeling Single Top Quark Backgrounds

4.1 Introduction

Having collected and pruned the dataset, it is necessary to accurately model its composition so that a single top signal can be extracted. The well established technique of Monte Carlo simulation is used to model most of the expected backgrounds in the data sample. From the hard scatter, through particle traversal of the detector, and electronics and trigger simulation, each event is painstakingly modeled. Generator simulations are based on the efforts of many particle physics theorists who improve the calculations, iteratively including many new measurements taken by the experimental community. Detector simulations accurately detail every detecting element and amount of material that particles traverse as they radiate from the interaction point. All that is left is to correctly identify the pertinent backgrounds in the analysis, and then begin the CPU-intensive task of generating the events.

4.2 Masters of Disguise

We have identified three main culprits in this analysis, craftily posing as single top: W +jets events, top quark pair production, and misreconstructed multijet events. Each has their own tactics for concealment in our data samples, and must be probed for its individual weaknesses.

W +jets events (Figure 4.1(a)) mimic our single top signal remarkably well, producing two or more jets, and a lepton and \cancel{E}_T from a W boson. Wbb events even pass the b quark jet requirement, and the sizeable cross section of Wjj events allows a number of events with jets misidentified as b 's to remain in the data. The only weaknesses are the lack of a real top quark (which can still be convincingly reconstructed in many events), and the angular correlations between physics objects inherent in a top quark decay, but not present here.

Top quark pair production or $t\bar{t}$ events pose quite a different problem (Figure 4.1(b)). Each event contains real top quarks, b quarks, and W bosons, and they must be filtered by the higher jet multiplicity, and differing \cancel{E}_T and jet p_T distributions in comparison to single top. The $l + jets$ events can be dealt with through their high jet multiplicity and jet p_T distributions. Dilepton events are more easily spotted through their extra lepton, and when that is not reconstructed, the \cancel{E}_T and invariant mass of the W boson are both discriminating factors.

Multijet events or “QCD” events (Figure 4.1(c)) make their way into our data sample through either the misreconstruction of a jet as an EM cluster (electron channel) or b quark pair production with one of the b 's decaying into a muon that swings wide of the jet or the jet is lost, and it thus fakes an isolated muon (muon channel). The exceedingly large production cross section for *multijet* events initially allows many events into the analysis, but the lack of any real top quarks and W bosons helps to beat back this overflowing tide.

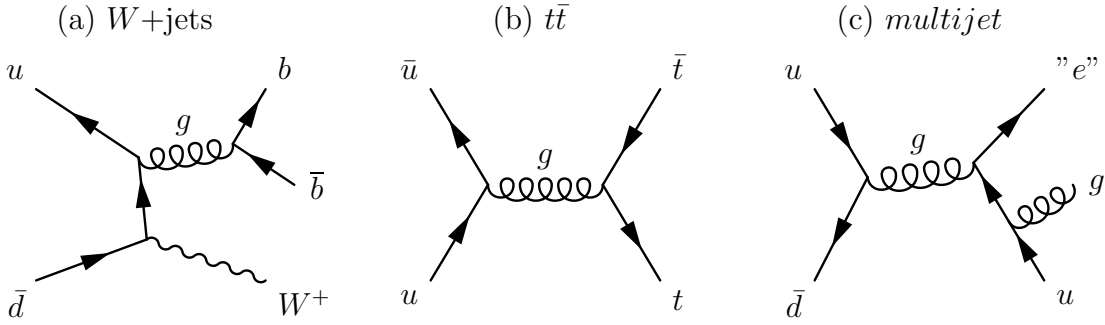


Figure 4.1: The main backgrounds to single top production. Shown are (a) a $Wb\bar{b}$ example from the $W + jets$ background, (b) top pair production, and (c) an example of the *multijet* background for the electron analysis.

In addition to the main backgrounds, there is a small contribution to both channels from diboson events, such as $WW \rightarrow l\nu jj$ and $WZ \rightarrow l\nu jj$.

All of the backgrounds except *multijets* are modeled through Monte Carlo simulations. Although the *multijet* background can be modeled well by Monte Carlo, it is more accurately modeled by real data and is extracted instead from a subset of the $l+jets$ data samples.

4.3 Monte Carlo Simulation

The first stage of Monte Carlo simulation (Figure 4.2) involves creating the hard scattered parton 4-vectors (E, p_x, p_y, p_z) . Based on calculations of the matrix elements for a process, and using the proton parton distribution functions generated by the CTEQ group [104], there are several packages available in the high energy physics community to create these vectors. The relevant ones are mentioned with each sample description in the following section. In addition to creating the parton 4-vectors, all initial and final state radiation is included in the model, as well as the hadronization of the hard scattered partons, and their subsequent 4-vectors. Additionally, the “underlying event”, or the remains of the fragmented proton and antiproton are modeled. Decays of heavier particles such as the b quark and τ lepton are also accurately modeled by other packages.

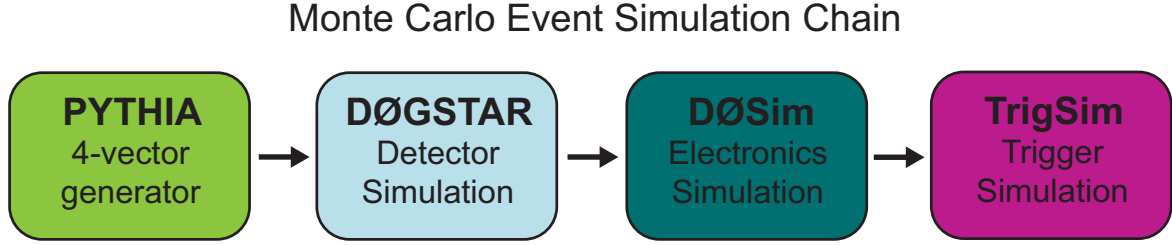


Figure 4.2: The chain of packages used to generate Monte Carlo events. PYTHIA is one example of many 4-vector generators, and in practice, TrigSim is replaced by a trigger efficiency calculation.

To account for events with multiple hard-interactions in the event, additional $p\bar{p}$ interactions are overlaid with the modeled background event, with the number of additional scatters following a poisson distribution with a mean chosen to match that seen in data for a given instantaneous luminosity.

Once the 4-vectors for all involved hadrons have been determined, the traversal of particles through a full detector simulation is begun. The showering of each hadron through all detector material, including the many liquid argon-uranium layers of the calorimeter, is modeled, as well as the energy deposition at each step. A simulation program known as DØGSTAR accomplishes this task, and another package known as DØSim simulates the detector electronics response to the deposited energy. In addition to the deposited energy, electronic noise is added for all detector systems.

Following the electronics simulation, the TrigSim package models the DØ trigger system. Recalling that the L1 system utilizes hardware triggers, and the L2 and L3 systems use software triggers, only the L1 trigger needs additional modeling for Monte Carlo simulation. Once the L1 system has been modeled, real data events and simulated Monte Carlo events have an identical structure, and can be passed through the real online and offline software. Triggered Monte Carlo events are marked, and finally passed through the offline reconstruction algorithms described in Section 3.3. While TrigSim is the ideal approach to finishing the Monte Carlo simulation, it has been difficult to execute in practice and an alternative method has been applied (Section 4.3.1.5).

4.3.1 Correction Factors

After the full simulation, data and Monte Carlo samples look indistinguishable to the offline reconstruction and ID algorithms, yet the generally cleaner and more precise Monte Carlo events must still be modified to better mimic true data events. The object efficiencies and associated correction factors are listed below.

4.3.1.1 Primary Vertex Correction Factors

There is a slight difference in the primary vertex reconstruction efficiency in Monte Carlo and data. Monte Carlo is therefore scaled by the following factor:

- Electron channel primary vertex factor, $\varepsilon_{PV} = 1.004 \pm 0.018$
- Muon channel primary vertex factor, $\varepsilon_{PV} = 0.991 \pm 0.004$

4.3.1.2 Electron Correction Factor

The efficiency to identify an electron candidate depends on the EM cluster's properties, its isolation, a track match, and a likelihood. Monte Carlo efficiencies exceed the data in all respects, and are therefore scaled accordingly. $Z \rightarrow ee$ events are used to derive the efficiencies through a standard tag-and-probe method. Selecting events in a narrow Z -mass window ($80 - 100$ GeV), one electron passing tight selection cuts is labeled the “tag” and the efficiency for the other electron, the “probe”, to pass the cut in question is measured. Background events in the measurement are taken into account by looking at the Z -mass sidebands ($60 - 70$ GeV and $110 - 120$ GeV) and using them for subtraction. The ratios of the data and Monte Carlo efficiencies are used to scale the Monte Carlo.

$$\begin{aligned}
 F_{e-ID} &= \frac{\varepsilon_{\text{Iso,EMF,Hm7}}^{\text{Data}}}{\varepsilon_{\text{Iso,EMF,Hm7}}^{\text{MC}}} \times \frac{\varepsilon_{\text{pT(Trk),}\Delta Z(\text{Trk,PV})}^{\text{Data}}}{\varepsilon_{\text{pT(Trk),}\Delta Z(\text{Trk,PV})}^{\text{MC}}} \times \frac{\varepsilon_{\text{Likelihood}}^{\text{Data}}}{\varepsilon_{\text{Likelihood}}^{\text{MC}}} \\
 F_{e-ID} (E_T^e < 30 \text{ GeV}) &= (0.996 \pm 0.03) \times (0.943 \pm 0.03) \times (0.902 \pm 0.02) \\
 &= 0.847 \pm 0.05
 \end{aligned}$$

$$\begin{aligned}
F_{e\text{-ID}} (E_T^e > 30 \text{ GeV}) &= (0.996 \pm 0.03) \times (0.943 \pm 0.05) \times (0.983 \pm 0.01) \\
&= 0.923 \pm 0.04
\end{aligned}$$

4.3.1.3 Muon Correction Factor

The efficiency to identify a muon candidate involves the following: satisfying the Muon ID requirements, locating a track match, passing a DCA significance cut, matching the primary vertex to the track, and passing tight isolation requirements. Monte Carlo simulation again shows higher efficiencies, and a scale factor is derived. The tag-and-probe method described in the previous section is used to derive the muon cut efficiencies, and the subsequent Monte Carlo scale factor is given below. We note that only one scale factor is given, and that it would have been more accurate to include two, for the p_T region less than 30 GeV and greater than 30 GeV. The subsequent effect on the analysis is within the assigned errors, and this will be addressed in future analyses.

$$\begin{aligned}
F_{\mu\text{-ID}} &= \frac{\varepsilon_{\text{Tracking}}^{\text{Data}}}{\varepsilon_{\text{Tracking}}^{\text{MC}}} \times \frac{\varepsilon_{\chi_{\text{track}}^2}^{\text{Data}}}{\varepsilon_{\chi_{\text{track}}^2}^{\text{MC}}} \times \frac{\varepsilon_{\text{DCA}}^{\text{Data}}}{\varepsilon_{\text{DCA}}^{\text{MC}}} \times \frac{\varepsilon_{\text{TightIso}}^{\text{Data}}}{\varepsilon_{\text{TightIso}}^{\text{MC}}} \times \frac{\varepsilon_{\text{PV}}^{\text{Data}}}{\varepsilon_{\text{PV}}^{\text{MC}}} \\
F_{\mu\text{-ID}} &= 0.988 \times 0.983 \times 0.997 \times 0.985 \times 0.998 \\
&= 0.953 \pm 0.050
\end{aligned}$$

4.3.1.4 Jet Corrections

Energy corrections and deviations between Monte Carlo and data are taken into account by the jet energy scale (JES) and jet energy resolution (JER) described in Section 3.4.4.

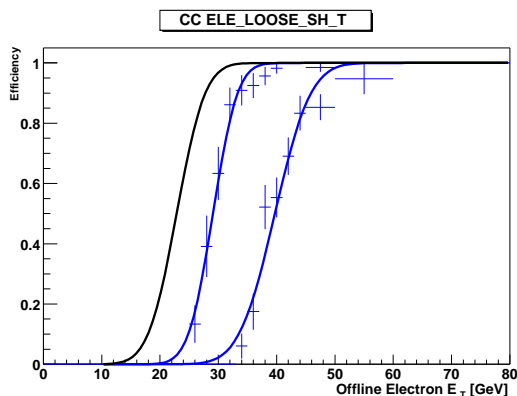


Figure 4.3: A sample L3 trigger turn-on curve for electrons. Data points and their fit are shown for a 25 and 35 GeV p_T threshold. The extrapolated curve for a 15 GeV threshold is also shown.

4.3.1.5 Trigger Efficiency

As an alternative to the TrigSim package for Monte Carlo trigger simulation, the single-object L1, L2, and L3 trigger efficiency curves are calculated as a function of p_T in unbiased data samples, and folded together to simulate the multi-object triggers used in the analysis. These efficiencies, or turn-on curves, are determined for each trigger list and used to weight each Monte Carlo event [105]. The probability for a single electron, muon, or jet to pass a trigger requirement is the product of the conditional probabilities of each trigger level:

$$P(L1, L2, L3) = P(L1) \cdot P(L2|L1) \cdot P(L3|L1, L2) \quad (4.1)$$

The probability for an event to fire a given trigger is the product of individual object probabilities:

$$P(object1, object2) = P(object1) \cdot P(object2) \quad (4.2)$$

For $e + jets$ triggers, the L1, L2, and L3 electron efficiencies are measured in data samples constructed from triggers involving at least one EM object. A subset of $Z \rightarrow ee$ events are then located in this sample, and the efficiencies are found using the “tag-and-

probe” method for the L1, L2, and L3 trigger conditions. The efficiencies are measured independently for each trigger-list version, and a luminosity-weighted average is applied to Monte Carlo events. At L1, the trigger efficiency is found to be a constant for $E_T > 15$ GeV, and nearly identical for all trigger-list versions, $\epsilon_{L1} = 0.99$. At L2, the trigger is fully efficient, $\epsilon_{L2} = 1.00$, and at L3, the trigger efficiency is found to have a sharp turn-on as a function of lepton transverse energy, $\epsilon_{L3}(E_T)$. With the restriction of all electrons in our data sample having $E_T > 15$ GeV, measuring the turn-on curve for a 15 GeV L3 trigger must be done by extrapolation. First, turn-on curves are derived for these electrons with L3 trigger thresholds of 35, 30, 25, and 20 GeV. The fit parameters from each trigger threshold are then used to extrapolate down to the values for a 15 GeV trigger. The electron L3 trigger efficiency takes the form:

$$\begin{aligned}\epsilon_{L3}(E_T) &= 0.5 A \times \left(1 + \text{Erf}\left(\frac{E_T - B}{\sqrt{2}C}\right)\right) \\ \text{Erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du\end{aligned}\tag{4.3}$$

where A , B , and C are parameters of the fit, and $\text{Erf}(x)$ is the standard error function. Example 35 GeV, 25 GeV, and the extrapolated 15 GeV turn-on curves are shown in Figure 4.3. On average, the L3 triggers reach a maximum efficiency of 0.97 at 2 GeV above their threshold setting.

For $\mu + jets$ triggers, the muon efficiencies are again measured from unbiased data samples, collected with EM and jet triggers. For muons above 15 GeV/ c , the L1 and L2 triggers are found to be constants as a function of p_T , but show a distinct detector η dependance. They are parameterized as shown in Equation 4.4, where a , b , c , and d are parameters of the fit, and the efficiency is set to zero for the uninstrumented regions at the base of the detector. The L1 efficiency peaks at 0.99 in the central and end caps of the calorimeter, and dips down to 0.90 in the inter-cryostat region. The L2 efficiency

peaks at 0.98 for forward detector η , and smoothly drops to 0.95 at detector $\eta = 0$. The L3 trigger efficiency is fully efficient for all triggers, $\epsilon_{L3} = 1.00$.

$$\begin{aligned}\epsilon_{L1,L2}(\eta) &= a + b e^{-c(\eta^2 - d^2)} \sin(\eta^2 - d^2) \\ &= 0 \quad \text{for } |\eta| < 1.1 \text{ and } 4.25 < \phi < 5.15\end{aligned}\tag{4.4}$$

Jet trigger efficiencies are measured in two data samples, one triggered by muons, and another with single electrons confirmed offline. The efficiency is measured in three detector η regions ($|\eta| < 0.8$, $0.8 \leq |\eta| < 1.5$, $|\eta| \geq 1.5$) as a function of jet p_T and trigger version. The same functional form as the L3 electron trigger efficiency is used (Equation 4.3). For jets with $p_T > 15$ GeV, the L1 trigger efficiency is about 0.99 for all regions, however, this peak value is not reached in the v12 EC until 20 GeV, and for earlier trigger versions, 23 GeV (CC/ICD) and ~ 30 GeV (EC). The L2 trigger efficiencies are nearly fully efficient, $\epsilon_{L2} \simeq 1.00$ for all regions. The L3 trigger efficiencies range from 0.96–0.99, and the maximum turn-ons are reached about 10 GeV above the L3 trigger threshold. The efficiency for a jet to pass an EM trigger is also measured, and used to correct the EM trigger probabilities. The fit to this efficiency takes one of two forms:

$$f(E_T) = (a + b \cdot E_T) \cdot \left(1 + \text{Erf}\left(\frac{E_T - c}{\sqrt{E_T} \cdot d + e}\right)\right)\tag{4.5}$$

$$f(E_T) = (a + b \cdot E_T) \cdot \left(1 + \text{Erf}\left(\frac{E_T - c + f \cdot E_T}{\sqrt{E_T} \cdot d + e}\right)\right)\tag{4.6}$$

where a–f are parameters of the fit.

The final trigger weight applied to a Monte Carlo event is thus a multiplicative factor of all object L1, L2, and L3 efficiencies, with care taken in adding the probability for a jet to trigger an EM trigger. To take into account the variations in the trigger efficiency in data from each trigger list, the luminosity-averaged values or averaged functions are

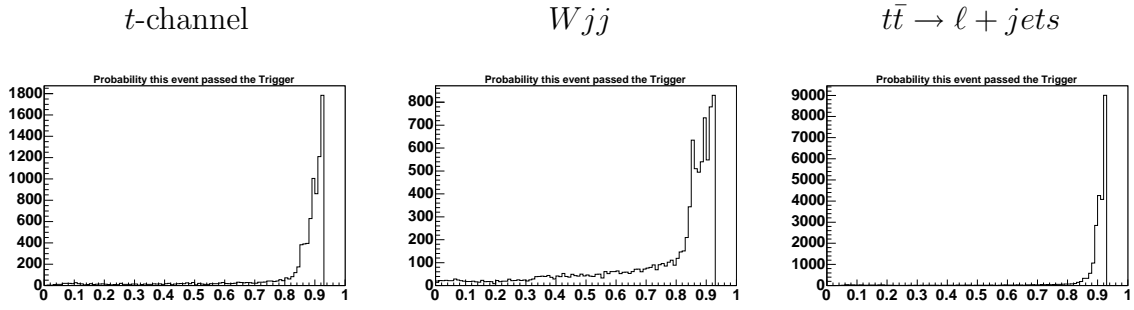


Figure 4.4: Sample distributions of the trigger weight applied to events in the electron channel for t -channel signal (left), Wjj (center), and $t\bar{t} \rightarrow \ell + jets$ (right) Monte Carlo samples.

applied to Monte Carlo. Sample event trigger weight distributions are shown in Figure 4.4 and the average event trigger weights for all samples are given in Table 4.1.

Average Event Trigger Weights		
	Electron	Muon
	Before tagging	Before tagging
tqb	0.83	0.83
tb	0.83	0.84
Wbb	0.78	0.78
Wjj	0.75	0.76
$t\bar{t} \rightarrow \ell + jets$	0.88	0.87
$t\bar{t} \rightarrow \ell\ell$	0.86	0.87
WW	0.79	0.83
WZ	0.80	0.84

Table 4.1: The average event trigger weights applied to Monte Carlo events in the electron and muon samples before tagging.

4.3.1.6 Tag-Rate Functions

The various b -taggers which compose the neural network tagger are not modeled well enough to directly compare Monte Carlo results with data — in turn, the neural network tagger inherits this feature. Monte Carlo tagging efficiencies are higher than data, show sample dependence, and variation as a function of jet E_T and detector η . To deal with these issues, Taggability Functions and Tag Rate Functions (TRFs) are derived from data and Monte Carlo samples to calculate the probability that each jet in an event

is tagged by the neural network tagger. Taggability Functions measure the probability that a jet has a *chance* of being tagged. A jet is “taggable” if it has a track-jet and a calorimeter-jet within $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.5$. TRFs are then used to calculate the probability that a taggable jet *is* tagged by the tagging algorithm. TRFs are derived for b , c and light quark jets. Each Monte Carlo Taggability Function and TRF is scaled to a data equivalent to make the samples directly comparable. While neural network tagging in data is a boolean parameter (tagged or untagged), Monte Carlo events are weighted by the probability that the event has one or two b -tagged jets. Thus, events passing all other event selection criteria are given the appropriate b -tag weight before use in the final analysis.

The data TRF is calculated from a μ +jets data sample, where the muon is found inside of a jet cone and is expected to have come from the decay of a b quark. These muon-jets are required to be “taggable,” and a taggability function is derived as a function of jet p_T and η (Figure 4.5). The TRF is then calculated. Since the true b -quark fractions in the data sample are not known, the tagging efficiencies are calculated by playing two independent b -tagging methods off of one another, the neural network tagger applied in this analysis, and a semi-leptonic b -tagger that looks for a muon from the b -quark decay to tag a jet. Through a system of eight equations and eight unknowns, the efficiencies of both taggers are solved by studying their tag rates in the $\mu + jets$ sample, and in a b -enriched subset of this sample with back-to-back b jets. The tagging efficiencies are further parameterized as a function of jet E_T and detector η separately.

Monte Carlo b -tagging efficiencies are derived from various samples that include b -quark jets ($b\bar{b}$, $t\bar{t}$, $Z \rightarrow b\bar{b}$, $Z \rightarrow b\bar{b} \rightarrow \mu X$). Taggable jets require the usual calorimeter jet and track jet quality cuts, and must also have a b quark within a radius $R = 0.5$ of the jet. The Taggability Function is derived as a function of the jet p_T and η (Figure 4.5). The neural network tagging algorithm is then directly applied and the TRF is calculated as a function of jet E_T and detector η separately (Figure 4.6).

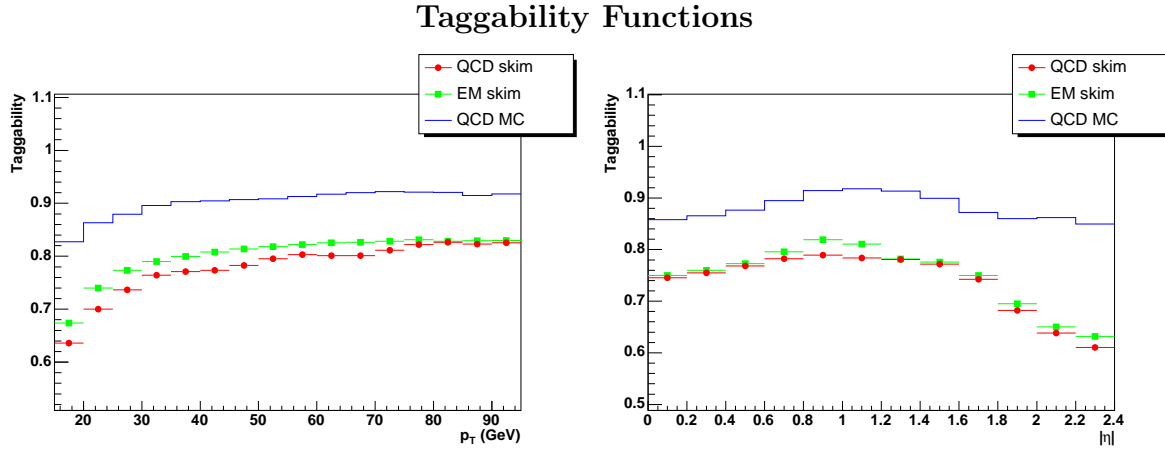


Figure 4.5: The taggability functions as a function of jet p_T (left) and jet η (right), shown for *multijet* (QCD) and “EM” data samples, and a Monte Carlo *multijet* (QCD) sample.

In addition to identifying b jets, the neural network algorithm correctly identifies some c quark decays. These decays typically have a shorter decay length, making them harder to resolve from the primary vertex, and lower in overall tagging efficiency. Again, the efficiency for taggable jets (this time with a matched c quark) to be tagged is calculated as a function of jet E_T and detector η (Figure 4.7).

Once all Taggability Functions and TRFs are derived, a two-dimensional parameterization is made for each one based on the observed E_T and detector η dependance, and the average overall efficiency of the sample. This parameterization takes the form:

$$\epsilon(p_T, \eta) = \frac{\epsilon(p_T) \epsilon(\eta)}{\epsilon_{\text{avg}}} \quad (4.7)$$

A 2D function is derived from the data sample, and an averaged 2D function is taken for all Monte Carlo samples that have been studied. To correct for Monte Carlo differences from data, a 2D function is derived to scale the monte carlo down to the data for both Taggability and Tag Rates, $F(E_T, \eta) = \frac{\epsilon_{\text{data}}(E_T, \eta)}{\epsilon_{\text{MC}}(E_T, \eta)}$. The product of these two final, scaled efficiency ratios, is then applied to each Monte Carlo jet. Monte Carlo jets encompassing a b quark have TRF_b applied, and those with a c quark have TRF_c applied.

b -quark Jet Tagging Efficiencies (TRF_b)

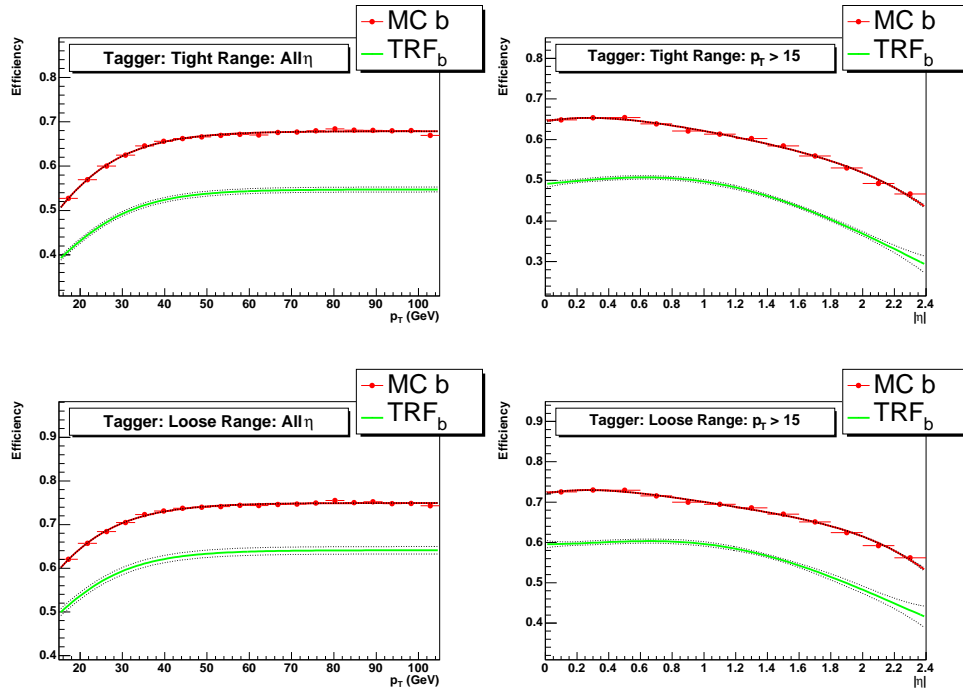


Figure 4.6: The b -quark jet tagging efficiency (TRF_b) applied to jets in Monte Carlo samples. The Monte Carlo TRF is shown before (red) and after (green) scaling to the data TRF. Both the E_T (left) and detector η (right) dependences are shown, as well as the tight (upper) and loose (lower) versions of TRF.

c -quark Jet Tagging Efficiencies (TRF_c)

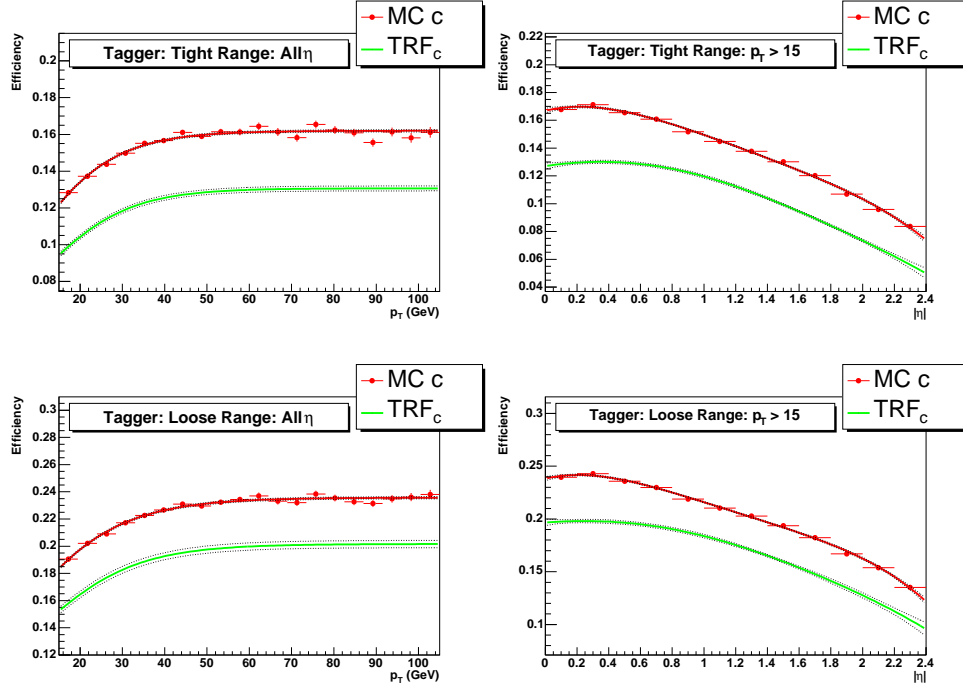


Figure 4.7: The c -quark jet tagging efficiency (TRF_c) applied to jets in Monte Carlo samples. The Monte Carlo TRF is shown before (red) and after (green) scaling to the data TRF. Both the p_T (left) and detector η (right) dependences are shown, as well as the tight (upper) and loose (lower) versions of TRF.

Fake ($light$ -jet) Tagging Efficiencies (TRF_{fake})

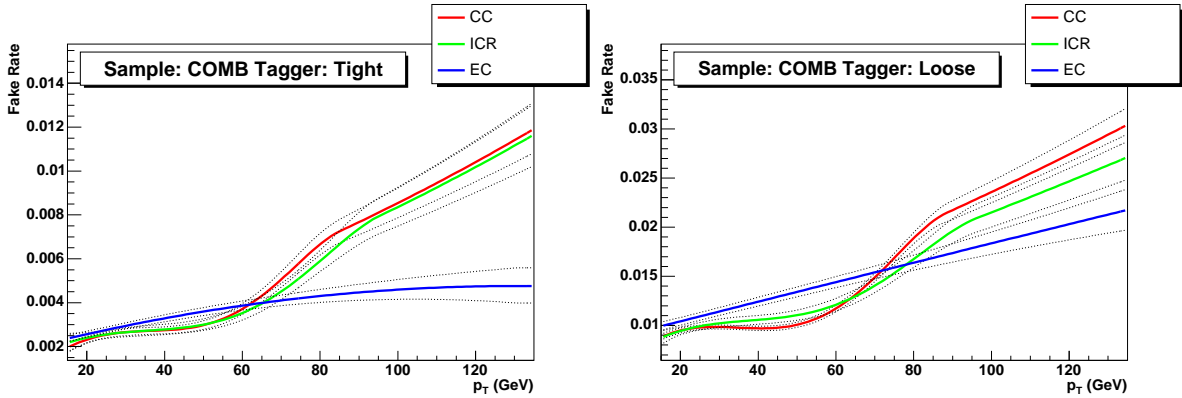


Figure 4.8: The fake, or $light$ -jet tagging efficiency (TRF_{fake}) applied to jets in Monte Carlo samples. The TRF is shown for the CC, ICR, and EC regions of the calorimeter as a function of jet E_T . The TIGHT (left) and LOOSE (right) versions of the TRF are shown.

There is a small fraction of light quark jets which, through detector resolution and noise effects, are mistagged by the neural network tagging algorithm. There is also a legitimate source of muons from strange meson decays (such as the $K^+ \rightarrow \mu^+ \nu_\mu$), but these are not often reconstructed because they are softer, and may not decay until they have passed beyond the inner tracking ($c\tau_{K^+} = 3.7$ m [48]). These strange meson decays are taken into account with the light jet mistag rate. The mistag rate is inferred from the expected symmetry of the signed decay length significance (Equation 3.11) about zero. This rate is derived from light quark jets in data. Biases from b - and c -quark contamination in the light quark jet sample are corrected for with a Monte Carlo based scale factor. The final parameterization is a function of jet E_T in three detector η regions, CC, ICR, EC (Figure 4.8). Mistag rates for the LOOSE and TIGHT versions of the tagger are given in Table 4.2.

Mistag Efficiency per jet (%)		
	LOOSE	TIGHT
<i>light</i> -jet TRF (Data)		
CC	1.03 ± 0.10	0.289 ± 0.04
ICR	1.06 ± 0.09	0.288 ± 0.03
EC	1.19 ± 0.06	0.299 ± 0.02

Table 4.2: The mistagging efficiencies for the LOOSE and TIGHT versions of the neural network tagger, shown for each calorimeter region.

4.3.2 Normalization

To allow for direct comparison between Monte Carlo and data, the Monte Carlo samples must be normalized appropriately. The Monte Carlo sample production cross section (σ), branching fraction (\mathcal{B}), and number of generated events in the initial sample (N_{initial}) all contribute to this factor. In addition, the sample must be normalized to the integrated luminosity (\mathcal{L}) of the data sample. Monte Carlo samples are normalized by the following factor:

$$F_{\text{norm}} = \mathcal{L}_{\text{data}} \times \frac{\sigma \mathcal{B}}{N_{\text{initial}}} \quad (4.8)$$

To better model the fraction of *multijet* and *W+jets* events that make up the data sample after all event selection (before *b*-tagging), we calculate the fractions of real and fake *W* bosons in the data using the so-called “Matrix Method” (Section 4.6). This method provides the sole normalization factor for the *multijet* sample and an additional one for the *W + jets* samples (*Wjj*, *Wbb*) on an event-by-event basis.

4.4 Monte Carlo Samples

4.4.1 Single Top Signal Samples

The single top samples were produced using a modified version of CompHEP[49] dubbed “SingleTop”. The model incorporates the CTEQ6M parton distribution functions, and sets $Q_W^2 = (M_t/2)^2$ for the *t* channel, and $Q_W^2 = M_t^2$ for the *s* channel. PYTHIA 6.221 [52] was used for ISR, FSR, the underlying event, and jet showering. EVTGEN 9.39.02 [53] was used for *b*-quark decays and TAUOLA 2.5.04 [54] was used for τ decays. PYTHIA was also used for the overlay of additional $p\bar{p}$ interactions, with a mean value of 0.5 for the poisson distribution. Sample statistics are given in Table 4.3.

4.4.2 Top Pair Production Samples

Two $t\bar{t}$ samples have been produced, a $t\bar{t} \rightarrow \ell + jets$ sample, and a $t\bar{t} \rightarrow \ell\ell$ sample. Both have been created by the ALPGEN 1.2.1 [55] package, utilizing the CTEQ6.1M parton distribution functions, and a $Q_W^2 = M_t^2$. PYTHIA 6.202 [52] was used for ISR, FSR, the underlying event, *t* and *W* decay, and jet showering. EVTGEN 9.39.02 [53] was used for

b decay and TAUOLA 2.5.04 [54] for τ decay. ALPGEN [55] further includes the top quark spin information for production and decay. PYTHIA was also used for the overlay of additional $p\bar{p}$ interactions, with a mean value of 0.8 for the poisson distribution. Sample statistics are given in Table 4.3.

4.4.3 W +jets Samples

A Wjj sample ($j = u, d, s, g, c$) and a Wbb sample have been produced for the analysis. The hard scatter was produced by the ALPGEN 1.1 [55] package using the CTEQ6.1M parton distribution function, and $Q_W^2 = (M_W/2)^2$. PYTHIA 6.202 [52] was applied for ISR, FSR, the underlying event, W decay, and jet showering (PYTHIA internally used CTEQ5L). EVTGEN 9.39.02 [53] was used for b quark decay and TAUOLA [54] for τ decay. Parton level cuts have been applied (for u,d,s,g,c,b), requiring $p_T > 12$ GeV/c and $|\eta| < 3.5$, $\Delta R(j, j) > 0.4$, lepton $p_T > 12$ GeV and detector $|\eta| < 2.7$, and neutrino $p_T > 8$ GeV. PYTHIA was also used for the overlay of additional $p\bar{p}$ interactions, with a mean value of 0.5 for the poisson distribution. Sample statistics are given in Table 4.3.

4.4.4 Diboson Samples

The diboson samples, $WW \rightarrow l\nu jj$ and $WZ \rightarrow l\nu jj$, were produced with the same packages and parameters as the W +jets samples. Sample statistics can be found in Table 4.3.

4.5 Data Samples

The *multijet* background has been derived from a data sample. In the electron channel, this background is derived from the $e + jets$ sample with events where the EM object

Monte Carlo Samples				
Event Type	Cross Section [pb]	Branching Fraction	Numbers of Events	Int. Lum. [fb ⁻¹]
Signals				
$tqb \rightarrow e + \text{jets}$	1.98 ± 0.30	0.1309 ± 0.0026	33,000	127
$tqb \rightarrow \mu + \text{jets}$	1.98 ± 0.30	0.1304 ± 0.0026	31,500	122
$tb \rightarrow e + \text{jets}$	0.88 ± 0.14	0.1309 ± 0.0026	32,500	282
$tb \rightarrow \mu + \text{jets}$	0.88 ± 0.14	0.1304 ± 0.0026	31,000	270
Backgrounds				
$t\bar{t} \rightarrow l + \text{jets}$	6.8 ± 1.2	0.4444 ± 0.0089	191,300	65
$t\bar{t} \rightarrow ll$	6.8 ± 1.2	0.1111 ± 0.0089	97,750	131
$Wbb \rightarrow e\nu bb$	30.2 ± 5.40	0.1111 ± 0.0022	99,500	30
$Wbb \rightarrow \mu\nu bb$	30.2 ± 5.40	0.1111 ± 0.0022	99,000	30
$Wjj \rightarrow e\nu jj$	$2,580 \pm 460$	0.1111 ± 0.0022	189,500	0.7
$Wjj \rightarrow \mu\nu jj$	$2,580 \pm 460$	0.1111 ± 0.0022	188,000	0.7
$WW \rightarrow l\nu jj$	6.80 ± 0.15	0.3928 ± 0.0079	23,000	9
$WZ \rightarrow l\nu jj$	2.10 ± 0.19	0.3928 ± 0.0079	23,000	28

Table 4.3: The cross sections, branching fractions, initial numbers of events, and integrated luminosities of the Monte Carlo event samples. A top quark mass of 175 GeV is used for all samples.

has a very low electron likelihood ($\mathcal{L} < 0.05$) and a poor or missing track, and $\cancel{E}_T < 15$ GeV. It is expected that these clusters are low energy or poorly reconstructed hadronic jets, and the low \cancel{E}_T signifies the lack of a real neutrino. The sample contains 24,741 events after selection, before b tagging. The muon channel derives its *multijet* background from the $\mu + \text{jets}$ data sample by removing the muon isolation requirement, requiring anti-isolation (Track Halo Isolation > 0.15 and Calorimeter Halo Isolation > 0.15), and requiring $\cancel{E}_T < 15$ GeV. The sample contains 43,552 events after selection, before b -tagging.

4.6 Background Fraction Estimation – The Matrix Method

To estimate the fraction of real and fake lepton events that make up the data, we apply the so-called “Matrix Method” [106, 107]. While we have acceptable models for both

the real lepton events (Monte Carlo), and fake lepton events (data), their fractions in the data sample must be determined. The matter is complicated by the fact that some real leptons can leak into the fake-lepton data sample, and vice versa. The lepton fractions are evaluated in the data sample before b -tagging to (a) take advantage of the increased statistics and (b) not bias or destroy our single top search capabilities in the final b -tagged samples. Only the W +jets and *multijet* samples are scaled by the matrix method, as the modeled normalizations of the other real lepton samples are sufficiently understood.

In both the e and μ channels, there is a single cut that defines a real lepton, and distinguishes the “loose” and “tight” samples (where tight means the complete event selection without b tagging, and Loose further removes the lepton ID cut). The probabilities for real (P_{real}) and fake (P_{fake}) leptons to pass this cut are measured in samples independent of our main data sample, and are then used to estimate the content in the data sample. The matrix method starts with a few basic ideas concerning the number of events in a sample:

$$N_{\text{loose}} = N_{\text{loose}}^{\text{fake}} + N_{\text{loose}}^{\text{real}} \quad (4.9)$$

$$\begin{aligned} N_{\text{tight}} &= N_{\text{tight}}^{\text{fake}} + N_{\text{tight}}^{\text{real}} \\ &= P_{\text{fake}} N_{\text{loose}}^{\text{fake}} + P_{\text{real}} N_{\text{loose}}^{\text{real}} \end{aligned} \quad (4.10)$$

Given these equations, one can solve for the number of events that pass the tight selection in the real and fake samples:

$$N_{\text{tight}}^{\text{fake}} = P_{\text{fake}} \frac{P_{\text{real}} N_{\text{loose}} - N_{\text{tight}}}{P_{\text{real}} - P_{\text{fake}}} \quad (4.11)$$

$$N_{\text{tight}}^{\text{real}} = P_{\text{real}} \frac{N_{\text{tight}} - P_{\text{fake}} N_{\text{loose}}}{P_{\text{real}} - P_{\text{fake}}} \quad (4.12)$$

With these equations, we are equipped to estimate the fractions of real and fake lepton events that make up our data sample, and apply the appropriate scale factors. This calculation is discussed separately for each lepton channel.

4.6.1 Electron Calculation

The cut separating the loose and tight samples is the electron likelihood, \mathcal{L} . N_{loose} and N_{tight} are obtained by simply counting the number of events in the loose and tight samples.

The probability for real electrons to pass the likelihood cut, P_{real} , is measured using the DIEM sample. This sample is expected to contain $Z \rightarrow ee$ events and is verified through the reconstructed invariant mass of the Z boson. A standard tag-and-probe method is used:

- Select events passing primary vertex quality cuts, and containing two electron candidates passing all requirements but the likelihood.
- Require one of the candidates to pass the likelihood requirement (tag) and calculate the efficiency for the other (probe) to pass the likelihood cut.
- To ensure an enriched sample of true $Z \rightarrow ee$ events, only events in an invariant mass window of 80 – 100 GeV are used. Events in the 60 – 70 GeV and 110 – 120 GeV side bands are used to estimate and subtract background events from the 80 – 100 GeV window.

The real-electron probabilities are parameterized as a function of electron E_T . Detector η of the electron and the number of jets in the event were also studied, and no strong dependance was found.

$$\begin{array}{ll} P_{\text{real}}(e) = 0.830 \pm 0.042 & E_T^e < 30 \text{ GeV} \\ P_{\text{real}}(e) = 0.826 + 0.003 \times E_T^e \pm 5\% & 30 \text{ GeV} < E_T^e < 45 \text{ GeV} \\ P_{\text{real}}(e) = 0.980 \pm 0.049 & E_T^e > 45 \text{ GeV} \end{array}$$

The fake-electron probability is measured in the $e + jets$ dataset, using a low- \cancel{E}_T sample orthogonal to the data sample. Requiring $\cancel{E}_T < 15$ GeV, we do not expect many real W bosons in these events, and therefore expect a highly enriched sample of fake

electrons. The probability for these events to pass the electron likelihood is then directly calculated. The fake-electron probabilities were parameterized as a function of electron E_T . The electron detector η and numbers of jets in the event were also studied, and no strong dependance was found.

Trigger List v8–11	$P_{\text{fake}}(e) = 0.492 - 0.018 \times E_T^e \pm 10\%$	$E_T^e < 20 \text{ GeV}$
	$P_{\text{fake}}(e) = 0.101 \pm 0.010$	$20 \text{ GeV} < E_T^e < 30 \text{ GeV}$
	$P_{\text{fake}}(e) = 0.190 \pm 0.019$	$E_T^e > 30 \text{ GeV}$
Trigger List v12–13	$P_{\text{fake}}(e) = 0.751 - 0.030 \times E_T^e \pm 10\%$	$E_T^e < 20 \text{ GeV}$
	$P_{\text{fake}}(e) = 0.151 \pm 0.015$	$20 \text{ GeV} < E_T^e < 30 \text{ GeV}$
	$P_{\text{fake}}(e) = 0.282 \pm 0.028$	$E_T^e > 30 \text{ GeV}$

To check the accuracy of our real and fake probabilities, we plot the W transverse mass in the data sample (before b -tagging), and the composition in each bin as predicted by the matrix method (Figure 4.9). As expected, fake electrons cluster towards a low, unphysical W boson mass, and real electrons peak about the true W boson mass.

With the measured probabilities, the matrix method is applied, and the following values attained:

<u>Electron Channel Matrix Method Numbers</u>			
	N_{loose}	$N_{\text{loose}}^{\text{real}}$	$N_{\text{loose}}^{\text{fake}}$
$E_T^e < 30 \text{ GeV}$	4422 ± 67	1117 ± 111	3305 ± 85
$E_T^e > 30 \text{ GeV}$	6827 ± 83	3181 ± 261	3646 ± 150
	N_{tight}	$N_{\text{tight}}^{\text{real}}$	$N_{\text{tight}}^{\text{fake}}$
$E_T^e < 30 \text{ GeV}$	1402 ± 37	927 ± 60	475 ± 46
$E_T^e > 30 \text{ GeV}$	3949 ± 63	3062 ± 143	887 ± 128

The matrix method scale factors listed below are applied to the *multijet* and W +jets samples. The *multijet* scale factor is left as a function of the electron fake efficiency. The yields (\mathcal{Y}) in the $W + jets$ scale factor are for samples after event selection but excluding b -tagging, and are defined in Section 4.7.2.

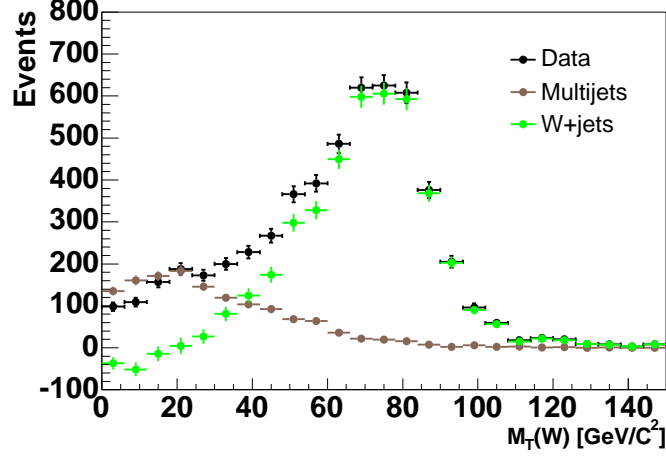


Figure 4.9: The reconstructed W transverse mass in electron channel data (before b -tagging), and the matrix method predicted composition in each bin.

$$\begin{aligned}
 F_{multijet} &= \frac{N_{\text{tight}}^{\text{fake}}}{N_{multijet}} & F_{W+jets} &= \frac{N_{\text{tight}}^{\text{real}} - (\mathcal{Y}_{ljets} + \mathcal{Y}_{ll} + \mathcal{Y}_{WW} + \mathcal{Y}_{WZ})}{\mathcal{Y}_{Wjj} + \mathcal{Y}_{Wbb}} \\
 &= \frac{P_{\text{fake}}(E_T, \text{trig}) N_{\text{loose}}^{\text{fake}}}{N_{multijet}}
 \end{aligned}$$

Electron Channel Matrix Method Scale Factors

Trigger List v8-11	$F_{multijet}$	$=$	$0.456 \times P_{\text{fake}}(e)$	$E_T^e < 30 \text{ GeV}$
	$F_{multijet}$	$=$	$0.208 \times P_{\text{fake}}(e)$	$E_T^e > 30 \text{ GeV}$
Trigger List v12-13	$F_{multijet}$	$=$	$0.439 \times P_{\text{fake}}(e)$	$E_T^e < 30 \text{ GeV}$
	$F_{multijet}$	$=$	$0.238 \times P_{\text{fake}}(e)$	$E_T^e > 30 \text{ GeV}$
Trigger List All	F_{W+jets}	$=$	1.196	All E_T^e

4.6.2 Muon Calculation

The cuts separating the muon loose and tight samples are the isolation cuts, Track Halo Isolation and Calorimeter Halo Isolation. N_{loose} and N_{tight} are obtained by simply counting the numbers of events in the loose and tight samples.

The probability for real muons to pass the isolation cuts, P_{real} , is measured using the DIMU data sample. This sample is expected to contain $Z \rightarrow \mu\mu$ events, which is verified by checking the reconstructed invariant mass of the Z boson. The same steps are taken as in the electron calculation. We find the following real muon probabilities as a function of the number of jets in the event:

$E_T^\mu < 30 \text{ GeV}$	$P_{\text{real}}(\mu) = 0.690 \pm 0.001$	2 Jet Events
	$P_{\text{real}}(\mu) = 0.651 \pm 0.002$	3 Jet Events
	$P_{\text{real}}(\mu) = 0.675 \pm 0.005$	4 Jet Events
$E_T^\mu > 30 \text{ GeV}$	$P_{\text{real}}(\mu) = 0.982 \pm 0.001$	2 Jet Events
	$P_{\text{real}}(\mu) = 0.978 \pm 0.002$	3 Jet Events
	$P_{\text{real}}(\mu) = 0.982 \pm 0.005$	4 Jet Events

The fake muon probability is measured in the $\mu + jets$ dataset, but using a low- \cancel{E}_T sample orthogonal to the main data sample. Requiring $\cancel{E}_T < 15 \text{ GeV}$, we do not expect many real W bosons in these events, and therefore expect an enriched sample of fake muons. The probability for these events to pass the muon isolation is then directly calculated. We find the fake muon efficiency as a function of muon p_T :

$P_{\text{fake}}(\mu) = 0.057 \pm 0.003$	$p_T^\mu < 20 \text{ GeV}/c$
$P_{\text{fake}}(\mu) = 0.108 \pm 0.006$	$20 \text{ GeV} \leq p_T^\mu < 30 \text{ GeV}/c$
$P_{\text{fake}}(\mu) = 0.43 \pm 0.02$	$30 \text{ GeV} \leq p_T^\mu < 40 \text{ GeV}/c$
$P_{\text{fake}}(\mu) = 0.59 \pm 0.03$	$40 \text{ GeV} \leq p_T^\mu < 50 \text{ GeV}/c$
$P_{\text{fake}}(\mu) = 0.77 \pm 0.04$	$p_T^\mu \geq 50 \text{ GeV}/c$

To ensure the accuracy of our real and fake probabilities, we plot the W transverse mass in the data sample (before b -tagging), and the predicted composition in each bin as predicted by the matrix method (Figure 4.10). As expected, fake muons cluster towards a low, unphysical W boson mass, and real muons peak about the true W boson mass.

With the measured probabilities, the matrix method is applied, and the following values attained:

<u>Muon Channel Matrix Method Numbers</u>		
N_{loose}	$N_{\text{loose}}^{\text{real}}$	$N_{\text{loose}}^{\text{fake}}$
7866 ± 88.7	3437.4 ± 85.7	4428.6 ± 274.4
N_{tight}	$N_{\text{tight}}^{\text{real}}$	$N_{\text{tight}}^{\text{fake}}$
4747 ± 68.9	854.4 ± 52.6	3892.6 ± 87.5

The following matrix method scale factors are thus applied to the *multijet* and *W + jets* samples:

$$F_{\text{multijet}} = \frac{N_{\text{tight}}^{\text{fake}}}{N_{\text{multijet}}} \quad F_{W+\text{jets}} = \frac{N_{\text{tight}}^{\text{real}} - (\mathcal{Y}_{l\text{jets}} + \mathcal{Y}_{ll} + \mathcal{Y}_{WW} + \mathcal{Y}_{WZ})}{\mathcal{Y}_{Wjj} + \mathcal{Y}_{Wbb}}$$

Muon Channel Matrix Method Scale Factors

$$\begin{aligned} F_{\text{multijet}} &= 0.066 \times P_{\text{fake}}(\mu) & p_T^\mu < 30 \text{ GeV}/c \\ F_{\text{multijet}} &= 0.145 \times P_{\text{fake}}(\mu) & p_T^\mu > 30 \text{ GeV}/c \\ F_{W+\text{jets}} &= 1.16 & \text{All } p_T^\mu \end{aligned}$$

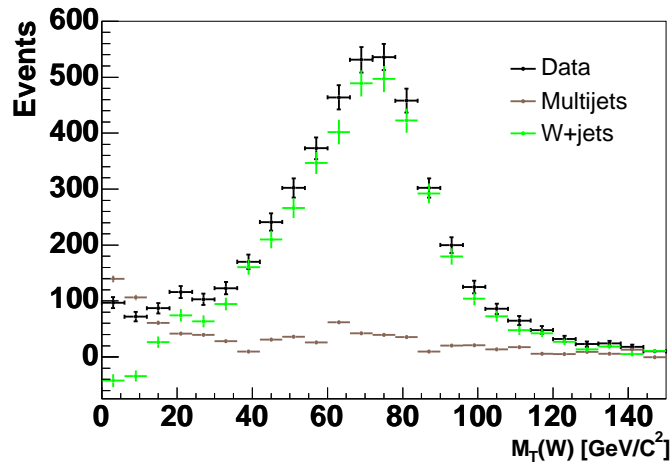


Figure 4.10: The reconstructed *W* transverse mass in muon channel data (before *b*-tagging), and the matrix method predicted composition in each bin.

4.7 Certifying the Model

After applying all scale factors and corrections to the background samples, we are left to test our model and how well it represents the data. It is essential that all physics object distributions in the background model are consistent with data (within the uncertainties of our model), or we must admit there is a crucial lack of understanding, and that subsequent results will not be reliable. In the following event selection plots before b -tagging (Figure 4.11), the p_T distributions for the lepton, leading jet, and second leading jet, as well as the number of jets, \cancel{E}_T and the reconstructed W transverse mass are shown. Plots are shown for the electron and muon channels separately. Within the errors of the background model and statistics of the data samples, all plots are found to be in agreement. In Appendix A, a more comprehensive set of plots is shown for all physics objects. To further gauge our understanding of the individual components of the background model, we create a $W + jets/multijet$ enriched sample, and a $t\bar{t}$ enriched sample to compare with data. These plots are also seen to be consistent with data and can be found in Appendix A.

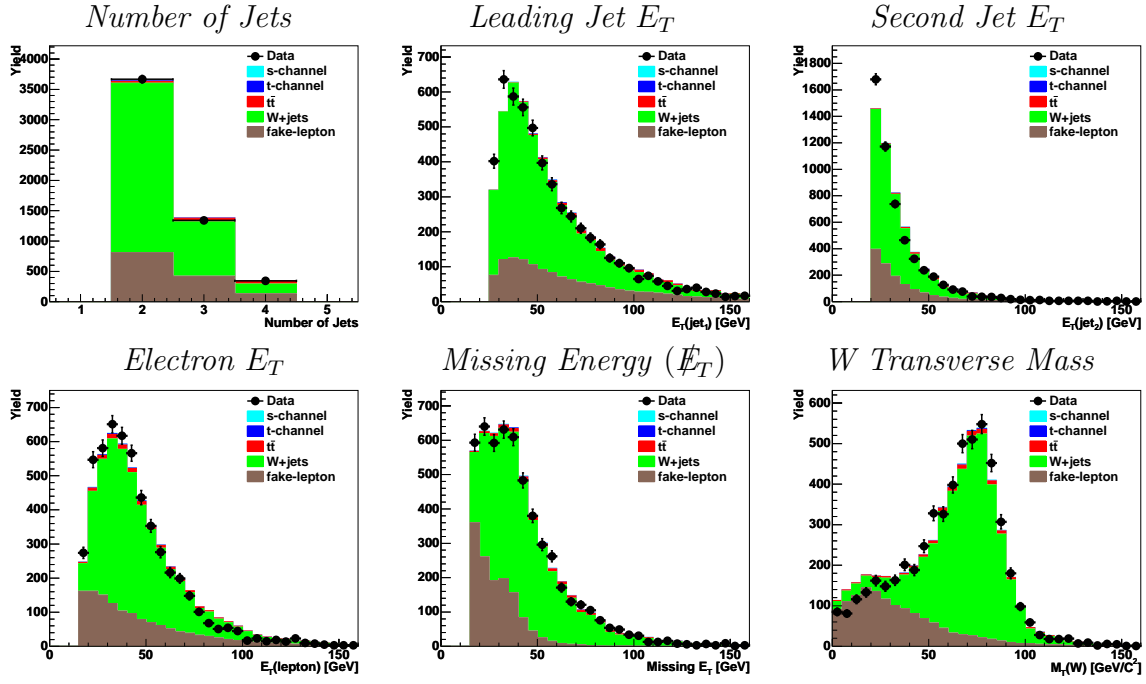
4.7.1 Acceptances

The fraction of events in a Monte Carlo sample that passes the event selection is called the acceptance. This value is later used in the calculation of sample yields (Section 4.7.2), and the calculation of final cross section limits (Chapter 6). The acceptance for a sample is defined as:

$$\mathcal{A}^{MC} = \frac{\mathcal{B}}{N_{initial}} \sum_{events} \varepsilon_{trigger} \varepsilon_{correction} \varepsilon_{b-tagging} \quad (4.13)$$

where $N_{initial}$ is the number of events in the Monte Carlo sample before any selection, \mathcal{B} is the branching fraction for a process to a specific decay mode, and ε_x are the efficiencies of

SELECTION BEFORE TAGGING — Electron Channel Distributions



SELECTION BEFORE TAGGING — Muon Channel Distributions

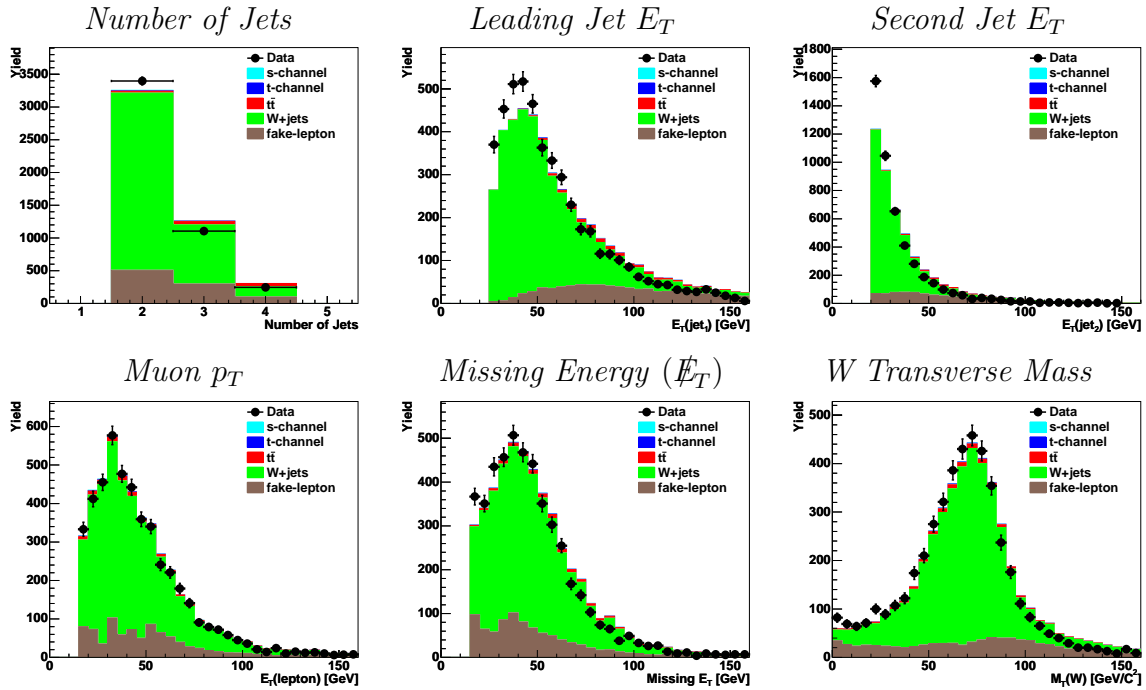


Figure 4.11: The data and background model comparison for all samples after event selection, excluding b -tagging. Each object p_T distribution is shown, as well as the number of jets in the event and reconstructed W transverse mass.

each event to pass the trigger (Section 4.3.1.5), Monte Carlo to data corrections (Section 4.3.1.1 – 4.3.1.4), and the b -tagging requirement (Section 4.3.1.6). Signal acceptances are given in Table 4.4.

	Percentage Acceptance for Signal		
	Before Tagging	1 Tag	2 Tag
Electron Channel	2.55	1.09	0.14
Muon Channel	2.39	1.00	0.13

Table 4.4: The percentage acceptance for t -channel single top in each lepton and tagging channel.

4.7.2 Yields

A sample’s yield is defined as the number of events predicted to pass event selection after all corrections have been applied, and the sample has been scaled to the appropriate integrated luminosity. For Monte Carlo samples, the yield is calculated as

$$\text{Yield} = \mathcal{Y}^{MC} = \mathcal{A}^{MC} \mathcal{L} \sigma \quad (4.14)$$

where \mathcal{A}^{MC} is the acceptance (Equation 4.13, \mathcal{L} is the integrated luminosity, and σ is the cross section of the sample in question. For the *multijet* sample, the yield is calculated as

$$\text{Yield} = \mathcal{Y}^{multijet} = \sum_{\text{events}} F_{multijet} \quad (4.15)$$

where $F_{multijet}$ is the matrix method scale factor for each event as described in Section 4.6. The yield for data is simply the number of events passing event selection.

Yields before and after b -tagging for all samples are summarized in Tables 4.7 and 4.8. It should be noted that the matrix method brings the background sum and data samples into close agreement before tagging by construction.

Sample	Before Tagging	Electron Channel Yields			
		1 Tag (YIELD)	1 Tag (TRAIN)	2 Tag (YIELD)	2 Tag (TRAIN)
t -channel	18.5	7.9	7.8	1.0	1.1
s -channel	8.4	3.5	3.5	0.5	0.5
$t\bar{t} \rightarrow \ell + jets$	113.9	46.6	46.6	23.1	22.8
$t\bar{t} \rightarrow \ell\ell$	26.9	11.4	11.5	3.0	2.8
Wbb	49.1	20.3	20.2	2.0	2.3
Wjj	3775.6	101.1	103.0	2.4	2.8
WW	17.6	0.9	—	0.0	—
WZ	5.8	0.8	—	0.0	—
$multijet$	1359.6	57.9	56.1	3.3	—
Background Sum					
without s -channel	5348.5	239.0	—	33.7	—
with s -channel	5356.9	242.5	—	34.2	—
Data	5351.0	270.0	—	53.0	—

Table 4.5: The electron sample yields before and after b -tagging. The background sum is given for comparison with data. The tagged samples are split into YIELD and TRAIN samples, used for calculating limits and neural network training respectively (Section 5.5). Samples not used for network training are not split.

Sample	Before Tagging	Muon Channel Yields			
		1 Tag (YIELD)	1 Tag (TRAIN)	2 Tag (YIELD)	2 Tag (TRAIN)
t -channel	17.2	7.2	7.1	0.9	0.9
s -channel	7.9	3.3	3.3	0.5	0.4
$t\bar{t} \rightarrow \ell + jets$	103.7	42.2	42.2	20.6	20.8
$t\bar{t} \rightarrow \ell\ell$	26.0	11.1	11.0	2.7	2.6
Wbb	43.7	17.9	18.1	1.8	1.7
Wjj	3692.5	98.5	91.3	2.7	3.0
WW	19.9	1.0	—	0.0	—
WZ	6.9	0.8	—	0.1	—
$multijet$	904.8	65.9	65.7	8.5	11.0
Background Sum					
without s -channel	4797.4	237.3	—	36.3	—
with s -channel	4805.3	240.6	—	36.8	—
Data	4747.0	225.0	—	32.0	—

Table 4.6: The muon sample yields before and after b -tagging. The background sum is given for comparison with data. The tagged samples are split into YIELD and TRAIN samples, used for calculating limits and neural network training respectively (Section 5.5). Samples not used for network training are not split.

4.7.3 Distributions

The final distributions for the electron and muon samples in the single and double tag channels are shown in Figures 4.12 and 4.13. The E_T distributions for the lepton, leading jet, and second leading jet, as well as the number of jets, \cancel{E}_T and the reconstructed W transverse mass are shown. All plots are in agreement within the errors of the model. A more complete set of plots is included in Appendix A.

Sample	Electron Channel Yields		
	Before Tagging	Single Tag	Double Tag
t -channel	18.5	7.9	1.0
s -channel	8.4	3.5	0.5
$t\bar{t} \rightarrow \ell + jets$	113.9	46.6	23.1
$t\bar{t} \rightarrow \ell\ell$	26.9	11.4	3.0
Wbb	49.1	20.3	2.0
Wjj	3775.6	101.1	2.4
WW	17.6	0.9	0.0
WZ	5.8	0.8	0.0
$multijet$	1359.6	57.9	3.3
Background Sum			
without s -channel	5348.5	239.0	33.7
with s -channel	5356.9	242.5	34.2
Data	5351.0	270.0	53.0

Table 4.7: The electron sample yields before and after b -tagging. The background sum is given for comparison with data. The tagged samples are split into YIELD and TRAIN samples, used for calculating limits and neural network training respectively (Section 5.5). Samples not used for network training are not split.

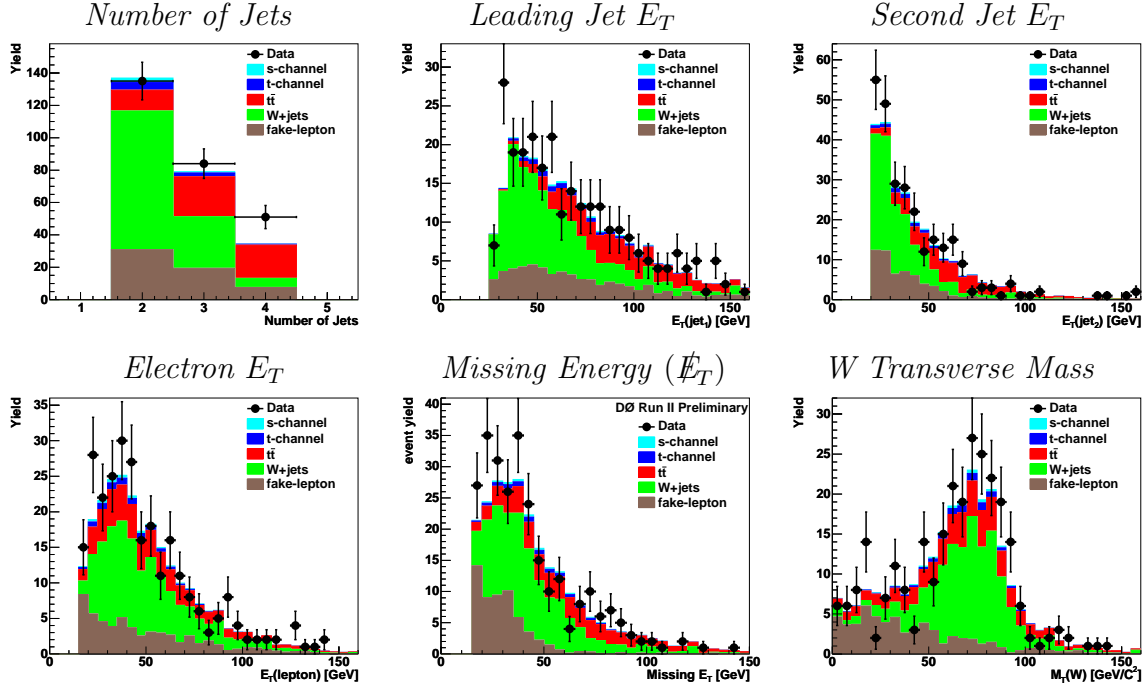
4.7.4 Systematic Uncertainties

4.7.4.1 Integrated Luminosity

The integrated luminosity uses a total inelastic cross section that is an average of results from the CDF and E811 experiments at Fermilab [108]. The uncertainty on the integrated luminosity is:

- Integrated luminosity error = $\pm 6.5\%$

Single Tag Electron Channel Distributions



Single Tag Muon Channel Distributions

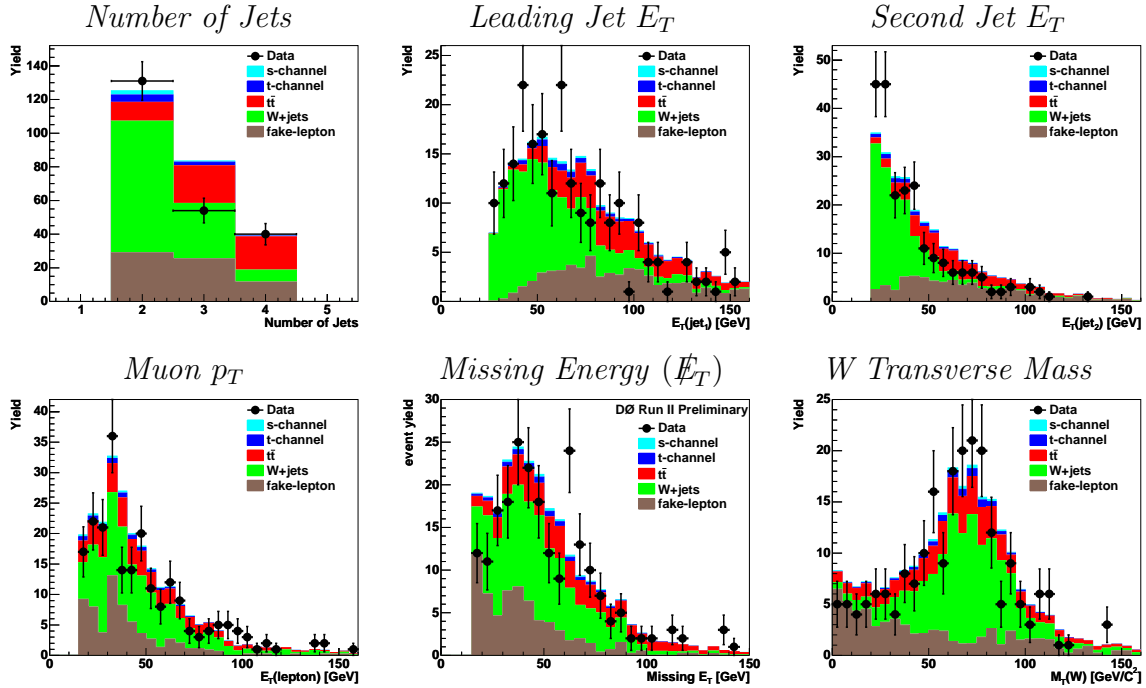
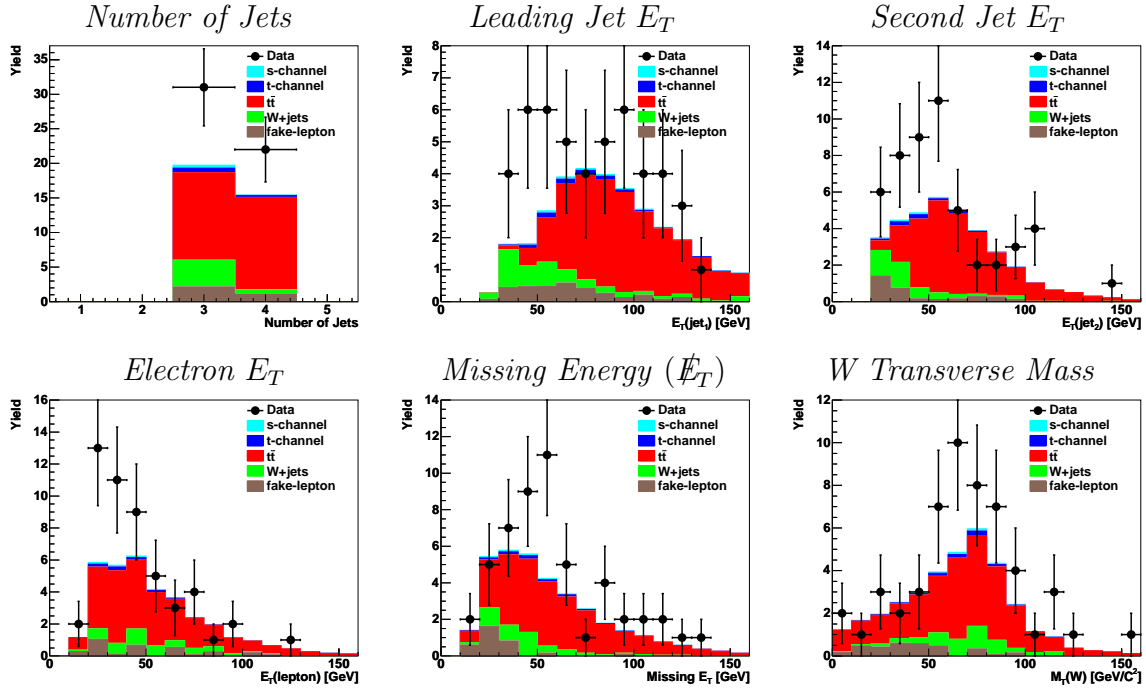


Figure 4.12: The data and background model comparison for the single tag samples after event selection. Each object E_T distribution is shown, as well as the number of jets in the event and the reconstructed W transverse mass.

Double Tag Electron Channel Distributions



Double Tag Muon Channel Distributions

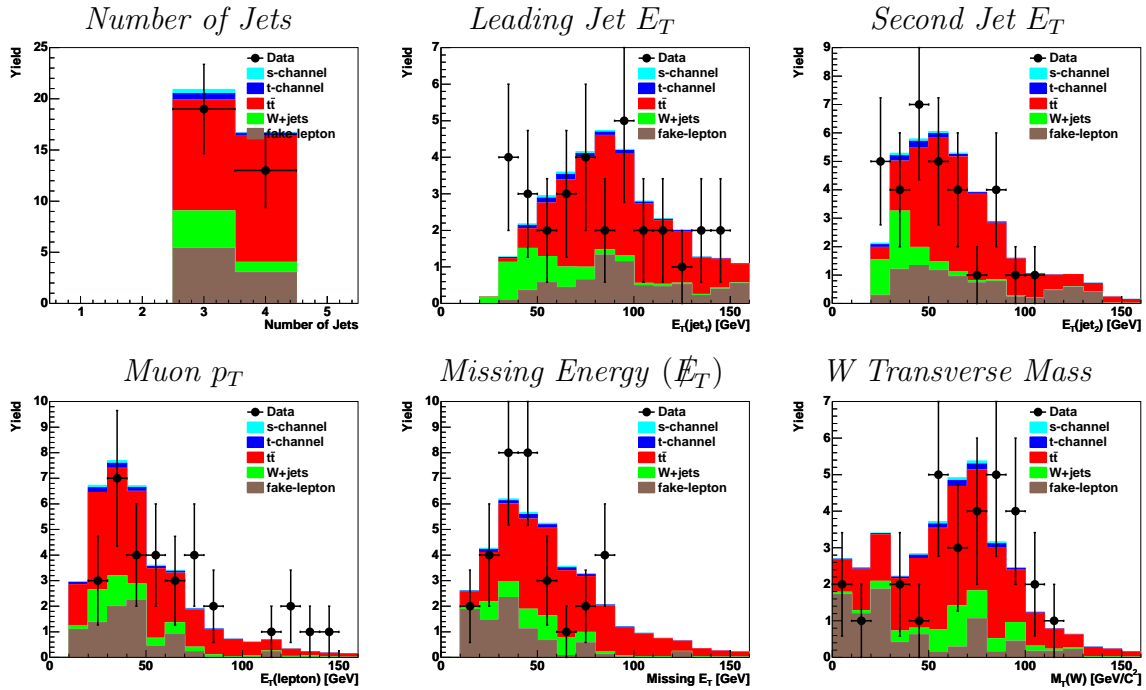


Figure 4.13: The data and background model comparison for the double tag samples after event selection. Each object E_T distribution is shown, as well as the number of jets in the event and the reconstructed W transverse mass.

Sample	Muon Channel Yields		
	Before Tagging	Single Tag	Double Tag
t -channel	17.2	7.2	0.9
s -channel	7.9	3.3	0.5
$t\bar{t} \rightarrow \ell + jets$	103.7	42.2	20.6
$t\bar{t} \rightarrow \ell\ell$	26.0	11.1	2.7
Wbb	43.7	17.9	1.8
Wjj	3692.5	98.5	2.7
WW	19.9	1.0	0.0
WZ	6.9	0.8	0.1
$multijet$	904.8	65.9	8.5
Background Sum			
without s -channel	4797.4	237.3	36.3
with s -channel	4805.3	240.6	36.8
Data	4747.0	225.0	32.0

Table 4.8: The muon sample yields before and after b -tagging. The background sum is given for comparison with data. The tagged samples are split into YIELD and TRAIN samples, used for calculating limits and neural network training respectively (Section 5.5). Samples not used for network training are not split.

4.7.4.2 Cross Sections

The single top, $t\bar{t}$, and diboson samples are normalized with the calculated theoretical cross sections. The errors on these values, including the error on the top quark mass, $\Delta M_t = 5.1$ GeV, are:

- Cross section errors

t -channel	$= \pm 16\%$	$t\bar{t} \rightarrow \ell + jets$	$= \pm 18\%$	WZ	$= \pm 8.9\%$
s -channel	$= \pm 15\%$	$t\bar{t} \rightarrow \ell\ell$	$= \pm 18\%$	WW	$= \pm 2.2\%$

4.7.4.3 Branching Fractions

The single top, $t\bar{t}$, and diboson samples have been forced to decay into the selected lepton channels. To calculate acceptances and final yields, the branching fractions are taken into account. The errors on these branching fractions are [48]:

- Branching fraction error $= \pm 2\%$

4.7.4.4 Triggers

The trigger efficiencies described in Section 4.3.1.5 are applied to all Monte Carlo samples. To measure the error, the acceptance and yields are remeasured in all samples while varying the trigger object turn-on curves by $\pm 1\sigma$. The percentage change in each sample is listed in Table 4.9, and these are used as the trigger errors.

Systematic Errors on Monte Carlo Trigger Calculation (%)				
Sample	<u>Electron Channel</u>		<u>Muon Channel</u>	
	Single Tag	Double Tag	Single Tag	Double Tag
t-channel	10.7	10.9	4.2	3.7
s-channel	10.0	9.8	3.9	3.8
$t\bar{t} \rightarrow \ell + jets$	7.9	8.0	3.4	3.4
$t\bar{t} \rightarrow \ell\ell$	8.1	7.8	3.4	3.3
Wbb	9.9	7.9	5.5	4.6
Wjj	9.7	8.6	5.7	6.1
WW	10.2	8.8	5.2	5.2
WZ	11.2	11.1	5.4	6.2

Table 4.9: The average percentage difference in sample yields after varying the trigger efficiency turn-on curves by $\pm 1\sigma$.

4.7.4.5 Primary Vertex Identification

Monte Carlo samples are adjusted for the slight difference in primary vertex reconstruction efficiency with comparison to data. The error on this correction factor is:

- Primary vertex error
Electron channel = $\pm 1.8\%$
Muon channel = $\pm 0.4\%$

4.7.4.6 Electron Identification

Monte Carlo samples are adjusted for the differences in electron cut efficiencies observed between Monte Carlo and data. The errors on these correction factors are:

- Electron identification errors

	$E_T < 30 \text{ GeV}$	$E_T > 30 \text{ GeV}$
Isolation, EMF, Hm7	$= \pm 3.0\%$	$= \pm 3.0\%$
$p_{T_{trk}}$	$= \pm 3.2\%$	$= \pm 3.2\%$
Likelihood	$= \pm 2.2\%$	$= \pm 1.0\%$
Combined	$= \pm 5.9\%$	$= \pm 4.3\%$

4.7.4.7 Muon Identification

Monte Carlo samples are adjusted for the differences in muon cut efficiencies observed between Monte Carlo and data. The error on this correction factor is:

- Muon identification error
All cuts combined $= \pm 5.2\%$

4.7.4.8 Jet Fragmentation

Differences are observed in the jet fragmentation models between various Monte Carlo generators, such as HERWIG and PYTHIA. The error associated with the jet fragmentation was studied in the Run I $t\bar{t} \rightarrow \text{alljets}$ analysis, and we use these errors as no current measurement has been performed yet.

- Jet fragmentation error

t -channel	$= \pm 5\%$	$t\bar{t} \rightarrow \ell + jets$	$= \pm 7\%$	WW	$= \pm 5\%$
s -channel	$= \pm 5\%$	$t\bar{t} \rightarrow \ell\ell$	$= \pm 5\%$	WZ	$= \pm 5\%$

4.7.4.9 Jet Reconstruction and Identification

The identification cuts used to define an acceptable hadronic jet have been shifted by $\pm 1\sigma$ to take into account the stability of our jet definition and its effects in sample acceptances and final yields. These uncertainties are given in Table 4.10.

4.7.4.10 Jet Energy Resolution

The systematic error is quoted from the percentage change in the yields after varying the jet energy resolution by $\pm 1\sigma$. The errors are listed in Table 4.11.

Systematic Errors on the Jet Identification (%)				
Sample	Electron Channel		Muon Channel	
	Single Tag	Double Tag	Single Tag	Double Tag
t-channel	6.8	7.7	5.5	10.2
s-channel	4.2	11.8	3.4	14.3
$t\bar{t} \rightarrow \ell + jets$	0.5	3.2	1.2	3.4
$t\bar{t} \rightarrow \ell\ell$	3.2	15.6	16.3	20.3
Wbb	7.3	20.4	7.0	15.3
Wjj	5.2	24.4	11.4	38.3
WW	17.3	19.5	23.5	37.0
WZ	7.3	38.0	17.9	9.3

Table 4.10: The percentage difference in sample yields after varying the jet energy identification by $\pm 1\sigma$. The larger of the two variations is quoted as the error.

Systematic Errors on the Jet Energy Resolution (%)				
Sample	Electron Channel		Muon Channel	
	Single Tag	Double Tag	Single Tag	Double Tag
t-channel	1.5	1.8	0.3	6.2
s-channel	0.6	1.6	1.7	4.7
$t\bar{t} \rightarrow \ell + jets$	0.5	1.4	0.6	1.9
$t\bar{t} \rightarrow \ell\ell$	0.8	4.9	14.0	13.5
Wbb	1.3	4.4	2.8	6.0
Wjj	4.4	43.8	3.3	14.1
WW	35.6	42.8	11.3	37.0
WZ	1.5	5.1	7.1	9.3

Table 4.11: The percentage difference in sample yields after varying the jet energy resolution by $\pm 1\sigma$. The larger of the two variations is quoted as the error.

4.7.4.11 Jet Energy Scale

The systematic error from the jet energy scale is derived by varying the jet energy scale by $\pm 1\sigma$ and calculating the percentage difference in the sample acceptances. Instead of averaging the acceptances for the varied σ 's, a conservative approach is taken, and the larger deviation is quoted as the error. The errors are listed in Table 4.12.

Sample	Systematic Errors on the Jet Energy Scale (%)			
	Electron Channel		Muon Channel	
	Single Tag	Double Tag	Single Tag	Double Tag
t-channel	5.8	9.4	3.7	8.7
s-channel	5.1	14.2	4.7	16.3
$t\bar{t} \rightarrow \ell + jets$	2.0	4.4	1.3	2.2
$t\bar{t} \rightarrow \ell\ell$	3.3	8.3	7.1	9.1
Wbb	12.7	17.7	14.5	15.0
Wjj	14.8	19.4	14.6	18.4
WW	12.0	23.1	9.9	16.0
WZ	10.6	24.9	10.0	23.4

Table 4.12: The percentage difference in sample yields after varying the jet energy scale by $\pm 1\sigma$. The larger of the two variations is quoted as the error.

4.7.4.12 Flavor-Dependant Tag Rate Functions

The flavor dependant TRFs involve systematic errors from several sources, including the errors on the taggability rate functions, the equations used to derive the data TRFs (System8), the data/MC correction, differences in varying Monte Carlo samples, and statistical uncertainties. To account for these errors, we shift the taggability and tag rate functions by $\pm 1\sigma$ and calculate the sample acceptances. The larger of the two acceptances from the nominal value is quoted as the error. These errors are listed in Table 4.13.

4.7.4.13 Matrix Method

The matrix method allows us to calculate the real and fake lepton fractions in our data sample before tagging. These fractions are in turn used to normalize our *multijet*, Wjj ,

Systematic Errors on the Tag Rate Functions(%)				
Sample	<u>Electron Channel</u>		<u>Muon Channel</u>	
	Single Tag	Double Tag	Single Tag	Double Tag
t-channel	1.8	4.2	1.7	3.4
s-channel	0.5	3.8	0.6	3.1
$t\bar{t} \rightarrow \ell + jets$	0.6	3.6	1.8	2.9
$t\bar{t} \rightarrow \ell\ell$	0.5	3.7	1.8	3.1
Wbb	1.0	3.9	1.3	3.2
Wjj	5.0	7.4	0.3	4.8
WW	3.7	8.7	0.1	7.9
WZ	1.2	3.8	1.5	3.2

Table 4.13: The percentage difference in sample yields after varying the taggability and tag rate functions by $\pm 1\sigma$. The larger of the two variations is quoted as the error.

and Wbb samples. The definition of and error on these normalization factors and their components are shown below.

Multijet Normalization Factor ($F_{multijet}$)

- $F_{multijet} = \frac{N_{tight}^{fake}}{N_{multijet}}$
- $\Delta F_{multijet} = F_{multijet} \sqrt{\left(\frac{\Delta N_{tight}^{fake}}{N_{tight}^{fake}}\right)^2 + \left(\frac{\Delta N_{multijet}}{N_{multijet}}\right)^2}$
- $N_{tight}^{fake} = \frac{\varepsilon_{real} N_{Loose} - N_{Tight}}{\varepsilon_{real} - \varepsilon_{fake}}$, the number of fake lepton events predicted in the tight sample by the matrix method (Section 4.6).
- ΔN_{tight}^{fake} = the uncertainty on the number of fake lepton events predicted in the data sample (before tagging) by the matrix method (Section 4.6). The uncertainty is calculated by standard error propagation, and the expression is given in its entirety in [106].
- $N_{multijet}$ = the number of events in the *multijet* background sample before tagging (Tables 4.7 and 4.8).
- $\Delta N_{multijet} = \sqrt{N_{multijet}}$, the uncertainty on the number of events in the *multijet* background sample before tagging (Tables 4.7 and 4.8).

Wjj and Wbb Normalization Factors ($F_{wjj, wbb}$)

- $F_{Wjj, Wbb} = \frac{N_{tight}^{real} - (\mathcal{Y}_{t\bar{t}} + \mathcal{Y}_{WW} + \mathcal{Y}_{WZ})}{\mathcal{Y}_{Wbb} + \mathcal{Y}_{Wjj}}$
- $\Delta F_{Wjj, Wbb} = F_{Wjj, Wbb} \sqrt{\frac{\Delta N_{tight}^{real2} + \Delta \mathcal{Y}_{t\bar{t}}^2 + \Delta \mathcal{Y}_{WW}^2 + \Delta \mathcal{Y}_{WZ}^2}{(N_{tight}^{real} - (\mathcal{Y}_{t\bar{t}} + \mathcal{Y}_{WW} + \mathcal{Y}_{WZ}))^2} + \frac{\Delta \mathcal{Y}_{Wbb}^2 + \Delta \mathcal{Y}_{Wjj}^2}{(\mathcal{Y}_{Wbb} + \Delta \mathcal{Y}_{Wbb}^2)^2}}$. The uncertainties on all yield errors are considered negligible compared to ΔN_{tight}^{real} , and so only this error is taken into account.
- $N_{tight}^{real} = \frac{N_{Tight} - \varepsilon_{fake} N_{Loose}}{\varepsilon_{real} - \varepsilon_{fake}}$, the number of real lepton events predicted in the tight sample by the matrix method (Section 4.6).
- ΔN_{tight}^{real} = the uncertainty on the number of real lepton events predicted in the data sample before tagging by the matrix method (Section 4.6). The uncertainty is calculated by standard error propagation, and the expression is given in full in [106].

The *multijet*, Wjj , and Wbb scale factors and errors shown above are given for completeness, but the relevant error with respect to the limit calculation is on the final tagged samples. The definitions of these yields in terms of the matrix method are given, as well as their associated errors:

Matrix Method Normalization Errors After b -Tagging

- $N_{sample}^{final} = f_{sample}^{final} N_{tight}^{real}$, the final sample yield, where sample is *multijet* or W +jets.
- $f_{sample}^{final} = \frac{f_{sample}^{final}}{f_{tight}^{final}}$, the fraction of events in the tight sample passing the final b -tagging requirement, where the sample is *multijet* or W +jets. This is determined by applying the neural network tagger (data) or tag-rate-function (monte carlo) to the sample.
- $\Delta N_{sample}^{final} = N_{sample}^{final} \sqrt{\left(\frac{\Delta f_{sample}^{final}}{f_{sample}^{final}}\right)^2 + \left(\frac{\Delta N_{tight}^{real}}{N_{tight}^{real}}\right)^2}$, the uncertainty on the final yield of the sample. It is calculated through standard error propagation, and correctly accounts for all correlations in the calculation of N_{tight}^{real} and f_{sample}^{final} . The expression is given in its entirety in [107].
- $\Delta f_{sample}^{final}$ = the uncertainty on the fraction of events in the tight sample that pass the final b -tagging requirement. This uncertainty is derived from Monte Carlo for the W +jets samples, and the *multijet* data sample directly. Standard error propagation is applied, and the complete expression is given in [107].

4.7.4.14 Final Uncertainties

The final uncertainties are given in Tables 4.14–4.17 for each lepton and tagging channel. These values and their correlations are properly taken into account in the calculation of cross section limits, as discussed in Chapter 6.

UNCERTAINTIES FOR SINGLE-TAGGED ANALYSES

		Single-Tagged Electron Channel Percentage Errors							
	tqb	tb	Wbb	Wjj	$ttlj$	ttl	WW	WZ	$multijet$
Components for Normalization									
Luminosity	6.5	6.5	—	—	6.5	6.5	6.5	6.5	—
Cross Section	16.0	15.0	—	—	18.0	18.0	2.2	8.9	—
Branching Fraction	2.0	2.0	—	—	2.0	2.0	2.0	2.0	—
Matrix Method	—	—	3.3	3.3	—	—	—	—	3.3
Primary Vertex	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	—
Electron ID	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	—
Jet Fragmentation	5.0	5.0	—	—	7.0	5.0	5.0	5.0	—
Components for Normalization and Shape									
Jet ID	6.8	4.2	—	—	0.8	3.2	17.3	7.3	—
Jet Energy Scale	5.8	5.1	—	—	2.0	3.3	12.0	10.6	—
Jet Energy Resolution	1.5	0.6	—	—	0.5	0.8	35.6	1.5	—
Trigger	10.7	10.0	—	—	7.9	8.1	10.2	11.2	—
Flavor-Dependant TRFs	1.8	0.5	1.0	5.0	0.6	0.5	3.7	1.2	—
Statistics	1.1	1.1	1.4	1.0	0.6	0.9	2.6	2.5	4.2
Combined									
Acceptance Uncertainty	14.2	—	—	—	—	—	—	—	—
Yield Uncertainty	22.9	22.4	6.4	8.0	22.7	22.6	44.0	21.8	14.1

Table 4.14: The percentage uncertainties for all modeling components and scale factors applied in the single tag electron analysis.

UNCERTAINTIES FOR DOUBLE-TAGGED ANALYSES

		Double-Tagged Electron Channel Percentage Errors							
	tqb	tb	Wbb	Wjj	$ttlj$	ttl	WW	WZ	$multijet$
Components for Normalization									
Luminosity	6.5	6.5	—	—	6.5	6.5	6.5	6.5	—
Cross Section	16.0	15.0	—	—	18.0	18.0	2.2	8.9	—
Branching Fraction	2.0	2.0	—	—	2.0	2.0	2.0	2.0	—
Matrix Method	—	—	9.1	9.1	—	—	—	—	9.1
Primary Vertex	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	—
Electron ID	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	—
Jet Fragmentation	5.0	5.0	—	—	7.0	5.0	5.0	5.0	—
Components for Normalization and Shape									
Jet ID	7.7	11.8	—	—	3.2	15.6	19.5	38.1	—
Jet Energy Scale	9.4	14.2	—	—	4.4	8.3	23.1	24.9	—
Jet Energy Resolution	6.2	1.6	—	—	1.4	4.9	53.0	5.1	—
Trigger	10.9	9.8	—	—	8.0	7.8	8.8	11.1	—
Flavor-Dependant TRFs	4.2	3.8	3.9	7.4	3.6	3.7	8.7	3.8	—
<u>Statistics</u>	1.1	1.1	1.4	1.0	0.6	0.9	2.6	2.5	18.3
<u>Combined</u>									
Acceptance Uncertainty	22.6	—	—	—	—	—	—	—	—
Yield Uncertainty	25.4	28.5	13.0	14.4	23.6	28.9	63.2	49.1	22.6

Table 4.15: The percentage uncertainties for all modeling components and scale factors applied in the double tag electron analysis.

UNCERTAINTIES FOR SINGLE-TAGGED ANALYSES

		Single-Tagged Muon Channel Percentage Errors							
	tqb	$t\bar{b}$	$Wb\bar{b}$	$Wj\bar{j}$	$t\bar{t}l\bar{j}$	$t\bar{t}l\bar{l}$	WW	WZ	$multijet$
<u>Components for Normalization</u>									
Luminosity	6.5	6.5	—	—	6.5	6.5	6.5	6.5	—
Cross Section	16.0	15.0	—	—	18.0	18.0	2.2	8.9	—
Branching Fraction	2.0	2.0	—	—	2.0	2.0	2.0	2.0	—
Matrix Method	—	—	4.3	4.3	—	—	—	—	4.3
Primary Vertex	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	—
Electron ID	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	—
Jet Fragmentation	5.0	5.0	—	—	7.0	5.0	5.0	5.0	—
<u>Components for Normalization and Shape</u>									
Jet ID	5.5	3.4	—	—	1.2	16.3	23.5	17.9	—
Jet Energy Scale	3.7	4.7	—	—	1.3	7.1	9.9	10.0	—
Jet Energy Resolution	0.3	1.7	—	—	0.6	14.0	11.3	7.1	—
Trigger	4.2	3.9	—	—	3.4	3.4	5.2	5.4	—
Flavor-Dependant TRFs	1.7	0.6	1.3	0.3	1.8	1.8	0.1	1.5	—
<u>Statistics</u>	1.2	1.2	1.4	1.0	0.7	1.0	2.6	2.4	2.1
<u>Combined</u>									
Acceptance Uncertainty	10.5	—	—	—	—	—	—	—	—
Yield Uncertainty	19.6	20.2	5.3	5.9	21.5	30.8	30.1	26.0	16.2

Table 4.16: The percentage uncertainties for all modeling components and scale factors applied in the single tag muon analysis.

UNCERTAINTIES FOR DOUBLE-TAGGED ANALYSES

		Double-Tagged Muon Channel Percentage Errors							
	tqb	$t\bar{b}$	$Wb\bar{b}$	$Wj\bar{j}$	$t\bar{t}l\bar{j}$	$t\bar{t}l\bar{l}$	WW	WZ	$multijet$
<u>Components for Normalization</u>									
Luminosity	6.5	6.5	—	—	6.5	6.5	6.5	6.5	—
Cross Section	16.0	15.0	—	—	18.0	18.0	2.2	8.9	—
Branching Fraction	2.0	2.0	—	—	2.0	2.0	2.0	2.0	—
Matrix Method	—	—	10.0	10.0	—	—	—	—	10.0
Primary Vertex	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	—
Electron ID	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	—
Jet Fragmentation	5.0	5.0	—	—	7.0	5.0	5.0	5.0	—
<u>Components for Normalization and Shape</u>									
Jet ID	10.2	14.3	—	—	3.4	20.3	37.0	9.3	—
Jet Energy Scale	8.7	16.3	—	—	2.2	9.1	16.0	23.4	—
Jet Energy Resolution	6.2	4.7	—	—	1.9	13.5	37.0	9.3	—
Trigger	3.7	3.8	—	—	3.4	3.3	5.2	6.2	—
Flavor-Dependant TRFs	3.4	3.1	—	—	2.9	3.1	7.9	3.2	—
<u>Statistics</u>	1.2	1.2	1.4	1.0	0.7	1.0	2.6	2.4	6.0
<u>Combined</u>									
Acceptance Uncertainty	24.0	—	—	—	—	—	—	—	—
Yield Uncertainty	23.8	29.5	10.9	11.5	22.0	33.4	76.3	59.1	16.6

Table 4.17: The percentage uncertainties for all modeling components and scale factors applied in the double tag muon analysis.

Chapter 5

Isolating Single Top Quarks From Background

5.1 Introduction

This is the point where I take a deep breath... now release. It is a sort of halfway point in the analysis — we have collected our data, trimmed it down, modeled it, and are satisfied with our understanding thus far. So the big question is: “Where are the single top quarks???” A multitude of paths lay open for anyone at this point. A large enough excess in the data and a few hints in the correct physics object distributions could beg for the calculation of a cross section using event counts alone, but that is not our case. A cut could be placed on a single discriminant variable that shows moderate separation between our t -channel single top signal and one or more of the dominant backgrounds. A limit based on event counts could then be derived, but only after removing a precious region of phase-space containing some signal. This is also unacceptable with a signal cross section that is very small. To maximize our signal acceptance, we avoid making any further cuts, and instead choose a multivariate approach that combines the discriminant power of several variables, while still retaining 100% of the signal sample after event

selection. This approach manifests in the form of neural networks. To give the main idea, the technique requires the selection of several discriminating variables, a sample of monte carlo signal events, and a sample of background events. The network “studies” the input samples, and produces a function that guides background events towards zero and signal events towards one. And that’s it. It sounded fantastic the first time around, but curiosity quickly ensued, leading me to wonder, “How does it work?”

5.2 Building A Brain

The human brain consists of about a 100 billion computing units known as neurons. These neurons are elaborately connected, and together give us the amazing powers of reasoning and memory that separate us from all other life on this planet. Although these neurons are functionally much slower than today’s modern computing processors, their design and connectivity allow the brain to run complex algorithms that easily outperform a computer. With the seemingly limitless ability to adapt and learn, it was reasoned that many problems could be more efficiently solved on computers by mimicking the designs of the brain, leading to the birth of neural networks.

The brain’s fundamental unit, the neuron, is composed of a cell body, several input lines known as dendrites, and a multi-branched output line, the axon (Figure 5.1(a)). The gap where axons from one neuron meet with the dendrite of another is called the synapse. Electrical signals are sent along each axon, and based on the inhibitory and excitatory responses of the neurotransmitters in the synapse, a signal may be passed on to the dendrite for processing in the next neuron. The two key components of this electrical signal processing are the summing of electrical signals from all input dendrites by the cell body, and the all-or-nothing transmission through the synaptic gap. An artificial neuron similarly receives multiple inputs, linearly combines them, and then processes this result through a mathematical function closely mimicking the on/off nature of the

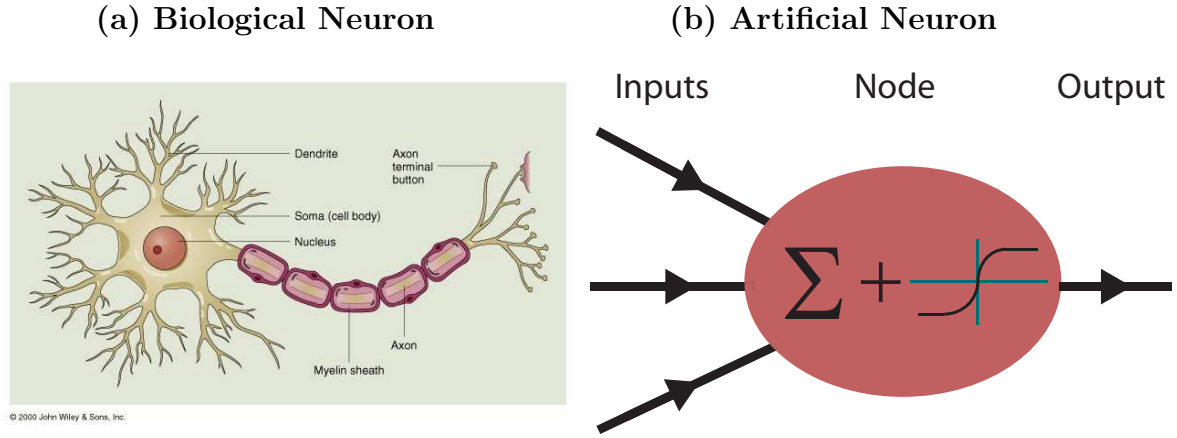


Figure 5.1: A schematic of a biological neuron (left) and an artificial neuron (right). The biological neuron shows many dendrites feeding into the cell body, where their signals are summed, carried along the axon, and possibly transmitted at its ends to the synaptic gap. Similarly, the artificial neuron has several inputs fed into a cell body, and a node that dually sums the signals and functionally mimics the synaptic gap for transmission of an output signal.

synaptic gap (Figure 5.1(b)). By connecting several layers of these artificial neurons, an artificial neural network is constructed.

In Figure 5.2, the architecture of a simple feed-forward neural network is shown. The network consists of two input variables or “input nodes” (x_i), three hidden nodes or “neurons” (n_j), and a single output node (O_1). The connectivity represents the flow of information from left to right, and each connection can be weighted (w_{ij} , z_{jk}). The hidden nodes contain the function modeling the synaptic gap, which process a linear combination of all input nodes and transmit a relatively large or small value, mimicking the desired on/off nature of a real neuron. A sigmoid is chosen for the functional form of the synaptic gap. It is defined as

$$\text{Sigmoid}(\vec{x}) = \frac{1}{1 + e^{-a - b\vec{x}}} = \frac{1}{1 + e^{-a - \vec{w} \cdot \vec{x}}}$$

where a is the threshold, a constant determining the turn-on point of the sigmoid, and b is a constant dictating how sharp the turn-on is (Figure 5.3). The sigmoid is

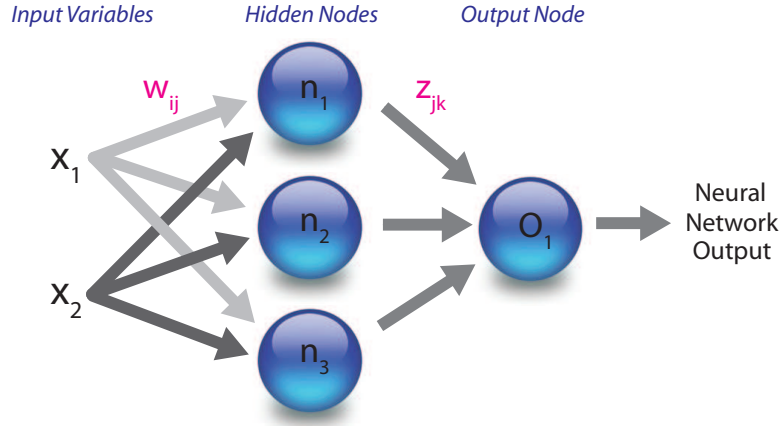


Figure 5.2: A simple, feed-forward neural network architecture involving two input variables, three hidden nodes, and a single output node. Information flows from left to right.

chosen from a class of several logistic functions, all fulfilling a theorem stating: a linear combination of logistic functions can approximate any continuous function made up of one or more variables [109]. This theorem assures us that a properly trained network can derive a function capable of separating signal from background, if such a function exists, and dictates that our output node combines the hidden nodes linearly, rather than sigmoidally.

For the example neural network given in Figure 5.2, the explicit functional form of the output is given below (where C is a parameter of the output node).

$$\begin{aligned}
 O(\vec{x}) &= C + z_{11}n_1 + z_{21}n_2 + z_{31}n_3 \\
 &= C + z_{11} \frac{1}{1 + e^{a_1 - (w_{11}x_1 + w_{21}x_2)}} + z_{21} \frac{1}{1 + e^{a_2 - (w_{12}x_1 + w_{22}x_2)}} + z_{31} \frac{1}{1 + e^{a_3 - (w_{13}x_1 + w_{23}x_2)}}
 \end{aligned}$$

While the output function is now completely defined, the weights connecting each node that ultimately allow for signal-background discrimination are yet to be determined. The flexibility of the method should become apparent, as you can design an infinite number of architectures, varying the number of nodes at any layer, and the number of connections from any node to any other node.

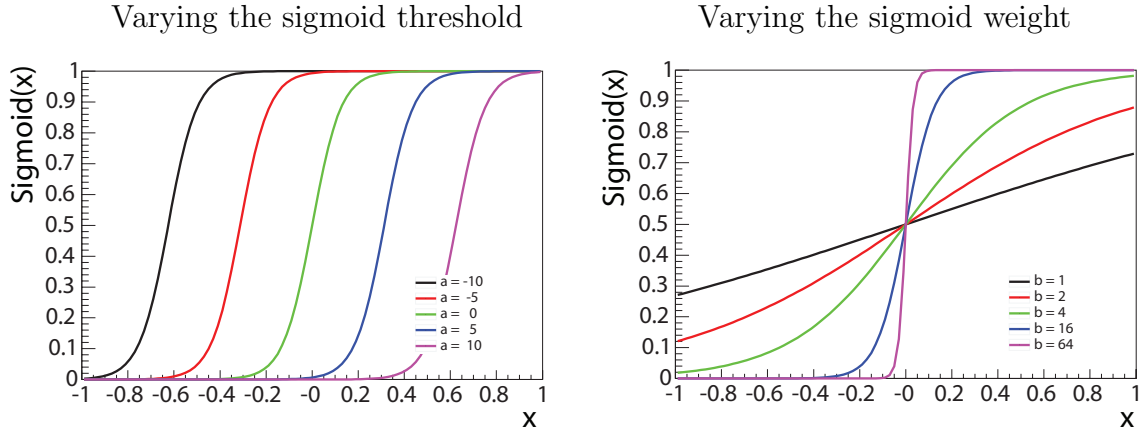


Figure 5.3: Illustration of the sigmoid function, and the effects of varying the exponential constants. Varying the threshold a , shifts the turn-on point of the curve left or right (left plot), and varying the value of the weight in the exponential b , affects how sharply the turn-on occurs (right plot). A large exponential weight approximates a step-function, a simple on/off switch.

5.3 Learning (In Theory)

As it stands, we have created a fixed model of a brain. All the components are there, but they define a static output function — there is no way to evolve. The essential ingredient is the ability for self-modification, to adapt to the information being processed, to produce a desired result. Seeking to produce a function that gives signal events a value close to one, and background events a value close to zero, we define an error function (E) to measure how well we are achieving this. The error function is defined as

$$E = \sum_p \frac{1}{2} \omega_p (O_p - T_p)^2 \quad (5.1)$$

where the summation is over all signal and background events p , and ω_p is the overall weight of an event. O_p and T_p are the actual output, and target output of the neural network for each event. For signal events, $T_p = 1$, and for background events, $T_p = 0$. The process of learning means minimization of the error function with respect to the network weights $(\vec{a}, \vec{w}, C, \vec{z})$. Several algorithms have been devised to this end

[110, 111, 112]; we use a conjugate gradient algorithm by Broyden, Fletcher, Goldfarb, and Shanno known as the BFGS method. The method involves computing a direction (\vec{s}) in error space based on the conjugate gradient. For each epoch, or iteration t of the algorithm, an α that minimizes $E(\vec{w}_t + \alpha\vec{s}_t)$ is computed, and all weights in epoch $t + 1$ are adjusted by this α , $\vec{w}_{t+1} = \vec{w}_t + \alpha\vec{s}_t$. The size of the weight adjustment, α , is a tunable parameter. Through successive iterations, the neural network output function is trained to separate the phase-space of signal and background events.

5.4 Learning (In Practice)

An open source neural network package called Multi Layer Perceptron Fit (MLPFit) [113], is used in this analysis. The same package was used in DØ's Run I single top quark search [66]. It uses the same feed-forward architecture described in Section 5.2 for our neural networks. The parameters we vary during optimization are the number and choice of input variables, the number of hidden nodes, and duration of training (number of epochs). Once a network architecture has been specified, the following training process begins:

- **Initialize** — each connective weight in the network is assigned a random value, $-0.5 < w_{ij} < 0.5$.
- **Calculate $E(\vec{x})$** — begin a training epoch, all signal and background events are passed through the network, and the error function $E(\vec{x})$ is calculated.
- **Minimize $E(\vec{x})$** — the BFGS method is used to minimize the error function and slightly adjust all weights in the network.
- **Repeat** — with the new weights, begin another training epoch.

The user decides when the training process is complete, or specifically how many training epochs they would like to perform. By construction, every consecutive epoch will reduce the error function and improve the performance of the network, but one

must be careful to avoid overtraining. The hope for the neural network is that it will find significant features, or areas of phasespace that discriminate between signal and background events. In the beginning epochs, this is effectively done, but as the number increases, the network may begin learning to discriminate against features unique to the samples themselves, and not the global properties of either signal or background. To avoid this pitfall, a testing epoch is performed after each training epoch. A testing epoch uses independent signal and background samples to the training ones, and calculates the error function. An effective neural network will successively decrease the value of the testing error function, but once over-training begins, this value will start to rise. At this minimum of the testing error function, we stop the training process. Example testing and training curves are shown in Figure 5.4.

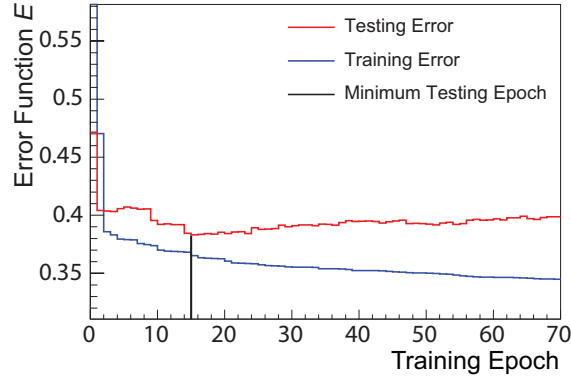


Figure 5.4: Example training (blue) and testing (red) curves for the neural network learning process. For each curve, the error function $E(\vec{x})$ is shown as a function of the number of network epochs. The vertical line at 15 epochs corresponds to the minimum of the testing error. This point signals that training is complete.

5.5 Our Neural Network Approach

In our search for t -channel single top, we have selected five backgrounds to discriminate against: Wjj ($j = u, d, s, c, g$), Wbb , $t\bar{t} \rightarrow l + jets$, $t\bar{t} \rightarrow ll$, and *multijet* events. The

relative fraction of the total background model for each is given in Table 5.1. Each is trained and optimized independently for the electron and muon channels. We further separate events into samples with exactly one b -tagged jet and those with two b -tagged jets. This approach leaves us with 20 independently trained and optimized networks: 2 lepton decay channels (electron, muon) \times 5 backgrounds (Wjj , Wbb , $t\bar{t} \rightarrow \ell + jets$, $t\bar{t} \rightarrow \ell\ell$, $multijet$) \times 2 tagging options (1 tag, 2 tags). In addition to these networks, a final step is taken to combine the discriminative power of the five background networks, into one “Super” network (Section 5.8).

	Individual Background Model Fractions (%)			
	<u>Electron Channel</u>		<u>Muon Channel</u>	
	Single Tag	Double Tag	Single Tag	Double Tag
Wjj	42	7	41	8
$multijet$	24	10	28	23
$t\bar{t} \rightarrow \ell + jets$	19	68	18	57
Wbb	9	6	7	5
$t\bar{t} \rightarrow \ell\ell$	5	9	5	7
WW/WZ	1	0	1	0

Table 5.1: The fraction of the total background model that each background accounts for. These are shown in percent for both the electron and muon channels.

In addition to the individual channels described above, our signal and background files have further been split into what we call TRAIN and YIELD samples (Figure 5.5). To avoid any bias in the training of the neural networks, and in the calculation of yields and final limit, these two aspects of the analysis have been completely decoupled. The TRAIN files are used for testing and training of the neural networks, and the YIELD files are used to calculate final yields, and are run through the optimized neural networks to derive cross section limits. The unweighted, raw statistics of each of these samples are given in Table 5.2. The files are not split exactly evenly, but rather odd numbered events go in one sample, and even numbered events go in the other. The number of odd and even events in a sample *after* event selection are not necessarily equal, but the statistical deviation is small and within the errors of the analysis. The “number” of

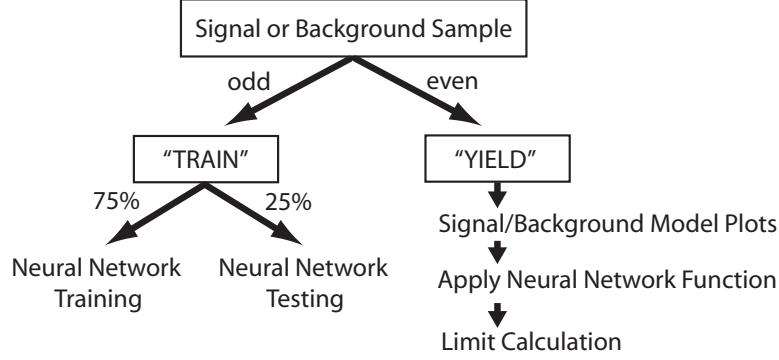


Figure 5.5: Signal and background samples are split into “TRAIN” and “YIELD” samples for the neural network and limit calculation respectively. They are split by odd and even numbered events *after* event selection, allowing for slight deviation from a true 50/50 split. The TRAIN sample is further split for neural network training and testing.

an event is the generation number for Monte Carlo events and the real event number for data samples. This system of splitting by odd and even number events (instead of a perfect 50/50 split) is applied so that the systematics can be derived on the correct set of events for the jet energy resolution and jet identification. The final yields of the YIELD and TRAIN samples are correctly adjusted to account for the sample splitting.

	Number of Events for Tagged Training Samples			
	<u>Electron Channel</u>		<u>Muon Channel</u>	
	TRAIN	YIELD	TRAIN	YIELD
$tq\bar{b}$	4,322	4,349	4,045	4,080
$t\bar{t} \rightarrow \ell + jets$	12,346	12,624	11,217	11,301
$t\bar{t} \rightarrow \ell\bar{\ell}$	6,092	6,137	5,901	6,022
$Wb\bar{b}$	2,724	2,609	3,408	3,475
Wjj	4,714	4,713	6,883	6,799
$multijet$ (single tag)	1,042	1,064	2,240	2,304
$multijet$ (double tag)	60	60	284	276

Table 5.2: The number of events in each TRAIN and YIELD sample for each background. The values are given for each lepton channel and further, each tagging channel. The single tag $multijet$ channel is trained with the statistics shown, but there are too few events for double tag training, and no network is trained.

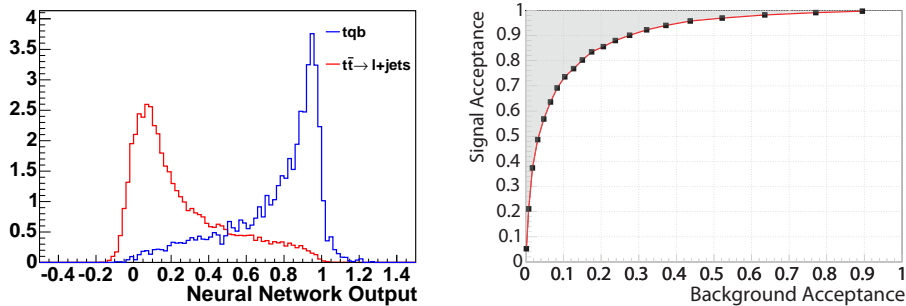


Figure 5.6: The output for a neural network trained against the $t\bar{t} \rightarrow \ell + jets$ background is shown (left). The signal vs background efficiency (right) is plotted for cut values along the neural network output axis. The integral of the shaded area above this efficiency curve is defined as our figure of merit.

5.5.1 Variable Optimization

With a search strategy in place, it remains to find an optimum set of variables with which to train each network. There is no predefined method for locating this set of variables, but there are some guidelines, and of course the quality of the final product. We have therefore experimented with many combinations of variables, and the total number of input variables. To directly compare the results of each trained network, we have developed a figure of merit. Using the final neural network output distribution, for example Figure 5.6, we plot the acceptance of signal versus the acceptance of background to pass a cut on the distribution. For each acceptance plot, we take 21 cut points along the axis of the neural network output distribution, in steps of 0.05 from 0 to 1. The figure of merit is then defined as the integral of the area above the acceptance curve. The smaller this area, the more effectively the network has separated signal and background.

Starting with an initial pool of 41 variables, it is essential to identify the most discriminating variables, and remove the weakest ones. While it is possible to train a neural network with all 41 variables and derive an effective output function, this is avoided for several reasons: (a) the computational time is enormous (b) there is a large redundancy in the phasespace offered by these variables, as many are correlated, and this serves

no helpful purpose in training the network (c) no physics understanding is gained by training with all 41 variable - you know some of the variables in the set were useful, but you don't know which ones. To begin eliminating variables from this pool, another multivariate technique known as a decision tree (Appendix C) is used. The decision tree provides a ranking system for the importance of all variables used in the technique. By feeding the entire list of variables into a decision tree, we select the most important variables to train our neural networks with. The ranking term is called the *gini* value, and we keep all variables with a gini value greater than 1% of the gini sum of all ranked variables. This seed set of variables is then used to optimize the neural networks. Starting with the highest ranked variable, variables are added one at a time to the neural network and the number of hidden nodes and training epochs are optimized. The figure of merit is calculated, and then the process is repeated with the next variable. After all variables have been added and the final network trained, the smallest figure of merit is found and the associated set of variables is now the optimized set.

It is important to note that the decision tree is simply a guide for the neural network optimization, and its ranking lists do not necessarily reflect the order of the most discriminating variables for the neural network, or single top production for that matter. We do use a large enough set of variables from the decision tree ranking list to allow the neural network to better probe the single top phasespace. Other algorithms for optimization could certainly be used, including a more brute force method of trying every combination of input variables and order of those input variables. The great advantage of the decision tree is that it allows us to automate the optimization process and provides an effective subset of the initial 41 variables, saving computational time.

A final concern lies with the total number of variables in each network, and moreover, the total number of independent variables in all five background networks for each channel. From first principle, the total number of free parameters in a single top process come from the energy-momentum 4-vectors of all involved final state physics objects. With

one lepton, one neutrino, and two, three, or four jets, we have 16, 20, or 24 free parameters. At first thought, it seems natural to just use these variables, a complete set of the available phase space, and train the network with them. In practice, detector resolution effects can weaken the strength of these variables, and so one seeks a complimentary set of variables, which through transformation, would map to the original 4-vectors. These variables may be more complex, such as the transverse mass of the reconstructed W boson or the sum of all jet transverse energies. Regardless of the mapping, there should still be a small number of variables which are capable of describing the single top phase space. To further trim down the number of variables in all the background networks for a given signal channel, we take the complete set of variables used in these optimized networks, and use the decision tree to rank them once more for each background. We then repeat the optimization process in their new ranking order and use the figure of merit to decide the optimized set of variables. This second round of optimization allows us to reduce the total number of variables in all five background networks and still retain the same discriminating power found in the first round of optimization.

5.6 Discriminating Variables

Through a careful study of the t -channel single top Feynman diagrams, and the dominant backgrounds, a set of 41 variables was created with which to train the neural networks. These variables fall into three categories:

- **Single-Object Kinematics** Transverse momenta of physics objects.
- **Event Kinematics** Properties such as transverse momentum, transverse mass, total transverse energy, and total energy of composite final state objects such as the W boson or the top quark, and total event energy.
- **Angular Variables** Angles between objects that take account of their origin from a polarized top quark or not.

5.6.1 Final State Object Reconstruction

In t -channel single top, one expects a final state top quark and a light quark; a low p_T b quark may also be present. The top quark decays into a W boson and a higher p_T b quark. The W boson is reconstructed from the identified lepton and the neutrino from its decay. The z component of the neutrino momentum is determined through a W mass constraint, and the smaller of the two $|p_z(\nu)|$ solutions is taken. This choice is correct 70% of the time. The calculation of $p_z(\nu)$ and the W transverse mass are given in Appendix B. The top quark is reconstructed from the highest p_T b -tagged jet in the event and the W boson. This choice of b -tagged jet is correct 90% of the time in t -channel production.

5.6.2 b -Tagged Jet Variables

Powerful background discrimination can be attained in a t -channel single top search by constructing variables based on the b -tagged jet. While the neural network tagging algorithm is applied directly to the data and *multijet* samples and there is no ambiguity in which jet has been tagged, a weight is applied to Monte Carlo events based on the probability of the *event* having one tagged jet or two tagged jets. This probability is composed of the probabilities for each of the individual jets being tagged, as derived in our tag-rate functions (Section 4.3.1.6). Instead of defining the jet in the event with the highest probability of being b -tagged as the b -quark jet, we create multiple copies of the same event, and allow all combinations of jets to be tagged. The weight of the original event is retained by properly weighting all copies such that their sum is equivalent to the original event. For example, an event with three jets would have the following tagging permutations:

Three Jet Event Tagging Permutations			
Jet1	Jet2	Jet3	Weight
—	—	—	0.114
tag	—	—	0.456
—	tag	—	0.076
—	—	tag	0.006
tag	tag	—	0.304
—	tag	tag	0.004
tag	—	tag	0.024
tag	tag	tag	0.016
Total			1.000

The tagging probability is a function of jet p_T and η , and for illustration I have assumed these jets to have tag-rate values of: Jet1 = 0.80, Jet2 = 0.40, Jet3 = 0.05. The event weight for a Jet1 tag only is thus $0.80 \times (1 - 0.40) \times (1 - 0.05) = 0.456$. In this three jet event, there is 1 untagged permutation, 3 single tag permutations, 3 double tag permutations, and 1 triple tag permutation. For the single tag channel, this event would be copied into the three single tag permutations, and their total weight would still add up to the weight of the original “single-tagged” event, 0.538. The double tag channel follows similarly. It is important to note that the same Monte Carlo events are used in the single and double tagged analyses, but are weighted by the appropriate tagging probabilities to differentiate the two channels.

5.6.3 Variables

The initial list of sensitive variables studied in this analysis is presented below. A description of each variable is given, and the optimized networks that use this variable, if any, are listed (s = single tag, d = double tag). Figures 5.7–5.12 show the background model and data comparison for those variables used in the final optimized networks. Accompanying each of these plots are the distributions for t -channel single top and the background(s) this variable is used to discriminate against. We note that the angular variables are not used in the muon channel analysis. These variables did not show good agreement in the background model and data comparison, specifically because of the *multijet* model. All other distributions show acceptable agreement from the *multijet*

sample. Furthermore, this is an isolated occurrence in the muon channel, as the electron channel has a very different source of *multijet* events. It was decided that omitting these variables was the best decision for the muon channel neural networks.

Single-Object Kinematics

1. **$E_T(\text{jet1})$** : Transverse energy of the leading jet. [d: $t\bar{t} \rightarrow \ell + jets$]
2. **$E_T(\text{jet2})$** : Transverse energy of the second leading jet. [d: Wbb, Wjj]
3. **$E_T(\text{tag1})$** : Transverse energy of the leading tagged jet.
4. **$E_T(\text{untag1})$** : Transverse energy of the leading untagged jet. This is the leading jet that is not b -tagged.
5. **\cancel{E}_T** : The transverse component of the the missing energy in the event. [s: $t\bar{t} \rightarrow \ell\ell$, *multijet*]

Event Kinematics

6. **$E_T(\text{jet1}, \text{jet2})$** : Transverse energy of the system of the leading two jets, i.e., the four-vector sum of the leading two jets. [s: $Wbb, Wjj, t\bar{t} \rightarrow \ell\ell$]
7. **$E_T(\text{alljets} - \text{tag1})$** : Transverse energy of the system of all jets except the leading tagged jet.
8. **$H_T(\text{alljets})$** : Scalar sum of the transverse energies of all jets in the event. [s: $Wjj, t\bar{t} \rightarrow \ell + jets$, d: $Wbb, Wjj, t\bar{t} \rightarrow \ell + jets$]
9. **$H_T(\text{alljets} - \text{tag1})$** : Scalar sum of the transverse energies of all jets except the leading tagged jet.
10. **$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$** : Scalar sum of the transverse energies of all jets, the lepton, and the \cancel{E}_T . [s: $t\bar{t} \rightarrow \ell + jets, t\bar{t} \rightarrow \ell\ell, \text{multijet}$, d: $t\bar{t} \rightarrow \ell + jets, \text{multijet}$]
11. **$H_T(\text{jet1}, \text{jet2})$** : Scalar sum of the transverse energies of the leading and second leading jets. [s: Wbb, Wjj]
12. **$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$** : Scalar sum of the transverse energies of the leading jet, second leading jet, lepton, and \cancel{E}_T . [s: $Wbb, t\bar{t} \rightarrow \ell + jets, t\bar{t} \rightarrow \ell\ell$, d: $Wjj, t\bar{t} \rightarrow \ell\ell$]
13. **$H(\text{alljets})$** : Scalar sum of the energy of all jets in the event. [d: $Wbb, t\bar{t} \rightarrow \ell\ell$]
14. **$H(\text{alljets} - \text{tag1})$** : Scalar sum of the energy of all jets except the leading tagged jet. [s: $Wbb, Wjj, t\bar{t} \rightarrow \ell + jets, t\bar{t} \rightarrow \ell\ell$]

15. **$H(\text{alljets}, \text{lepton}, \cancel{E}_T)$** : Scalar sum of the energy of all jets, lepton, and \cancel{E}_T . [d: Wbb]
16. **$H(\text{jet1}, \text{jet2})$** : Scalar sum of the energy of the leading and second leading jets.
17. **$H(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$** : Scalar sum of the energy of the leading jet, second leading jet, lepton, and \cancel{E}_T .
18. **$M_T(\text{jet1}, \text{jet2})$** : Transverse mass ($= \sqrt{(E_1 + E_2)^2 - (p_{x,1} + p_{x,2})^2 - (p_{y,1} + p_{y,2})^2}$) of the system of the leading two jets.
19. **$M_T(W)$** : Transverse mass ($= \sqrt{(E_\ell + E_\nu)^2 - (p_{x,\ell} + p_{x,\nu})^2 - (p_{y,\ell} + p_{y,\nu})^2}$) of the W boson, constructed from the lepton and \cancel{E}_T . [s: $t\bar{t} \rightarrow \ell\ell$, *multijet*, d: $t\bar{t} \rightarrow \ell\ell$, *multijet*]
20. **$M(\text{alljets})$** : Invariant mass of the system of all jets, i.e., the four-vector sum of all jets in the event.
21. **$M(\text{alljets} - \text{tag1})$** : Invariant mass of the system of all the jets except the leading tagged jet. [s: Wbb , $t\bar{t} \rightarrow \ell + \text{jets}$]
22. **$M(\text{alljets}, \text{lepton}, \cancel{E}_T)$** : Invariant mass of the system of all jets, lepton, and \cancel{E}_T .
23. **$M(\text{jet1}, \text{jet2})$** : Invariant mass of the system of the leading and second leading jets.
24. **$M(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$** : Invariant mass of the system of the leading and second leading jets, lepton, and \cancel{E}_T .
25. **$M(W, \text{tag1})$** : Invariant mass of the system of the W boson and the leading tagged jet (the reconstructed tagged top quark mass). [s: Wjj]
26. **$\sqrt{\hat{s}}$** : Invariant mass of the system of all objects in the event, $\sqrt{(\sum E_i)^2 - (\sum \vec{p}_i)^2}$, where the sums go over the objects in the event. Equivalent to the center of mass energy. [d: Wbb , $t\bar{t} \rightarrow \ell + \text{jets}$]
27. ***Aplanarity* (\mathcal{A})**: a variable used to describe the momentum flow of jets in the event. It is based on the smallest of the three eigenvalues of the momentum tensor M_{ab} of the jets in the event. The eigenvalues of planar events would be: $Q_1 = Q_2 = 0$ and $Q_3 = 1$. $\mathcal{A} = \frac{3}{2}Q_1$
28. ***Sphericity* (\mathcal{S})**: a variable used to describe the momentum flow of jets in the event. It is based on the smaller of two eigenvalues of the momentum tensor M_{ab} of the jets in the event. The eigenvalue of spherical events would be: $Q_1 = Q_2 = Q_3 = \frac{1}{3}$. $\mathcal{S} = \frac{3}{2}(Q_1 + Q_2)$

Angular Variables

29. **$\Delta R(\text{jet1}, \text{jet2})$** : Angular separation in η, ϕ between the leading two jets. [s: Wbb , $t\bar{t} \rightarrow \ell + \text{jets}$, $t\bar{t} \rightarrow \ell\ell$, d: Wjj]
30. **$Q(\text{lepton}) \times \eta(\text{untag1})$** : Charge of the lepton multiplied by the pseudorapidity of the leading untagged jet. In the t -channel, the final state d quark produced with the top quark tends to go along the incoming proton direction. Similarly, the \bar{d} quark produced with the anti-top quark goes along the anti-proton direction. We take this CP symmetry into account by multiplying the η distribution by the charge of the lepton, which reflects the charge of the top quark.
31. **$\cos(\text{jet1}, \text{lepton})_{\text{lab}}$** : Cosine of the angle between the leading jet and the lepton in the laboratory rest frame. [d: Wbb]
32. **$\cos(\text{jet2}, \text{lepton})_{\text{lab}}$** : Cosine of the angle between the second leading jet and the lepton in the laboratory rest frame. [s: Wjj]
33. **$\cos(\text{tag1}, \text{lepton})_{\text{lab}}$** : Cosine of the angle between the leading tagged jet and the lepton in the laboratory rest frame.
34. **$\cos(\text{jet1}, \text{alljets})_{\text{alljets}}$** : Cosine of the angle between the leading jet and the alljets system in the alljets reference frame.
35. **$\cos(\text{jet2}, \text{alljets})_{\text{alljets}}$** : Cosine of the angle between the second leading jet and the alljets system in the alljets reference frame.
36. **$\cos(\text{tag1}, \text{alljets})_{\text{alljets}}$** : Cosine of the angle between the leading tagged jet and the alljets system in the alljets rest frame.
37. **$\cos(\text{untag1}, \text{alljets})_{\text{alljets}}$** : Cosine of the angle between the leading untagged jet and the alljets system in the alljets reference frame.
38. **$\cos(\text{jet1}, \text{lepton})_{\text{tag1top}}$** : Cosine of the angle between the leading jet and the lepton in the rest frame of the top quark reconstructed with the leading tagged jet. [s: Wbb , Wjj]
39. **$\cos(\text{jet2}, \text{lepton})_{\text{tag1top}}$** : Cosine of the angle between the second leading jet and the lepton in the rest frame of the top quark reconstructed with the leading tagged jet.
40. **$\cos(\text{tag1}, \text{lepton})_{\text{tag1top}}$** : Cosine of the angle between the leading tagged jet and the lepton in the rest frame of the top quark reconstructed with the leading tagged jet. [s: *multijet*]
41. **$\cos(\text{untag1}, \text{lepton})_{\text{tag1top}}$** : Cosine of the angle between the leading untagged jet and the lepton in the rest frame of the top quark reconstructed with the leading tagged jet.

NEURAL NETWORK INPUT VARIABLES

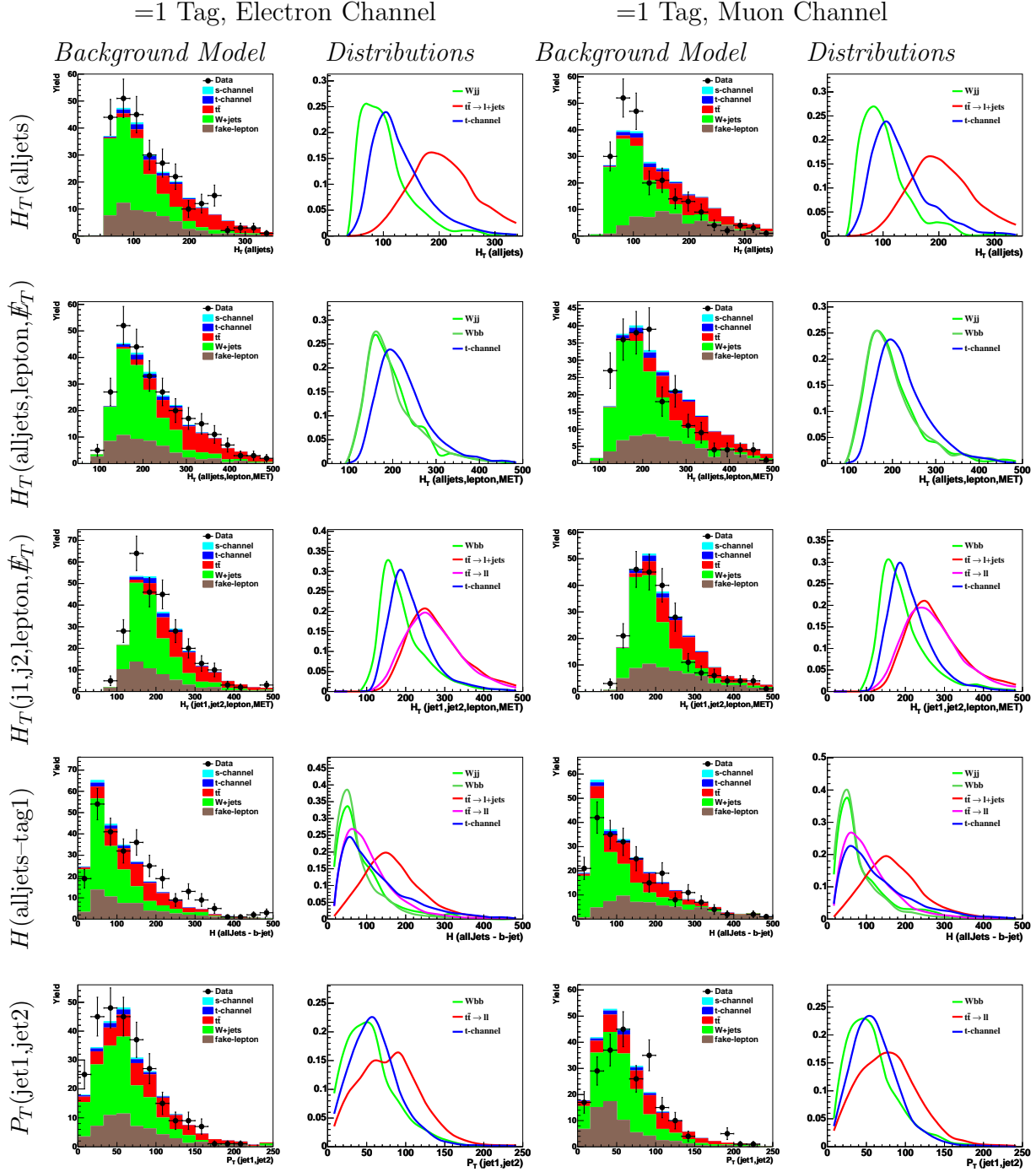


Figure 5.7: Distributions for input variables in the single tag neural networks. For each lepton channel, we show the background model compared to data (left column) and the unit-normalized distributions for each network that uses the variable (right column). Starting with the top row, we show $H_T(\text{alljets})$, $H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$, $H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$, $H(\text{alljets} - \text{tag1})$, and $P_T(\text{jet1}, \text{jet2})$.

NEURAL NETWORK INPUT VARIABLES

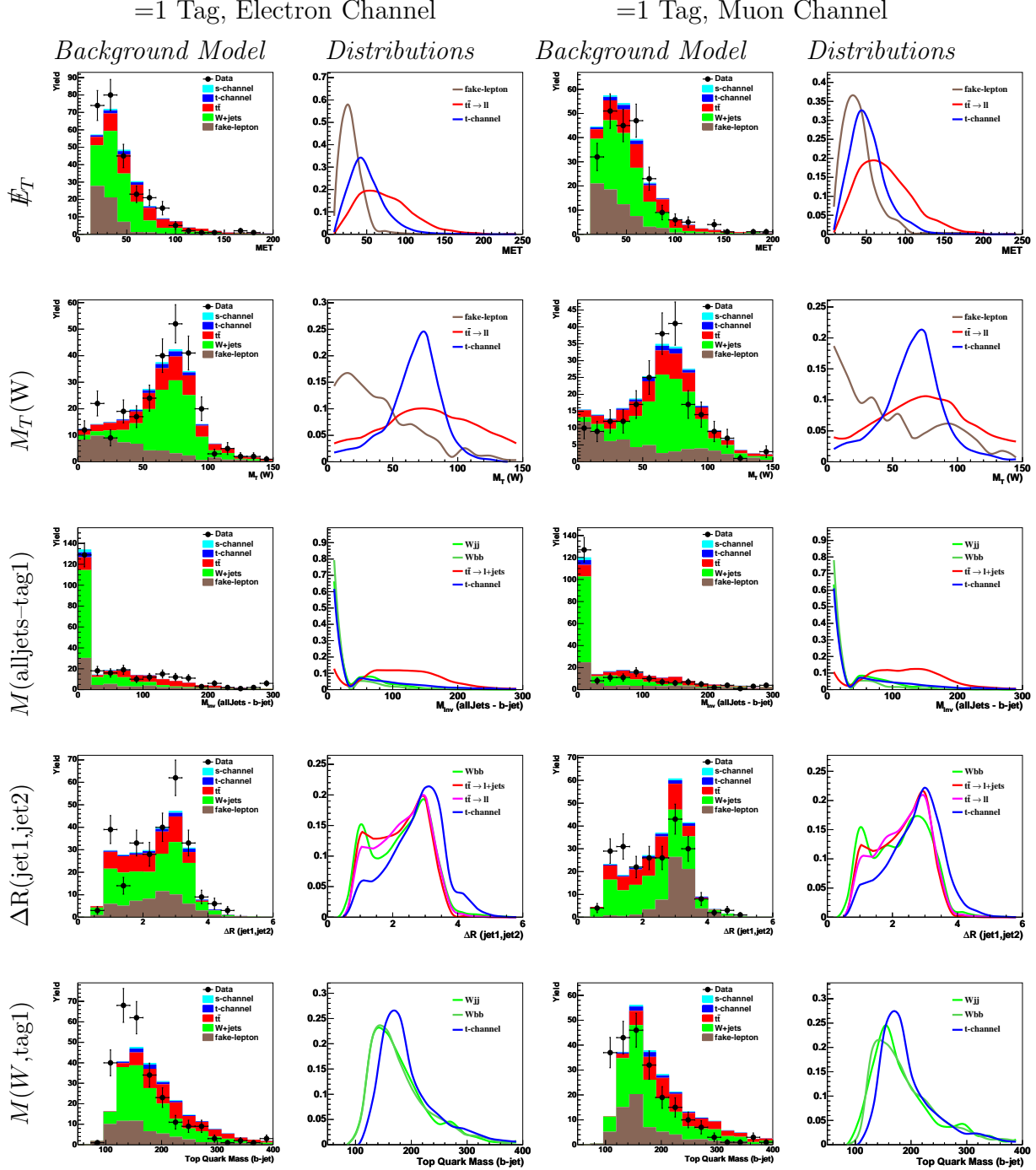


Figure 5.8: Distributions for input variables in the single tag neural networks. For each lepton channel, we show the background model vs data (left column) and the unit-normalized distributions for each network that uses the variable (right column). Starting with the top row, we show \cancel{E}_T , $M_T(W)$, $M(\text{alljets}-\text{tag1})$, $\Delta R(\text{jet1}, \text{jet2})$, and $M(W, \text{tag1})$.

NEURAL NETWORK INPUT VARIABLES

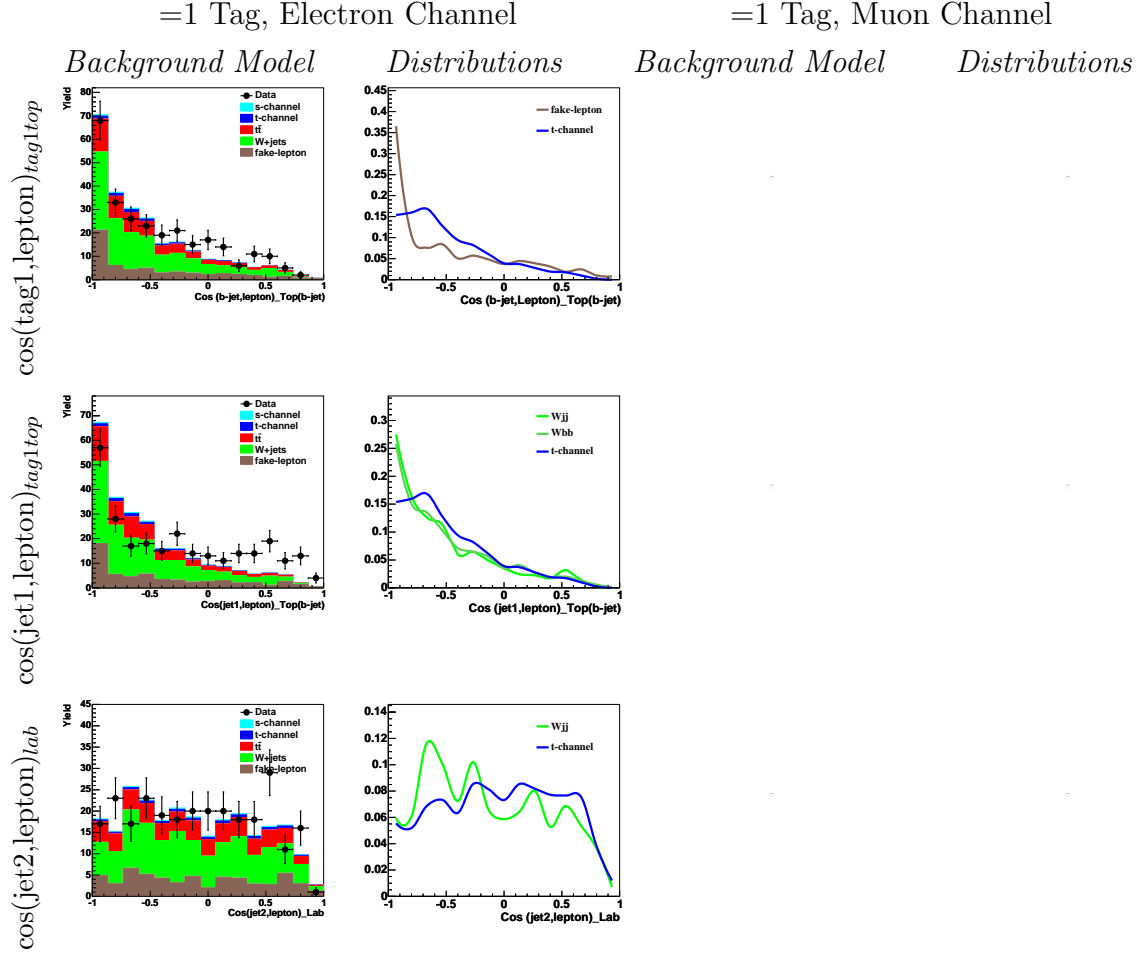


Figure 5.9: Distributions for input variables in the single tag neural networks. For each lepton channel, we show the background model vs data (left column) and the unit normalized distributions for each network that uses the variable (right column). Starting with the top row, we show the angular variables $\cos(\text{tag1,lepton})_{\text{tag1top}}$, $\cos(\text{jet1,lepton})_{\text{tag1top}}$, and $\cos(\text{jet2,lepton})_{\text{lab}}$. These angular variables are not used in the muon channel networks.

NEURAL NETWORK INPUT VARIABLES

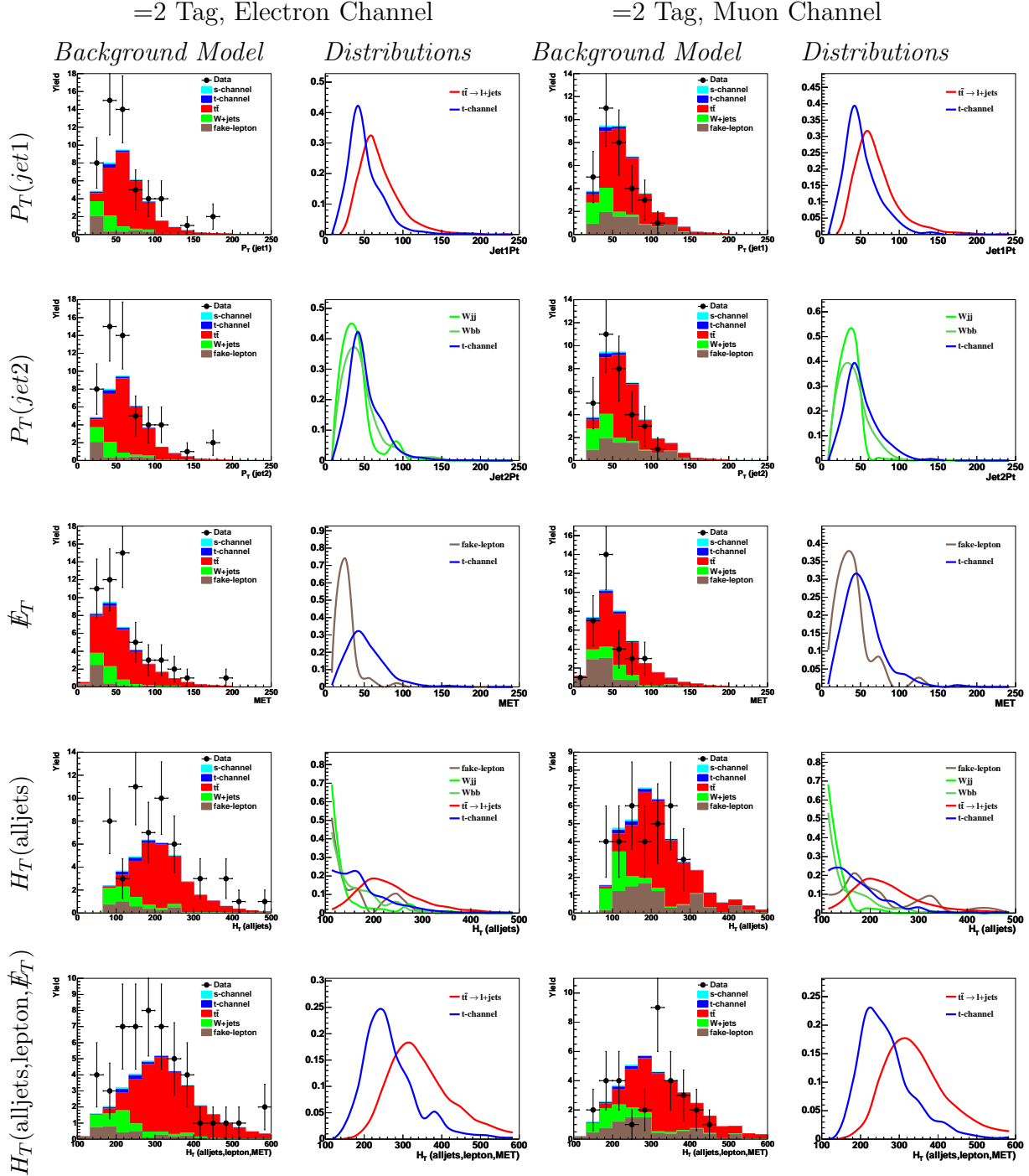


Figure 5.10: Distributions for input variables in the double tag neural networks. For each lepton channel, we show the background model compared to data (left column) and the unit-normalized distributions for each network that uses the variable (right column). Starting with the top row, we show $P_T(jet1)$, $P_T(jet2)$, \cancel{E}_T , $H_T(\text{alljets})$, and $H_T(\text{alljets,lepton}, \cancel{E}_T)$.

NEURAL NETWORK INPUT VARIABLES

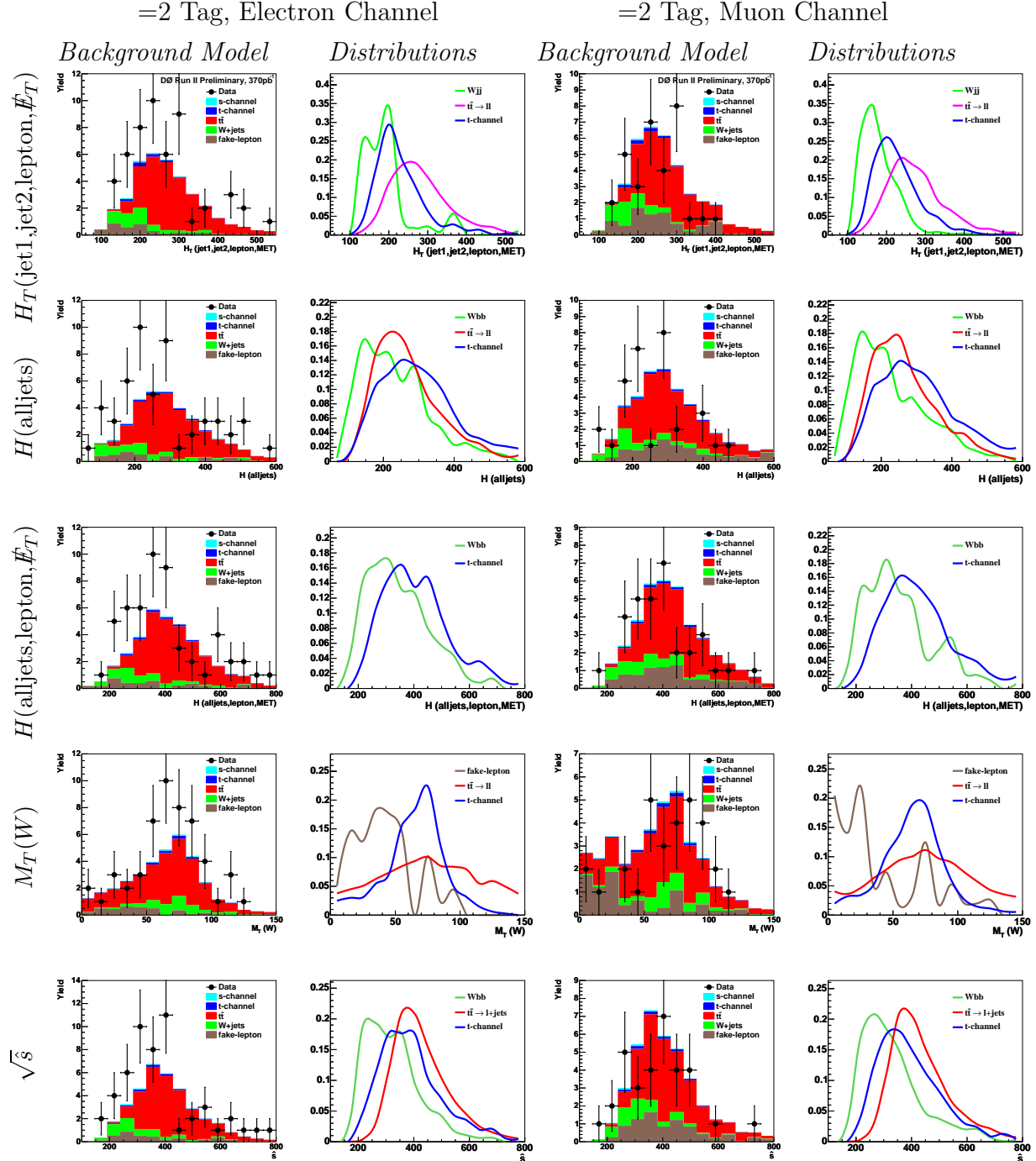


Figure 5.11: Distributions for input variables in the double tag neural networks. For each lepton channel, we show the background model compared to data (left column) and the unit-normalized distributions for each network that uses the variable (right column). Starting with the top row, we show $H_T(\text{jet1, jet2, lepton}, \cancel{E}_T)$, $H(\text{alljets})$, $H(\text{alljets, lepton}, \cancel{E}_T)$, $M_T(W)$, and $\sqrt{\hat{s}}$.

NEURAL NETWORK INPUT VARIABLES

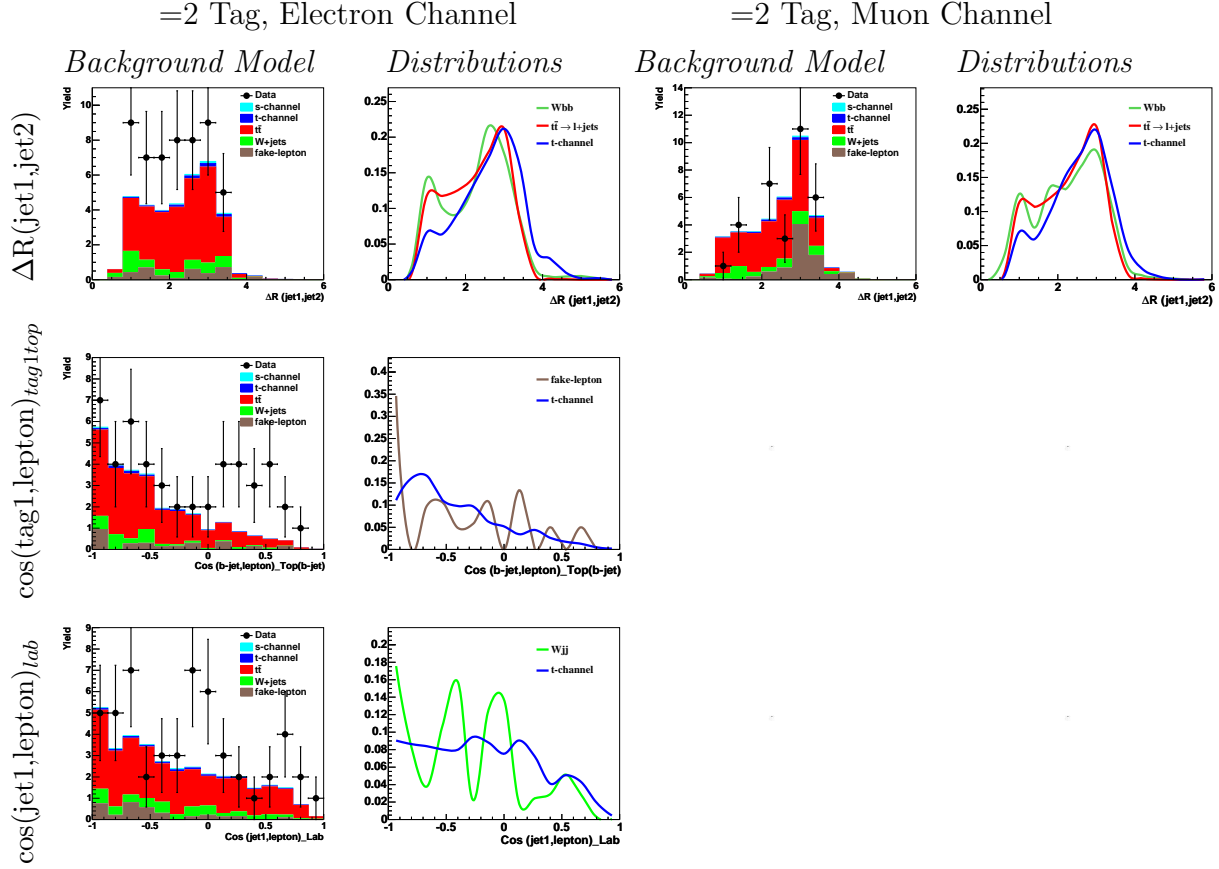


Figure 5.12: Distributions for input variables in the double tag neural networks. For each lepton channel, we show the background model compared to data (left column) and the unit-normalized distributions for each network that uses the variable (right column). Starting with the top row, we show $\Delta R(\text{jet1}, \text{jet2})$, $\cos(\text{tag1}, \text{lepton})_{\text{tag1top}}$, $\cos(\text{jet1}, \text{lepton})_{\text{lab}}$. These cosine variables are not used in the muon channel networks.

5.7 Optimized Neural Networks

Following the procedure given in Section 5.5, we have trained and optimized all 20 single background networks. The electron channel was mainly used in the optimization process, and the final variable choices found in the electron channel were then applied to the muon channel. The muon network architecture is still optimized on the muon samples,

and one or two modifications were made for obvious differences in the discrimination power achieved by blindly using the electron findings. Starting with the input variable optimization for each network, the decision tree ranking order is shown for pass 1 and pass 2 variables for each background and each tagging channel in the electron samples (Tables 5.3–5.7). Generally, one or two variables dominate the *gini* ranking, but we find that optimizing with further variables produces greater discrimination in the final networks. This is attributed to the fact that decision trees and neural networks are fundamentally different multivariate techniques, and while the decision tree can guide us in variable selection, it will not necessarily find the same optimum set as a neural network. As much as it guides, its use further helps to automate the optimization process. Automation is important since this analysis will be repeated in the future using larger datasets. Using the pass 2 list given in these tables, the optimized set of variables for each network is selected by adding the variables one at a time, in ranking order. The figure of merit is calculated for each subset of variables, and the lowest figure of merit defines the optimum set. Figure 5.13 shows the figure of merit as a function of the number of input variables, added in the ranking order shown in the *gini* ranking tables. We found that some higher ranked variables were detracting from the power of the lower ranked variables, and so removed these from the list and repeated the exercise. The figures of merit as a function of the final set of network input variables are shown in Figure 5.14. The associated set of variables and their order in these figure of merit graphs are given in Table 5.8. The muon channel uses the same set of variables, with the omission of all angular cosine variables. The architecture and training parameters for the final networks are summarized in Tables 5.11 and 5.12. The signal and background efficiency for the final neural networks are given in Figure 5.15. The final neural network output for each signal-background pair is given in Figure 5.16.

Comparing the network outputs, it is evident that the $t\bar{t}$ and *multijet* backgrounds are more easily discriminated against than the $W+jets$ samples. The most important

background in the single tag channel is the Wjj background and it has proved the most difficult to remove. I believe further studies could find a more optimal network, but the Wbb network provides some discriminative power over the Wjj background, as shown in the super neural network, and helps to deal with this background. The limited statistics in the double tag *multijet* sample required application of the single tag network to the double tag sample. This choice proved effective in the double tag channel as shown by the network output. The background model and data comparison for each neural network is in good agreement, and illustrates correct modeling of the data and effective network discrimination.

NEURAL NETWORK INPUT VARIABLES — RANKING ORDER

<i>Wbb</i> Single-Tag Network Input Variables			
Pass 1 Variables	Rank	Pass 2 Variables	Rank
$M(\text{alljets})$	22.00	$H(\text{alljets-tag1})$	14.80
$\cos(\text{jet1,lepton})_{\text{tag1top}}$	2.95	$\Delta R(\text{jet1,jet2})$	3.43
$M_T(W)$	1.51	$H_T(\text{jet1,jet2})$	3.31
$E_T(\text{jet1})$	1.25	$\cos(\text{jet1,lepton})_{\text{tag1top}}$	2.87
$H_T(\text{jet1,jet2,lepton},\cancel{E}_T)$	1.06	$E_T(\text{jet1,jet2})$	1.58
$E_T(\text{jet1,jet2})$	0.73	$H_T(\text{jet1,jet2,lepton},\cancel{E}_T)$	0.72
$M(W,\text{tag1})$	0.67	$M(\text{alljets-tag1})$	0.71
$M(\text{alljets-tag1})$	0.60	$M(W,\text{tag1})$	0.51
$E_T(\text{alljets-tag1})$	0.44	$H_T(\text{alljets,lepton},\cancel{E}_T)$	0.29
$\cos(\text{jet1,alljets})_{\text{alljets}}$	0.43		
$M(\text{jet1,jet2,lepton},\cancel{E}_T)$	0.40		
$H(\text{alljets})$	0.40		
$\cos(\text{jet1,lepton})_{\text{lab}}$	0.37		
Sum of all Rankings	34.74	Sum of all Rankings	28.22

<i>Wbb</i> Double-Tag Network Input Variables			
Pass 1 Variables	Rank	Pass 2 Variables	Rank
$M(\text{alljets})$	17.40	$H(\text{alljets})$	15.30
$M_T(W)$	5.73	$H_T(\text{alljets})$	6.86
$H(\text{alljets,lepton},\cancel{E}_T)$	5.23	$\Delta R(\text{jet1,jet2})$	6.59
$\cos(\text{jet1,lepton})_{\text{tag1top}}$	3.65	$E_T(\text{jet2})$	4.42
$\Delta R(\text{jet1,jet2})$	3.09	\sqrt{s}	4.15
\cancel{E}_T	2.11	$H(\text{alljets,lepton},\cancel{E}_T)$	3.98
$H((\text{jet1,jet2,lepton},\cancel{E}_T)\text{-tag1})$	1.55	$\cos(\text{jet1,lepton})_{\text{lab}}$	2.33
$\cos(\text{jet1,lepton})_{\text{lab}}$	1.47	$E_T(\text{jet1})$	2.31
<i>Aplanarity</i>	1.22	$H_T(\text{jet1,jet2,lepton},\cancel{E}_T)$	1.76
<i>Sphericity</i>	1.16	$H_T(\text{alljets,lepton},\cancel{E}_T)$	1.72
$H(\text{jet1,jet2})$	1.06		
$E_T(\text{jet1,jet2})$	1.00		
$E_T(\text{jet1})$	0.99		
$\cos(\text{jet2,lepton})_{\text{lab}}$	0.79		
$\cos(\text{jet1,alljets})_{\text{alljets}}$	0.70		
$M(\text{jet1,jet2,lepton},\cancel{E}_T)$	0.54		
Sum of all Rankings	49.08	Sum of all Rankings	49.42

Table 5.3: The decision tree ranking for variables considered in the *Wbb* networks. Pass 1 variables derive from the set of 41 variables, and Pass 2 variables derive from the combined set of optimum Pass 1 variables for all networks. Variables with a ranking greater than 1% of the sum of all rankings are considered for neural network training, and only these variables are shown.

NEURAL NETWORK INPUT VARIABLES — RANKING ORDER

<i>Wjj</i> Single-Tag Network Input Variables			
Pass 1 Variables	Rank	Pass 2 Variables	Rank
$H_T(\text{jet1}, \text{jet2})$	13.50	$H_T(\text{alljets})$	13.80
$M(\text{alljets})$	3.62	$H_T(\text{jet1}, \text{jet2})$	2.18
$\cos(\text{jet1}, \text{lepton})_{\text{tag1top}}$	1.51	$M(\text{alljets}-\text{tag1})$	2.10
$H(\text{alljets}-\text{tag1})$	1.25	$M(W, \text{tag1})$	1.82
$E_T(\text{alljets}-\text{tag1})$	1.13	$\cos(\text{jet1}, \text{lepton})_{\text{tag1top}}$	1.37
$H_T(\text{alljets}-\text{tag1})$	1.12	$H(\text{alljets}-\text{tag1})$	0.96
$Q \times \eta$	0.84	$E_T(\text{jet1}, \text{jet2})$	0.74
$\sqrt{\hat{s}}$	0.84	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	0.52
$\cos(\text{untag1}, \text{lepton})_{\text{tag1top}}$	0.80		
$M(W, \text{tag1})$	0.71		
$E_T(\text{jet1})$	0.69		
$E_T(\text{jet2})$	0.65		
$E_T(\text{jet1}, \text{jet2})$	0.54		
$\cos(\text{jet2}, \text{lepton})_{\text{lab}}$	0.51		
<i>Sphericity</i>	0.40		
$M_T(W)$	0.40		
$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	0.32		
Sum of all Rankings	30.29	Sum of all Rankings	23.79

<i>Wjj</i> Double-Tag Network Input Variables			
Pass 1 Variables	Rank	Pass 2 Variables	Rank
$H_T(\text{alljets})$	11.00	$H_T(\text{alljets})$	12.3
$E_T(\text{jet1})$	3.55	$E_T(\text{jet2})$	4.06
$E_T(\text{jet2})$	3.18	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	3.91
$M_T(W)$	3.13	$\cos(\text{jet1}, \text{lepton})_{\text{lab}}$	3.87
$\cos(\text{jet1}, \text{lepton})_{\text{lab}}$	2.03	$E_T(\text{jet1})$	3.66
$H(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	1.81	$H(\text{alljets})$	3.48
$H(\text{alljets})$	1.78	$\Delta R(\text{jet1}, \text{jet2})$	1.96
$M(\text{alljets})$	1.73	$\sqrt{\hat{s}}$	1.64
$E_T(\text{jet1}, \text{jet2})$	1.57	$H(\text{alljets}, \text{lepton}, \cancel{E}_T)$	1.06
<i>Sphericity</i>	1.33	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	0.58
$\cos(\text{jet1}, \text{alljets})_{\text{alljets}}$	1.21		
$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	1.16		
$M(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	1.06		
$M(\text{jet1}, \text{jet2})$	1.06		
$E_T(\cancel{E}_T)$	1.02		
$H(\text{jet1}, \text{jet2})$	1.00		
<i>Aplanarity</i>	0.84		
$\Delta R(\text{jet1}, \text{jet2})$	0.80		
$\cos(\text{jet2}, \text{lepton})_{\text{lab}}$	0.78		
$H_T(\text{jet1}, \text{jet2})$	0.60		
$M(\text{alljets}-\text{tag1})$	0.46		
$\sqrt{\hat{s}}$	0.46		
Sum of all Rankings	43.07	Sum of all Rankings	36.74

Table 5.4: The decision tree ranking for variables considered in the *Wjj* networks. Pass 1 variables derive from the set of 41 variables, and Pass 2 variables derive from the combined set of optimum Pass 1 variables for all networks. Variables with a ranking greater than 1% of the sum of all rankings are considered for neural network training, and only these variables are shown.

NEURAL NETWORK INPUT VARIABLES — RANKING ORDER

$t\bar{t} \rightarrow \ell + jets$ Single-Tag Network Input Variables			
Pass 1 Variables	Rank	Pass 2 Variables	Rank
$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	30.70	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	30.6
$M(\text{alljets}-\text{tag1})$	6.57	$M(\text{alljets}-\text{tag1})$	7.31
$H_T(\text{alljets}-\text{tag1})$	5.30	$H_T(\text{alljets})$	4.88
$\Delta R(\text{jet1}, \text{jet2})$	3.52	$\Delta R(\text{jet1}, \text{jet2})$	4.19
$H(\text{jet1}, \text{jet2})$	2.43	$H(\text{alljets}-\text{tag1})$	1.21
$H_T(\text{alljets})$	2.15	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	0.731
$M(\text{alljets})$	1.59	$E_T(\cancel{E}_T)$	0.685
\sqrt{hats}	0.94		
Sum of all Rankings	57.51	Sum of all Rankings	51.45

$t\bar{t} \rightarrow \ell + jets$ Double-Tag Network Input Variables			
Pass 1 Variables	Rank	Pass 2 Variables	Rank
$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	24.30	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	28.40
$\Delta R(\text{jet1}, \text{jet2})$	6.58	$\sqrt{\hat{s}}$	8.24
$H_T(\text{alljets})$	4.09	$\Delta R(\text{jet1}, \text{jet2})$	8.08
$H_T(\text{alljets}-\text{tag1})$	3.45	$H_T(\text{alljets})$	4.08
$\sqrt{\hat{s}}$	3.26	$E_T(\text{jet1})$	3.73
$M(\text{alljets})$	3.03	$H(\text{alljets})$	3.53
$M(\text{alljets}-\text{tag1})$	1.64	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	2.78
$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	1.55	$H(\text{alljets}, \text{lepton}, \cancel{E}_T)$	2.26
$H_T(\text{jet1}, \text{jet2})$	1.21	$\cos(\text{jet1}, \text{lepton})_{lab}$	1.79
$E_T(\cancel{E}_T)$	1.21	$M_T(W)$	1.76
$\cos(\text{jet1}, \text{alljets})_{alljets}$	1.19	$E_T(\text{jet2})$	1.41
$M_T(W)$	1.19	$H_T(\text{alljets}-\text{tag1})$	1.38
$M(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	1.17		
$E_T(\text{jet1})$	1.09		
$Sphericity$	1.07		
$H(\text{jet1}, \text{jet2})$	1.04		
$Aplanarity$	1.02		
$M_T(\text{jet1}, \text{jet2})$	1.01		
$H(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	0.92		
$H((\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)-\text{tag1})$	0.86		
$H(\text{alljets})$	0.68		
Sum of all Rankings	65.60	Sum of all Rankings	68.82

Table 5.5: The decision tree ranking for variables considered in the $t\bar{t} \rightarrow \ell + jets$ networks. Pass 1 variables derive from the set of 41 variables, and Pass 2 variables derive from the combined set of optimum Pass 1 variables for all networks. Variables with a ranking greater than 1% of the sum of all rankings are considered for neural network training, and only these variables are shown.

NEURAL NETWORK INPUT VARIABLES — RANKING ORDER

$t\bar{t} \rightarrow \ell\ell$ Single-Tag Network Input Variables			
Pass 1 Variables	Rank	Pass 2 Variables	Rank
$M_T(W)$	20.80	$M_T(W)$	20.20
$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	8.34	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	7.05
$\Delta R(\text{jet1}, \text{jet2})$	3.75	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	2.13
$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	1.90	$\Delta R(\text{jet1}, \text{jet2})$	1.96
$H(\text{jet1}, \text{jet2})$	1.06	$E_T(\text{jet1}, \text{jet2})$	1.37
$H(\text{alljets})$	0.91	$E_T(\cancel{E}_T)$	1.17
$H(\text{alljets}, \text{lepton}, \cancel{E}_T)$	0.54	$H(\text{alljets}-\text{tag1})$	1.13
$\cos(\text{untag1}, \text{alljets})_{\text{alljets}}$	0.48	$H(\text{jet1}, \text{jet2})$	1.11
$E_T(\text{jet1})$	0.45	$M(\text{alljets}-\text{tag1})$	0.64
		$H_T(\text{alljets})$	0.56
Sum of all Rankings	42.09	Sum of all Rankings	37.55

$t\bar{t} \rightarrow \ell\ell$ Double-Tag Network Input Variables			
Pass 1 Variables	Rank	Pass 2 Variables	Rank
$M_T(W)$	19.90	$M_T(W)$	19.7
$H_T(\text{je1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	6.40	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	7.37
$H(\text{alljets})$	3.47	$H(\text{alljets})$	6.52
$\Delta R(\text{jet1}, \text{jet2})$	3.05	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	3.21
$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	2.78	$H(\text{alljets}, \text{lepton}, \cancel{E}_T)$	2.40
\sqrt{s}	2.21	$E_T(\text{jet2})$	2.05
$H_T(\text{alljets})$	1.43	$\Delta R(\text{jet1}, \text{jet2})$	1.95
$\cos(\text{jet1}, \text{lepton})_{\text{lab}}$	1.20	$H_T(\text{alljets})$	1.54
$H(\text{alljets}, \text{lepton}, \cancel{E}_T)$	1.20	\sqrt{s}	1.51
$A_{\text{planarity}}$	1.17	$\cos(\text{jet1}, \text{lepton})_{\text{lab}}$	0.88
$S_{\text{sphericity}}$	0.99	$E_T(\text{jet1})$	0.78
$H(\text{jet1}, \text{jet2})$	0.88		
$E_T(\text{jet1}, \text{jet2})$	0.86		
$E_T(\cancel{E}_T)$	0.83		
$\cos(\text{jet1}, \text{alljets})_{\text{alljets}}$	0.69		
Sum of all Rankings	51.15	Sum of all Rankings	48.44

Table 5.6: The decision tree ranking for variables considered in the $t\bar{t} \rightarrow \ell\ell$ networks. Pass 1 variables derive from the set of 41 variables, and Pass 2 variables derive from the combined set of optimum Pass 1 variables for all networks. Variables with a ranking greater than 1% of the sum of all rankings are considered for neural network training, and only these variables are shown.

NEURAL NETWORK INPUT VARIABLES — RANKING ORDER

<i>multijet</i> Single-Tag Network Input Variables			
Pass 1 Variables	Rank	Pass 2 Variables	Rank
$E_T(\cancel{E}_T)$	28.50	$E_T(\cancel{E}_T)$	29.80
$M_T(W)$	9.28	$M_T(W)$	8.62
$E_T(\text{tag1})$	3.87	$\cos(\text{tag1}, \text{lepton})_{\text{tag1top}}$	6.29
$\cos(\text{tag1}, \text{lepton})_{\text{tag1top}}$	3.60	$E_T(\text{tag1})$	4.15
$E_T(\text{jet1})$	2.07	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	1.99
$M(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	1.54	$\Delta R(\text{jet1}, \text{jet2})$	0.90
$E_T(\text{untag1})$	0.86		
$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	0.61		
$\Delta R(\text{jet1}, \text{jet2})$	0.53		
$Q \times \eta$	0.53		
Sum of all Rankings	51.79	Sum of all Rankings	51.75

Table 5.7: The decision tree ranking for variables considered in the *multijet* networks. Pass 1 variables derive from the set of 41 variables, and Pass 2 variables derive from the combined set of optimum Pass 1 variables for all networks. Variables with a ranking greater than 1% of the sum of all rankings are considered for neural network training, and only these variables are shown. Low statistics did not allow for double-tag *multijet* training.

NEURAL NETWORK OUTPUT — FIGURE OF MERIT

Electron Channel Samples

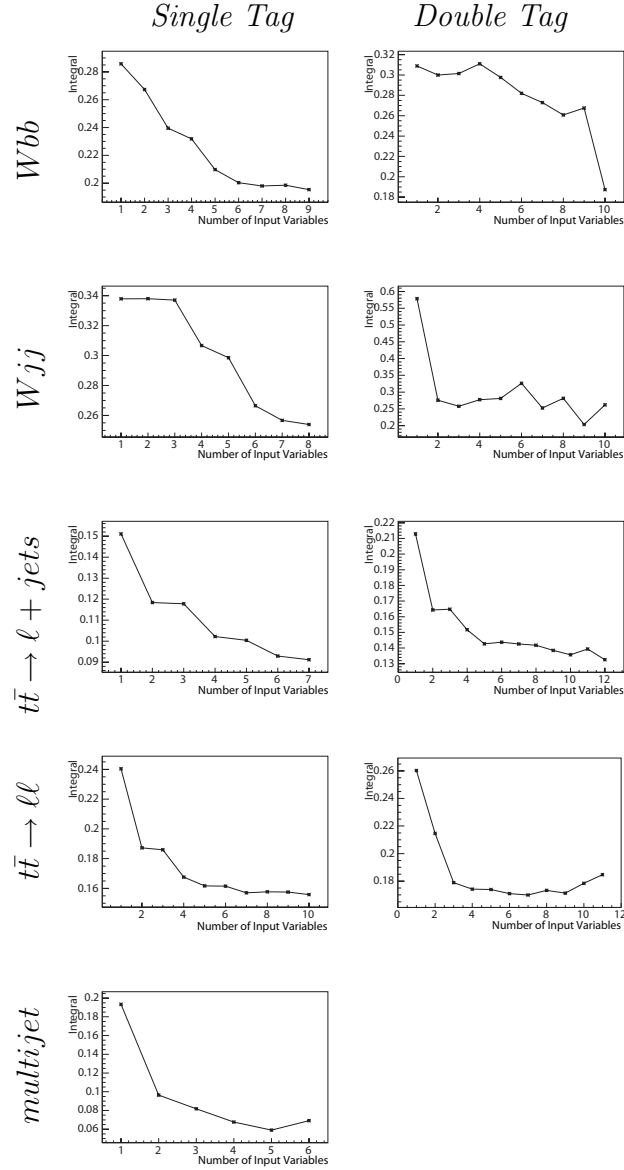


Figure 5.13: The figure of merit vs the number of input variables in the pass 2 list. Variables are added in the order of their decision tree ranking (given in Figures 5.3-5.7). Each column shows the optimization for single and double tag electron samples.

NEURAL NETWORK OUTPUT — FIGURE OF MERIT

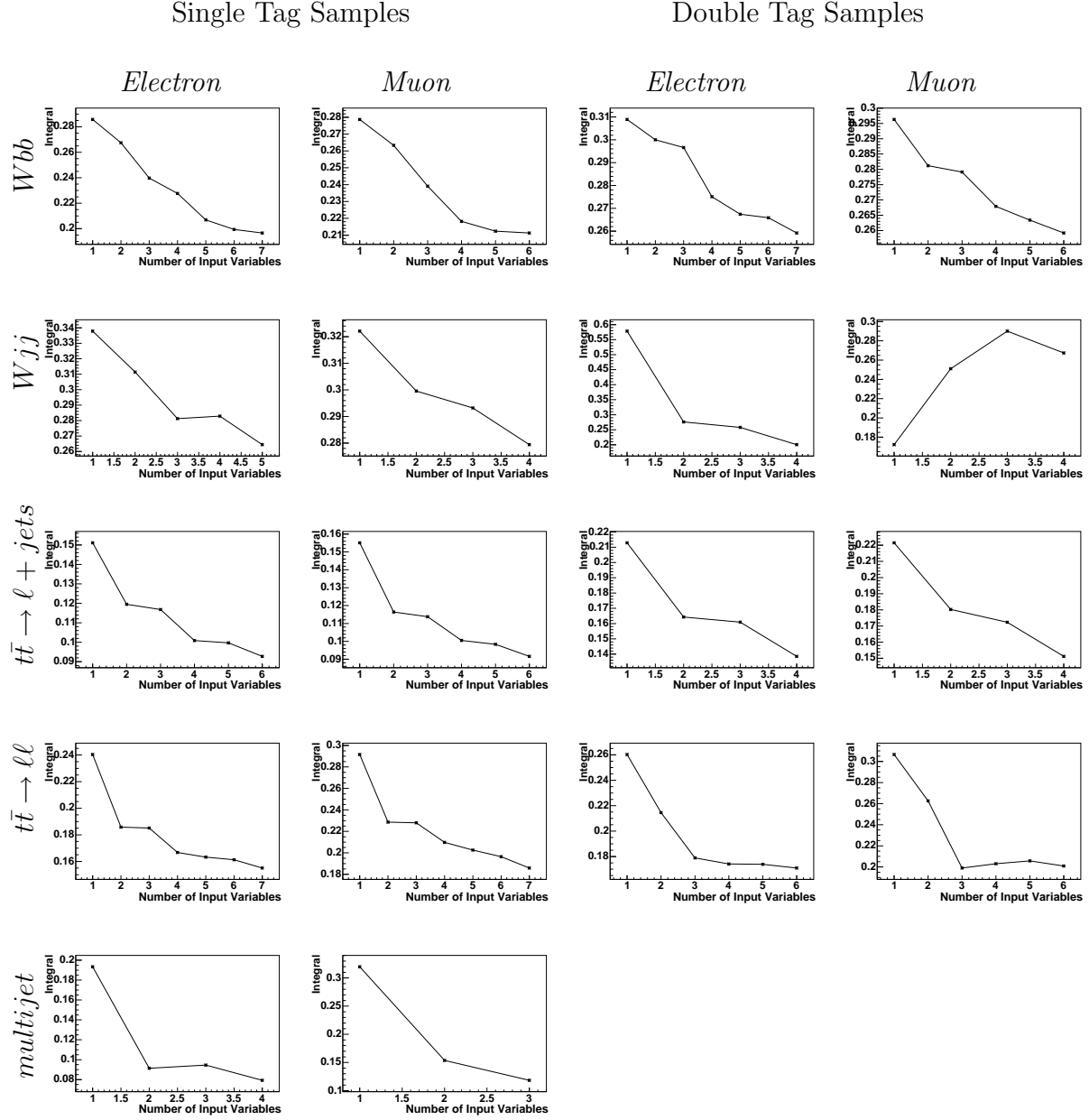


Figure 5.14: The figure of merit vs the number of input variables for the final set of variables. Variables are added one at a time, as given in Table 5.8. Each column shows the optimization for each lepton and tagging channel. Each row shows the optimization for each background. The double tag $multijet$ channel applies the single tag network.

FINAL NEURAL NETWORK INPUT VARIABLES

		Single Tag	Double Tag
Wbb	1	$H(\text{alljets} - \text{tag1})$	$H(\text{alljets})$
	2	$\Delta R(\text{jet1}, \text{jet2})$	$H_T(\text{alljets})$
	3	$H_T(\text{jet1}, \text{jet2})$	$\sqrt{\hat{s}}$
	4	$\text{Cos}(\text{jet1}, \text{lepton})_{\text{tag1top}}$	$H(\text{alljets}, \text{lepton}, \cancel{E}_T)$
	5	$E_T(\text{jet1}, \text{jet2})$	$\text{cos}(\text{jet1}, \text{lepton})_{\text{lab}}$
	6	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	$E_T(\text{jet1})$
	7	$M(\text{alljets} - \text{tag1})$	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$
Wjj	1	$H_T(\text{alljets})$	$H_T(\text{alljets})$
	2	$M(W, \text{tag1})$	$E_T(\text{jet2})$
	3	$\text{cos}(\text{jet1}, \text{lepton})_{\text{tag1top}}$	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$
	4	$H(\text{alljets} - \text{tag1})$	$\Delta R(\text{jet1}, \text{jet2})$
	5	$E_T(\text{jet1}, \text{jet2})$	
$t\bar{t} \rightarrow \ell + \text{jets}$	1	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$
	2	$M(\text{alljets} - \text{tag1})$	$\sqrt{\hat{s}}$
	3	$H_T(\text{alljets})$	$H_T(\text{alljets})$
	4	$\Delta R(\text{jet1}, \text{jet2})$	$E_T(\text{jet1})$
	5	$H(\text{alljets} - \text{tag1})$	
	6	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	
$t\bar{t} \rightarrow \ell\ell$	1	$M_T(W)$	$M_T(W)$
	2	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	$H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$
	3	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	$H(\text{alljets})$
	4	$\Delta R(\text{jet1}, \text{jet2})$	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$
	5	$E_T(\text{jet1}, \text{jet2})$	$H(\text{alljets}, \text{lepton}, \cancel{E}_T)$
	6	\cancel{E}_T	$E_T(\text{jet2})$
	7	$H(\text{alljets} - \text{tag1})$	
multijet	1	\cancel{E}_T	\cancel{E}_T
	2	$M_T(W)$	$M_T(W)$
	3	$\text{cos}(\text{tag1}, \text{lepton})_{\text{tag1top}}$	$\text{cos}(\text{tag1}, \text{lepton})_{\text{tag1top}}$
	4	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	$H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$

Table 5.8: The final list of input variables for each neural network. The order is the same as displayed in the figure-of-merit plots in Figure 5.14. The electron lists are given, and the muon lists are identical, apart from the removal of all angular cosine variables.

FINAL NEURAL NETWORK INPUT VARIABLES

	Optimized Single Tag Variables				
	Wbb	Wjj	$t\bar{t} \rightarrow l + \text{jets}$	$t\bar{t} \rightarrow ll$	$multijet$
1. $E_T(\text{jet1})$	—	—	—	—	—
2. $E_T(\text{jet2})$	—	—	—	—	—
3. \cancel{E}_T	—	—	—	✓	✓
4. $E_T(\text{jet1}, \text{jet2})$	✓	✓	—	✓	—
5. $H_T(\text{alljets})$	—	✓	✓	—	—
6. $H_T(\text{alljets}, l, \cancel{E}_T)$	—	—	✓	✓	✓
7. $H_T(\text{jet1}, \text{jet2})$	✓	—	—	—	—
8. $H_T(\text{jet1}, \text{jet2}, l, \cancel{E}_T)$	✓	—	✓	✓	—
9. $H(\text{alljets})$	—	—	—	—	—
10. $H(\text{alljets}, \text{lepton}, \cancel{E}_T)$	—	—	—	—	—
11. $H(\text{alljets} - \text{tag1})$	✓	✓	✓	✓	—
12. $M_T(W)$	—	—	—	✓	✓
13. $M(\text{alljets} - \text{tag1})$	✓	—	✓	—	—
14. $M(W, \text{tag1})$	—	✓	—	—	—
15. $\sqrt{\hat{s}}$	—	—	—	—	—
16. $\Delta R(\text{jet1}, \text{jet2})$	✓	—	✓	✓	—
17. $\cos(\text{jet1}, \text{lepton})_{\text{tag1}top}$	e	e	—	—	—
18. $\cos(\text{tag1}, \text{lepton})_{\text{tag1}top}$	—	—	—	—	e
19. $\cos(\text{jet1}, \text{lepton})_{lab}$	—	—	—	—	—

Table 5.9: The optimized set of neural network input variables for the single-tag analysis, after the first and second round of selection. There are a total of 13 unique variables for all networks combined. An “e” denotes the variable was only used in the electron networks, a “✓” for both lepton networks.

FINAL NEURAL NETWORK INPUT VARIABLES

	Optimized Double Tag Variables				
	Wbb	Wjj	$t\bar{t} \rightarrow l+\text{jets}$	$t\bar{t} \rightarrow ll$	$multijet$
1. $E_T(\text{jet1})$	—	—	✓	—	—
2. $E_T(\text{jet2})$	✓	✓	—	—	—
3. \cancel{E}_T	—	—	—	—	✓
4. $E_T(\text{jet1}, \text{jet2})$	—	—	—	—	—
5. $H_T(\text{alljets})$	✓	✓	✓	—	—
6. $H_T(\text{alljets}, \text{lepton}, \cancel{E}_T)$	—	—	✓	—	✓
7. $H_T(\text{jet1}, \text{jet2})$	—	—	—	—	—
8. $H_T(\text{jet1}, \text{jet2}, \text{lepton}, \cancel{E}_T)$	—	✓	—	✓	—
9. $H(\text{alljets})$	✓	—	—	✓	—
10. $H(\text{alljets}, \text{lepton}, \cancel{E}_T)$	✓	—	—	—	—
11. $H(\text{alljets}-\text{tag1})$	—	—	—	—	—
12. $M_T(W)$	—	—	—	✓	✓
13. $M(\text{alljets}-\text{tag1})$	—	—	—	—	—
14. $M(W, \text{tag1})$	—	—	—	—	—
15. $\sqrt{\hat{s}}$	✓	—	✓	—	—
16. $\Delta R(\text{jet1}, \text{jet2})$	—	✓	—	—	—
17. $\cos(\text{jet1}, \text{lepton})_{\text{tag1top}}$	—	—	—	—	—
18. $\cos(\text{tag1}, \text{lepton})_{\text{tag1top}}$	—	—	—	—	e
19. $\cos(\text{jet1}, \text{lepton})_{\text{lab}}$	e	—	—	—	—

Table 5.10: The final optimized set of neural network input variables for the double-tag analysis, after the first and second round of selection. There are a total of 13 unique variables for all networks combined. There were not enough statistics to train the *multijet* network, and so the final single tag *multijet* filter function is applied to the double tag *multijet* sample. An “e” denotes the variable was only used in the electron networks, a “✓” for both lepton networks.

<u>t-channel, Electron Network Configurations</u>						
Network	Single Tag			Double Tag		
	Number of Variables	Hidden Nodes	Best Epoch	Number of Variables	Hidden Nodes	Best Epoch
$tqb-Wbb$	7	14	84	7	12	37
$tqb-t\bar{t}(\ell+\text{jets})$	6	12	81	4	8	68
$tqb-t\bar{t}(\ell\ell)$	7	12	83	6	6	66
$tqb-Wjj$	5	8	82	4	6	50
$tqb-multijet$	4	8	49	4	8	49

Table 5.11: The final optimized values of the t -channel electron networks for the number of variables, number of hidden nodes, and number of training epochs. There were not enough statistics to train the double tag $multijet$ network, and so the single tag network is applied to the double tag analysis.

<u>t-channel, Muon Network Configurations</u>						
Network	Single Tag			Double Tag		
	Number of Variables	Hidden Nodes	Best Epoch	Number of Variables	Hidden Nodes	Best Epoch
$tqb-Wbb$	6	11	69	6	10	27
$tqb-t\bar{t}(\ell+\text{jets})$	6	9	69	4	7	64
$tqb-t\bar{t}(\ell\ell)$	7	14	69	6	12	69
$tqb-Wjj$	4	4	24	4	6	68
$tqb-multijet$	3	4	61	3	4	61

Table 5.12: The final optimized values of the t -channel muon networks for the number of variables, number of hidden nodes, and number of training epochs. There were not enough statistics to train the double tag $multijet$ network, and so the single tag network is applied to the double tag analysis.

NEURAL NETWORK OUTPUT — EFFICIENCY

Single Tag Channels

Double Tag Channels

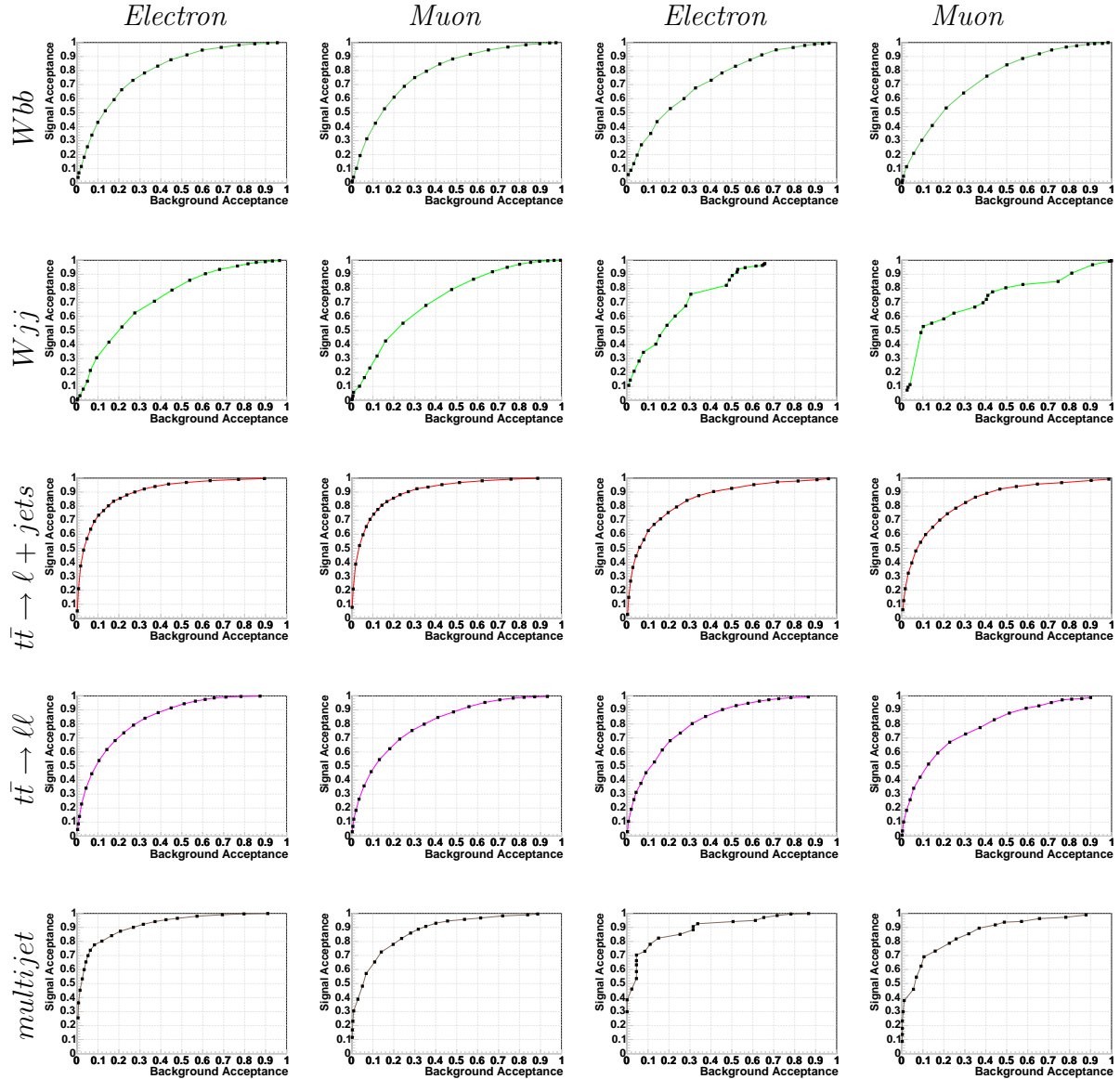


Figure 5.15: The signal versus background acceptance curves based on cut points on the neural network output distribution. Columns are organized by lepton and tagging channel. Rows are organized by the background discriminated against. The figure of merit is defined as the integral of the area above each of curve.

NEURAL NETWORK OUTPUT — BACKGROUND SEPARATION

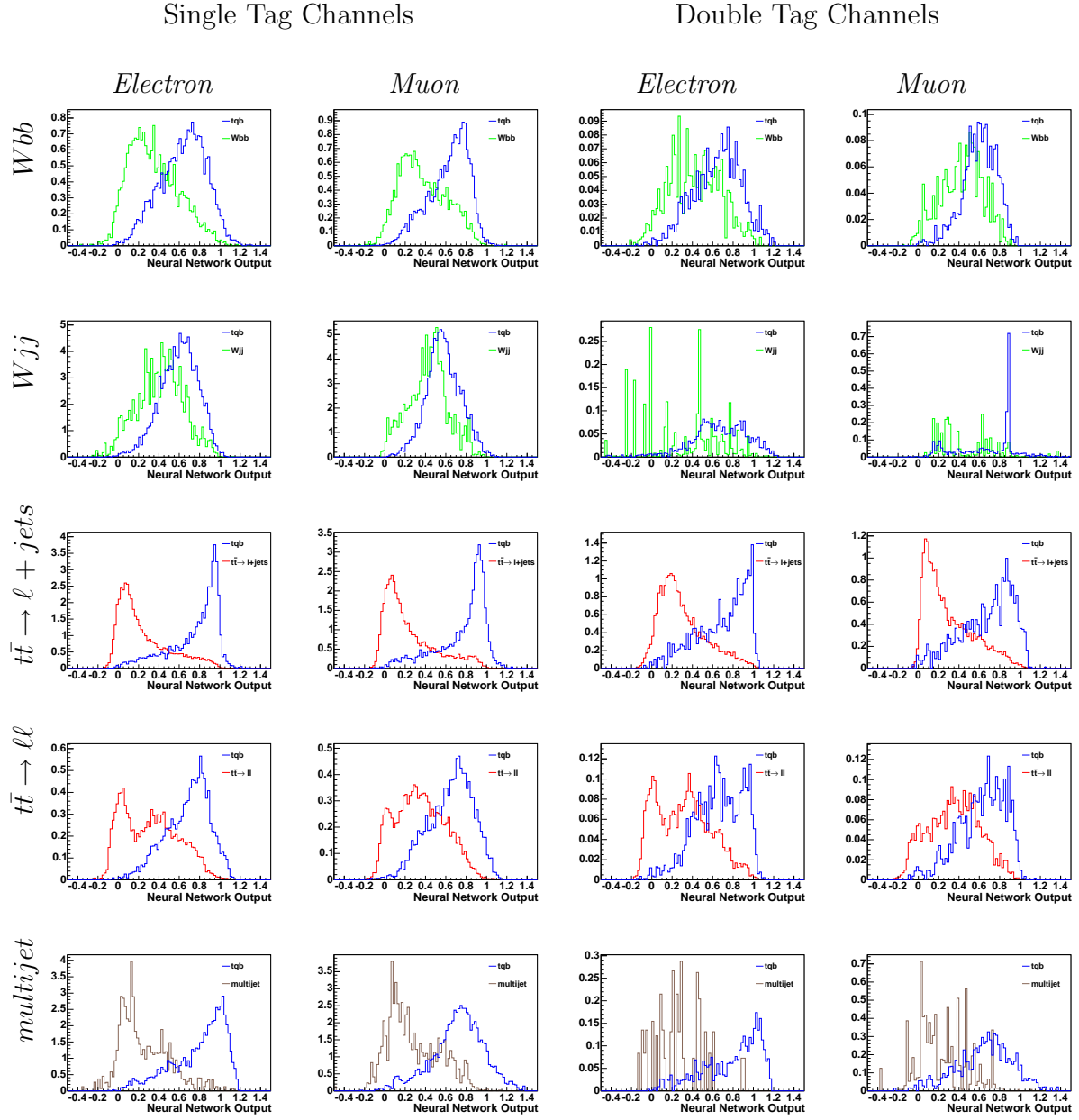


Figure 5.16: The final optimized neural network output for each lepton and tagging channel is shown. Each row shows the output for a specific background: Wbb , Wjj , $t\bar{t} \rightarrow \ell + jets$, $t\bar{t} \rightarrow \ell\ell$, and $multijet$. In each distribution, the background peaks around zero and the signal peaks around one.

NEURAL NETWORK OUTPUT — BACKGROUND MODEL AND DATA

Single Tag Channel

Double Tag Channel

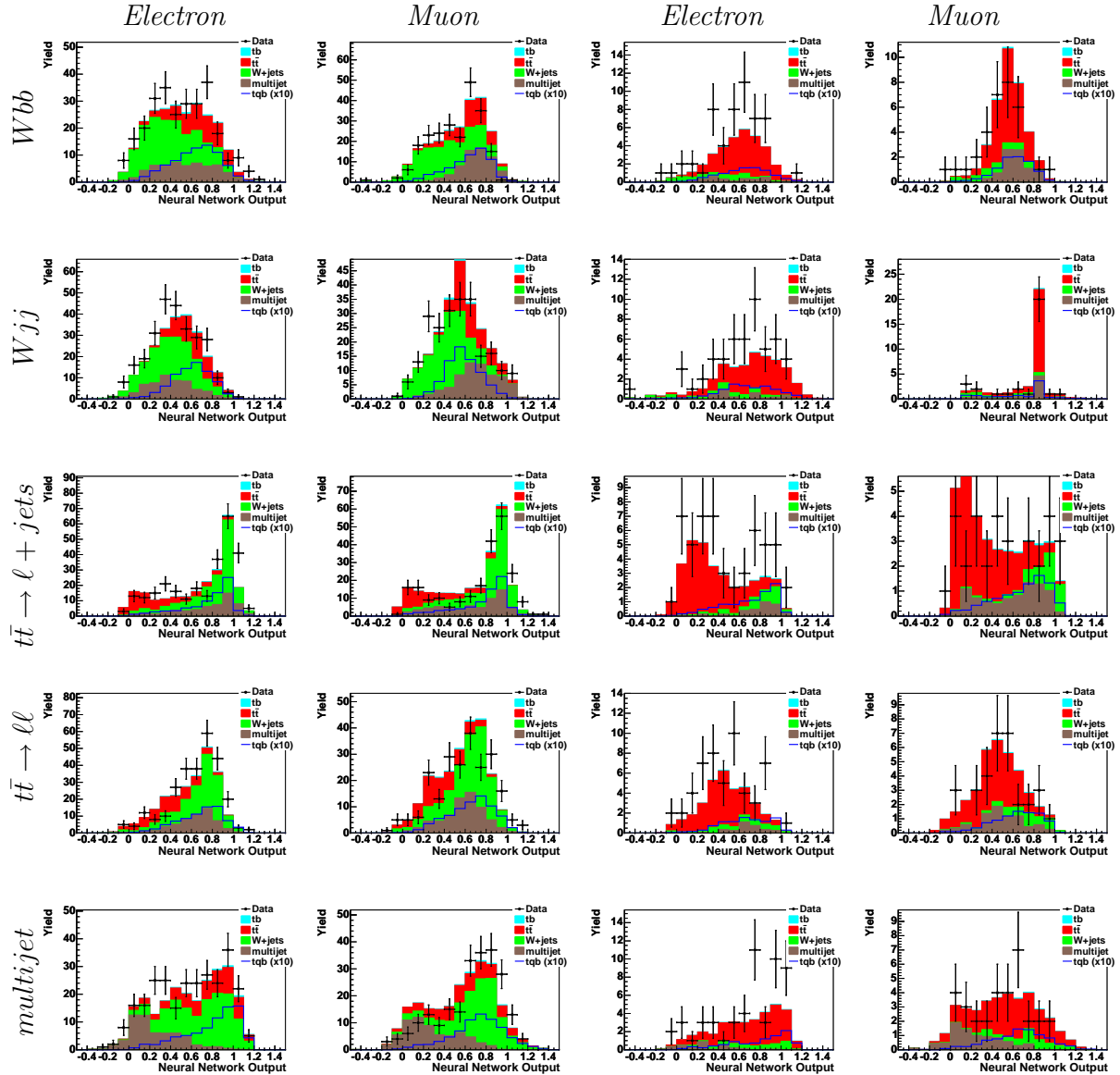


Figure 5.17: The background model vs data is shown for each neural network. The columns are split by tagging and lepton channel, and the rows by each background discriminated against: Wbb , Wjj , $t\bar{t} \rightarrow \ell + jets$, $t\bar{t} \rightarrow \ell\ell$, and $multijet$.

5.8 Optimized Super Neural Network

The final step in our neural network approach is combining the discriminative power of all five signal-background pair networks, into one so-called “super” neural network. While our goal remains to derive the best cross section limit, we also seek clarity in our approach and the results we share with others. With this aim, the primary goal of the super neural network is not to gain new discriminative power, but to summarize the sum total in a single distribution that can be used to derive limits. This is a more comprehensible approach, though no more correct, than deriving limits in a five dimensional neural network space.

In order to derive the best cross section limit, we optimize the set of networks used as inputs to the super neural network, and use the expected limit as our figure of merit. The final set of input networks was optimized in the electron channel, and then applied to the muon channel, but both channels are still trained on their respective samples. The final optimized network parameters are given in Table 5.13, and the final super neural network output distributions for each signal and background are given in Figure 5.18. The background model compared to data for the super neural network output distribution is given in Figure 5.19.

First looking at the optimized set of inputs, the single tag networks excluded the use of the Wjj network. This is due to the poor discrimination power attained by this network. As mentioned before, further studies with this network could enhance its power, and thus enhance the final limit — this should be looked at in the future. In the double tag channel, all five networks helped to improve the final limit, and were used in the super neural network. For both tagging channels, the super network outputs show that the larger backgrounds were correctly targeted during training, and the others are either ignored, or discriminated against slightly. As expected, results in both lepton channels are similar. The background model and data comparison shows the super

neural network is still correctly modeling the data. The background model and the data are similar in shape, with fluctuations due to the differences in the data and background model yields (where the electron channel shows a deficit in the model, and the muon channel shows an excess).

	Electron		Muon	
	=1 Tag	=2 Tag	=1 Tag	=2 Tag
input variables	4	5	4	5
hidden nodes	8	8	8	8
training epochs	24	13	24	14

Table 5.13: The final optimized values of the super neural networks for the number of input variables, number of hidden nodes, and number of training epochs. Results are shown for the electron and muon channels for both single and double networks.

FINAL SUPER NETWORK INPUT VARIABLES

	Single Tag	Double Tag
1. Wbb	✓	✓
2. Wjj	—	✓
3. $t\bar{t} \rightarrow \ell + jets$	✓	✓
4. $t\bar{t} \rightarrow \ell\ell$	✓	✓
5. $multijet$	✓	✓

Table 5.14: The final set of inputs used in the single and double tag super neural networks. The same inputs are used for both lepton channels.

SUPER NEURAL NETWORK OUTPUT — BACKGROUND SEPARATION

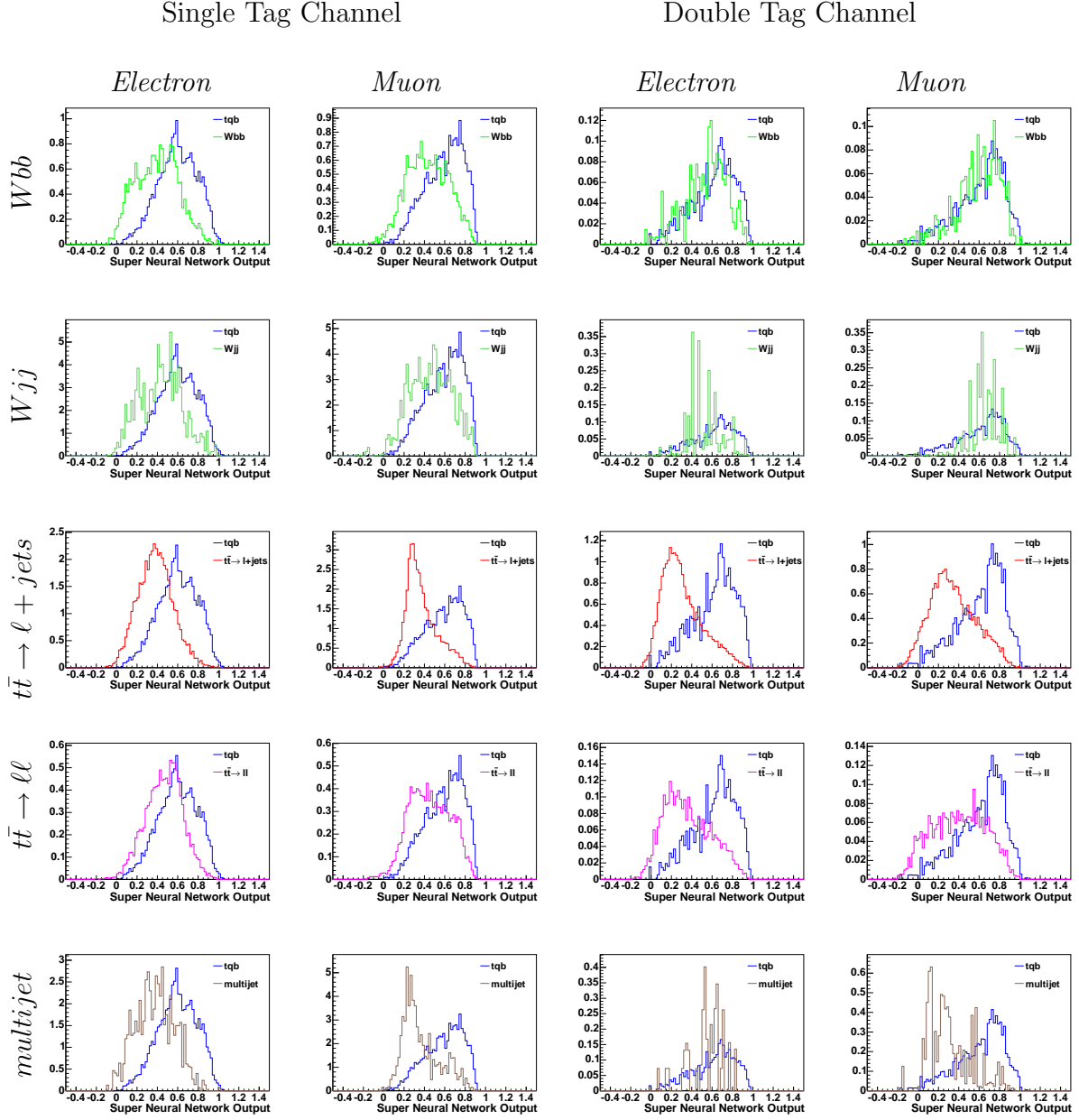
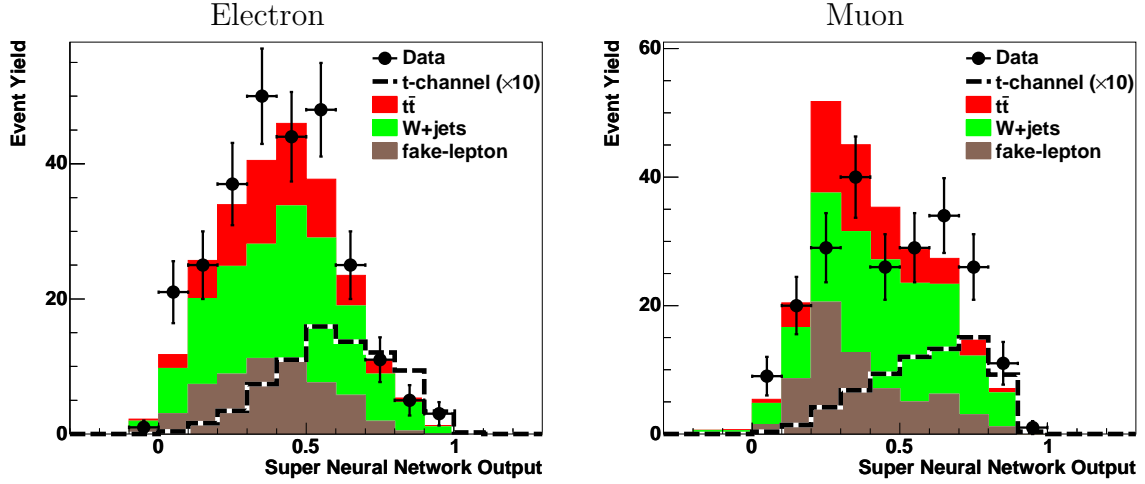


Figure 5.18: Super neural network outputs are shown for the final optimized networks. Each row shows the output for a specific background: Wbb , Wjj , $t\bar{t} \rightarrow l + \text{jets}$, $t\bar{t} \rightarrow ll$, and multijet . The signal distribution is identical for all distributions in a given column. The columns are separated by lepton and number of tags: Electron Single Tag, Muon Single Tag, Electron Double Tag, Muon Double Tag. In each distribution, the background peaks towards zero and the signal peaks towards one.

SUPER NEURAL NETWORK OUTPUT — BACKGROUND AND DATA

Single Tag



Double Tag

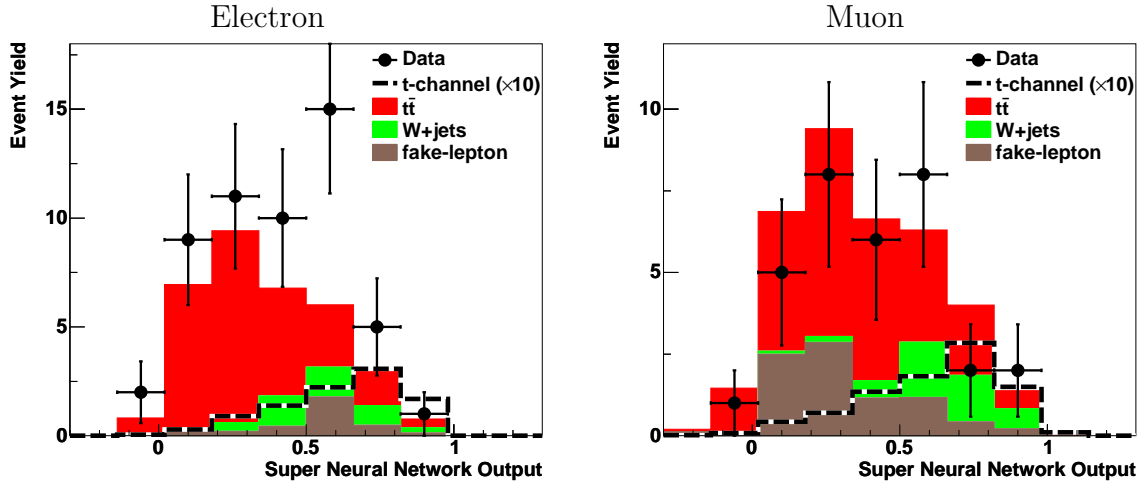


Figure 5.19: Super neural network output distributions are shown, comparing data and the background model for each channel. The signal distribution is multiplied by a factor of 10 to clearly show its distribution. In each histogram, the background model peaks towards zero and the signal peaks towards one.

Chapter 6

Setting Limits on Single Top Quark Production

6.1 Introduction

Now we must venture deep into the seedy world of statistics and probabilities. Through back doors, and down alley ways, we find the beady eyes of Frequentists and Bayesianists, each trying to lure us in with their wares, and hypnotize us with their statistical mystique. It is too late to turn back now. We have fought hard for the dataset at hand, dabbled with the dark arts of Monte Carlo generation, and poured blood and sweat into the final super neural network distributions we grip ever-so-tightly. And so we carry on.

We are ultimately after a single top quark production cross section measurement. With a still-limited data set, and many uncertainties damping the power of our search, we must refocus, and set our sights on a cross section limit. A typical first approach is to derive a limit based on total event counts, using the observed count, the predicted background yield, and predicted signal acceptance, to derive a 95% confidence level cross section limit. Such an approach ignores the plethora of information contained in the events themselves, and an improved approach takes into account the shapes of the

data, background model, and signal in the distribution of a discriminating variable. We have used neural networks to push the idea to an extreme, and conjured up our best attempt at the *ultimate* discriminating variable, taking into account all features of the signal and background phase space. Using the super neural network output distribution, we derive cross section limits based on its shape. Limits are derived separately for each lepton and tagging channel, with and without uncertainties, and finally combined.

6.2 Bayesian Approach to Limits

We are applying Bayesian statistics in our final analysis [114]. They are governed by the fundamental theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{Bayes' Theorem} \quad (6.1)$$

where the component probabilities are defined as:

- $P(A|B)$ = the probability of event A occurring, given event B
- $P(B|A)$ = the probability of event B occurring, given event A
- $P(A)$ = the probability of event A occurring (the prior for A)
- $P(B)$ = the probability of event B occurring (the prior for B)

We must now define the relevant parameters in our analysis. The observed number of events in our data sample D has been measured, and the predicted mean yield d for all backgrounds and signal has been modeled:

$$d = s + \sum_{i=1}^N b_i \quad (6.2)$$

$$= \sigma(\mathcal{AL}) + \sum_{i=1}^N b_i \quad (6.3)$$

where s is the predicted signal yield, and b_i is the predicted background yield for each of the N backgrounds we expect. The predicted signal is expressed in terms of the cross section (σ), signal acceptance (\mathcal{A}), and integrated luminosity (\mathcal{L}). For ease of notation, we define the vector of backgrounds $\mathbf{b} = (b_1, b_2, \dots)$, and the effective luminosity for the signal, $a \equiv \mathcal{A}\mathcal{L}$. We assume that the likelihood for our measurement D is proportional to the probability of D modeled by a Poisson distribution about the mean count d :

$$L(D|d) \propto P(D|d) = \frac{\exp(-d)d^D}{\Gamma(D+1)} \quad (6.4)$$

where the Gamma function $\Gamma(D+1) = D!$. We can then use Bayes' theorem as follows:

$$P(\mathbf{d}|D) = P(\sigma, a, \mathbf{b}|D) \equiv \frac{L(D|\mathbf{d})\pi(\mathbf{d})}{\mathcal{N}} \quad (6.5)$$

$$\pi(\mathbf{d}) = \pi(\sigma)\pi(a, \mathbf{b}) \quad (6.6)$$

where the probability to measure a mean yield d given D , $P(d|D)$, is a function of the likelihood defined in Equation 6.4, the prior probability for d , $P(d)$, and a normalization factor \mathcal{N} which facilitates unit probability, $\int P(\sigma|D)d\sigma = 1$. The cross section and yields are assumed independent, and the prior is factored into two components. We assume a flat prior for the cross section, where

$$\pi(\sigma) = \frac{1}{\sigma_{max}}, 0 < \sigma < \sigma_{max} \quad (6.7)$$

$$= 0, \quad \text{otherwise} \quad (6.8)$$

and σ_{max} is some value for which the posterior probability for $\sigma > \sigma_{max}$ is negligible. The choice of prior implies a maximal ignorance of any preferred value for the signal cross section, and is best regarded as a convention within our field. With these provisions, the probability of measuring a cross section given the observed number of events is:

$$P(\sigma|D) = \frac{1}{\mathcal{N}} \int \int L(D|\sigma, a, \mathbf{b}) \pi(a, \mathbf{b}) da d\mathbf{b} \quad (6.9)$$

A Bayesian upper limit σ_{CL} at a confidence level CL is a solution of Equation 6.10, while a cross section *measurement* is either the mode or mean of $P(\sigma|D)$.

$$\int_0^{\sigma_{CL}} P(\sigma|D) d\sigma = CL \quad (6.10)$$

In practice, the integral in Equation 6.9 is calculated numerically using Monte Carlo importance sampling. The integral is transformed into a swarm of points (a_k, \mathbf{b}_k) that represent the prior density $\pi(a, \mathbf{b})$, and we estimate the posterior using:

$$\frac{1}{\mathcal{N}} \int \int L(D|\sigma, a, \mathbf{b}) \pi(a, \mathbf{b}) da d\mathbf{b} = \frac{1}{K} \sum_{k=1}^K L(D|\sigma, a_k, \mathbf{b}_k) \quad (6.11)$$

For each sampled set of points, (a_k, \mathbf{b}_k) , the likelihood is then directly calculated (Equation 6.4). The most important step in this process is evaluating the prior, $\pi(a, \mathbf{b})$, correctly. Here we take into account the uncertainties on each yield from all sources. Each point a and \mathbf{b} is sampled from a multivariate Gaussian with a mean based on the modeled yield, and a width based upon a covariance matrix from all associated uncertainties, and taking into account all possible correlations.

The limit calculation described applies to total events and yields in the final sample. The calculation is easily adapted to take into account the shape of a discriminant variable distribution, by transforming the likelihood into a product of likelihoods calculated for each bin in the distribution:

$$L_{shape}(D|\mathbf{d}) = \prod_{bin=1}^M L(D_{bin}|\mathbf{d}_{bin}) \quad (6.12)$$

The uncertainties in the multivariate Gaussians now take on two components — the first being global uncertainties that apply to all bins equally, and the second being those that alter the shape of the distribution, such as the jet energy scale, jet energy resolution, jet identification, tag rate functions, and trigger modeling. The Gaussian sampling for global uncertainties is handled as before, using the signal and background yields as the mean, and uncertainties defining the width. The shape-altering uncertainties are then taken into account one at a time. The jet energy scale, jet energy resolution, jet identification, and so on, are each adjusted by plus or minus one standard deviation with respect to their nominal value, and the signal and background yields are calculated in each bin. Each systematic is then sampled from a Gaussian distribution with the nominal yield as the mean, and plus or minus yield as the width (1 standard deviation). If the random sampling gives a positive number, we use the plus yield as the width and sample from it, and if the number is negative, we use the negative yield as the width and sample from it. This is repeated for each shape-changing systematic, and the relative shifts from the nominal yields are added linearly. This is repeated for each bin, and the likelihood is then calculated, correctly taking into account all systematic uncertainties throughout the distributions.

Extending the limit calculation once more, we can combine results from various orthogonal channels through the product of their likelihoods:

$$L_{combined}(\mathbf{D}|\mathbf{d}) = \prod_{channel=1}^{N_c} L_{shape}(D_{channel}|\mathbf{d}_{channel}) \quad (6.13)$$

6.3 Cross Section Limits

Expected and measured limits have been calculated for a variety of circumstances, providing evidence for the increasing discriminant power of our final analysis. An expected limit provides a gauge of the quality and discrimination power of an analysis. In calcu-

lating the expected limit, one artificially sets the observed number of events D equal to the background yield. This “ideal” experiment is then used to tell you how well you are isolating your signal and how significant your signal is. For a measured limit, one sets D equal to the observed number of events in the data sample, and calculates the limit. A measured limit thus takes into account how well you have modeled your data sample in addition to the significance of a signal. An observation or discovery is commonly claimed when the peak of the measured posterior density probability distribution (Equation 6.9) is found several standard deviations away from zero.

We have first derived count-based limits with and without systematic uncertainties, as shown in Table 6.1. One will first notice the sizeable difference between the expected and measured limits. This is a direct result of the excess (deficit) in data with respect to the background model in the electron (muon) analysis, observed in both tagging channels. It is also noted that limits in the double tag channel are significantly worse than the single tag channel. This is expected, as t -channel single top only produces a second b -quark jet in NLO diagrams, and this generally softer jet may not be reconstructed by the detector. The tiny acceptance and overall available data in the double tag channel thus leave much less information with which to derive a limit, but we still pursue this channel as it enhances our final combined limits. A final observation one may make is that some of the muon channel limits actually improve with the addition of systematic uncertainties. This is found to be a feature of the jet-energy related systematics, particularly the jet energy scale and jet energy resolution. A $+(-)1\sigma$ shift in the systematic does not necessarily denote an increase(decrease) in the number of events for a sample in the output distribution, and so the limit can fluctuate either way.

Expected limits have also been derived based on the shape and separation power of the single-background neural networks shown in Section 5.7. The results of this binned likelihood calculation are shown in Table 6.2. Looking at the single tag electron channel, the single background networks show a 1–2 pb gain over the 10.0 pb count-

based expected limit. The double tag electron channel shows a larger 3–9 pb gain over the 32.7 pb count-based expected limit, and the muon channel shows similar results.

As a comparison to the previously published $D\bar{O}$ single top paper [67] that used 230 pb^{-1} , the expected limit with systematic uncertainties, derived from a combination of the Wbb and $t\bar{t} \rightarrow \ell + jets$ networks was 9.3 (10.1) pb in the single tag electron (muon) channel and 61.2 (60.5) pb in the double tag electron (muon) channel. Both the increase in integrated luminosity and further optimization of our neural networks have lead to the improvements over the previous analysis.

Finally, we have derived expected and measured limits based on the shape of the super neural network. The inputs to the super neural network were first optimized to yield the best expected limit with systematic uncertainties included. Following this optimization, the corresponding measured limits were then calculated. We probed various combinations of the five input background neural networks, and found that the use of four (five) networks provided the best limits in the single (double) tag channel. Several trials involving various combinations of super neural network inputs are shown in Table 6.4. The final binned likelihood calculation based on the shape of the super neural networks is performed, and final limits are shown in Table 6.3. Our final expected/measured 95% CL limit with systematics and both lepton and tagging channels combined is 3.5/8.0 pb. The posterior probability density distribution for the electron, muon and combined channels, with and without systematic uncertainties, are shown in Figure 6.1. In light of the higher signal to background ratio in the super network output distribution near a value of one, we wondered if limits would be improved by first making a hard cut on the super network distribution. Shown in Figure 6.2 is a plot of the expected limit in the electron channel as a function of a cut on the super network. It is clear that it is better to use all the information available in the shape of the entire distribution than just the region with the highest signal to background ratio. Again, comparing to the previous 230 pb^{-1} analysis, our analysis has improved on the previous combined expected limit

of 5.8 pb, but not the measured limit of 5.0 pb. This shows that the discrimination power of our analysis has improved beyond that given by statistics alone, but we can still improve our understanding of the background model.

TOTAL COUNT-BASED LIMITS

95% Confidence Level Expected/Measured Cross Section Limits			
	Single Tag	Double Tag	Combined
No Uncertainties			
Electron Channel	8.1/14.2	27.4/66.9	7.7/16.6
Muon Channel	8.8/6.3	30.7/23.5	8.4/5.7
Combined	5.9/7.2	19.6/34.9	5.6/8.2
With Uncertainties			
Electron Channel	10.0/16.1	32.7/77.8	9.8/17.7
Muon Channel	8.4/6.1	31.0/24.0	7.9/5.4
Combined	6.1/7.1	22.4/38.9	5.9/7.8

Table 6.1: The 95% confidence level cross section limits for t -channel single top production, based on totals counts only.

SINGLE BACKGROUND NETWORK SHAPE-BASED LIMITS

95% Confidence Level Expected Cross Section Limits				
Network	Electron Channel		Muon Channel	
	Single Tag	Double Tag	Single Tag	Double Tag
With Uncertainties				
Wbb	7.8	27.8	8.9	28.7
$t\bar{t} \rightarrow \ell + jets$	8.7	23.9	8.2	26.6
$t\bar{t} \rightarrow \ell\ell$	9.1	27.7	8.4	27.6
Wjj	8.5	29.4	8.4	29.7
$multijet$	8.1	29.0	8.1	27.7

Table 6.2: The 95% confidence level expected cross section limits for t -channel single top production in the individual single background neural networks.

SUPER NEURAL NETWORK SHAPE-BASED LIMITS

95% Confidence Level Expected/Measured Cross Section Limits			
	Single Tag	Double Tag	Combined
No Uncertainties			
Electron Channel	5.6/9.7	19.3/51.1	5.3/11.6
Muon Channel	6.5/10.9	22.8/20.2	6.1/9.7
Combined	4.1/8.5	13.5/27.2	3.8/9.0
With Uncertainties			
Electron Channel	5.8/10.0	22.3/51.9	5.5/11.4
Muon Channel	5.9/8.1	24.8/22.1	5.7/7.5
Combined	3.9/7.3	16.1/27.6	3.5/8.0

Table 6.3: The 95% confidence level cross section limits for t -channel single top production, derived from the shape of the super neural network output. The final, combined result of this analysis is highlighted in bold.

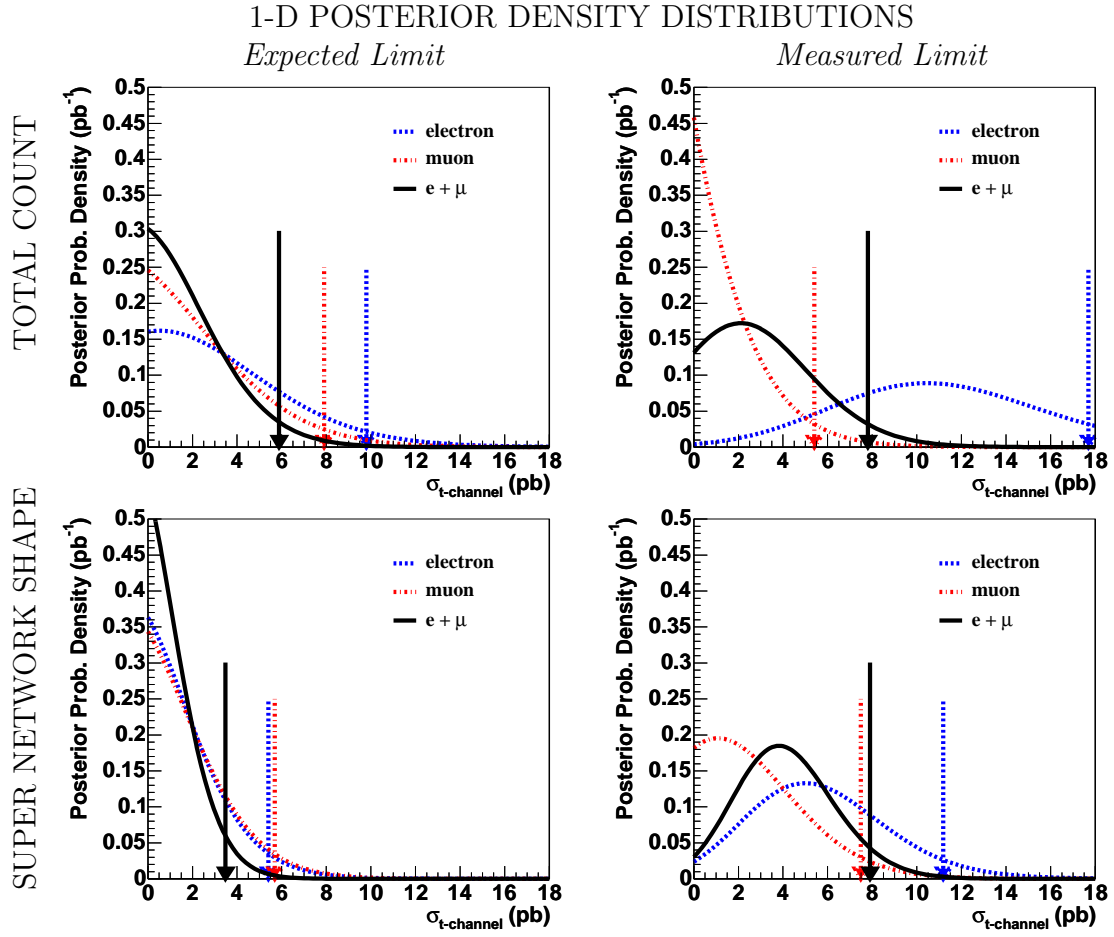


Figure 6.1: The posterior probability density distribution are shown for the combined tagging samples in the electron, muon and combined lepton channels. All distributions include systematic uncertainties and expected(measured) distributions are shown left(right). Limits derived on total event counts (above) and the shape of the super network output (below) are shown.

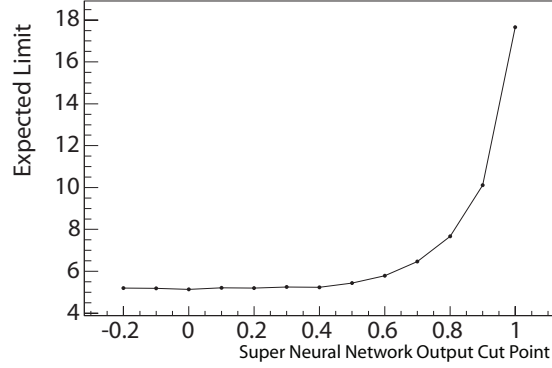


Figure 6.2: The 95% confidence level expected limit on the t -channel single top cross section as a function of a cut on the super neural network output. This was a preliminary version of the analysis, and not the final set, but the results still apply.

SUPER NEURAL NETWORK INPUT OPTIMIZATION

95% Confidence Level Expected Cross Section Limits (pb)			
Order	Limit	Order	Limit
<i>Trial 1</i>		<i>Trial 2</i>	
Wjj	—	$t\bar{t} \rightarrow \ell\ell$	—
+ $multijet$	8.0	+ Wbb	6.5
+ $t\bar{t} \rightarrow \ell + jets$	8.2	+ $t\bar{t} \rightarrow \ell + jets$	6.4
+ Wbb	6.0	+ $multijet$	5.8
+ $t\bar{t} \rightarrow \ell\ell$	6.0	+ Wjj	6.4
<i>Trial 3</i>		<i>Trial 4</i>	
$t\bar{t} \rightarrow \ell\ell$	—	$t\bar{t} \rightarrow \ell + jets$	—
+ $multijet$	7.6	+ $t\bar{t} \rightarrow \ell\ell$	8.1
+ Wjj	6.5	+ Wjj	6.7
+ Wbb	6.5	+ Wbb	6.5
+ $t\bar{t} \rightarrow \ell + jets$	6.3	+ $multijet$	6.0
<i>Trial 5</i>			
$t\bar{t} \rightarrow \ell\ell$	—		
+ $multijet$	7.6		
+ Wbb	5.9		
+ $t\bar{t} \rightarrow \ell + jets$	5.8		
+ Wjj	6.5		

Table 6.4: To optimize the super neural network inputs, the order and number of input networks were varied. These trials are for the expected limits with systematics for the single tag electron channel, and the results are applied to the muon channel. Each trial shows networks added consecutively, and the corresponding expected limit. In *Trial 1*, networks are added by background fraction, largest to smallest. *Trial 2* probes the converse, smallest to largest. *Trial 3* and *Trial 4* show two other orders, and *Trial 5* is the optimized combination used in the final analysis, using four networks and yielding a limit of 5.8 pb.

Chapter 7

Summary

I have performed a search for the rare process of t -channel single top quark production. This required an analysis of the electroweak production mechanism for single top, the dominant backgrounds masking the channel, and accurately modeling each. I designed an event selection criteria to maximize the acceptance of single top quark events collected by the DØ detector, while reducing the background processes. To further isolate our signal, artificial neural networks were designed and trained to counter each of the dominant backgrounds and the discrimination powers of these neural networks were combined into a single “super” neural network. The shape of the super neural network distribution was used to set limits on the signal production. Limits were derived individually for the electron and muon channels in samples with one or two identified b -quark jets, and were later combined. Searching a data set of 365 pb^{-1} , I am reporting a final expected/measured 95% confidence level limit for t -channel single top quark production of $3.5/8.0 \text{ pb}$, to be compared with the standard model prediction of $1.98 \pm 0.08 \text{ pb}$. The fluctuations in data have left us with a measured limit slightly worse than our previously published analysis which had 5.0 pb [67], yet the expected limit is better than any result published thus far (5.8 pb [67]) and demonstrates the improved sensitivity of the analysis.

Chapter 8

What's to Come

8.1 Observation

Observation of single top quark production, as predicted by the standard model of particle physics, is quickly approaching. While my analysis has moved us closer to the theoretical boundary, only a limit has been set. A single top analysis is already in progress with a dataset three times larger than my own, and by the end of RunII in 2009, we expect more than an order of magnitude increase, or about $4\text{--}8\text{ fb}^{-1}$ of data. So when can we expect to observe single top quark production? To answer this, we define the significance as the quantity of merit:

$$\text{Significance} = \frac{\sigma_{\text{max}}}{\Delta\sigma_{\text{max}}}$$

where σ_{max} is the maximum of the cross section posterior probability density (for example Figure 6.1), and $\Delta\sigma_{\text{max}}$ is the uncertainty on σ_{max} . The significance becomes of interest when it is greater than one, and can allow us to identify what integrated luminosity will yield a cross section measurement three and five standard deviations away from zero. To perform these tests, we create an imaginary data set that is equal

to the background model and predicted signal for a given integrated luminosity. We then calculate the measured limit as described in Section 6.2 at integer multiples of the 0.366 fb^{-1} integrated luminosity used in this analysis. A plot of the significance versus the integrated luminosity is shown in Figure 8.1. One can see that our current analysis would have a cross section measurement 1σ away from zero with a 1 fb^{-1} dataset and 2σ at 5 fb^{-1} . While these values show a direct extrapolation of the discrimination power of the current analysis, we can also expect improvements in several areas of the analysis which will lower the necessary integrated luminosity for an observation. Among these improvements are: better performance in the b -tagging algorithm, further optimization of the neural networks or other signal-background separation technique, increasing signal acceptance by adjusting our event selection criteria, and a general reduction in systematic and statistical errors in all Monte Carlo correction factors and the matrix method scale factors. There is also a large portion of signal still available if one were to perform an analysis using untagged events, include events with only one jet, or look at events with forward electrons ($\eta > 1.5$). Although more difficult, one could also look for events where the W boson (from the top quark decay) decays hadronically, or into a τ , which then decays hadronically. Pursuing any of these options could lead to a 10% reduction in the necessary integrated luminosity, and pursuing several of them could have a dramatic effect. A 5σ observation of single top quark production seems well within reach during the life of the Tevatron.

8.2 A New Window — The Large Hadron Collider

While we expect to observe single top quark production at the Fermilab Tevatron, whether by the standard or non-standard model, there is a new collider on the horizon that can greatly aid the search, or at least provide unprecedented precision – the Large Hadron Collider (LHC) at the European Center for Nuclear Research (CERN).

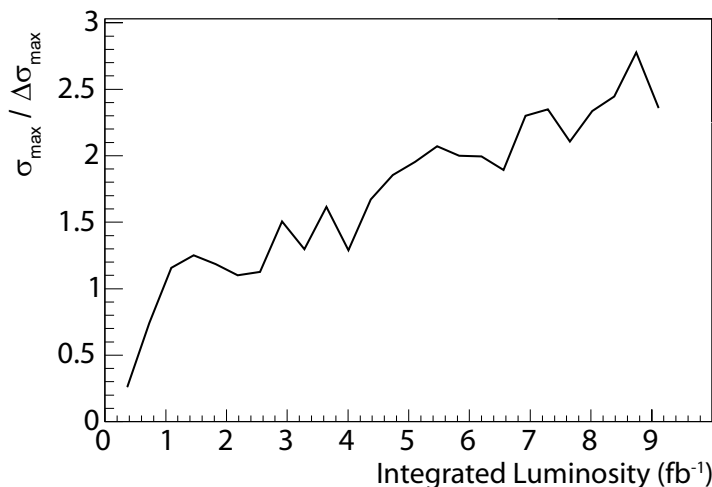


Figure 8.1: The cross section significance ($\sigma_{\max}/\Delta\sigma_{\max}$) is shown as a function of the integrated luminosity for an extrapolation of the current analysis.

The LHC is a proton-proton collider being built in the existing tunnel of the famous Large Electron-Positron (LEP) collider. It has been designed to accelerate each beam of protons to 7 TeV, for a center of mass energy of 14 TeV [115]. In addition to an order of magnitude increase in energy over the Tevatron, the instantaneous luminosity will shatter current values by a 100-fold, reaching $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. While the increased energy and luminosity will greatly enhance the electroweak production of top quarks, many competing processes will also increase, and isolating a signal may become an even greater challenge than that posed by the Tevatron. Theoretical calculations for the t - and s -channel production cross sections give 155.9 pb and 6.56 pb for t processes, and 90.7 pb and 4.09 pb for \bar{t} processes [116]. The difference in t and \bar{t} production values stems from requiring an incoming valence quark versus an incoming sea anti-quark. The tW process becomes significantly more important than at the Tevatron as more energy is available to produce the final state W boson. The cross sections for both tW and $\bar{t}W$ production are identical, calculated to be 31 pb each [117]. Within the first few

months of running, the LHC will gather the same amount of data as collected in the entire Tevatron data-taking period, and quickly gain an order of magnitude over that. A simple Monte Carlo analysis by the ATLAS experiment based on two-jet events and a top mass window cut predicts 7000 t -channel single top events to be collected with a dataset of 30 fb^{-1} [118]. They further conclude a signal:background ratio of three and a statistical uncertainty of $\sqrt{S+B}/B = 1.4\%$. Obvious improvements in event selection and multivariate signal isolation will result in even higher precision. With this example analysis, and countless other possible analyses, we can expect single top quark production to be experimentally measured and thoroughly understood in the years of the LHC, scheduled to begin at the end of 2007.

Bibliography

- [1] J.J. Thompson, “Cathode Rays,” *Philosophical Magazine* 44, 293 (1897).
- [2] J.-E. *et al.*, “Discovery of a Narrow Resonance in $e^+ e^-$ Annihilation,” *Phys. Rev. Lett.* 33, 1406-1408 (1974).
- [3] M.L. Perl *et al.*, “Evidence for Anomalous Lepton Production in $e^+ e^-$ Annihilation,” *Phys. Rev. Lett.* 35, 1489-1492 (1975).
- [4] J.J. Aubert *et al.*, “Experimental Observation of a Heavy Particle J ,” *Phys. Rev. Lett.* 33, 1404-1406 (1974).
- [5] J.H. Christenson, J.W. Cronin, V.L. Fitch, and R. Turlay, “Evidence for the 2π Decay of the K_2^0 Meson,” *Phys. Rev. Lett.* 13, 138-140 (1964).
- [6] G. Danby *et al.*, “Observations of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos,” *Phys. Rev. Lett.* 9, 36-44 (1962).
- [7] G. Arnison *et al.*, “Experimental Observation of Isolated Large Transverse Energy Electrons With Associated Missing Energy at $\sqrt{s}=540$ GeV,” *Phys. Lett. B* 122, 103-116 (1983).
- [8] G. Arnison *et al.*, “Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/ c^2 at the CERN SPS Collider,” *Phys. Lett. B* 126, 398-410 (1983).
- [9] M.Z. Akrawy *et al.*, “Measurement of the Z^0 Mass and Width with the Opal Detector at LEP,” *Phys. Lett. B* 231, 530 (1989).
- [10] M.Z. Akrawy *et al.*, “Measurement of the Decay of the Z^0 Into Lepton Pairs,” *Phys. Lett. B* 235, 379 (1990).
- [11] K. Ackerstaff *et al.*, “Measurement of the Mass of the W Boson in $e^+ e^-$ Collisions at \sqrt{s} 161 GeV,” *Phys. Lett. B* 389, 416-428 (1996).
- [12] Y. Fukuda *et al.*, “Evidence for Oscillation of Atmospheric Neutrinos,” *Phys. Rev. Lett.* 81, 1562 (1998).
- [13] Q.R. Ahmad *et al.*, “Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory,” *Phys. Rev. Lett.* 89, 011301 (2002).

- [14] B.P. Roe, “Particle Physics at the New Millenium,” p. 2, Springer Verlag (1996).
- [15] Richard P. Feynman, “QED: The Strange Theory of Light and Matter,” p. 115-119, Princeton University Press (1985).
- [16] T. Nakano *et al.*, “Evidence for a Narrow $S = +1$ Baryon Resonance in Photoproduction from the Neutron,” *Phys. Rev. Lett.* 91, 012002 (2003).
- [17] C. Alt *et al.*, “Evidence for an Exotic $S = -2$, $Q = -2$ Baryon Resonance in Proton-Proton Collisions at the CERN SPS,” *Phys. Rev. Lett.* 92, 042003 (2004).
- [18] S.W. Herb *et al.*, “Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton-Nucleus Collisions,” *Phys. Rev. Lett.* 39, 252 (1977).
- [19] S. Abachi *et al.*, (DØ Collaboration), “Observation of the Top Quark,” *Phys. Rev. Lett.* 74, 2632 (1995).
- [20] F. Abe *et al.*, (CDF Collaboration), “Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab,” *Phys. Rev. Lett.* 74, 2626 (1995).
- [21] R. Bonciani, S. Catani, M. Mangano, and P. Nason, “NLL Resummation of the Heavy-Quark Hadroproduction Cross-Section,” *Nucl. Phys. B* 529, 424 (1998).
- [22] N. Kidonakis and R. Vogt, “Next-to-Next-to-Leading Order Soft-Gluon Corrections in Top Quark Hadroproduction,” *Phys. Rev. D* 68, 114014 (2003).
- [23] K.G. Chetyrkin, R. Harlander, T. Seidensticker, and M. Steinhauser, “Second Order QCD Corrections to $\Gamma(t \rightarrow Wb)$,” *Phys. Rev. D* 60, 114015 (1999).
- [24] A. Denner and T. Sack, “The Top Width,” *Nucl. Phys. B* 238, 46 (1991).
- [25] The Tevatron Electroweak Working Group, “Combining of CDF and DØ Results on the Mass of the Top Quark,” FERMILAB-TM-2347-E, hep-ex/0603039 (2006).
- [26] The DØ Collaboration, “Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using b -tagged Lepton+Jets Events,” Summer Conference Note, DØNote 4888 (2005).
- [27] Sarah Budd *et al.*, CDF Collaboration, “Top Pair Production Cross Section in Lepton+Jets Channel With Displaced Vertex,” [updated 23 March 2006; cited 23 March 2006] Available at http://www-cdf.fnal.gov/physics/new/top/2006/xs_ljetsvx/public.html.
- [28] V.M. Abazov *et al.*, (DØ Collaboration), “Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV in Dilepton Final States,” *Phys. Lett. B* 626, 265-276 (2005).

- [29] V.M. Abazov *et al.*, (DØ Collaboration), “Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using Lepton + Jets Events with Lifetime b -Tagging,” *Phys. Lett. B* 626, 35 (2005).
- [30] V.M. Abazov *et al.*, (DØ Collaboration), “Measurement of the $t\bar{t}$ Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV using Kinematic Characteristics of Lepton + Jets Events,” *Phys. Lett. B* 626, 45 (2005).
- [31] V.M. Abazov *et al.*, (DØ Collaboration), “ $t\bar{t}$ Production Cross Section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV,” *Phys. Rev. D* 67, 012004 (2003).
- [32] V.M. Abazov *et al.*, (DØ Collaboration), “Measurement of the Top Quark Mass in All-Jet Events,” *Phys. Lett. B* 606, 25 (2005).
- [33] A. Abulencia *et al.*, (CDF Collaboration), “Top Quark Mass Measurement Using the Template Method in the Lepton + Jets Channel at CDF II,” *Phys. Rev. D* 72, 032003 (2006).
- [34] D. Dicus and S. Willenbrock, “Production of Heavy Quarks from W -Gluon Fusion,” *Phys. Rev. D* 34, 155 (1986).
- [35] C.-P. Yuan, “New Method to Detect a Heavy Top Quark at the Fermilab Tevatron,” *Phys. Rev. D* 41, 42 (1990).
- [36] S. Cortese and R. Petronzio, “The Single Top Production Channel at Tevatron Energies,” *Phys. Lett. B* 253, 494 (1991).
- [37] R. K. Ellis and S. Parke, “Top-Quark Production by W -gluon Fusion,” *Phys. Rev. D* 46, 3785 (1992).
- [38] D. O. Carlson and C.-P. Yuan, “Studying the Top Quark via the W -gluon Fusion Process,” *Phys. Lett. B* 306, 386 (1993).
- [39] G. Bordes and B. van Eijk, “Calculating QCD Corrections to Single Top Production in Hadronic Interactions,” *Nucl. Phys. B* 435, 23 (1995).
- [40] D.O. Carlson, E. Malkawi, and C.-P. Yuan, “Probing the Couplings of the Top Quark to Gauge Bosons,” *Phys. Lett. B* 337, 145 (1994).
- [41] T. Stelzer and S. Willenbrock, “Single-Top-Quark Production via $q\bar{q} \rightarrow t\bar{b}$,” *Phys. Lett. B* 357, 125 (1995).
- [42] M. Smith and S. Willenbrock, “QCD and Yukawa Corrections to Single-Top-Quark Production via $q\bar{q} \rightarrow t\bar{b}$,” *Phys. Rev. D* 54, 6696 (1996).
- [43] T. Stelzer, Z. Sullivan, and S. Willenbrock, “Single-Top-Quark Production via W -Gluon Fusion at Next-to-Leading Order,” *Phys. Rev. D* 56, 5919 (1997).

- [44] A.P. Heinson, A.S. Belyaev, and E.E. Boos, “Single Top Quarks at the Fermilab Tevatron,” *Phys. Rev. D* 56, 3114 (1997).
- [45] A. Belyaev, E. Boos, and L. Dudko, “Single Top Quark and Light Higgs Boson at Tevatron,” *Phys. Rev. D* 59, 075001 (1999).
- [46] S. Zhu, “Complete Next-to-Leading Order QCD Corrections to Charged Higgs Boson Associated Production with Top Quark at the CERN Large Hadron Collider,” *Phys. Lett. B* 524, 283 (2002).
- [47] E.E. Boos and A.V. Sherstnev, “Spin Effects in Processes of Single Top Quark Production at Hadron Colliders,” *Phys. Lett. B* 534, 97 (2002).
- [48] S. Eidelman *et al.*, “The Review of Particle Physics,” Chapter 10, Electroweak Model and Constraints on New Physics, *Phys. Lett. B* 592, 1 (2004).
- [49] E.E. Boos *et al.*, (CompHEPCollaboration), “CompHEP4.4: Automatic Computations From Lagrangians to Events,” *Nucl. Instrum. Methods A* 534, 250 (2004).
- [50] E.E. Boos, V.E. Bunichev, L.V. Dudko, V.I. Savrin, and A.V. Sherstnev, “A Simulation Method of the Electroweak Top Quark Production Events in the NLO Approximation. Monte-Carlo Generator *SingleTop*,” DØ Note 4999 (2006).
- [51] Z. Sullivan, “Understanding Single-Top-Quark Production and Jets at Hadron Colliders,” *Phys. Rev. D* 70, 114012 (2004).
- [52] T. Sjöstrand, L. Lönnblad, S. Mrenna, and P. Skands, “PYTHIA 6.3 Physics and Manual,” hep-ph/0308153 (2003).
- [53] Stanford Linear Accelerator Center, “The EvtGen Package Home Page” [cited 22 January 2006] Available from <http://www.slac.stanford.edu/~lange/EvtGen/>.
- [54] CERN, “Programs and Papers by Zbigniew Was” [cited 22 January 2006] Available from <http://wasm.home.cern.ch/wasm/goodies.html>.
- [55] CERN, “ALPGENv2.0” [cited 22 January 2006] Available from <http://m.home.cern.ch/m/mlm/www/alpgen/>.
- [56] G. Mahlon and S. Parke, “Improved Spin Basis for Angular Correlation Studies in Single Top Quark Production at the Fermilab Tevatron,” *Phys. Rev. D* 55, 7249 (1997).
- [57] G. Mahlon and S. Parke, “Observing Spin Correlations in Single Top Production and Decay,” *Phys. Lett. B* 476, 323 (2000).
- [58] T.M.P. Tait, C.-P. Yuan, “Single Top Quark Production as a Window to Physics Beyond the Standard Model,” *Phys. Rev. D* 63, 014018 (2001).

- [59] M. Klein, H. Pietschmann and H. Rupertsberger, “On the Decay of the Top Quark,” *Phys. Lett. B* 153, 341 (1985).
- [60] A. Czarnecki and M. Jezabek, “Distributions of Leptons in Decays of Polarized Heavy Quarks,” *Nucl. Phys. B* 427, 3 (1994).
- [61] A. Czarnecki, M. Jezabek, and J.H. Kuhn, “Lepton Spectra From Decays of Polarized Top Quarks,” *Nucl. Phys. B* 351, 70 (1991).
- [62] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, “Associated Production of Higgs and Single Top at Hadron Colliders,” *Phys. Rev. D* 65, 073005 (2002).
- [63] B.W. Harris, E. Laenen, L. Phaf, Z. Sullivan, and S. Weinzierl, “Fully Differential Single-Top-Quark Cross Section in Next-to-Leading Order QCD,” *Phys. Rev. D* 66, 054024 (2002).
- [64] T.M.P. Tait, “The tW^- Mode of Single Top Production,” *Phys. Rev. D* 61, 034001 (2000).
- [65] B. Abbott *et al.*, (DØ Collaboration), “Search for Electroweak Production of Single Top Quarks in $p\bar{p}$ Collisions,” *Phys. Rev. D* 63, 031101 (2000).
- [66] V.M. Abazov *et al.*, (DØ Collaboration), “Search for Single Top Quark Production at DØ Using Neural Networks,” *Phys. Lett. B* 517, p. 282 (2001).
- [67] V.M. Abazov *et al.*, (DØ Collaboration), “Search for Single Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV,” *Phys. Lett. B* 622, p. 265 (2005).
- [68] D. Acosta *et al.*, (CDF Collaboration), “Search for Single-Top-Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8'$ TeV,” *Phys. Rev. D* 65, 091102 (2002).
- [69] D. Acosta *et al.*, (CDF Collaboration), “Optimized Search for Single-Top-Quark Production at the Fermilab Tevatron,” *Phys. Rev. D* 69, 052003 (2004).
- [70] D. Acosta *et al.*, (CDF Collaboration), “Search for Electroweak Single-Top-Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV,” *Phys. Rev. D* 71, 012005 (2005).
- [71] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, P. Nadolsky, W.K. Tung, “New Generation of Parton Distributions with Uncertainties from Global QCD Analysis,” *JHEP* 0207, 012 (2002).
- [72] Fermilab Beams Division, “Fermilab’s Chain of Accelerators” [updated 5 September 2000; cited 20 October 2005] Available from <http://www-bd.fnal.gov/public/chain.html>.
- [73] Fermilab, “Accelerator Report No. 2: The Fermi Accelerator Complex” [updated 15 September 2004; cited 20 October 2005] Available from http://www.fnal.gov/pub/news04/update_archive/update_9-15.html.

- [74] Fermilab Beams Division, “How Fast are the Protons Traveling?” [updated 5 September 2000; cited 20 October 2005] Available from <http://www-bd.fnal.gov/public/speed.html>.
- [75] Fermilab Proton Source Department, “Proton Source Department” [updated 2004; cited 20 Oct. 2005] Available from <http://www-bd.fnal.gov/proton/proton.html>.
- [76] Fermilab Main Injector Department, “The Fermilab Main Injector – Technical Design Report” [cited 20 October 2005] Available from http://www-fmi.fnal.gov/fmiinternal/MI_Technical_Design/index.html.
- [77] V.M. Abazov *et al.*, (DØ Collaboration), “The Upgraded DØ Detector,” *Nucl. Instrum. Methods A* (Submitted), Fermilab-Pub-05-341-E.
- [78] V.M. Abazov *et al.*, (DØ Collaboration), “The Muon System of the Run II DØ Detector,” Fermilab-PUB-05-034-E (2005).
- [79] D. Mohl, G. Petrucci, L. Thorndahl, and S. van der Meer, *Phys. Rep.* 58, 76 (1980).
- [80] R.D. Pecci and X. Zhang, “Dynamical Symmetry Breaking and Universality Breakdown,” *Nucl. Phys. B* 337, 269 (1990).
- [81] G. Borissov, “Ordering a Chaos or... Technical Details of AA Tracking” [updated 28 February 2003; cited 16 January 2006] Available from http://www-d0.fnal.gov/global_tracking/talks/20030228/talk-adm-030228.ps.
- [82] A. Khanov, “HTF: Histogramming Method for Finding Tracks. The Algorithm Description,” DØ Note 3778 (2000).
- [83] R. Frühwirth, “Application of Kalman Filtering to Track and Vertex Fitting,” *Nucl. Instrum. Methods A* 262 444 (1987).
- [84] H. Greenlee, “The DØ Kalman Track Fit,” DØ Note 4303 (2004).
- [85] A. García-Bellido, S. Lager, F. Rizatdinova, A. Schwartzman, G. Watts, “Primary Vertex certification in p14,” DØ Note 4320 (2004).
- [86] M. Agelou *et al.*, “DØ Top Analysis and Data Sample for the Winter Conferences 2004,” DØ Note 4419 (2004).
- [87] C. Clément, F. Deliot, T. Golling, K. Haganaki, B. Leonhardt, M. Mulders, E. Nurse, S. Söldner-Remboldt, J. Stark, “Muon ID Certification for p14,” DØ Note 4350 (2004).
- [88] G.C. Blazey *et al.*, “Run II Jet Physics,” DØ Note 3750 (2000).
- [89] E. Busato and B. Andrieu, “Jet Algorithms in the DØ Run II Software: Description and User’s Guide,” DØ Note 4457 (2004).

- [90] J. Vlimant, U. Bassler, G. Bernardi, and S. Trincz-Duvoid, “Technical description of the T42 Algorithm for the Calorimeter Noise Suppression,” DØ Note 4146 (2003).
- [91] G. Bernardi, E. Busato, and J. Vlimant, “Improvements from the t42 Algorithm on Calorimeter Objects Reconstruction,” DØ Note 4335 (2004).
- [92] A. Schwartzman and M. Narain, “ b -quark Jet Identification via Secondary Vertex Reconstruction,” DØ Note 4080 (2003).
- [93] R. Demina, A. Khanov, and F. Rizatdinova, “ b -Tagging with Counting Signed Impact Parameter Method,” DØ Note 4049 (2002).
- [94] D. Bloch, B. Clément, D. Gelé, S. Greder, and I. Ripp-Baudot, “Performance of the JLIP b -tagger in p14,” DØ Note 4348 (2004).
- [95] D. Bloch, B. Clément, “Update of the JLIP b -Tagger Performance in p14/pass2 with JES 5.3,” DØ Note 4824 (2005).
- [96] L. Feligioni, M. Narain, P. Schieferdecker, and A. Schwartzman, “Update of b -Quark Jet Identification with Secondary Vertex Reconstruction Using DØ Reco Version p14,” DØ Note 4414 (2004).
- [97] T. Scanlon, “A Neural Network b -Tagging Tool,” DØ Note 4889 (2005).
- [98] T. Scanlon and M. Anastasoie, “Performance of the NN b -tagging Tool on Pass 2 p14 Data,” DØ Note 4890 (2005).
- [99] DØ EM-ID Group, “EM-ID Certification Results, Version 5.0 for p14” [updated 3 December 2003; cited 17 January 2006] Available from http://www-d0.fnal.gov/phys_id/emid/d0_private/certification/main_v5_0.html.
- [100] DØ EM-ID Group, “EM Particle Documentation” [updated 30 September 2003; cited 17 January 2006] Available from http://www-d0.fnal.gov/phys_id/emid/d0_private/EM_Particle_Documentation_EMID.html.
- [101] J. Kozminski, R. Kehoe, H. Weerts, S. Park, A. Quadt, J. Gardner, and S. Jabeen, “Electron Likelihood in p14,” DØ Note 4449 (2004).
- [102] J. Agram *et al.*, “Jet Energy Scale at DØ Run II,” DØ Note 4720 (2005).
- [103] Y. Fisyak and J. Womersley, “DØGSTAR DØ GEANT Simulation of the Total Apparatus Response,” DØ Note 3191 (1997).
- [104] E.L. Berger *et al.*, “The Coordinated Theoretical-Experimental Project on QCD,” [updated 1 Mar 2006; cited 1 Mar 2006] Available from <http://www.phys.psu.edu/~cteq/>

- [105] M. Agelou *et al.*, “Top Trigger Efficiency Measurements and the `top_trigger` Package,” DØ Note 4512 (2004).
- [106] E. Barberis, T. Golling, I. Iashvili, A. Juste, A. Quadt, and P. Schieferdecker, “The Matrix Method and its Error Calculation,” DØ Note 4564 (2004).
- [107] R. Schwienhorst and E. Perez, “Combined Uncertainty on Tagged W +jets and QCD Combined Background,” DØ Note 4597 (2004).
- [108] T. Edwards *et al.*, “The Updated DØ Luminosity Determination – Short Summary,” DØ Note 4328 (2004).
- [109] K. Hornik *et al.*, “Multilayer Feedforward Networks are Universal Approximators,” Neural Networks, vol. 2 359 (1989).
- [110] H. Robbins and S. Monro, “A Stochastic Approximation Method,” Annals of Math. Stat. 22, 400 (1951).
- [111] S.E. Fahlman, “An Empirical Study of Learning Speed in Back-Propagation Networks,” Tech. Rep., Carnegie Mellon University (1988).
- [112] R. Fletcher, “Practical Methods of Optimization,” second edition, Wiley (1987).
- [113] Multi-Layer Perceptron Fit Home Page, “MLPfit: a tool for Multi-Layer Perceptrons” [updated 31 May 2000; cited 15 February 2006] Available from <http://schwind.home.cern.ch/schwind/MLPfit.html>.
- [114] S. Jain, D. Kau, H.B. Prosper, R. Schwienhorst, B. Vachon, “Limits Using a Bayesian Approach in the Package `top_statistics`,” DØ Note (not submitted).
- [115] LHC Design Report Volume I, “Beam Parameters and Definitions” [cited 17 March 2006] Available from <http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html>
- [116] Z. Sullivan, “Understanding Single-Top-Quark Production and Jets at Hadron Colliders,” *Phys. Rev. D* 70, 114012 (2004).
- [117] A. Belyaev and E. Boos, “Single Top Quark $tW + X$ Production at the LHC: A Closer Look,” *Phys. Rev. D* 63, 034012 (2001).
- [118] A. Giammanco, CMS and ATLAS Collaborations, “Single Top Production at LHC,” CMS Conference Report (2006).
- [119] L. Breidman, J. Friedman, R. Olshen, and C. Stone, “Classification and Regression Trees,” Wadsworth (1984).
- [120] B.P. Roe, H.J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, “Boosted Decision Trees as an Alternative to Artificial Neural Networks for Particle Identification,” to appear in Nuclear Instruments and Methods in Physics Research A, arXiv:physics/0408124 (August 2004).

Appendix A

Physics Object Distributions

In this appendix, a more thorough selection of physics object distributions is shown, comparing the background model with the data sample. Figures A.1–A.3 show object distributions for the electron samples before tagging and after tagging in the single-tag and double-tag channels. The same distributions are shown for the muon channel in Figures A.9–A.11.

Figures A.4–A.8 show object distributions in a W +jets/*multijet*-enhanced sample and in a $t\bar{t}$ -enhanced sample in the electron channel. The same plots are shown for the muon channel in Figures A.12–A.16. The W +jets/*multijet* sample applies all cuts in the event selection (Section 3.5) and further requires a cut on $H_T(\text{jet1, jet2, lepton}, \cancel{E}_T) < 200$ GeV and only two jets in the event. The $t\bar{t}$ enriched sample also requires all event selection cuts, as well as $H_T(\text{alljets, lepton}, \cancel{E}_T) > 250$ GeV and exactly four jets in the event.

ELECTRON CHANNEL DISTRIBUTIONS

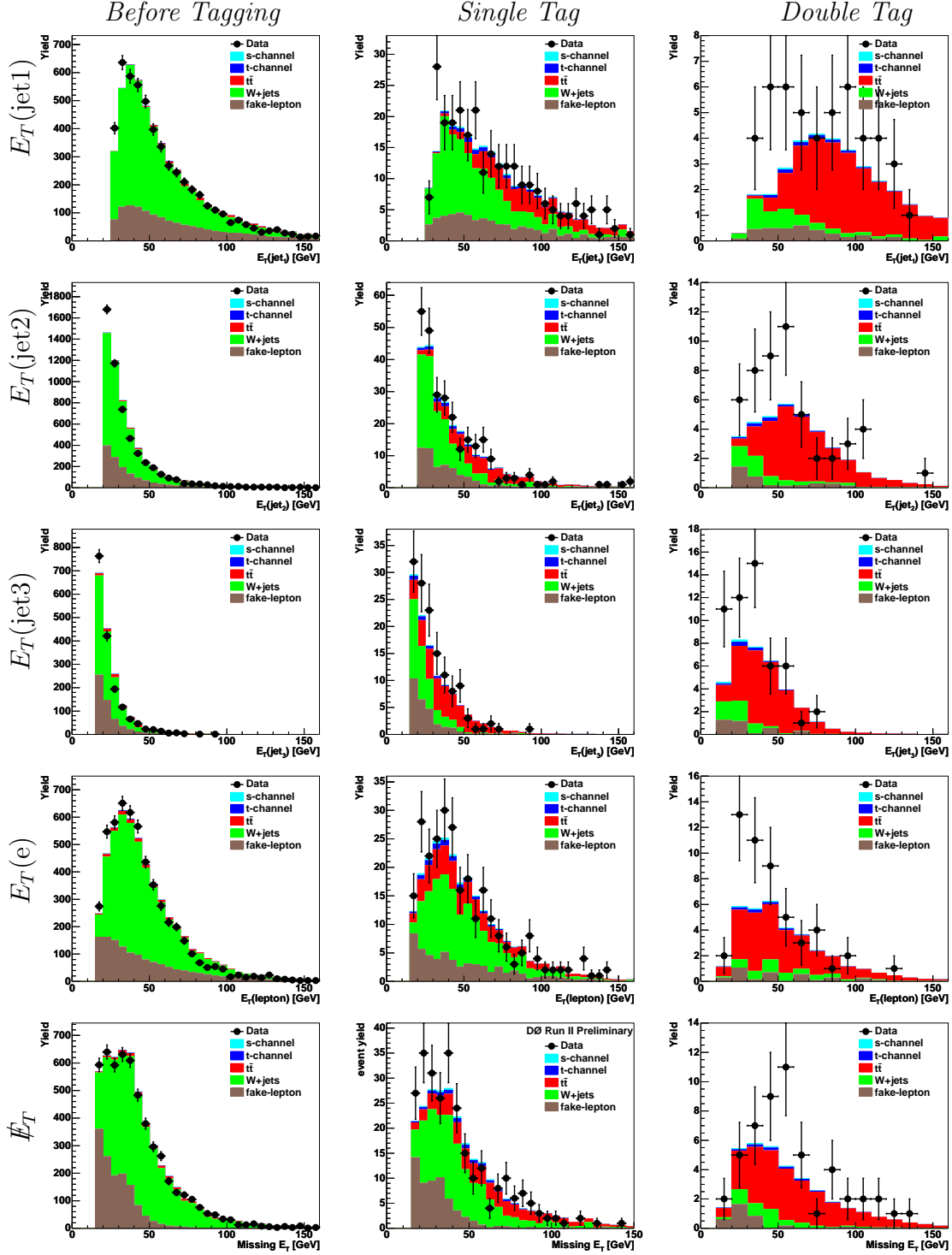


Figure A.1: Electron channel distributions before tagging (left), and in the single (middle) and double (right) tag channels. The data and background model are compared for $E_T(\text{jet1})$ (first row), $E_T(\text{jet2})$ (second row), $E_T(\text{jet3})$ (third row), $E_T(e)$ (fourth row), and the E_T (fifth row).

ELECTRON CHANNEL DISTRIBUTIONS

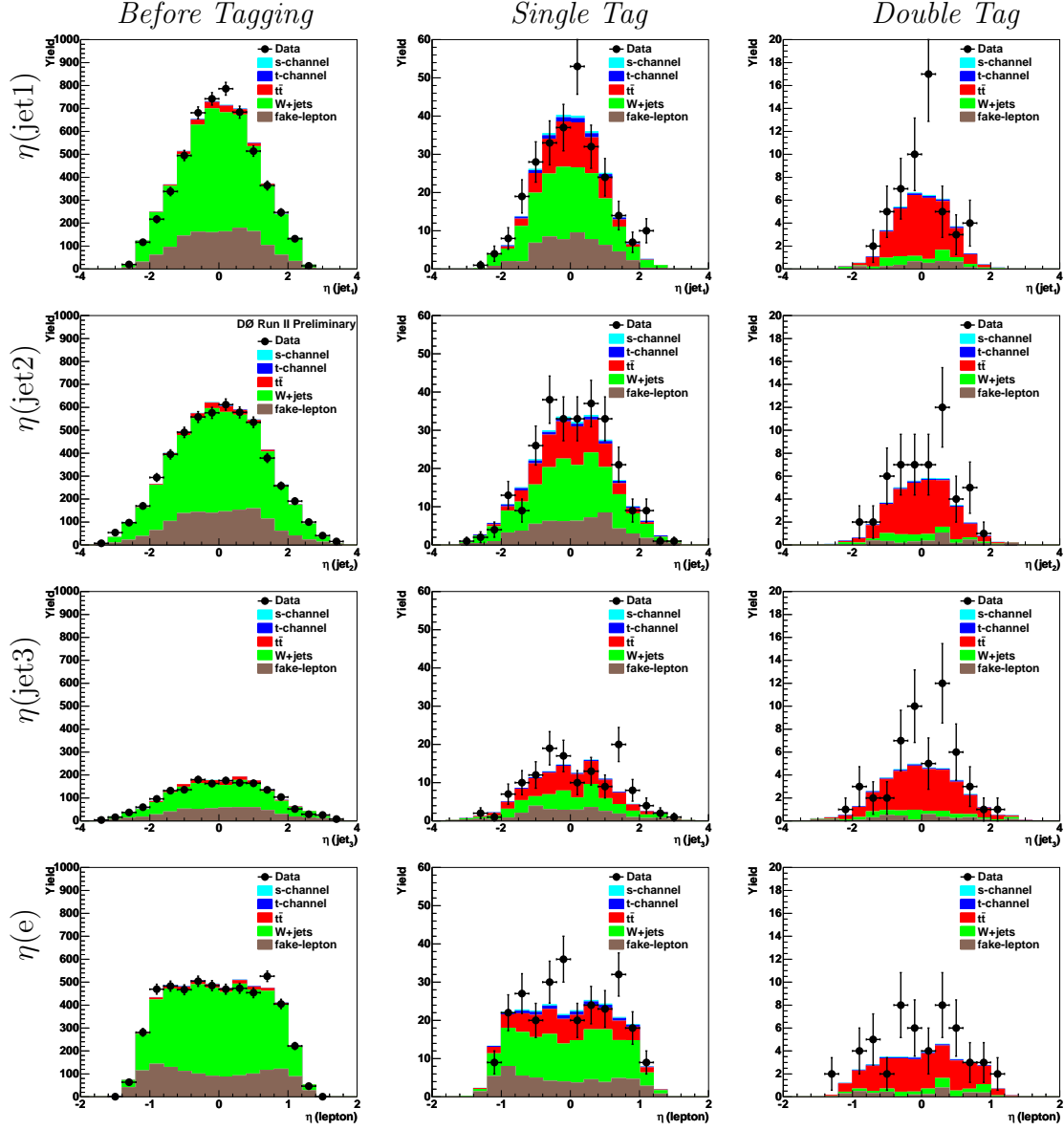


Figure A.2: Electron channel distributions before tagging (left), and in the single (middle) and double (right) tag channels. The data and background model are compared for $\eta(\text{jet1})$ (first row), $\eta(\text{jet2})$ (second row), $\eta(\text{jet3})$ (third row), and the $\eta(e)$ (fourth row)

ELECTRON CHANNEL DISTRIBUTIONS

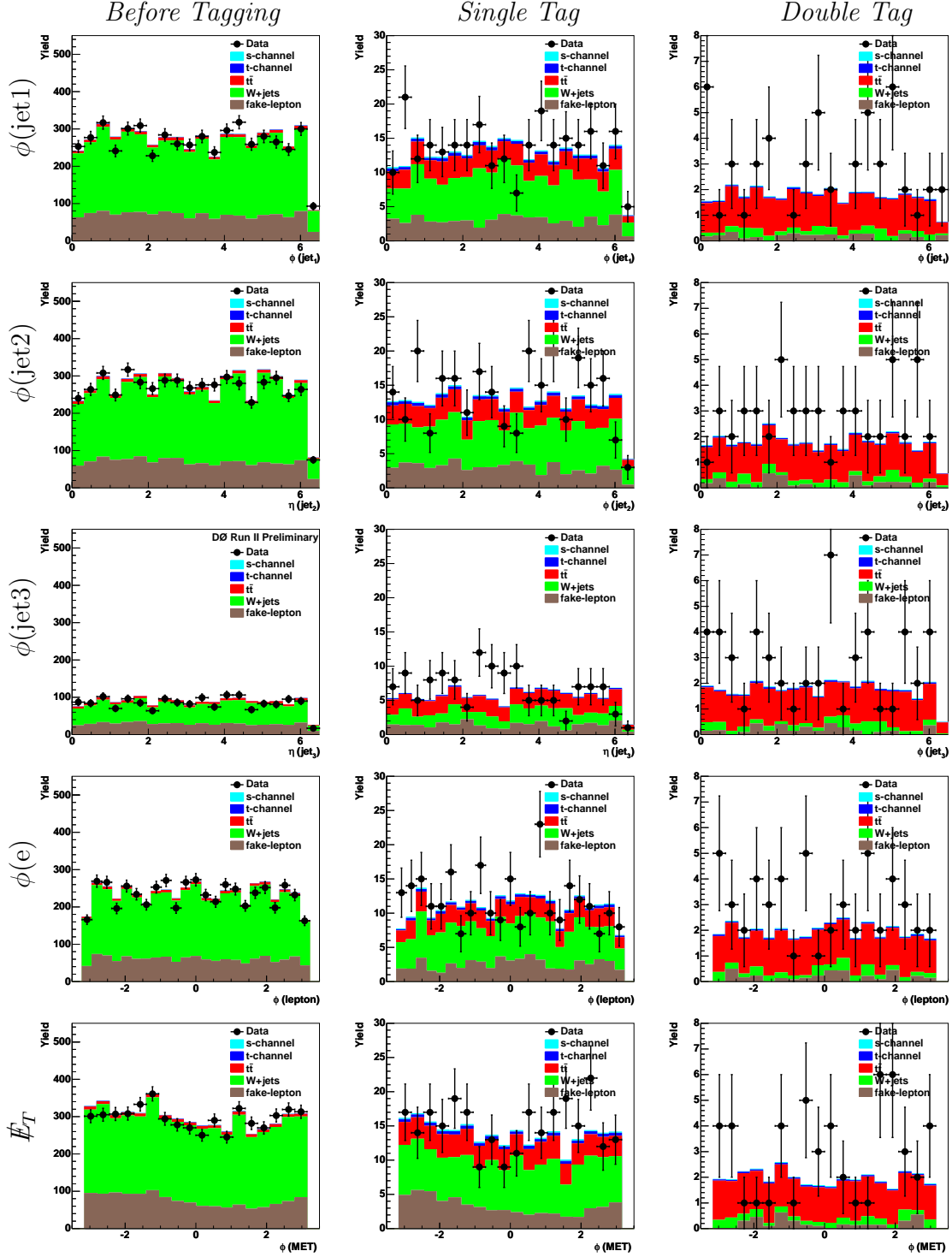


Figure A.3: Electron channel distributions before tagging (left), and in the single (middle) and double (right) tag channels. The data and background model are compared for $\phi(\text{jet1})$ (first row), $\phi(\text{jet2})$ (second row), $\phi(\text{jet3})$ (third row), $\phi(e)$ (fourth row), and the E_T (fifth row).

ELECTRON CHANNEL DISTRIBUTIONS

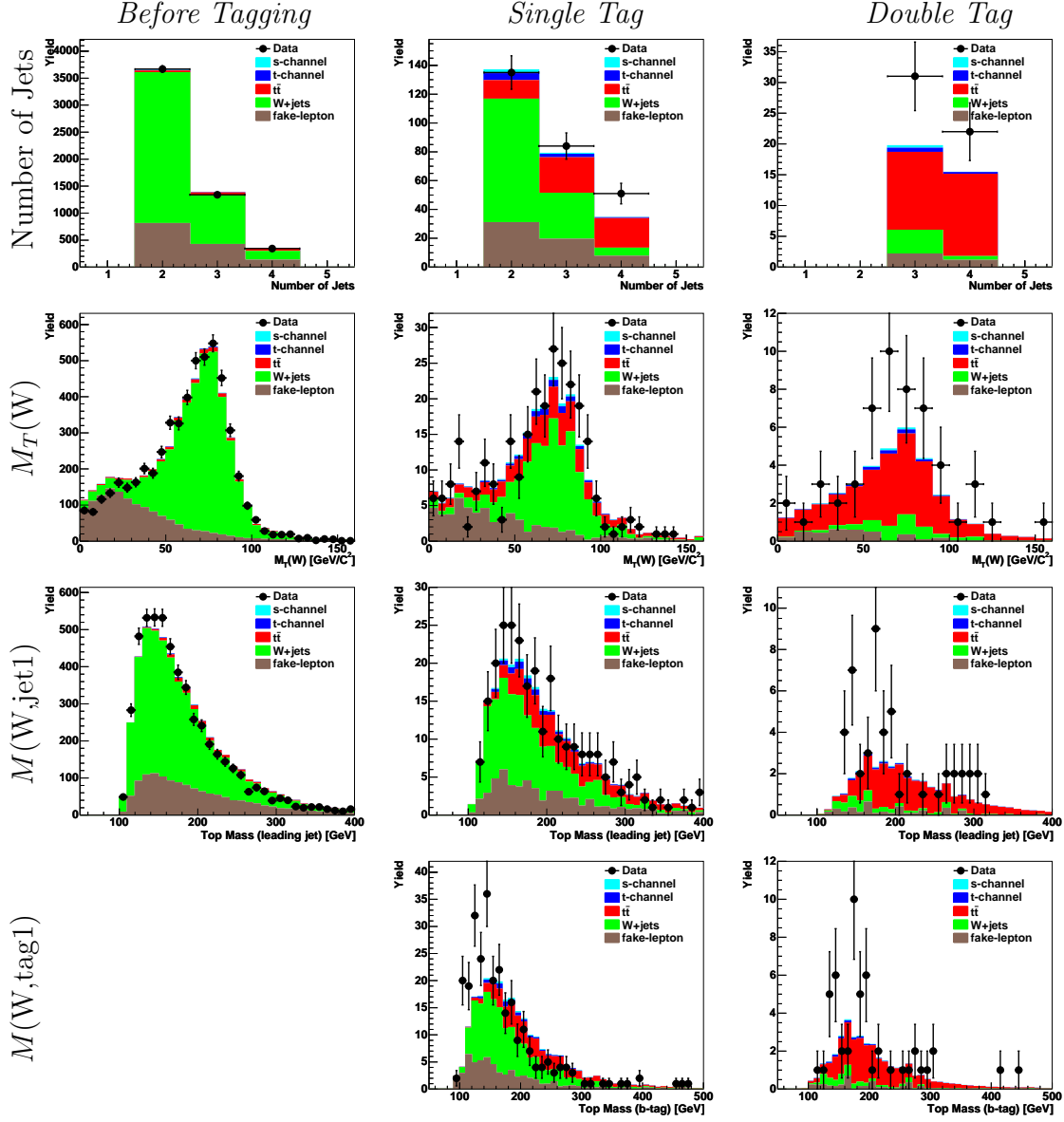


Figure A.4: Electron channel distributions before tagging (left), and in the single (middle) and double (right) tag channels. The data and background model are compared for the Number of Jets (first row), $M_T(W)$ (second row), $M(W, \text{jet1})$ (third row), and $M(W, \text{tag1})$ (fourth row).

ELECTRON CHANNEL DISTRIBUTIONS

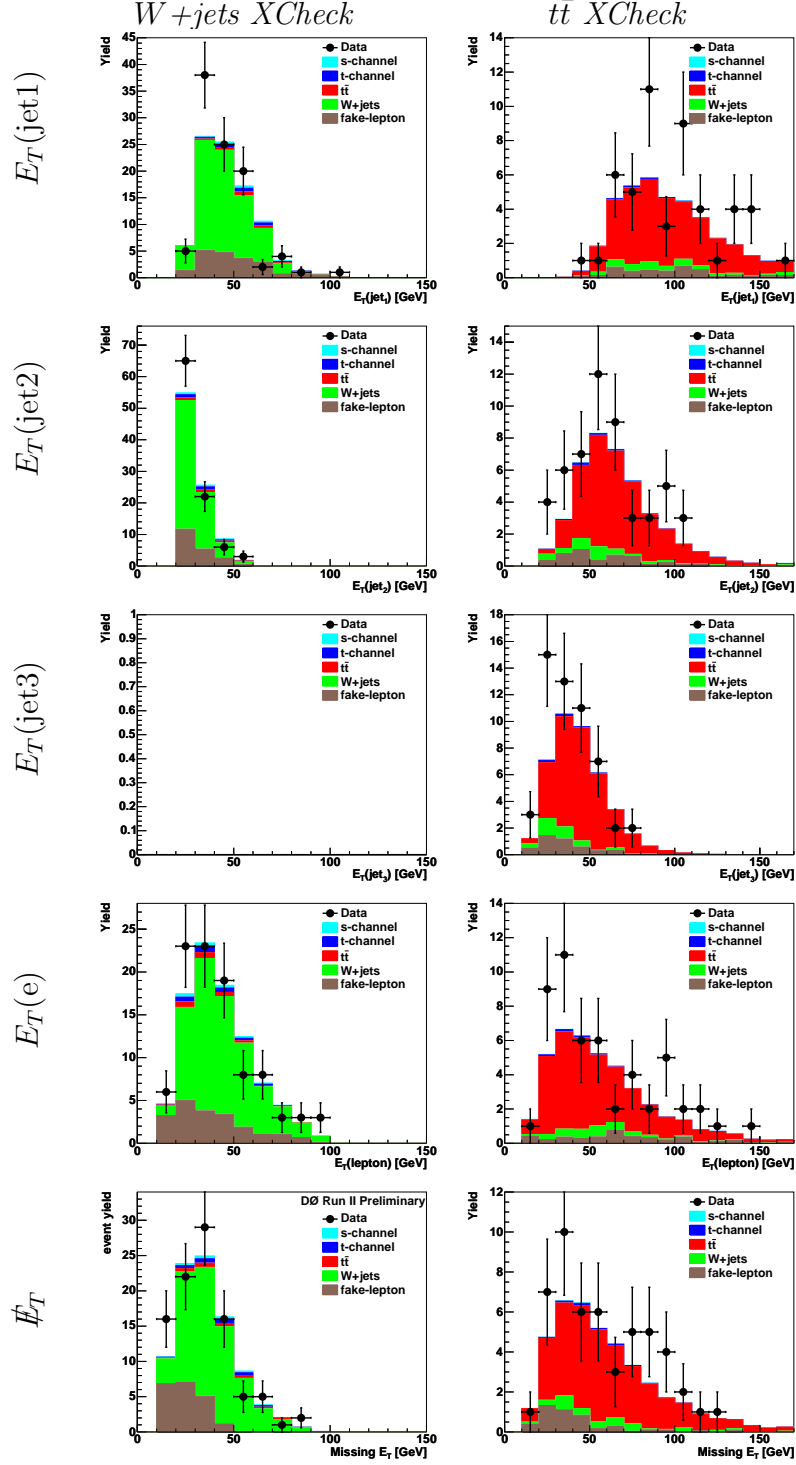


Figure A.5: Electron channel distributions in the $W+\text{jets}$ and $t\bar{t}$ enhanced samples. The data and background model are compared for $E_T(\text{jet1})$ (first row), $E_T(\text{jet2})$ (second row), $E_T(\text{jet3})$ (third row), $E_T(e)$ (fourth row), and the $E_T(\text{missing})$ (fifth row).

ELECTRON CHANNEL DISTRIBUTIONS

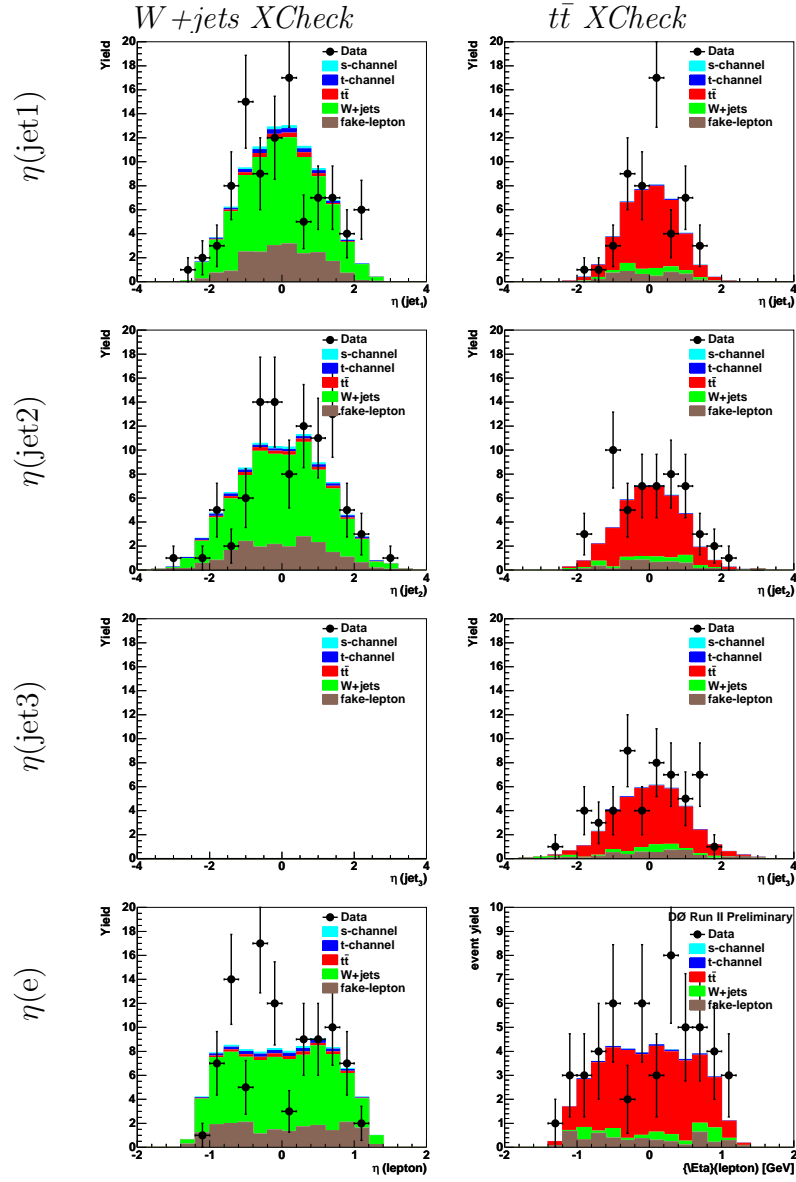


Figure A.6: Electron channel distributions in the W +jets and $t\bar{t}$ enhanced samples. The data and background model are compared for $\eta(\text{jet1})$ (first row), $\eta(\text{jet2})$ (second row), $\eta(\text{jet3})$ (third row), and the $\eta(e)$ (fourth row).

ELECTRON CHANNEL DISTRIBUTIONS

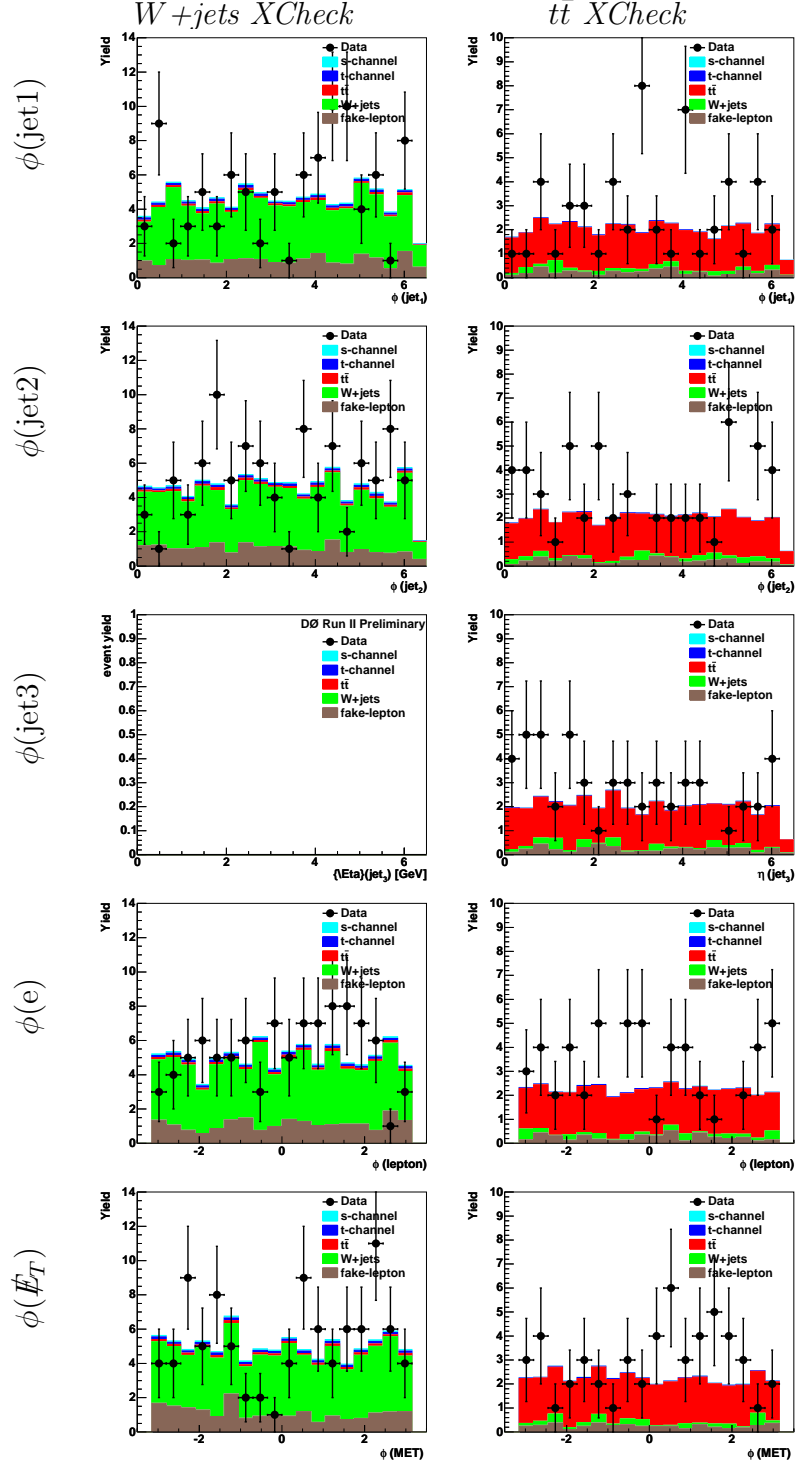


Figure A.7: Electron channel distributions in the $W + \text{jets}$ and $t\bar{t}$ enhanced samples. The data and background model are compared for $\phi(\text{jet1})$ (first row), $\phi(\text{jet2})$ (second row), $\phi(\text{jet3})$ (third row), $\phi(e)$ (fourth row), and the $\phi(\cancel{E}_T)$ (fifth row).

ELECTRON CHANNEL DISTRIBUTIONS

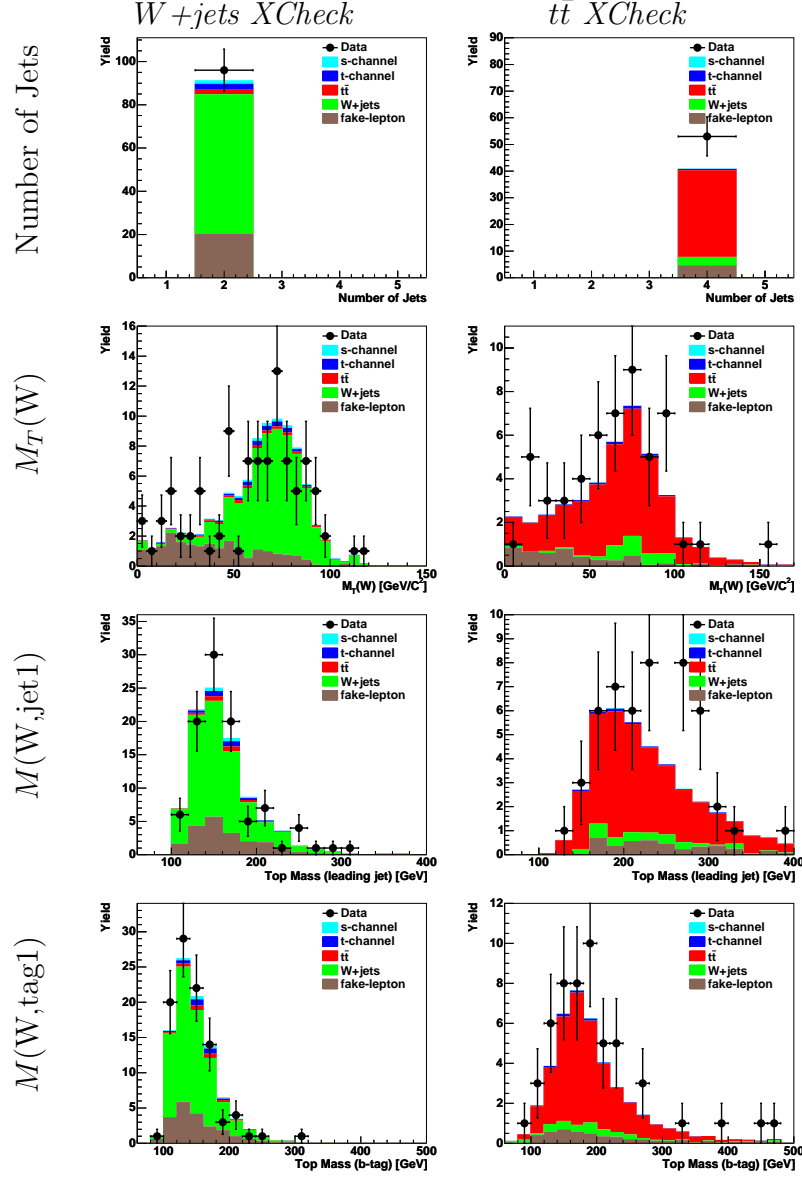


Figure A.8: Electron channel distributions in the $W + \text{jets}$ and $t\bar{t}$ enhanced samples. The data and background model are compared for the Number of Jets (first row), $M_T(W)$ (second row), $M(W, \text{jet1})$ (third row), and $M(W, \text{tag1})$ (fourth row).

MUON CHANNEL DISTRIBUTIONS

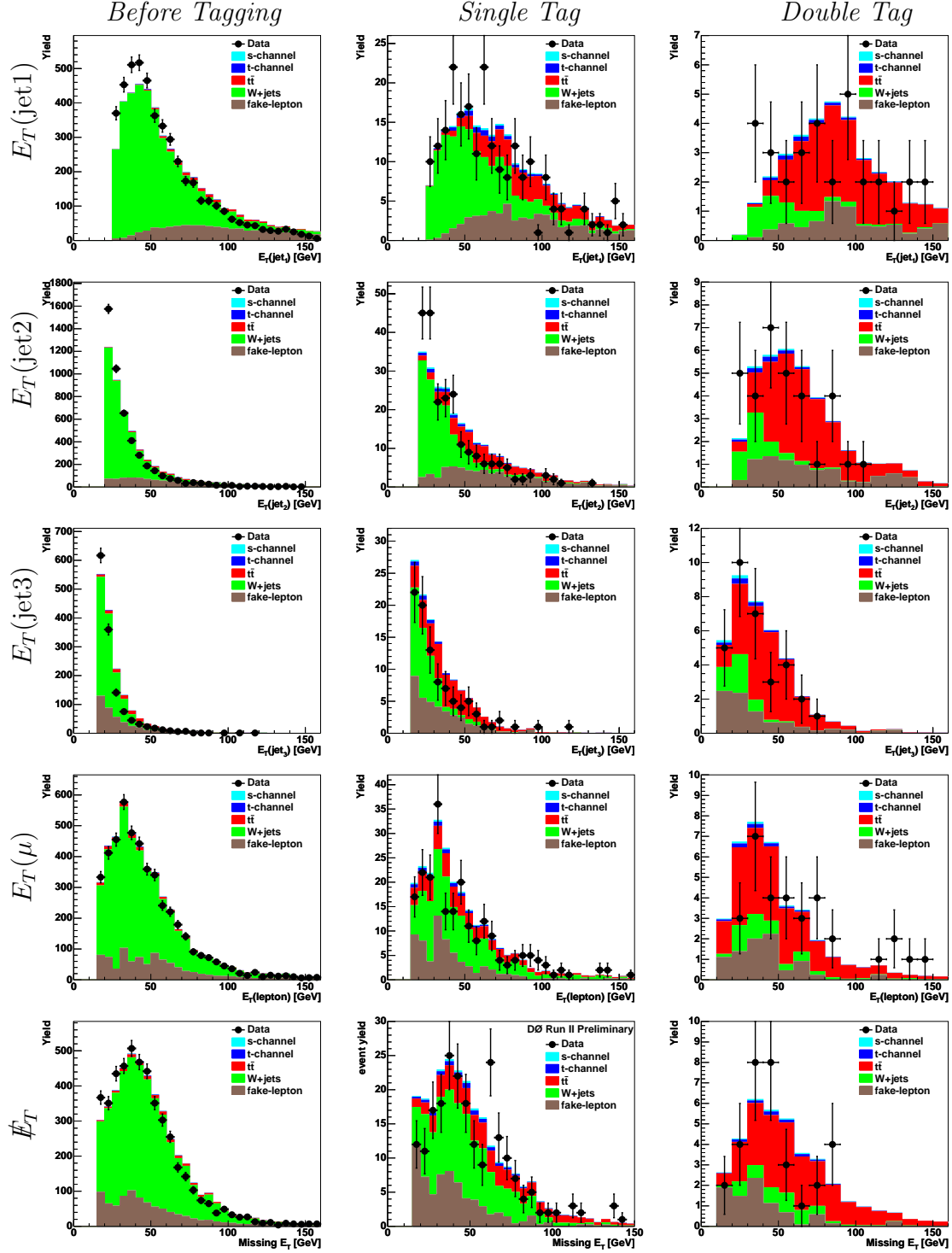


Figure A.9: Muon channel distributions before tagging (left), and in the single (middle) and double (right) tag channels. The data and background model are compared for $E_T(\text{jet1})$ (first row), $E_T(\text{jet2})$ (second row), $E_T(\text{jet3})$ (third row), $E_T(\mu)$ (fourth row), and the \cancel{E}_T (fifth row).

MUON CHANNEL DISTRIBUTIONS

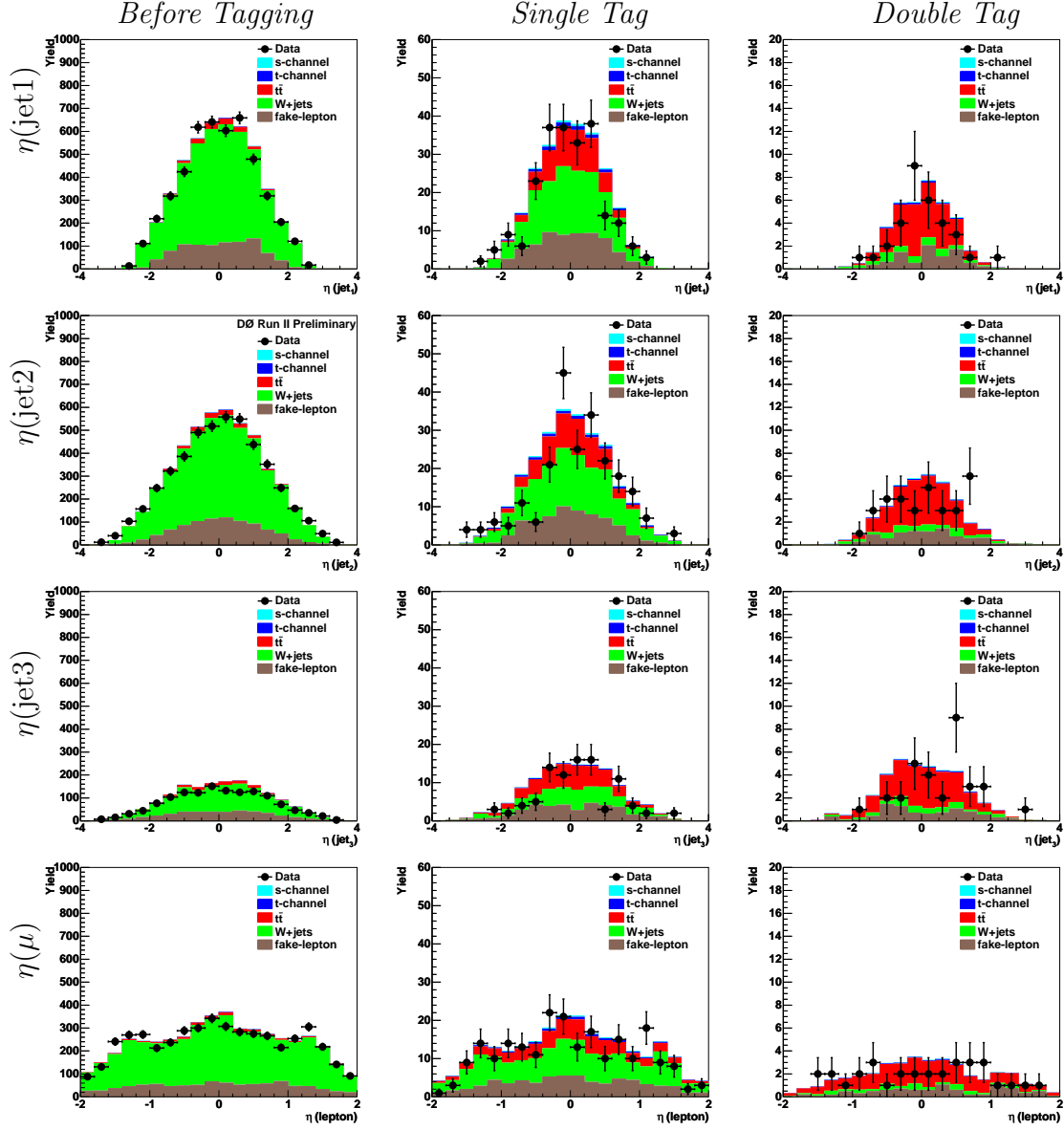


Figure A.10: Muon channel distributions before tagging (left), and in the single (middle) and double (right) tag channels. The data and background model are compared for $\eta(\text{jet1})$ (first row), $\eta(\text{jet2})$ (second row), $\eta(\text{jet3})$ (third row), and the $\eta(\mu)$ (fourth row)

MUON CHANNEL DISTRIBUTIONS

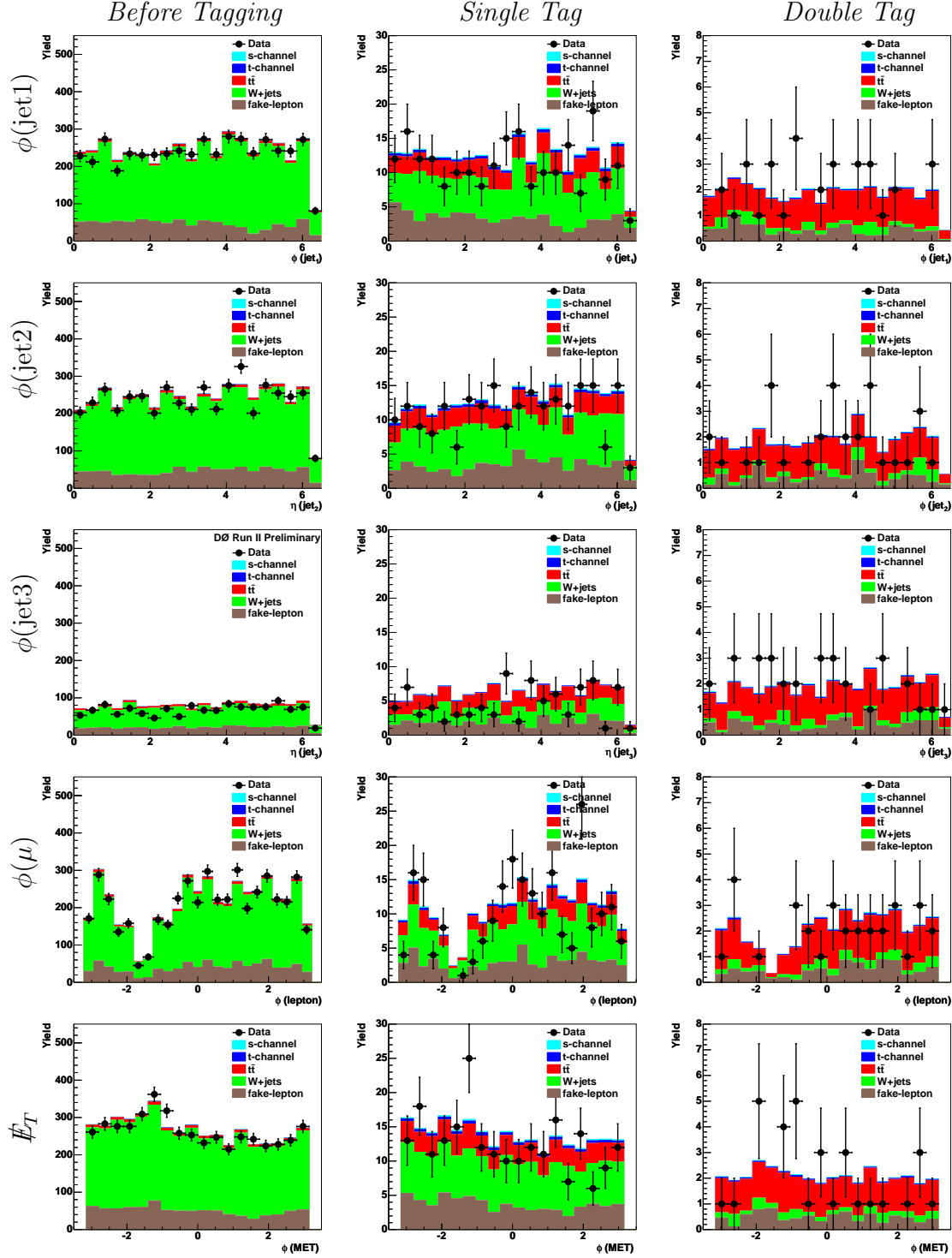


Figure A.11: Muon channel distributions before tagging (left), and in the single (middle) and double (right) tag channels. The data and background model are compared for $\phi(\text{jet1})$ (first row), $\phi(\text{jet2})$ (second row), $\phi(\text{jet3})$ (third row), $\phi(\mu)$ (fourth row), and the E_T (fifth row).

MUON CHANNEL DISTRIBUTIONS

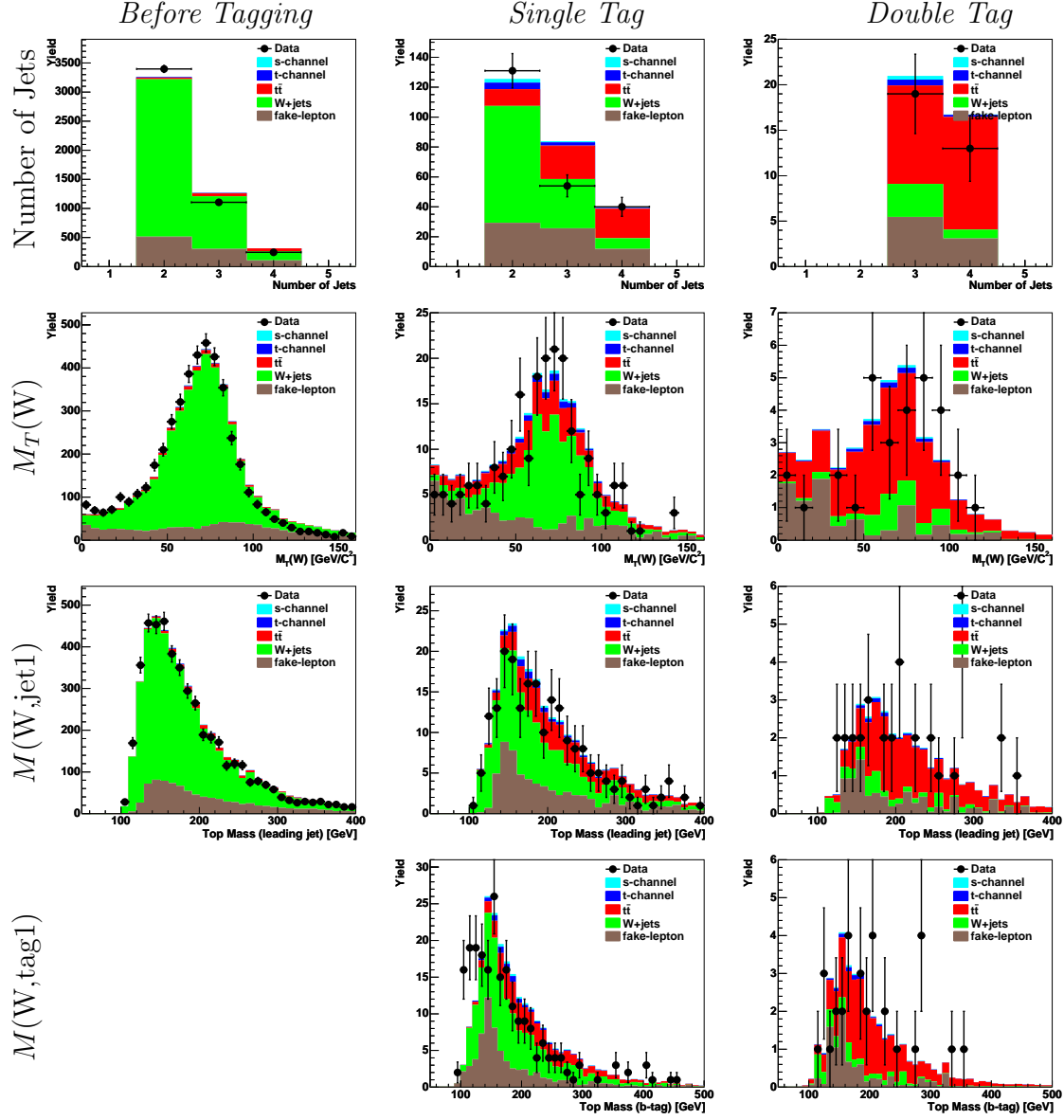


Figure A.12: Muon channel distributions before tagging (left), and in the single (middle) and double (right) tag channels. The data and background model are compared for the Number of Jets (first row), $M_T(W)$ (second row), $M(W, \text{jet1})$ (third row), and $M(W, \text{tag1})$ (fourth row).

MUON CHANNEL DISTRIBUTIONS

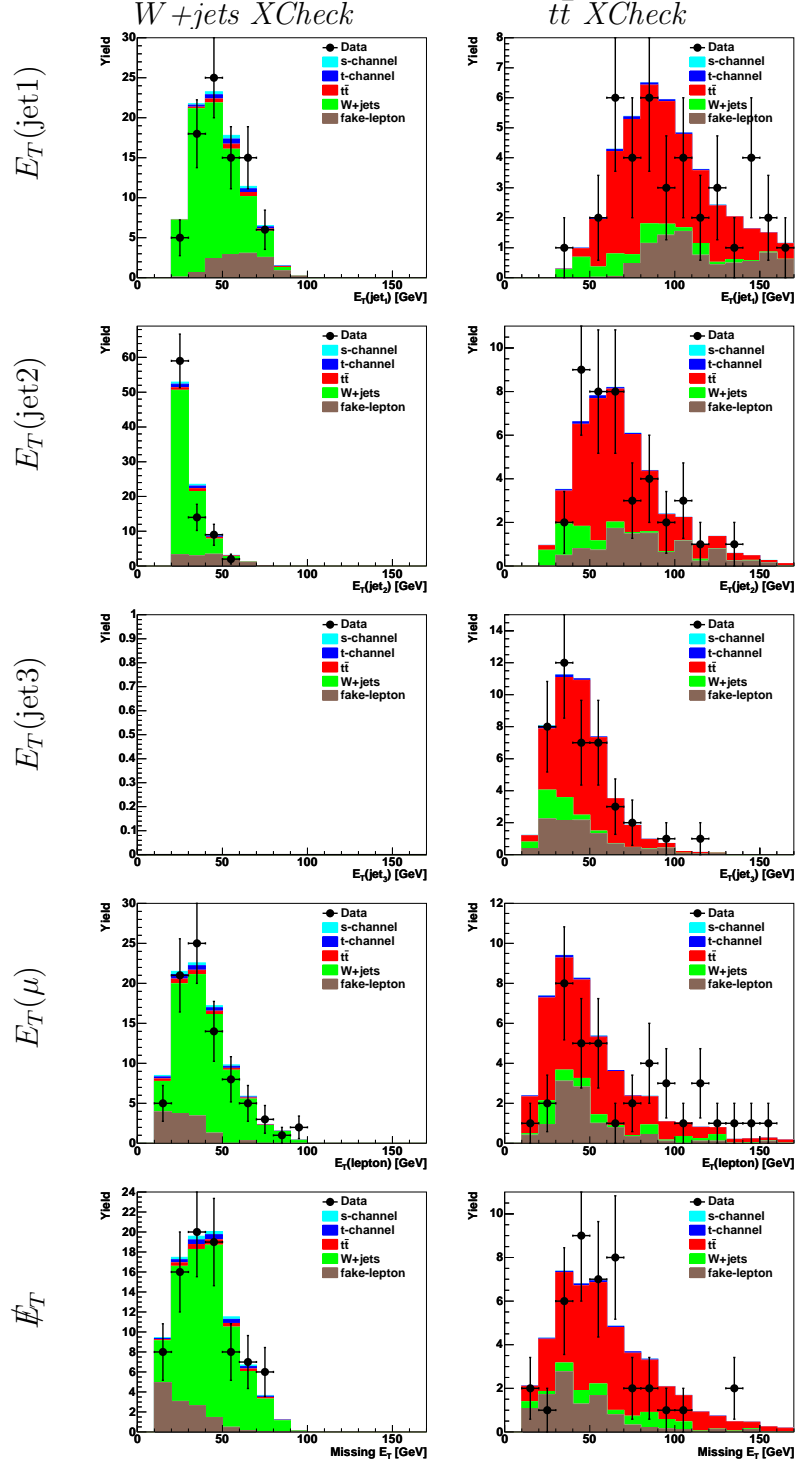


Figure A.13: Muon channel distributions in the W +jets and $t\bar{t}$ enhanced samples. The data and background model are compared for $E_T(\text{jet1})$ (first row), $E_T(\text{jet2})$ (second row), $E_T(\text{jet3})$ (third row), $E_T(\mu)$ (fourth row), and the $\#_T$ (fifth row).

MUON CHANNEL DISTRIBUTIONS

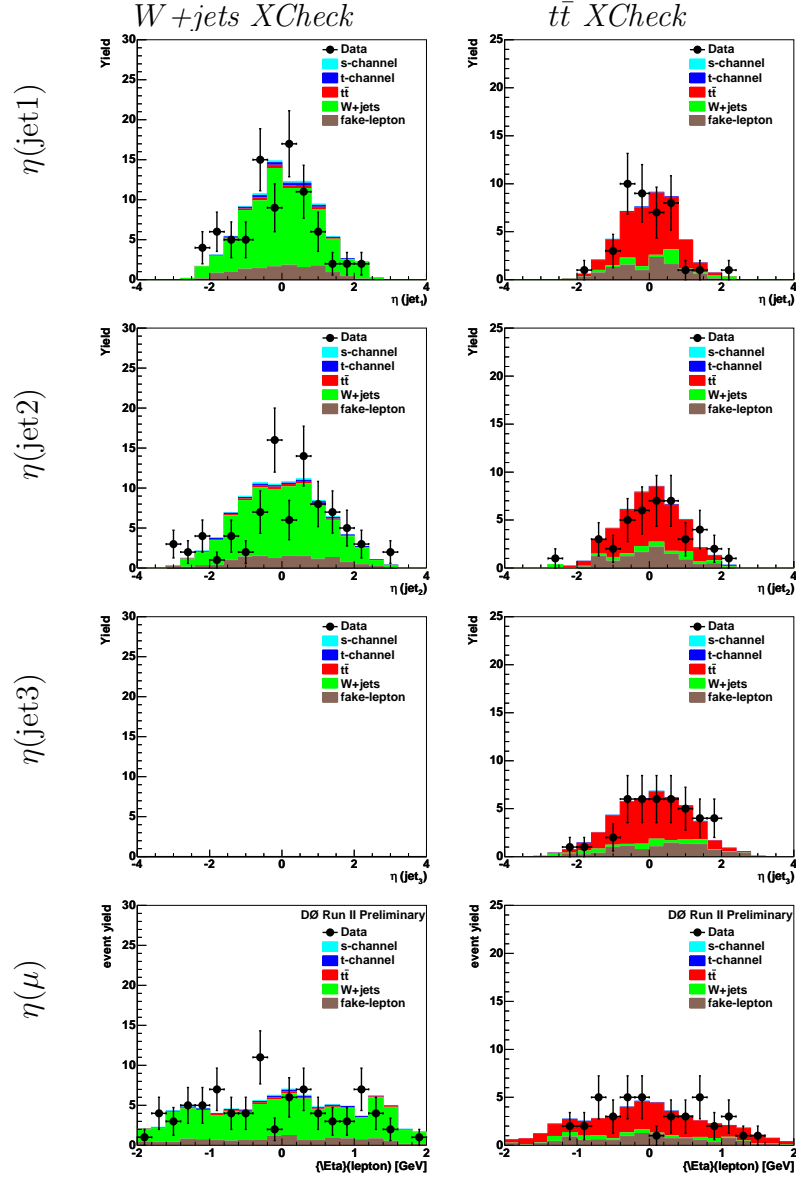


Figure A.14: Muon channel distributions in the $W + \text{jets}$ and $t\bar{t}$ enhanced samples. The data and background model are compared for $\eta(\text{jet1})$ (first row), $\eta(\text{jet2})$ (second row), $\eta(\text{jet3})$ (third row), and the $\eta(\mu)$ (fourth row).

MUON CHANNEL DISTRIBUTIONS

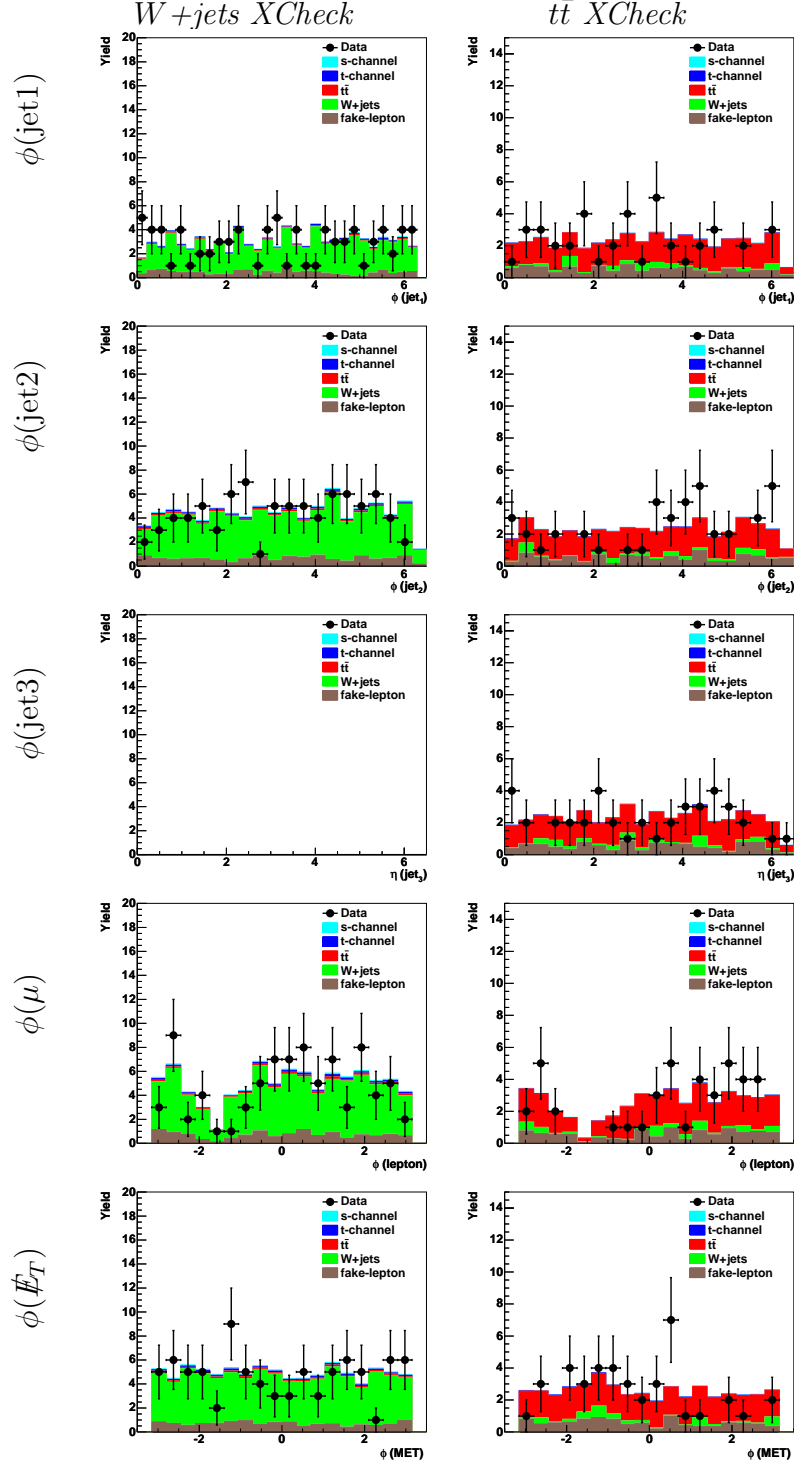


Figure A.15: Muon channel distributions in the W +jets and $t\bar{t}$ enhanced samples. The data and background model are compared for $\phi(\text{jet1})$ (first row), $\phi(\text{jet2})$ (second row), $\phi(\text{jet3})$ (third row), $\phi(\mu)$ (fourth row), and the $\phi(\cancel{E}_T)$ (fifth row).

MUON CHANNEL DISTRIBUTIONS

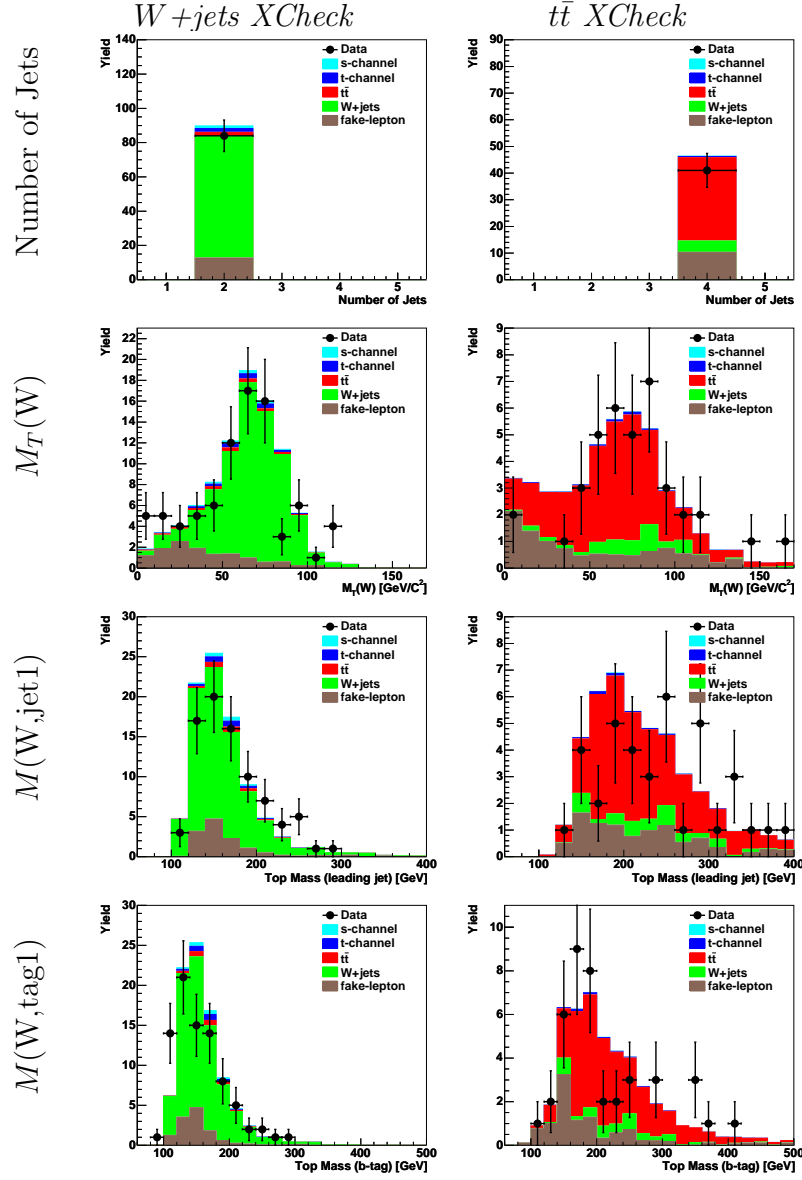


Figure A.16: Muon channel distributions in the W +jets and $t\bar{t}$ enhanced samples. The data and background model are compared for the Number of Jets (first row), $M_T(W)$ (second row), $M(W, \text{jet1})$ (third row), and $M(W, \text{tag1})$ (fourth row).

Appendix B

Calculation of Reconstructed Physics Objects

To measure properties of the W boson and the top quark in our single top events, and later use these to aid in isolating our signal, we must reconstruct the four-momentum of each particle. The W boson is constructed from the lepton and neutrino from its decay. The four-momentum of the lepton is completely measured by the detector, but the noninteracting neutrino is measured through the missing transverse energy (\cancel{E}_T) in the event. This leaves an ambiguity in the z component of the neutrino momentum. To recover this quantity, we constrain both four-vectors to reconstruct the W boson mass:

$$M_W^2 = E_W^2 - \vec{p}_W^2 \quad (\text{B.1})$$

$$= (E_\ell + E_\nu)^2 - (\vec{p}_\ell + \vec{p}_\nu)^2 \quad (\text{B.2})$$

$$= 2(E_\ell \sqrt{p_{z,\nu}^2 + \cancel{E}_T^2} - \vec{p}_{T,\ell} \cdot \vec{p}_T - p_{z,\ell} p_{z,\nu}) \quad (\text{B.3})$$

Solving the quadratic equation for $p_{z,\nu}$, we find

$$p_{z,\nu} = \frac{\beta p_{z,\ell} \pm \sqrt{E_\ell^2(\beta^2 - \cancel{E}_T^2 p_{T,\ell}^2)}}{p_{T,\ell}^2} \quad \text{and} \quad \beta = \frac{M_W^2}{2} + \vec{p}_{T,\ell} \cdot \vec{p}_T \quad (\text{B.4})$$

There remain two difficulties: there are two solutions to $p_{z,\nu}$, and a non-trivial fraction of events have no real solution (the value under the square-root is negative because the \cancel{E}_T is too large and results in a M_W greater than the known value). To deal with these issues, we first force the measured transverse mass of the W , $M_T(W)$, to be less than or equal to the true W boson mass ($M_W = 80.43$ GeV), scaling the components of the transverse neutrino momentum as necessary. If the value under the square root still remains negative, we set it to zero, and use the single remaining solution. Otherwise, we choose the smaller of the two $|p_{z,\nu}|$ solutions.

For completeness, the W transverse mass is defined as:

$$M_T(W) = \sqrt{(E_T^l + E_T^\nu)^2 - (\vec{p}_T^l + \vec{p}_T^\nu)^2} \quad (\text{B.5})$$

$$= \sqrt{2E_T^l E_T^\nu (1 - \cos(\phi^l - \phi^\nu))} \quad (\text{B.6})$$

and one expects to see a kinematical edge or “Jacobian edge” in the $M_T(W)$ distribution at the true mass of the W boson, and this is indeed observed in our analysis (see Figures A.4 and A.12).

To reconstruct the top quark in our events, we must use the W boson and the b -quark jet from the top decay. Since we do not definitively know which jet is the correct b -quark jet, we can reconstruct the top quark with the leading jet or the leading b -tagged jet, and the W boson. The four-momentum and top quark mass are then defined as:

$$\vec{p}_{top} = \vec{p}_W + \vec{p}_{jet} \quad (\text{B.7})$$

$$M_{top} = \sqrt{(E_W + E_{jet})^2 - (\vec{p}_W + \vec{p}_{jet})^2} \quad (\text{B.8})$$

Appendix C

Decision Trees

Another multivariate technique that can be used to discriminate between signal and background in a data sample is the decision tree. While we do not apply this technique for our final background discrimination, we do make use of its ranking feature for all input variables. A basic description of how a decision tree works, and an explanation of the ranking value that we use for neural network input variable optimization will be given.

A decision tree starts with a single node that contains all signal and background events. Selecting from the pool of input variables, this root node adds on a branch node. The branch node represents an input variable, and an associated cut on that variable, for example the “HT_AllJets > 96.82” branch in Figure C.1. Events are passed through this node, and if they pass the cut condition, they are sent to the right or “true” branch, and if they fail it, they are sent to the left or “false” branch. After all events have been passed through a branch node, the purity and the *gini* value for the true and false branches can be calculated. The purity and *gini* value are defined as:

$$\text{Purity} = P = \frac{\sum_s W_s}{\sum_s W_s + \sum_b W_b} \quad (\text{C.1})$$

$$gini = \Sigma(W_s + W_b)P(1 - P) \quad (\text{C.2})$$

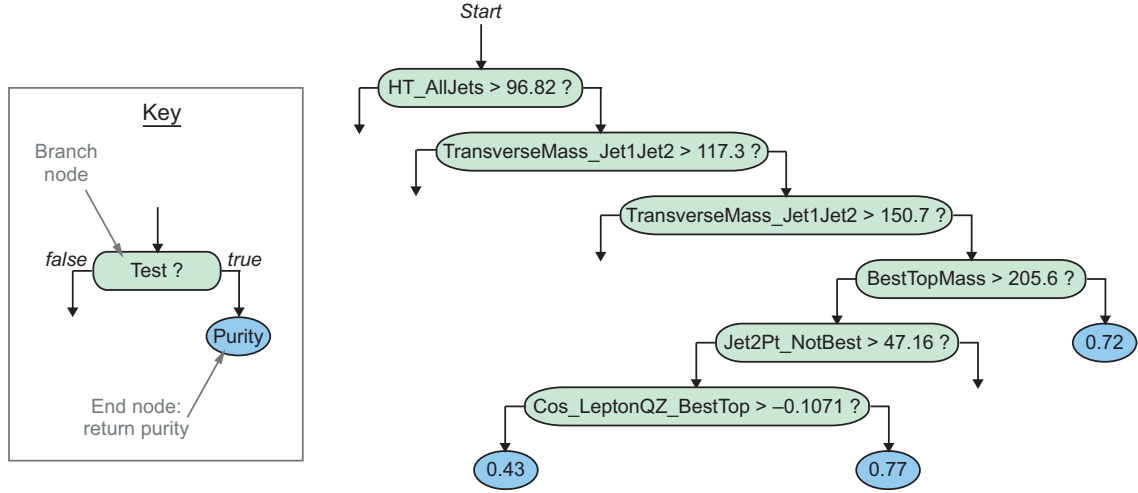


Figure C.1: An example path through a decision tree. The root node is the start, and events are successively passed through boolean conditions at each branch node until the desired purity is reached in a leaf, or there are too few events left to continue further branching.

where W_s and W_b are the weight of the signal and background events respectively (the sum of all signal events and background events are each normalized to unity for decision tree training). The purity gives the fraction of signal events after a given branch node, and the gini value takes both signal and background events into account after a node, and becomes smaller as discrimination power increases. The variable and cut value at each branch node are thus chosen to yield the greatest improvement in the *gini* value. Branch nodes are iteratively created until a branch either has a purity of 1.0 or there are less than 100 events in the branch. The termination point thus becomes a leaf, and the purity of this leaf is the output of a decision tree function. At each branch node in the tree, all variables are available for use, and the same variable can be used multiple times throughout the tree if it yields the greatest *gini* improvement. The variable ranking values used for the neural network optimization (Tables 5.3–5.7) are $100 \times (\text{gini improvement})$ by that variable. As shown in our ranking tables, the first variable has the largest *gini* improvement value from the starting value in the root node of 0.5, and further variables provide incremental improvements. Further details of the

decision tree algorithm can be found in the literature [119], and a current high energy physics analysis applying the method at the MiniBooNE experiment [120].