UNIVERSITY of CALIFORNIA
Santa Barbara

A Measurement of the Top Pair Production Cross-Section in the Dilepton Channel using Lepton plus Track Selection

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

Doctor of Philosophy

in

Physics

by

Corrinne Mills

Committee in charge:

Professor Joseph Incandela, Chair
Professor Claudio Campagnari
Professor Mark Srednicki

June 2007

The dissertation of Corrinne Mills is approved:

Professor Claudio Campagnari

Professor Mark Srednicki

Professor Joseph Incandela, Chair

June 2007

To whom it may concern.

Acknowledgements

During my time as a graduate student, I have had the great privilege of living in two places, in Santa Barbara and at Fermilab, where I have had fantastic colleagues and friends, many of whom are both.

First I thank my adviser Joe Incandela, for his support and advice, and particularly for encouraging me to pursue the fake lepton problem. Over the years, he has helped me keep my focus on the big picture, while also encouraging me to build up my experience in the different aspects of experimental particle physics. I also owe a great debt to Chris Hill, previously a postdoc in our group, under whose guidance I learned important, if sometimes frustrating, lessons about what it means to be a good scientist. It takes a village to raise a grad student, as they say, and I've benefited from the expertise of many people here at UCSB. I would like to thank Claudio Campagnari and David Stuart for many impromptu physics and statistics consultations. Jeff Richman, Harry Nelson, Tony Affolder, Puneeth Kalavase, Jake Ribnik, Ray Bunker, Tae Hong, Ford Garberson, and so many others have helped make this an interesting, challenging, and always entertaining place to get a Ph.D. I would also like to thank Debbie Ceder, our administrative assistant, for her ability to Make Things Happen.

At Fermilab, I had the opportunity to work with a group of really great people, from whom I inherited the lepton plus track analysis. Helping bring that analysis to publication so early in my research career was an amazing baptism by fire. I thank Avi Yagil for many great arguments (I mean that), Pierre Savard for his ongoing intellectual guidance (see below), Peter Wittich for his code (and the amusing editorial commentary therein), and Julia Thom for her patience and advice while I was trying to debug my fake rate code. Petra Merkel and Reda Tafirout have also been important guides along the way. I would never have survived my graduate career without the support and advice of the Fermilab branch of the UCSB group: Ben Brau, Cigdem Issever, Adam Scott, and Antonio Boveia. I thank them and the rest of the silicon group at the time for many excellent lunch and coffee breaks. Kathy Copic, Mitch Soderberg, Michael Kirby, Bo Jayatilaka, Marcel Stanitzki, Farukh Azfar, Nate Goldschmidt, and many others are excellent friends.

These acknowledgments would be incomplete without mention of my colleagues in Toronto, who have had such a large role in my analysis and my life, from a city to which I have never been. Pierre Savard has been a source of wisdom on physics, analysis, and life. I thank him for his guidance and his sense of humor. When it comes to analysis, my fellow graduate student Teresa Spreitzer has been the other half of my brain for the last year or so. She is

an irreplaceable colleague and friend, and has gotten me through a few rough patches, in the way that only someone who understands both the analysis and me can.

Graduate school would have been a lot less fun without the amazing group of friends I have made over the years at Santa Barbara. I thank Henriette Elvang for many coffee breaks and for staying friends with me even though I once called her native country "Daneland". Mel McLaurin and Mike Hermele, as co-tenants of Beer House with Dave and I, are great conversationalists, fine cooks, and have been constant friends. For giving me something to look forward to every week, I thank them and the rest of the Wine Night / Happy Hour crew: Jesse Berezovsky, Felix Mendoza, Jean-Luc Fraiken, Alelia Parenteau, Chris McKinney, Mike Requa, Loren Swenson, and many other good people.

Finally, I am grateful to my husband David Wood, who has been by my side for the entirety of my graduate career, first as a friend, then as more, and now as a spouse. Thank you, Dave, for waiting for me to move back from Fermilab, and for making graduate school the best years of my life so far.

Curriculum Vitæ

Corrinne Mills

Education

2007	PhD in Physics, University of California, Santa Barbara
2004	MA in Physics, University of California, Santa Barbara
2001	BA in Physics, Harvard University

Publications

"Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV using Dilepton Events", D. Acosta *et al.* [CDF Collaboration], Phys. Rev. Lett. **93**, 142001 (2004).

"SNEWS: The SuperNova Early Warning System", P. Antonioli *et al.*, New J. Phys. **6**, 114 (2004).

Abstract

A Measurement of the Top Pair Production Cross-Section in the Dilepton Channel using Lepton plus Track Selection

by

Corrinne Mills

Using 1.1 fb^{-1} of data collected by the Collider Detector at Fermilab (CDF) from Run II of the Fermilab Tevatron, we measure the $t\bar{t}$ production cross section in events with two leptons, significant missing transverse energy, and ≥ 2 jets. As the Run II dataset grows, more stringent tests of Standard Model predictions for the top quark sector are becoming possible. The dilepton channel, where both top quarks decay $t \to Wb \to \ell\nu b$, is of particular interest due to its high purity even in the absence of a b jet "tagging" requirement. Use of an isolated track as the second lepton significantly increases the dilepton acceptance, at the price of some increase in background, particularly from W + jets events where one of the jets is identified as a lepton. With the amount of data available, it has been possible to improve the estimate of the contribution from that background, reflected in a reduced systematic uncertainty. Assuming a branching ratio of

$BR(W \to \ell\nu) = 10.8\,\%$ and a top mass of $m_t = 175\,\mathrm{GeV}/c^2$, the measured cross-section is $\sigma(p\bar{p} \to t\bar{t}) = 8.3 \pm 1.3(\mathrm{stat.}) \pm 0.7(\mathrm{syst.}) \pm 0.5(\mathrm{lumi.})$ pb. The result is consistent with the Standard Model prediction of $6.7^{+0.7}_{-0.9}$ pb and represents a significant improvement in precision over previous results using this selection.

Contents

Contents x

List of Figures xiv

List of Tables xviii

1 Introduction 1
- 1.1 The Standard Model of Particle Physics . . . 2
- 1.2 Particle Physics at Hadron Colliders . . . 4
- 1.3 Selecting Top Quark Pairs . . . 9
- 1.4 Measuring the Top Quark Cross-section . . . 10

2 Relevant Theory 17
- 2.1 Top Pair Production at the Tevatron . . . 19
  - 2.1.1 The Tree-Level QCD Process . . . 19
  - 2.1.2 Parton Distribution Functions . . . 22
  - 2.1.3 Higher-Order Corrections . . . 24
  - 2.1.4 Predicted Cross-Section . . . 26
- 2.2 Top Quark Decay Modes . . . 27
- 2.3 Backgrounds . . . 29
  - 2.3.1 Drell-Yan . . . 29
  - 2.3.2 W+jets . . . 30
  - 2.3.3 Diboson . . . 35
- 2.4 The Connection to Physics Beyond the Standard Model . . . 37
- 2.5 Summary . . . 41

3 The Tevatron and the CDF Detector **42**
3.1 The Tevatron at Fermilab . . . 43
3.2 The CDF Detector . . . 51
3.2.1 Tracking . . . 53
3.2.2 Calorimetry . . . 60
3.2.3 Muon Detectors . . . 65
3.2.4 Luminosity Measurement . . . 67
3.3 Event Triggers and Data Acquisition . . . 68
3.3.1 The CDF Trigger System . . . 69
3.3.2 Triggers Used in this Analysis . . . 70
3.4 Summary . . . 72

4 Event Selection **74**
4.1 Lepton Definitions . . . 75
4.1.1 Electron Selection . . . 77
4.1.2 Muon Selection . . . 84
4.1.3 Track Lepton Selection . . . 88
4.2 Jet Definition . . . 90
4.2.1 Jet Clustering . . . 91
4.2.2 Jet Energy Corrections . . . 91
4.2.3 Jets for Analysis . . . 92
4.3 Missing Transverse Energy Reconstruction . . . 93
4.3.1 Basic $\not\!\!E_T$ Calculation . . . 94
4.3.2 Track E/P corrections . . . 94
4.3.3 Jet Corrections . . . 95
4.4 Event Selection . . . 95
4.4.1 Basic Selection . . . 96
4.4.2 "Delta Phi" Selection . . . 98
4.4.3 Drell-Yan Veto . . . 100
4.4.4 Candidate Event Classification . . . 100
4.5 Data Sample . . . 103
4.6 Summary . . . 103

5 Top Dilepton Acceptance **105**
5.1 Simulated $t\bar{t}$ Sample . . . 106
5.2 Corrections to Simulation . . . 107
5.2.1 Primary Lepton Scale Factors . . . 107
5.2.2 Track Lepton Scale Factors . . . 109
5.2.3 Track χ^2 Probability Cut Efficiency . . . 110

5.2.4 Event Trigger Efficiencies . . . 111
5.3 Acceptance for Top Dilepton Events . . . 114
5.4 Systematic Uncertainties on the Acceptance . . . 114
5.4.1 Primary Lepton Identification Efficiency . . . 115
5.4.2 Track Lepton Identification Efficiency . . . 117
5.4.3 Jet Energy Scale . . . 117
5.4.4 Initial and Final State Radiation . . . 118
5.4.5 Parton Distribution Functions . . . 120
5.4.6 Monte Carlo Generator . . . 121
5.5 Summary . . . 122

6 Dilepton Backgrounds **123**
6.1 Diboson . . . 124
6.1.1 Diboson Acceptance . . . 125
6.1.2 Correcting for the Number of Jets . . . 127
6.1.3 Systematic Uncertainties . . . 128
6.2 W + jets with a Fake Lepton . . . 130
6.2.1 Definition of the Fake Rate . . . 131
6.2.2 Jets from Quarks, Jets from Gluons . . . 132
6.2.3 Testing the Fake Rate . . . 139
6.2.4 Normalization from W + jets Data . . . 141
6.2.5 Contribution from Fake Fully Reconstructed Leptons . . . 147
6.2.6 Efficiency of the Z Veto and Track Lepton $\Delta\varphi$ Criteria for Events with a Fake Lepton . . . 149
6.2.7 Opposite-Sign Fraction . . . 151
6.2.8 Systematic Uncertainties . . . 158
6.3 Drell-Yan . . . 159
6.3.1 $Z/\gamma^\star \to \tau\tau$ . . . 159
6.3.2 $Z/\gamma^\star \to ee/\mu\mu$ . . . 160
6.3.3 Uncertainties on the $Z/\gamma^\star \to ee/\mu\mu$ Contribution . . . 165
6.4 Summary . . . 167

7 Results **169**
7.1 Main Result and Cross-section Calculation . . . 170
7.2 Results with Other Kinematic Thresholds . . . 171
7.3 Cross-section with Two Fully Reconstructed Leptons . . . 176
7.4 Kinematic Features of the Candidate Event Sample . . . 178

8 Conclusion **184**

A Glossary of Acronyms **189**

Bibliography **192**

List of Figures

1.1 Missing transverse energy of $t\bar{t}$ dilepton candidates at CDF in Run I, compared to the expectation for combined signal and background. Taken from Ref. [1]. 14

2.1 Leading order diagrams for top quark pair production at the Tevatron. 21
2.2 Momentum fraction of partons within the proton, for two momentum scales. Taken from Ref. [2]. 23
2.3 Representative subset of diagrams for the next to leading order contribution to top quark pair production at the Tevatron. The double line is the top-antitop pair, and the single line is a light (u or d) quark. 25
2.4 Drell-Yan ($q\bar{q} \to Z/\gamma^* \to \ell^+\ell^-$) cross-section vs. invariant mass. Invariant mass distribution taken from generator-level information from PYTHIA [3], normalized to an integrated cross-section of 251 pb in the interval 66 GeV/c^2 < M < 116 GeV/c^2 [4]. . . 31
2.5 Number of tracks with $p_T > 10$ GeV/c in jets with reconstructed $E_T > 20$ GeV, in W + jets data simulated using PYTHIA. Jets produced by quarks tend to have a smaller number of tracks, and jets produced by gluons are more likely to have either zero tracks above threshold or a larger number of tracks. 33
2.6 Number of tracks with $p_T > 10$ GeV/c in jets with reconstructed $E_T > 20$ GeV, in simulated W + jets data. Jets produced by quarks tend to have a smaller number of tracks, and jets produced by gluons are more likely to have either zero tracks above threshold or a larger number of tracks. 36

2.7 Leading diagrams for W/γ production at the Tevatron. The diagram on the left is the leading contribution because of the relative strength of the gluon PDF at low x. 37

3.1 Mechanism of acceleration in a linear accelerator. The lengths of the accelerating RF cavities are tuned so that the particle bunch enters a cavity as the field is switching to the right direction. Taken from Ref. [5]. 45
3.2 Particles entering the accelerating cavity are exposed to different accelerating potentials. Particles entering early are exposed to a weaker field, and those entering earlier are exposed to a stronger field. Those entering the cavity at dramatically different times are accelerated in the wrong direction or not at all. The net effect of this is to group a charged particle beam into bunches. 46
3.3 Geometry of the CDF detector. 52
3.4 R-z view of the silicon tracking system geometry. The detector is symmetric around $z = 0$. Layer 00, the innermost detector, is in green. The SVX II, in blue, is the five layers above it. The ISL are the layers in magenta at large radius. 55
3.5 Impact parameter (distance of closest approach to the beamline) for tracks with and without silicon. Tracks are from fully-reconstructed central electron W candidates in data, which have an electron with $E_T > 20$ GeV and > 25 GeV of $\not\!\!E_T$. 57
3.6 Geometry of the tracking system, including the COT and the silicon tracking detectors, with the solenoid and endplug calorimeter shown as well. The central calorimeter begins at a radius just beyond the solenoid. The detector is symmetric about $z = 0$ and is approximately azimuthally symmetric. 59

4.1 Kinematically allowed decay modes of the tau lepton. The τ is more massive than the estimated mass of the c quark, but no charmed meson is lighter than the τ, so $\bar{c}s$ and $\bar{c}d$ decays are not allowed. 76
4.2 Cartoon of the mechanism by which the decay in flight of a medium- to low-p_T kaon can be reconstructed as a high-p_T muon. 87
4.3 Missing transverse energy in $Z/\gamma^\star \to ee/\mu\mu$events with $\not\!\!E_T >$ 25 GeV, compared to the distribution from dilepton $t\bar{t}$. Distributions taken from events simulated using the PYTHIA generator. 97

4.4 Angle between jets and $\not\!\!E_T$ in $Z/\gamma^\star \to ee/\mu\mu$ events with $\not\!\!E_T$ > 25 GeV, compared to the distribution from dilepton $t\bar{t}$. Distributions taken from events simulated using the PYTHIA generator. 101
4.5 Dilepton invariant mass of $Z/\gamma^\star \to ee/\mu\mu$ events with $\not\!\!E_T >$ 25 GeV, compared to the distribution from dilepton $t\bar{t}$. Distributions taken from events simulated using the PYTHIA generator. 102

6.1 Comparison of the number of jets in Z candidate events between data and simulation. 128
6.2 Leading diagrams for medium-p_T (20-50 GeV/c) multijet production at the Tevatron. These dominate because of the relative strength of the gluon PDF at low x and because the matrix elements are larger than those including quarks. 134
6.3 Track lepton fake rates in simulation, for different jet samples. as a function of jet E_T. In the upper left plot it is apparent that the fake rate from generic QCD multijet events (blue) disagrees with the fake rate for jets associated with a W. In the upper right, we show the difference between the fake rates for jets matched to a quark and jets matched to a gluon. The lower two plots check the matching of the quark and gluon jet fake rates between samples separately. 135
6.4 The fraction of jets matched to a quark in simulated W, photon, and multijet events. For any jet multiplicity, the fraction of quark jets is highest in the photon sample, followed by the W sample and the multijet sample. One can also see that in the W and photon samples, the fraction of quark jets falls off with increasing jet multiplicity. 136
6.5 Track lepton fake rates from photon data selected using an 80 GeV photon, as a function of E_T and $|\eta|$. 138
6.6 Fake rates for fully reconstructed leptons from photon data selected using an 80 GeV photon, as a function of E_T and $|\eta|$. . . 139
6.7 Predicted and observed number of isolated tracks in simulated W + jets events, as a function of the number of jets in the event. The fake rate used is derived from photon+jets and (for events reconstructed with zero jets) Z+jets simulation. The over-prediction for zero jet events is incorporated in the systematic uncertainty on the fake lepton background. 142

6.8 Predicted and observed number of isolated tracks in Z + jets data, as a function of the number of jets in the event. The fake rate used is derived from photon+jets data. The shape appears to be not well-predicted for events with three or more jets, but there are only a handful of events in this category, and it seems likely that the distribution fluctuated. 143
6.9 Testing the applicability of the 80 GeV photon + jets fake rate to fake leptons from top in the lepton + jets mode, in simulation. No fake leptons are observed from events with one jet. 147
6.10 Testing the applicability of the using the same fake rate for from top in the lepton + jets mode as is used to predict fake leptons from W + jets events, by using a fake rate derived from the W + jets sample, in simulation. No fake leptons are observed from events with one jet. 148

7.1 Number of predicted events compared to the number observed in the data. The cross-hatched areas show the (1 σ) uncertainties on the predicted numbers (statistical and systematic). 180
7.2 Missing transverse energy of predicted and candidate events. The measured cross-section is used to normalize the $t\bar{t}$ distribution and the prediction is scaled so that the predicted number of events is equal to the observed. 181
7.3 Invariant mass of the lepton pair in predicted and candidate events. The measured cross-section is used to normalize the $t\bar{t}$ distribution and the prediction is scaled so that the predicted number of events is equal to the observed. 182
7.4 Transverse momentum of both lepton candidates in predicted and candidate events. The measured cross-section is used to normalize the $t\bar{t}$ distribution and the prediction is scaled so that the predicted number of events is equal to the observed. 183

List of Tables

1.1 Known fermions and their properties, from [6]. Note that, because they interact strongly and (except the top quark) are not observed in isolation, quark masses are only well-defined for a choice of renormalization scheme. 5

2.1 Range of theoretical predictions for the top quark pair production cross-section at $\sqrt{s} = 1.96$ TeV. 27

4.1 Fraction (in %) of the (opposite-sign, ≥ 2 jet) acceptance that is attributable to each possible generated lepton pair. The first two rows show the events divided according to whether they were reconstructed with an electron or muon as the primary lepton. The majority of events are accepted as electron plus track because there is more geometric acceptance for electrons, and because of way in which events are categorized when both leptons can be fully reconstructed. 77

4.2 Integrated luminosity for each primary lepton type. 104

5.1 Efficiencies for identifying the different primary lepton types in data and simulation, and the ratio used to correct the acceptance. 109

5.2 Single lepton and full per-event trigger efficiencies for the four primary lepton types. Numbers in bold are inputs to the cross-section measurement. Uncertainties are negligible compared to others uncertainties on the acceptance. 113

5.3 Top dilepton signal acceptance, in percent, for $m_t = 175\text{GeV}/c^2$, measured in PYTHIA, corrected for all known discrepancies between data simulation. Uncertainty is statistical only. 114

5.4 Summary of systematic uncertainties on the signal acceptance. The total systematic uncertainty is 3%. 115

6.1 Diboson acceptances, in percent, as measured with PYTHIA simulated events, categorized by primary lepton type and number of jets, corrected for all known discrepancies between data and simulation. Uncertainty is statistical only. 126
6.2 Fraction of events with a given jet multiplicity and scale factor correcting simulation to the level of data. 128
6.3 Predicted and observed number of fake leptons in the data and simulation cross-checks. The only statistically significant discrepancy observed is in the 1 jet category in the simulation cross-check, which is part of the basis of the systematic uncertainty on this background estimate. 141
6.4 Fake predictions split by primary lepton type, jet bin, and fully reconstructed vs. track lepton contributions. 150
6.5 Fraction of lepton + isolated track events with opposite sign in simulated W + jets samples. 156
6.6 Derivation of opposite-sign fraction for fakes from data. Attributing all data counts in excess of the diboson and Drell-Yan contributions to events with fake leptons, the opposite-sign fractions measured are $209.5/305.6 = .69 \pm .05$ in all zero jet data and $134.5/199.4 = .67 \pm .03$ in data where the track lepton is matched to a significant energy deposit in the hadronic calorimeter. . . . 157
6.7 Predicted and observed same-sign candidates, where the opposite-sign fraction is derived from ALPGEN simulation or from zero jet data. 157
6.8 $Z/\gamma^\star \to \tau\tau$ acceptance, in percent, as measured in PYTHIA Monte Carlo simulation, by primary lepton type and number of jets, corrected for all known discrepancies between data and simulation. Uncertainty is statistical only. 160
6.9 Counts from the data used to normalize the $Z/\gamma^\star \to ee/\mu\mu$ background estimate, and the expected contribution from other processes ("feeddown"). 163
6.10 Inputs to the Drell-Yan estimate from simulated $Z/\gamma^\star \to ee/\mu\mu$ events. 164
6.11 Summary table of systematic uncertainties on the background estimate. 167

7.1 Predicted and observed events in 1.1 fb^{-1}, with details of the background contributions. Systematic uncertainties are included in the prediction numbers. 172
7.2 Numbers of events expected from signal and background compared to the number of observed events in 1.1 fb^{-1} of data, for events where the primary lepton is a fully reconstructed electron. Recall that events are selected first as CEM + track, then as CMUP + track, then CMX + track, then PHX + track. Systematic uncertainties are included. 173
7.3 Numbers of events expected from signal and background compared to the number of observed events in 1.1 fb^{-1} of data, for events where the primary lepton is a fully reconstructed muon. Recall that events are selected first as CEM + track, then as CMUP + track, then CMX + track, then PHX + track. Systematic uncertainties are included. 174
7.4 The cross-section as calculated at several input top masses. Theoretical prediction from Ref. [7]. A common uncertainty of 6%, due to the uncertainty on the integrated luminosity, is omitted. 174
7.5 Predicted and observed events for the 1.1 fb^{-1} lepton + track sample, where the track lepton and jet p_T thresholds are both at 15 GeV/c instead of the default 20 GeV/c. 175
7.6 Predicted and observed events for the 1.1 fb^{-1} lepton + track sample, where the track lepton and jet p_T thresholds are both at 25 GeV/c instead of the default 20 GeV/c. 176
7.7 Predicted and observed events in the subsample where the track lepton is matched to a fully reconstructed lepton. Uncertainties are statistical only. 178

Chapter 1

Introduction

The goal of particle physics is the determination of the fundamental constituents of the universe and the nature of their interactions. In the last fifty years, much progress has been made in uncovering these constituent particles and elucidating the rules governing their interactions. All observed phenomena, with the exception of the observed neutrino masses, are described by what has come to be called the Standard Model of particle physics. Since the observation of the W and Z bosons predicted by the Standard Model, experimental particle physics has focused on searches for the remaining particles predicted by the Standard Model and measurements of their interactions.

With the observation of the top quark in 1995 [8, 9] and the tau neutrino in 2000 [10], nearly the full catalog of Standard Model particles has been directly

observed, with only the Higgs boson evading detection. Despite the great success of the Standard Model in describing the particles and interactions observed up to now, there is good theoretical motivation to believe that a new framework must come into play at approximately the TeV energy scale. Increasingly stringent measurements of the properties of particles and their interactions at the highest available energies, coupled with direct searches for phenomena not described by the Standard Model, is the project of of modern experimental particle physics. A significant part of that project is the elucidation of the properties of the top quark.

1.1 The Standard Model of Particle Physics

The Standard Model of particle physics categorizes the particles in the universe as spin-1/2 fermions which interact via three fundamental forces (gravity is excluded) mediated by spin-1 bosons [11, 12]. The three interactions are electromagnetic, strong, and weak. The massless, neutral photon (γ) mediates the electromagnetic force, a long-range force and the one most familiar from everyday experience. The gluon (g), also massless and neutral, mediates the strong force, and itself carries the corresponding charge (color). Strong interactions bind quarks together into nucleons, and the nucleons into atomic nuclei. The

massive charged W ($M_W = 80.4$ GeV) and neutral Z ($M_Z = 91.2$ GeV) mediate the weak force. Outside of high-energy particle physics, the weak force shows up primarily through the decay of radioactive nuclei. Each of these particles acts not only as force carrier but can also be produced and detected as a real particle, at least indirectly.

The fundamental fermions, the primary constituents of matter, are further divided into quarks and leptons. Quarks carry color as well as electrical charge and participate in all interactions. Quarks of charge +2/3 are grouped with quarks of charge -1/3 into SU(2) doublets: weak interactions can change the "flavor", or type, of the quarks, but other interactions leave it intact. The strength of the strong interaction increases with the separation between colored objects, which results in the *confinement* of quarks and gluons into compound objects, called hadrons. Leptons are colorless, and charged leptons are paired with neutral leptons, called neutrinos, into SU(2) doublets. The neutrinos, which interact only weakly, are identified by the type of the charged lepton that they are coupled to through the weak interaction.

There are three copies of the prototype quark and lepton doublets, called generations. The particles of the first generation, the up and down quarks and the electron (as well as the electron neutrino), are the constituents of ordinary matter. Their observed stability is a result of their being the lightest particles

of their kind and the apparent conservation of lepton and baryon number. Any kinematically allowed decays would require a violation of what are believed to be fundamental symmetries of Nature. The other two generations are successively more massive and are unstable; they exist only as the product of collisions by cosmic rays and at man-made particle accelerators. The three generations of fermions and their basic properties are given in Table 1.1.

For each particle there is a corresponding antiparticle, identical to the original but with all of the opposite quantum numbers. For reasons also still unknown, the universe in which we live is composed almost entirely of matter, but at colliders, matter and antimatter are produced, and generally treated, equally.

The character of the Standard Model at the energies necessary for top quark production, as it applies to the results of proton-antiproton collisions, is described in Chapter 2.

1.2 Particle Physics at Hadron Colliders

Hadron colliders, typically proton-proton or proton-antiproton, provide much useful data for the study of fundamental particles. These colliders run with the highest beam energies currently available, because the beams do not lose as

Quarks

generation	name	symbol	charge	mass (MeV)
1	up	u	+ 2/3	1.5 - 3.0
	down	d	- 1/3	3 - 7
2	charm	c	+ 2/3	1250 ± 90
	strange	s	- 1/3	95 ± 25
3	top	t	+ 2/3	$(1.74 \pm 0.03) \times 10^5$
	bottom	b	- 1/3	$(4.20 \pm 0.07) \times 10^3$

Leptons

generation	name	symbol	charge	mass (MeV)
1	electron	e	-1	0.511
	electron neutrino	ν_e	0	$< 2 \times 10^{-6}$
2	muon	μ	-1	106
	muon neutrino	ν_μ	0	$< 2 \times 10^{-6}$
3	tau	τ	-1	1780
	tau neutrino	ν_τ	0	$< 2 \times 10^{-6}$

Table 1.1: Known fermions and their properties, from [6]. Note that, because they interact strongly and (except the top quark) are not observed in isolation, quark masses are only well-defined for a choice of renormalization scheme.

much energy by radiation when held in a ring as electron beams. Inelastic collisions between the partons (quarks and gluons) within the hadrons occur at a broad spectrum of energies, producing all particles kinematically allowed by the beam energy, at rates depending on the particle masses and interactions governing their production.

At the Tevatron at Fermilab, protons and antiprotons held in a storage ring collide at a center-of-mass energy of 1.96 TeV. It is currently the world's highest-energy collider, and the only one with sufficient energy to produce the very massive top quark. The existence of the top quark was first established at the Tevatron, and measurements of its properties are an important part of the ongoing physics program at the collider.

The trade-off for this rich variety of events produced is a limited knowledge of the initial state of the interaction. The practice of particle physics at hadron colliders is fundamentally shaped by what is *not* known about the interactions. We do not know the fraction of momentum carried by each of the colliding partons in any given event, so that the total momentum of the resulting particles along the beam axis is not known. Even if this component of the momentum for most particles in the final state can be measured, neutrinos and particles traveling down the beam pipe make a full accounting impossible. Therefore the kinematics of interactions are described in terms of variables *transverse* to the

beam axis. Typically, instead of energy and momentum, the transverse momentum $p_T = p \sin\theta$ and transverse energy $E_T = E \sin\theta$ are used. Transverse energy is a somewhat counterintuitive concept because energy is not a vector, but it is convenient to treat only the fraction of energy which is attributable to motion transverse to the beam, which is much less dependent on the fraction of the proton momentum carried by the colliding parton.

Part of the convenience of the concept of transverse energy is that energy and momentum are conserved in the transverse directions. That is, if *every particle in the event is detected*, the summed momentum or energy in the two transverse directions (x and y) should be zero. Therefore, in a perfect detector, if the sum is not zero, not every particle in the event has been detected. If a particle has escaped, the x and y components of its momentum can be inferred from the momentum apparently missing from the event. In practice, this is the case for neutrinos, and no other currently known particles, so this *missing transverse energy*, or $\not\!\!E_T$, is the distinctive signature of one or more neutrinos in the final state of an event.

Another limit to the understanding of interactions is that quarks and gluons cannot be detected directly since they do not exist in isolation. The strength of the strong force increases with the separation distance between two colored objects, until the amount of energy required to pair-produce more colored par-

ticles is less than the amount of energy required to hold the particles apart. Therefore any quark or gluon created with significant p_T in an event is connected to the rest of the event by a color field, and as they fly away from the interaction point, this color field breaks and additional quarks and gluons are formed. The quarks and gluons reassemble into colorless hadrons, but retain a boost in the direction of the original parton. What we actually reconstruct when an energetic parton has been produced is a collimated *jet* of particles, all traveling in approximately the same direction as the original parton. Jets are a useful concept because of the directional correlation between it and the parton and because the summed transverse momentum of the particles in the jet is a good measure of its transverse momentum.

High-p_T physics processes are reconstructed through their expected signatures in the detector. The most general model of a hadron collider detector is a cylindrically symmetric detector, composed of several layers, each with a different structure and intent, surrounding the interaction point. The innermost layers are made of low-mass materials and it is expected that generated particles will pass through them, with charged particles leaving a trail of ionization in their path. This ionization can be detected and used to reconstruct the trajectory of the particle through this part of the detector. This reconstructed trajectory is referred to as a "track". Outside of the detector components used

to reconstruct particle tracks, layers of high-mass material, collectively referred to as a calorimeter, stop most particles produced by the interaction. With only two important exceptions, particles interact with the matter in the calorimeter, depositing all of their energy in the material, allowing the particle energy to be measured. The fact that muons and neutrinos are not stopped by the calorimeter is key to their identification.

The Tevatron accelerator and the components of the CDF detector relevant to this analysis are described more thoroughly in Chapter 3.

1.3 Selecting Top Quark Pairs

Particle signatures are reconstructed from information from the tracking and calorimeter detectors. Because of their stability, and because they do not participate in the strong interactions, electrons, muons and photons can often be completely reconstructed. Tau leptons can be at least partially reconstructed, depending on the decay mode of the τ. Quarks and gluons can be detected as jets, although in general the type of the parton cannot be determined.[1] $\not\!\!E_T$, which flags the presence of one or more neutrinos, completes the set of objects upon which event selection is based.

[1] The b quark, and to some extent the c quark, are exceptions; at the Tevatron they can be identified reasonably well because of the long lifetime of the hadrons they form and/or the decay of those hadrons to electrons or muons

The top quark decays weakly, into a b quark and a W boson. This analysis is based on identifying top quark decay events where both of the Ws have decayed into a charged lepton and a neutrino. The basic signature of such events is two charged leptons, two jets from the b quarks, and $\not\!\!E_T$ from the neutrinos. One of the charged leptons is fully reconstructed as an electron or muon. To increase the number of top events passing the selection, the second lepton is identified only as a track, a feature common to nearly all lepton final states. Details of the event selection, and further motivation for the choice of identifying the second lepton as a track, are given in Chapter 4.

1.4 Measuring the Top Quark Cross-section

Since the discovery of the b quark, experimentalists searched for the positively charged SU(2) partner required by the Standard Model. Because of the mass of this partner (≈ 175 GeV/c^2) turned out to be much larger than anticipated, it was not until 1995 that the existence of the top quark was established using data from Run I of the Tevatron at Fermilab [8, 9]. Because of its large mass, the cross-section to produce top pairs is extremely small compared to the total $p\bar{p}$ cross-section, or even the cross-section for electroweak processes like W and Z boson production. The cross-section for a process describes the number

of expected events given a certain amount of data. It is expressed in terms of an area, to give a number which is independent of the beam intensity. The customary units are *barns*, where 1 barn is $10^{-28}\mathrm{m}^2$. The barn was originally intended to represent the approximate area of a cross-section of a nucleus; total proton-antiproton cross-sections are of the order of microbarns, and the scale for electroweak processes and top quark production is picobarns.

The pair production cross-section is one of the conceptually simplest measurements that can be made for a given particle. For this reason the cross-section, along with the top quark mass, were the properties reported in the top quark discovery papers. All that is required to measure a cross-section is a count of the number of events produced, corrected for the number of events from other sources, called *backgrounds*, that will also pass the event selection. The cross-section is defined as

$$\sigma = \frac{N_{\mathrm{obs}} - N_{\mathrm{bkgnd}}}{A \cdot \int \mathscr{L} dt}$$

N_{obs} is the number of events observed, and N_{bkgnd} is the calculated expected number of background events. A, the *acceptance*, is the ratio of the number of events passing the selection to the number produced. Like the background, it depends on the event selection. $\int \mathscr{L} dt$ is the *integrated luminosity*, a measure of the amount of data collected. The integrated luminosity is often given in *inverse*

picobarns, so that it can readily be converted to a cross-section given a number of events, and vice versa. We describe the acceptance for the selection used in this measurement in Chapter 5, and the background calculations in Chapter 6.

The small $t\bar{t}$ production cross-section meant that in Run I the discovery could only be made by combining information from different decay channels. Each top quark is expected to decay to a b and a real W, and the W decays in turn. When both Ws decay to leptons, the event is referred to as a "dilepton" event in reference to the two charged leptons in the final state. When one decays to leptons and one to a quark pair, that final state is called "lepton plus jets". Events from both of these channels were included in establishing the existence of the top quark. Events where both Ws decay to jets have been used to measure properties of the top quark, particularly the mass, but were not used in the discovery due to the large amount of background in the channel.

The "discovery" papers were written with 67 pb^{-1} (50 pb^{-1}) of data from CDF (D0). At the end of Run I, final cross-section papers were written using the full dataset of over just over 100 pb^{-1} . The dilepton cross-section, as measured by CDF using the full Run I dataset, was $8.2^{+4.4}_{-3.4}$ pb at a top mass of 175 GeV/c^2. That is higher, but not significantly so, than the theoretical prediction for that top mass and $\sqrt{s} = 1.8$ TeV, $5.2^{+0.5}_{-0.7}$ pb [7, 13]. The result might not have had much impact but for the distribution of the events: a larger fraction (seven of

nine) of the candidate events were in the $e\mu$ final state compared to the half (4.5) that would be expected, and those events had more $\not\!\!E_T$ than would be expected. Figure 1.1 shows the $\not\!\!E_T$ of the observed candidate events compared to the combined expectation for signal and background. With approximately the same integrated luminosity, the D0 collaboration measured a cross-section closer to expectation, although the result in the $e\mu$ channel was higher than the ones in the ee or $\mu\mu$ channels (but again, not significantly) [14].

The results from Run I generated speculation that there was something other than just top quarks in the event excess being ascribed to top quark production, particularly since supersymmetry may look much like $t\bar{t}$ production except with more $\not\!\!E_T$. Such possibilities for the influence of physics beyond the Standard Model in the dilepton sample are discussed in more depth in Chapter 2. When reliable data started to accumulate in Run II, one of the highest-priority measurements planned was to see if the Run I dilepton numbers could be duplicated. The first published results used ≈ 200 pb^{-1} of data. Both the CDF and D0 collaborations found results in good agreement with the Standard Model prediction $6.7^{+0.7}_{-0.9}$ pb [7]. D0 used events with fully-reconstructed electrons and/or muons in the final state and measured $8.6^{+3.2}_{-2.7}$(stat.)$\pm$1.1(sys.) for a top mass of 175 GeV/c^2. The CDF result combined two measurements, one using primarily fully reconstructed leptons, and the other, the basis for the measurement presented

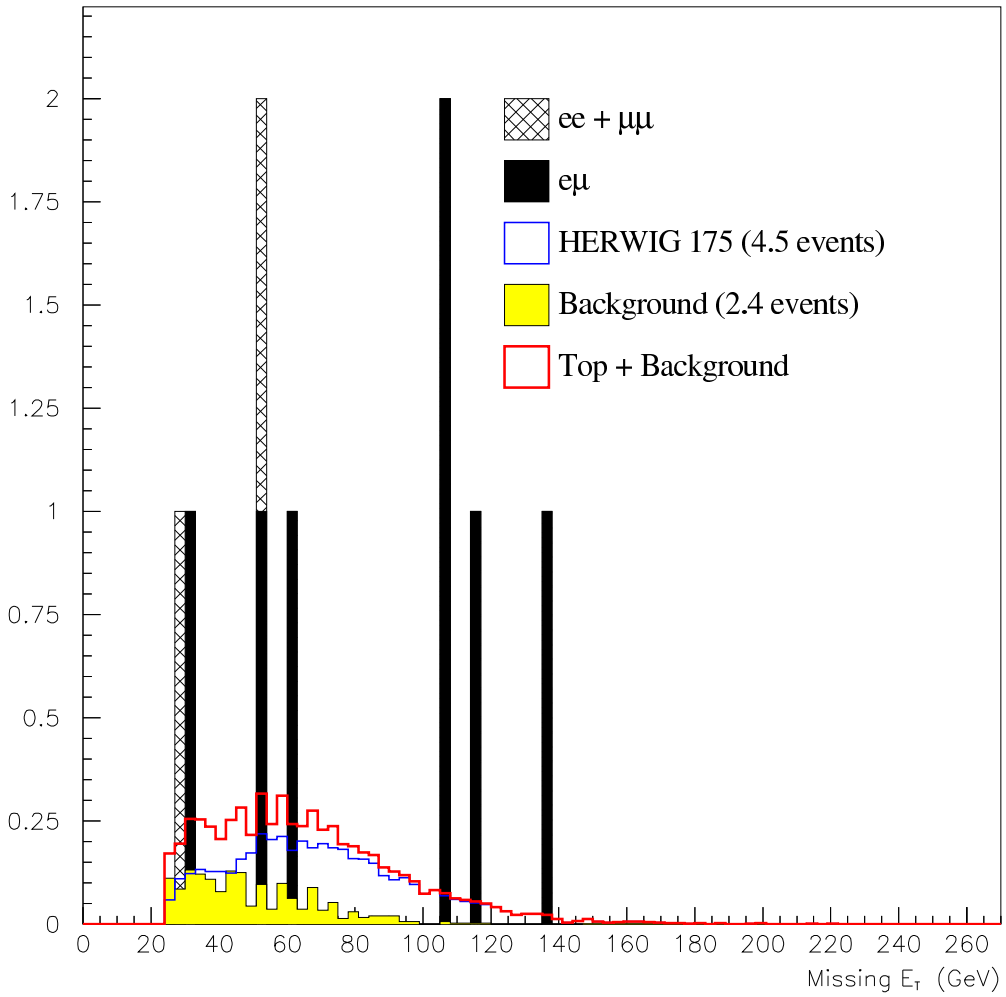

Figure 1.1: Missing transverse energy of $t\bar{t}$ dilepton candidates at CDF in Run I, compared to the expectation for combined signal and background. Taken from Ref. [1].

in this document, allowing one lepton to be only partially reconstructed. The combined result for $m_t = 175$ GeV/c^2 was $7.0^{+2.4}_{-2.1}$(stat.)$^{+1.6}_{-1.1}$(sys.). These measurements, while they improved significantly on the Run I results, still lacked the precision to make a clear statement about the content of the dilepton + jets + $\not\!\!E_T$ sample.

Since the publication of those results, the luminosity capabilities of the Tevatron have increased and CDF and D0 have rapidly integrated datasets in excess of 1 fb^{-1} . The analysis presented here is an update of the published CDF analysis which used lepton + track identification for the lepton pair. This version uses over 1 fb^{-1} of data and reflects improved understanding of the background composition in the dilepton sample. With ten times the amount of data used in Run I, detailed studies of the top quark sector are becoming possible. This is of fundamental interest in its own right, since the top quark is a unique member of the Standard Model, and one of the most newly discovered. In the context of pushing our understanding of particle physics beyond its current level, the motivation for study of the top quark has added layers. Inconsistencies with the Standard Model predictions for the production rate and decay mechanisms would indicate something other than just top quarks in the data. Recent results do not rule out that something not yet understood lurks in the sample, but with a signature less dramatic than the Run I results allowed. Even in the absence of

indications of new physics, solid understanding of the physics of the top quark sector (the process itself, and the composition of the multilepton + multijet + $\not\!\!E_T$ sample) will be a necessary prerequisite to discovering and understanding new physics if it is to be found at still higher energies.

Chapter 2

Relevant Theory

Understanding the signal and background of top quark production and decay requires a broad overview of high-p_T Standard Model physics. Top quark-antiquark pairs are produced at the Tevatron through the strong interaction, via the collision of either quark-antiquark or gluon-gluon pairs. The frequency of such interactions is determined by the strong interaction (quantum chromodynamics, or QCD) and the momentum distribution among the partons within the proton. The top quarks then both decay weakly, to a b quark and a real W boson. The Ws each decay in turn to either a quark and antiquark, where one is an up-type (u or c) quark and the other is down-type (d, s, or, rarely, b), or a charged lepton and its associated neutrino, with probabilities determined by the electroweak theory. Backgrounds to dilepton $t\bar{t}$ decays all arise from the

production and decay of one or two massive vector bosons, along with additional jets produced by the strong interactions. Therefore solid understanding of backgrounds to top quark production requires understanding not only of the electroweak production and decay mechanisms for the vector bosons but also of the production and characteristics of the additional jets.

The top quark also has potentially strong connections to physics beyond the Standard Model. One of the most unusual features of the top quark is its large mass, approximately double that of the massive vector bosons and well over ten times that of any other quark. Because the proposed Higgs boson couples to fermions in proportion to their masses, the top quark will likely have a significant role in the Higgs sector, if the Higgs mechanism works as it is expected to in the Standard Model. Another potential connection lies in the unique and complex signature of top quark pair decays with at least one lepton in the final state. No other leading-order Standard Model process generates such a large multiplicity of particles with large transverse momentum which includes charged leptons and $\not\!\!E_T$, but that kind of final state, particularly the significant $\not\!\!E_T$, is expected from supersymmetry and other extensions of the Standard Model. Even if top quarks are not produced as part of the decay chain, understanding the top quark can be a valuable training ground for the pursuit of new physics signals.

2.1 Top Pair Production at the Tevatron

2.1.1 The Tree-Level QCD Process

At the Tevatron, $t\bar{t}$ pairs are produced through the collision of either a $q\bar{q}$ or gluon-gluon pair. It is possible to calculate the inclusive pair production cross-section using perturbative QCD because of the large top quark mass, which is much bigger than the QCD scale ($\Lambda_{\rm QCD} \approx 1$ GeV). The cross-section is evaluated as a perturbation series in powers of the strong coupling constant α_s, where α_s is evaluated at the relevant energy scale, the top quark mass. The leading-order ($\mathcal{O}(\alpha_s^2)$) diagrams are shown in Figure 2.1. The corresponding squared matrix elements, averaged over initial spin and color states, and summed over final spin and color states, are

$$\overline{\sum} |\mathcal{M}_{q\bar{q}}|^2 = \frac{64}{9}\pi^2\alpha_s^2 \left(\frac{1}{1+\cosh(\Delta y)}\right)\left(\cosh(\Delta y) + \frac{m^2}{m^2+p_T^2}\right) \qquad (2.2.1)$$

for the $q\bar{q}$ contribution and

$$\begin{aligned}\overline{\sum} |\mathcal{M}_{gg}|^2 &= \frac{4}{6}\pi^2\alpha_s^2 \left(\frac{8\cosh(\Delta y)-1}{1+\cosh(\Delta y)}\right) \\ &\quad \cdot \left(\cosh(\Delta y) + 2\left(\frac{m^2}{m^2+p_T^2}\right) - 2\left(\frac{m^2}{m^2+p_T^2}\right)^2\right) \qquad (2.2.2)\end{aligned}$$

for the gg contribution [15].[1] The matrix elements are shown as a function of m, the top quark mass; p_T, the transverse momentum of the top quarks produced;

[1] The discussion in this section is largely drawn from the description of heavy quark production in chapter 10 of this reference.

and Δy, the rapidity difference between the quark and the antiquark, where $y = \frac{1}{2}\ln((E+p_z)/(E-p_z))$ is the rapidity.

Using the matrix elements above, the differential cross-section may be written

$$\frac{d\sigma}{dy_3 dy_4 d^2p_T} = \frac{1}{64\pi^2(m^2+p_T^2)^2(1+\cosh(\Delta y))^2} \cdot \sum_{ij} x_1 f_i(x_1,\mu^2) x_2 f_j(x_2,\mu^2) \overline{\sum} |\mathcal{M}_{ij}|^2 \qquad (2.2.3)$$

The subscripts 3 and 4 refer to the outgoing top quarks, and subscripts 1 and 2 to the incoming partons. So x_1 and x_2 refer to the momentum fractions of the incoming partons relative to the proton (or antiproton) momentum, but the sum over i and j refers to a sum over all possible combinations of incoming partons that can contribute. $f_i(x,\mu^2)$ is the parton distribution function, which describes the probability for a parton of type i to be found with a momentum fraction x given a *factorization scale* μ. The factorization scale is the momentum scale above which interactions and particles are treated as part of the perturbative calculation and below which they are treated as part of the parton distribution functions. This scale is (usually) set to be the same scale at which α_s is evaluated.

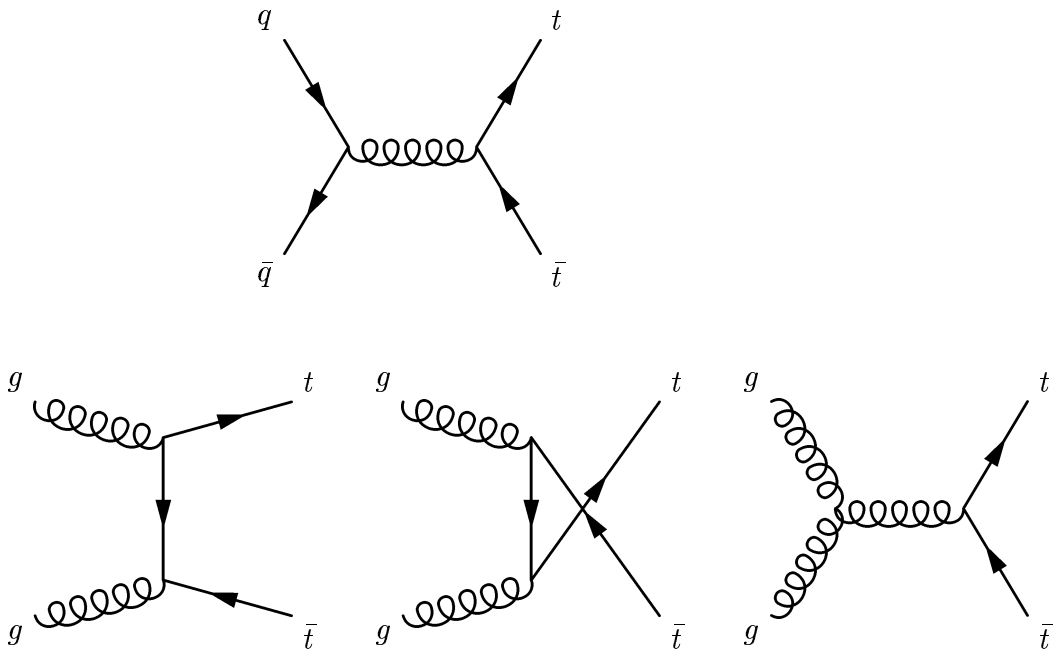

Figure 2.1: Leading order diagrams for top quark pair production at the Tevatron.

2.1.2 Parton Distribution Functions

Production cross-sections calculated for hadron colliders depend both on the perturbatively calculated parton-parton cross-sections, and the distribution of the hadron's momentum among its constituent partons. That momentum distribution is described by parton distribution functions, or PDFs, and is universal to all calculations made for a given collider. These functions describe QCD in its nonperturbative domain, and cannot at present be calculated from first principles, but must be determined by experimental data. Data from a wide variety of experiments are used, including the deep inelastic scattering of electrons, muons, neutrinos, and protons on fixed targets of differing nuclear content, as well as collider data from both electron-proton collisions and proton-antiproton collisions [2]. The distributions are shown for scales of 2 GeV and 100 GeV in Figure 2.2. The distributions on the right, for 100 GeV, are closer to the relevant ones for top quark production, or for W or Z production. As would be expected intuitively, up quarks carry most of the momentum, followed by down quarks. Less obvious is the large fraction carried by gluons below about $x \approx 0.1$. This shift has important consequences for understanding the dominance of different QCD diagrams for a given process.

A few general features of top quark pair production can be inferred. Top

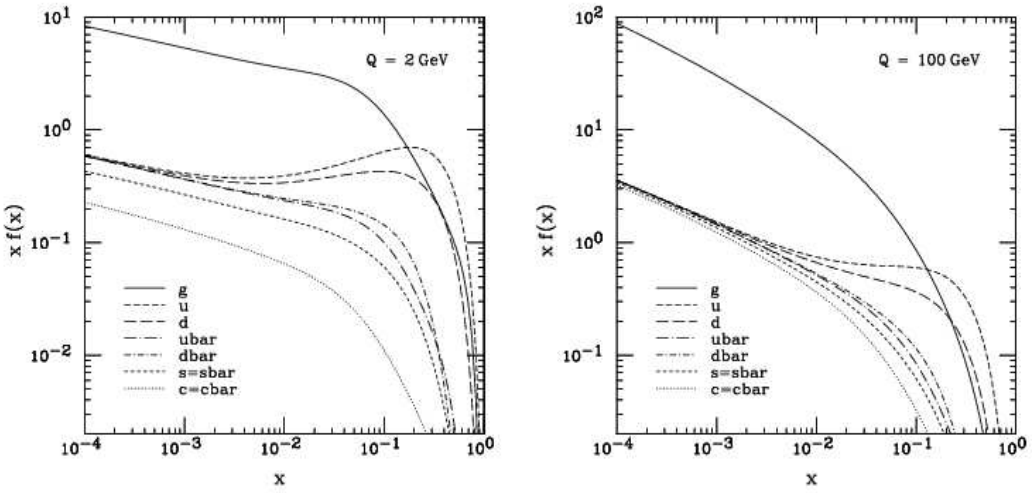

Figure 2.2: Momentum fraction of partons within the proton, for two momentum scales. Taken from Ref. [2].

quark pairs tend to be produced with similar rapidity. For large Δy and fixed top quark p_T, the $q\bar{q}$ matrix element (2.2.1) converges on a constant value, and the gg contribution (2.2.2) converges to $e^{\Delta y}$. Therefore the $1/(1+\cosh(\Delta y))^2$ dependence in the overall cross-section (2.2.3) favors configurations where the $t\bar{t}$ pair have similar rapidity. The $t\bar{t}$ pair is also typically produced near threshold, or almost at rest. This is a result both of the $1/(m^2+p_T^2)$ dependence of the cross-section, which favors production of pairs with $p_T \lesssim m$, and of the rapidly falling PDFs at the large momentum fractions necessary to create a $t\bar{t}$ pair at the Tevatron. Rapidly falling PDFs also mean that most top pairs are produced centrally, because the probability to sample from the PDF at a much higher x for one quark than the other is small. Finally, top quark pairs are

primarily made through $q\bar{q}$ collisions at the Tevatron. This is a consequence of the matrix elements, since the $q\bar{q}$ matrix element is larger in the kinematically favored configuration where p_T and Δy are close to zero, and of the larger quark contribution to the PDFs at large x.

2.1.3 Higher-Order Corrections

The leading-order process outlined above is only the beginning of a perturbative expansion in powers of α_s. In principle, the series must be calculated to all orders to determine the cross-section exactly. To date, the cross-section has been calculated to next to leading order (including all diagrams of order α_s^2 or less), including the next-to-leading-logarithm contributions, which has been sufficient to predict the cross-section with uncertainty less than 15%[16].

The next-to-leading-order (NLO) calculations account for the virtual and real emission of gluons. Representative contributing diagrams are shown in Figure 2.3. The full calculations were first done in [17] and [18]; a a more complete set of contributing diagrams are shown in [18]. Including the NLO contributions increases the predicted cross-section by $\approx 30\%$ at $\sqrt{s} = 1.96$ TeV [19].

An important prediction of the next-to-leading-order calculation is the possibility of additional hard partons produced in addition to the top quark pair. If

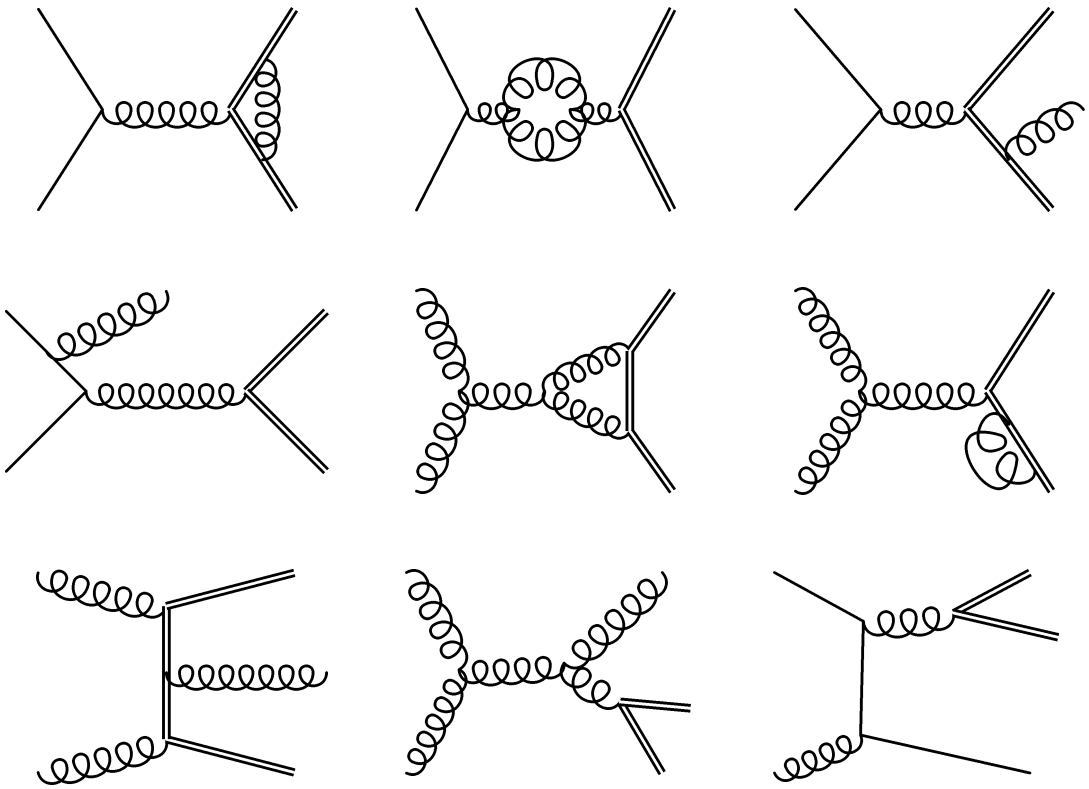

Figure 2.3: Representative subset of diagrams for the next to leading order contribution to top quark pair production at the Tevatron. The double line is the top-antitop pair, and the single line is a light (u or d) quark.

one is produced with sufficiently high transverse momentum and large angular separation from other objects in the event, it is detected as a jet.

The next-to-leading-logarithm contributions have been shown to be small, typically less than 10% depending on the chosen renormalization/factorization scale [7].

2.1.4 Predicted Cross-Section

The most up-to-date calculations of the pair production cross-section include assessments of the theoretical uncertainties. The most important uncertainties are the variation of the cross-section with the chosen renormalization/factorization scale, which contributes about 5%, and the variation due to the uncertainty on the PDFs, which contributes about 6-7% [7]. Most of the PDF uncertainty comes from the lack of knowledge about the gluon distributions at large x (i.e., how often a gluon is carrying a significant fraction of the proton momentum). Gluon-gluon fusion is expected to contribute between 11 and 21% of the total cross-section, but the fraction is not known to better precision. The authors of [7] combine the scale and PDF uncertainties linearly instead of in quadrature, since the scale uncertainty is not a "systematic error in the strict sense".

The predicted cross-section varies strongly with the input top mass, decreasing with increasing top mass, since top pairs are produced predominantly near the kinematic threshold at the Tevatron. Table 2.1 shows the predicted cross-section for three values of the mass in a range bracketing the usual range of measured values (from [16], as updated in Ref. [7]). The PDFs used to generate these numbers are taken from Ref. [2].

m_t (GeV/c^2)	σ (pb)
170	$7.8^{+0.9}_{-1.0}$
175	$6.7^{+0.7}_{-0.9}$
180	$5.8^{+0.6}_{-0.8}$

Table 2.1: Range of theoretical predictions for the top quark pair production cross-section at $\sqrt{s} = 1.96$ TeV.

Another calculation of the cross-section, to the same order as Ref. [7], is given in Ref. [20]. Using the same PDFs as that reference, and a top mass of 175 GeV/c^2, they predict a range of 6.08 - 7.21 pb for the cross-section, which is consistent with the $6.7^{+0.7}_{-0.9}$ from Ref. [7]. This calculation is not cited as often as the one in [7] because the quoted range does not include the uncertainty from the PDFs.

2.2 Top Quark Decay Modes

The top quark is unique among quarks in that it decays so quickly that it does not hadronize first. Its short ($\approx 10^{-25}$ s) lifetime, in spite of the fact that it decays weakly, is a direct consequence of its large mass [15]. The top quark is in fact sufficiently massive that it decays, before hadronizing, to a real W boson in addition to a b quark.

In the Standard Model, the top quark is expected to decay to a real W boson

and a b quark in 99.9% of decays. This branching fraction is not measured experimentally to better than 20% accuracy [21, 22], but can be predicted using other measurements and certain Standard Model assumptions. The Cabibbo-Kobayashi-Maskawa matrix describes the couplings among the quarks in the electroweak interaction. Other couplings have been measured very well, so using a global fit to data and constraining the matrix to be unitary (meaning that quarks do not turn into something else), the predicted value of the branching ratio is $0.999100^{+.000034}_{-.000004}$ [6].

The complete final state of the top quark pair is the product of the decays of the two Ws and bs. The b quarks hadronize and decay, and are detected as hadronic jets. Each W boson can decay either to a charged lepton and a neutrino or to a pair of quarks. Top events are categorized as "all-hadronic", "lepton+jets", or "dilepton", based on whether neither, one, or both of the Ws decays to leptons. This analysis uses events in the dilepton final state.

At leading order, all possible W decays, with three possibilities for each quark pair (for three colors), are predicted to occur with equal probability, or $1/9 \approx .111$. At higher orders, the QCD corrections to the partial widths to quarks enhance the branching fraction to quarks, causing the branching fractions to leptons to decrease [23]. In agreement with this prediction, the measured value of the branching ratio for $W \to \ell\nu$ is $.1080 \pm .0009$ [6]. This analysis

measures the total cross-section using the dilepton decay channel, so a value for the branching fraction must be assumed. We use the measured value.

2.3 Backgrounds

Backgrounds to top quark pair production all involve the production of one or more massive vector bosons, at least one of which subsequently decays to leptons. This includes single Z/γ^* production (Drell-Yan), single W production, and diboson (WW, WZ, or ZZ) production. In each case additional jets must be produced through the strong interaction in order for the event to pass the analysis selection.

2.3.1 Drell-Yan

The Drell-Yan process ($q\bar{q} \to Z/\gamma^\star \to \ell^+\ell^-$) is one of the best-understood real processes involving interacting quarks, because the main interaction is electroweak. The QCD corrections to this process have been calculated to next-to-next-to-leading order, or α_s^2 [24, 25, 26, 27]. Lepton pairs are produced with an invariant mass distribution that falls roughly as $1/M^2$, where M is the invariant mass of the lepton pair, except near the Z mass (91 GeV/c^2), where the differential cross-section has a resonant peak. The cross-section as a function of

invariant mass is shown in Figure 2.4. The increased cross-section at the Z mass peak is particularly important for this analysis since Z/γ^* decays in that mass range will produce opposite-sign lepton pairs with sufficient transverse momentum to pass the $t\bar{t}$ kinematic selection. The large cross-section at low invariant mass is less of a concern since less energetic leptons are produced. The integral cross-section times the branching ratio to leptons in the invariant mass interval 66 GeV/c^2 $< M <$ 116 GeV/c^2 is 251 $\pm$ 5 pb [4].

2.3.2 W+jets

W production at the Tevatron is analogous to Z production, except that the particle produced is charged, slightly less massive (80 GeV/c^2), and there is no interference with virtual photons. Like the Z, the W can decay into quarks or leptons. The difference for experimental analysis is that the leptonic decay of a W produces one charged lepton and one neutrino rather than two charged leptons. Therefore the event will have legitimate $\not\!\!E_T$ but only one charged lepton, and can only pass the dilepton event selection if something in the event is identified as a second lepton. For this analysis, the most probable way for this to happen is for a hadronic jet to be identified as an isolated track, the second lepton. Therefore to pass all the analysis selection the W must be produced in association with at least three jets.

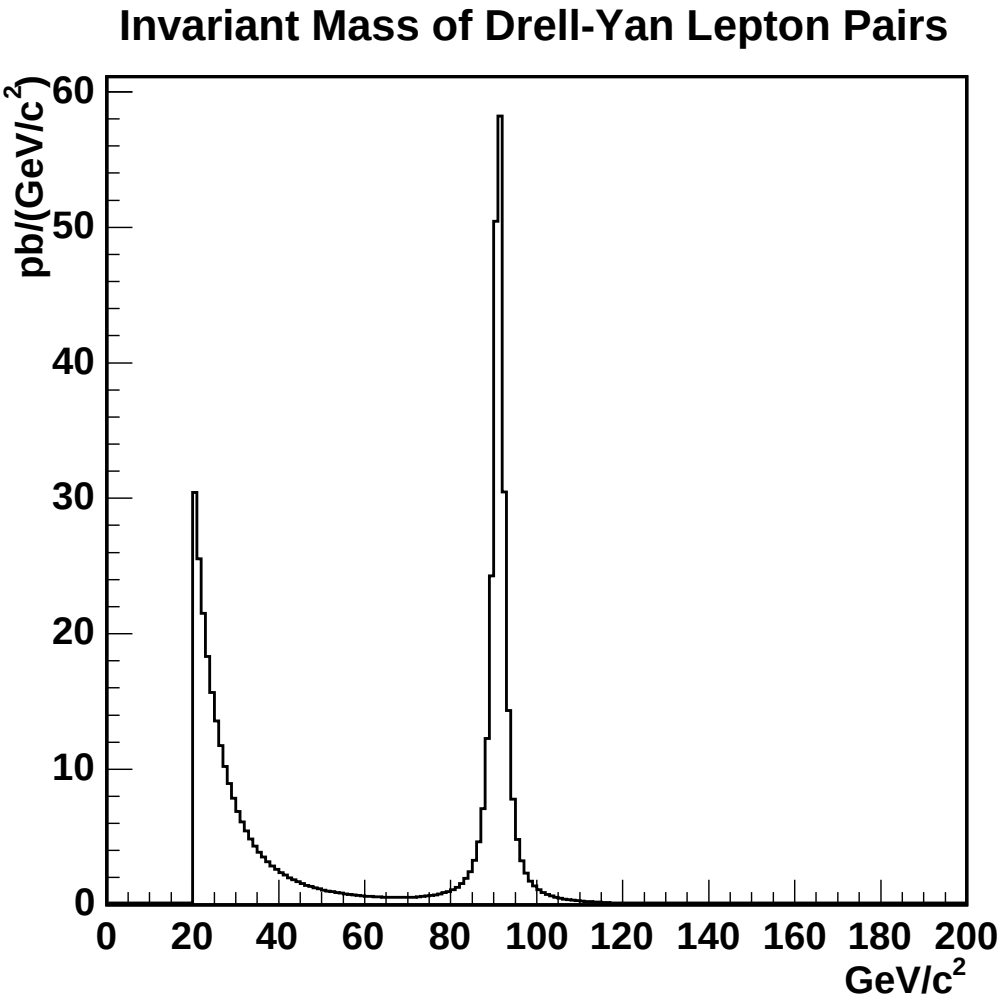

Figure 2.4: Drell-Yan ($q\bar{q} \rightarrow Z/\gamma^* \rightarrow \ell^+\ell^-$) cross-section vs. invariant mass. Invariant mass distribution taken from generator-level information from PYTHIA [3], normalized to an integrated cross-section of 251 pb in the interval 66 GeV/$c^2 < M <$ 116 GeV/c^2 [4].

The precise cross-section to produce a W with so many associated jets is not well-understood because it requires calculation to many orders in QCD. Fortunately, we can base the estimate of the number of background events from this source almost entirely on data (the technique is described in Chapter 6). Nevertheless, an understanding of the higher-order processes that produce the additional jets is important to accurately estimating the contribution from this background.

QCD, Fragmentation, and Fake Leptons

In hadron collider events, quarks or gluons with large transverse momentum are often produced by the hard interaction. They carry color and interact strongly, so as they pull away from the event, they produce a collimated shower of particles, or *jet* (see also Chapter 1 Section 1.2). Because the energy scale for the production of the extra particles is small compared to the momentum of the original parton, it is not currently possible to calculate, even perturbatively, features of jet formation such as the number and energy of the hadrons produced.

Phenomenological models for the formation of a jet from a parton do exist, and some qualitative features of jets can be predicted. Quark fragmentation proceeds through production of quark-antiquark pairs connected to the quark through radiated gluons. Figure 2.5 shows a (necessarily) simplified illustration

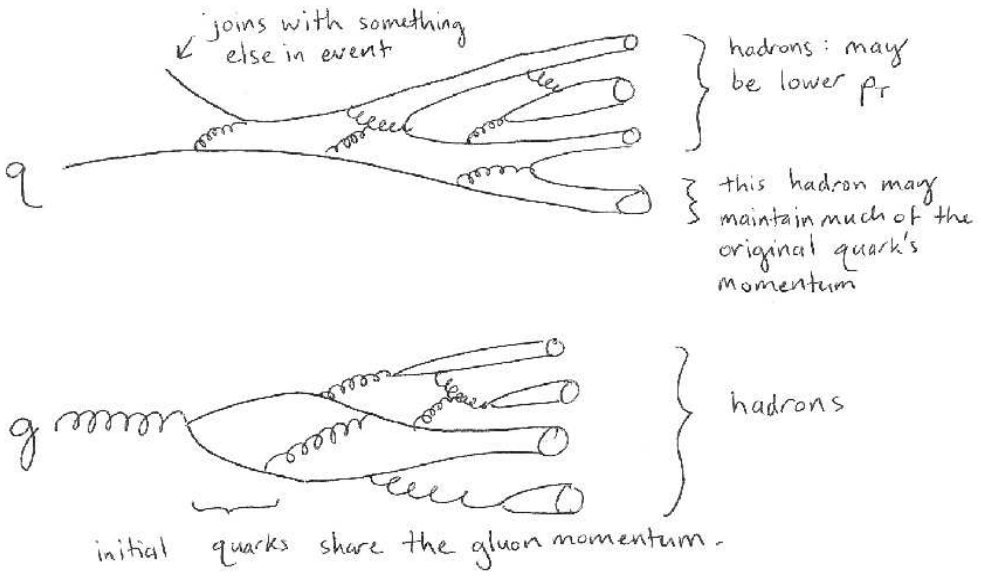

Figure 2.5: Number of tracks with $p_T > 10$ GeV/c in jets with reconstructed $E_T > 20$ GeV, in W + jets data simulated using PYTHIA. Jets produced by quarks tend to have a smaller number of tracks, and jets produced by gluons are more likely to have either zero tracks above threshold or a larger number of tracks.

of quark and gluon fragmentation. Relative to the particles that coalesce into hadrons with the original quark, the extra particles produced may have a small amount of momentum in the lab frame. In contrast, a necessary step in gluon fragmentation is the splitting of the gluon into either a quark-antiquark pair or two more gluons, which themselves split. In either case, both resulting particles retain the boost of the original. The process is iterative, for both quarks and gluons.

The crucial observation for our purposes is that the first step of gluon fragmentation typically results in two copies of the first step in quark fragmentation. This prediction about jets that originate from quarks as opposed to those that originate from gluons. Gluon jets have on average a larger number of particles (and therefore tracks) than quark jets. This behavior has been verified experimentally [28], and can be seen in simulation (Figure 2.6). The different distributions of the number of particles produced means that it is more likely that a quark jet will contain a single charged track than that a gluon jet will contain a single charged track

If the different multiplicity distributions for quark and gluon jets result in higher lepton fake rates for quark jets, it is important to understand the relative frequency of quark and gluon jets produced in association with the W. The two necessary components of this understanding are the possible diagrams for W + jet production and the PDFs for the incoming partons. Diagrams for $W + 1$ jet production are shown in Figure 2.7 One diagram has two incoming quarks and produces a gluon jet. The other has an incoming quark and an incoming gluon and produces a quark jet. The matrix elements corresponding to these diagrams will be similar in magnitude since they are related by swapping an incoming and outgoing particle. But at the Tevatron, the process with an incoming gluon is strongly favored. The amount of energy required to produce

a W boson at threshold is less than .05 of the total collision energy of 1.96 TeV. Even with an additional jet or two this process still occurs at fairly low x, where gluons carry a large fraction of the proton's momentum. Since the diagram with the incoming gluon is favored, most jets in $W + 1$ jet events will be quark jets. Although the effect diminishes somewhat with increasing jet multiplicity, attributable to contributions with incoming and outgoing gluons, quark jets remain the majority component.

2.3.3 Diboson

Diboson production occurs at the Tevatron with cross-sections on the order of a few picobarns, similar to $t\bar{t}$ production. These events can readily pass the $t\bar{t}$ event selection if both of the bosons decay leptonically and additional jets from QCD are present. Diboson production cross-sections have been calculated for Run II of the Tevatron at full next-to-leading order; WW is produced with a cross-section of 13.5 pb, WZ at 4.0 pb, and ZZ at 1.6 pb [29].

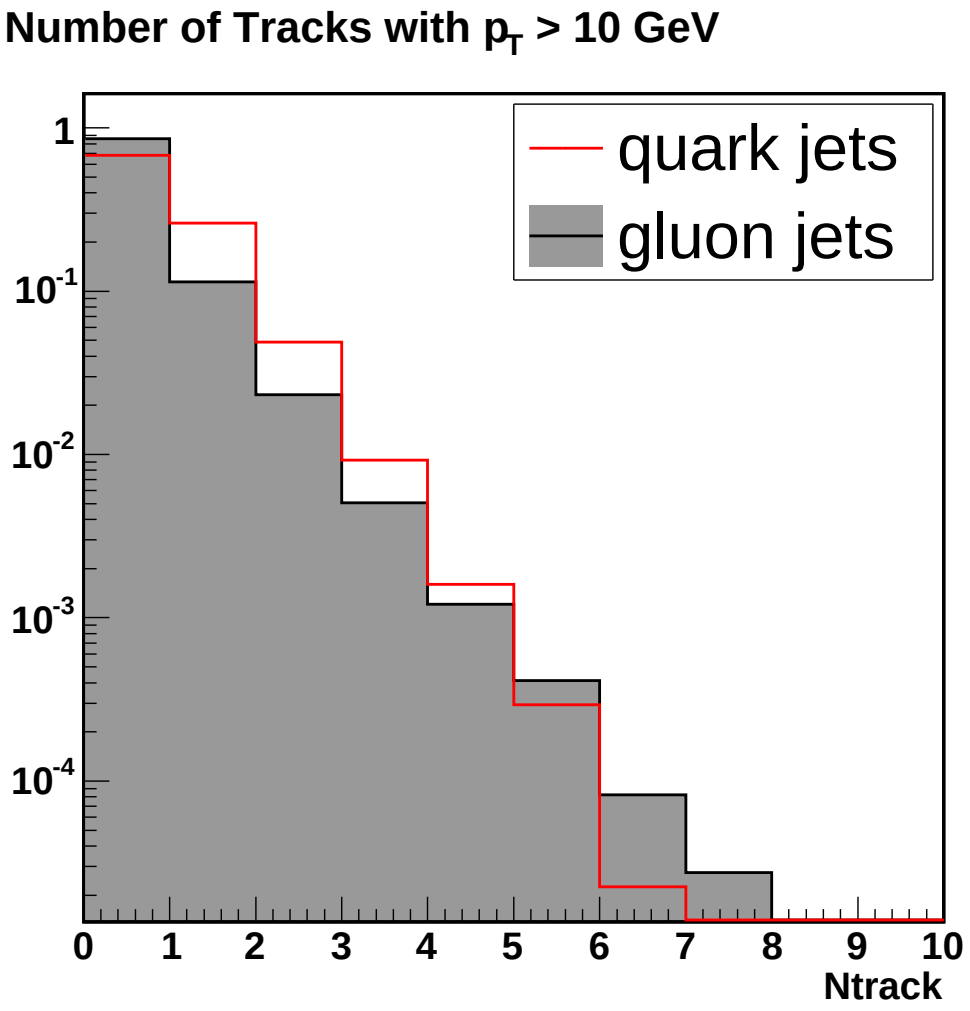

Figure 2.6: Number of tracks with $p_T > 10$ GeV/c in jets with reconstructed $E_T > 20$ GeV, in simulated W + jets data. Jets produced by quarks tend to have a smaller number of tracks, and jets produced by gluons are more likely to have either zero tracks above threshold or a larger number of tracks.

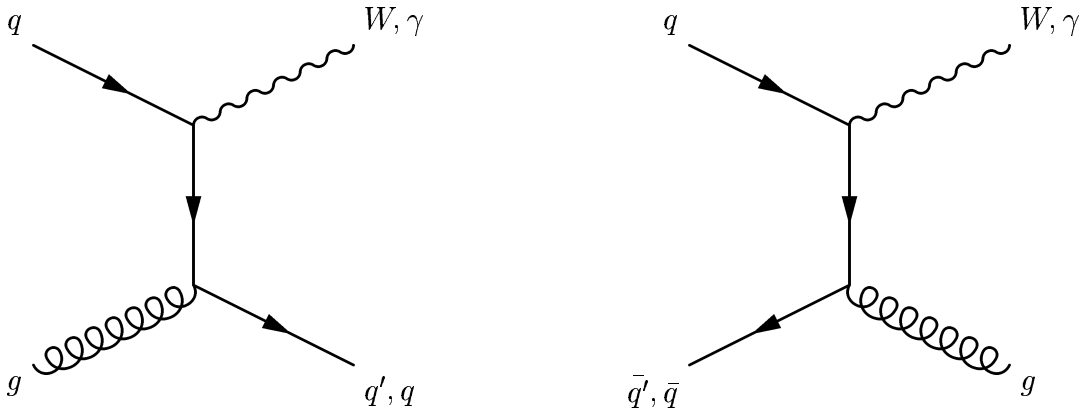

Figure 2.7: Leading diagrams for W/γ production at the Tevatron. The diagram on the left is the leading contribution because of the relative strength of the gluon PDF at low x.

2.4 The Connection to Physics Beyond the Standard Model

Top quark pair production is unique among the Standard Model processes observed at hadron colliders, because of the amount of energy required to produce $t\bar{t}$ pairs and the large number and variety of particles in the final state. This, combined with the relative lack of data on such events has led to extensive speculation about and searches for physics not described by the Standard Model in association with the top quark. The dilepton final state contains hadronic jets as well as two charged and two neutral leptons, with the large amount of $\not\!\!E_T$ being a particularly distinctive feature. Since new stable neutral particles (which would generate $\not\!\!E_T$) are by necessity part any model for physics beyond

the Standard Model, the dilepton channel is particularly relevant.

A large amount of energy is required to produce $t\bar{t}$ pairs because of the large mass of the top quark. At ≈ 170 GeV/c^2, it is the most massive fundamental particle ever studied, with a mass more typical of the electroweak scale ($M_Z = 91$ GeV/c^2, $M_W = 80$ GeV/c^2, M_h estimated to be $\approx \mathcal{O}(100)$ GeV/c^2) than of the other quarks [30]. The proposed Higgs boson couples to particles in proportion to their masses, so the top quark would have the largest coupling to the Higgs of any Standard Model particle. This means at a minimum that top quark physics will be relevant to understanding physics at these scales, and it could indicate that the top quark is more immediately related to electroweak symmetry breaking.

New physics could manifest itself in top quark measurements through imitation of the top quark signature or top quarks being among the decay products of an even heavier particle, enhancing the measured cross-section in either case. The decays of new particles need exactly duplicate the top dilepton final state to be included in the candidate event sample. The data sample used in this (and most) top quark measurements is defined in an "inclusive" way, referring to the minimum number of objects of a certain type but not excluding the possibility of more. The basic events selection requires two *or more* leptons, two *or more* jets, and $\not\!\!E_T$ which might have been generated by any number of neutral,

weakly-interacting particles. In practice this can accept a wide variety of signatures, including many of those expected by models extending the Standard Model. The classic example of such a theory is R-parity-conserving supersymmetry, where the partners of Standard Model particles decay to a final state containing non-supersymmetric particles and the lightest stable supersymmetric particle [31]. Enhancement to the cross-section from very massive particles decaying to top quarks could occur through either resonant or nonresonant production. Proposed mechanisms include a new heavy gauge boson ("Z'") or a technicolor-like scheme where the Higgs is actually a composite particle [32].

The dilepton decay mode of $t\bar{t}$ decays has a particularly strong association with new physics because of the common feature of significant $\not\!\!E_T$, more than is typically generated by the decay of a single W. The viability of any theory for physics beyond the Standard Model requires the existence of a dark matter candidate, which is neutral, stable, and interacts at most weakly (and perhaps only through gravity). The necessity of the existence of dark matter is almost certain and seems to require the existence of such a particle with a mass on the scale of 100 GeV, effectively ruling out neutrinos as candidates [6]. If such particles exist, and interact weakly, it should be possible to produce them at accelerators and, since they are stable, they will be part of the final state of the event. Such particles will escape the detector like neutrinos, and their presence

can only be inferred through $\not\!\!E_T$ in an event. If extraordinary processes were occurring and being accepted as "top quark" events, an excess of $\not\!\!E_T$, generated by the dark matter candidate, could be a flag for their existence.

These arguments are no less true than they were at the beginning of Run II. But, as measurements of the top quark at the Tevatron become more precise, they remain consistent with Standard Model predictions. With the imminent commissioning of the Large Hadron Collider (LHC), it is worth considering possible practical connections of the top quark to new physics. The common features of top quark decays, particularly in the dilepton channel, and the decays of proposed particles not described by the Standard Model, imply that understanding the top quark will be a prerequisite for studying physics beyond the Standard Model at the LHC. Production of $t\bar{t}$ pairs, for which the cross-section is ≈ 800 pb at the LHC [16], will be a significant background to any new physics process producing this type of signature. Through the combined effect of the large cross-section and accessible instantaneous luminosities over 10 times what is possible at the Tevatron, there will be a large sample of $t\bar{t}$ events. It will therefore also have the potential to be a useful calibration sample, since top event identification allows the simultaneous testing of several high-level objects: lepton identification, jet (particularly b-jet) identification, and $\not\!\!E_T$ reconstruction.

2.5 Summary

Top quark physics is inextricably tied to the electroweak sector through its large mass and resulting decay to a b quark and real W, but it is based on the dynamics of the strong interaction. Top quark pair production is a purely QCD process, and understanding the background rates also depends heavily on QCD in both the perturbative and nonperturbative regimes. Backgrounds to top all arise from the production of one or more massive vector bosons, but can only pass the event selection if there is additional hard scattering, producing partons with large transverse momentum which fragment, hadronize, and are identified experimentally as jets. Also, the details of jet fragmentation determine the probability for an object of hadronic origin to be identified as a lepton.

The physics of top quark production and decay connects to every aspect of modern collider physics, including a potentially significant role in the understanding of physics beyond the Standard Model. Therefore it is of great interest to study it in detail experimentally.

Chapter 3

The Tevatron and the CDF Detector

This analysis uses data produced by the collisions of protons and antiprotons at 1.96 TeV at the Tevatron accelerator at Fermilab. The Collider Detector at Fermilab (CDF) surrounds one of the two interaction points on the Tevatron. Closest to the interaction point are low-mass tracking detectors inside of a solenoid, which act as a spectrometer, recording the momenta and trajectories of charged particles. Outside the tracking and the magnet is the calorimeter, a mass of metal interleaved with scintillator detectors designed to stop nearly all the particles emerging from the interaction and record the energy deposited in the process. Surrounding all of that are muon detectors, which record the

signatures of charged particles exiting the detector. Far more collisions are produced than can be recorded, so a trigger system performs fast reconstruction of objects in the detector and flags those that are most interesting according to a pre-defined set of criteria. This measurement uses those events which seem to contain electrons or muons with large transverse momenta.

3.1 The Tevatron at Fermilab

The Tevatron is the most energetic particle accelerator in the world, until the commissioning of the Large Hadron Collider at CERN is completed in 2008. Five stages of acceleration transform protons from the nuclei of hydrogen atoms to counter-circulating beams of protons and antiprotons that each have an energy of 980 GeV [33, 34].

The first stage of the accelerator is a Cockroft-Walton accelerator, an electrostatic accelerator that boosts H^- ions from thermal energies to 750 keV. H^- ions are made by flowing hydrogen gas between a metallic cathode and anode [35, 5]. An electrical pulse generates an arc between the cathode and anode, liberating electrons from the metal surface. A magnetic field points parallel to the cathode surface, causing the electrons to travel in tight spiral. The electrons ionize the hydrogen gas. H^+ ions accelerate toward the cathode, where they may

pick up electrons from the plasma near the metal surface. The resulting H^- ions accelerate back out through gaps in the anode. The entire ion source assembly is inside a metal dome, kept at an electrical potential of -750 kV. Exiting the source, the ions accelerate toward a wall held at ground, accumulating 750 keV of energy.

From there the ions enter a *linear accelerator*, which brings their energies to 400 MeV. This and all subsequent stages of acceleration use radio frequency (RF) electromagnetic fields, rather than static electric fields. Electrostatic accelerators are limited by the intensity of the field that can be applied without causing electrical breakdown in the surrounding material. RF accelerators expose particles to a series of moderate accelerating fields, each one imparting more energy to the beam. A sinusoidally varying electrical field provides the accelerating kick, but the beam must be shielded from the field while it points in the direction opposite the desired one. The first part of the linear accelerator accomplishes this using *drift tubes*, metal tubes held at ground potential. The length and placement of the tubes are calculated based on the field timing and expected velocity of the particles so that they will only be exposed to accelerating fields in their direction of travel. The second part of the linear accelerator is a sequence of accelerating cavities, where the field in each cavity is $\pi/2$ out of phase with the field in the two adjacent cavities. The lengths of the cavities

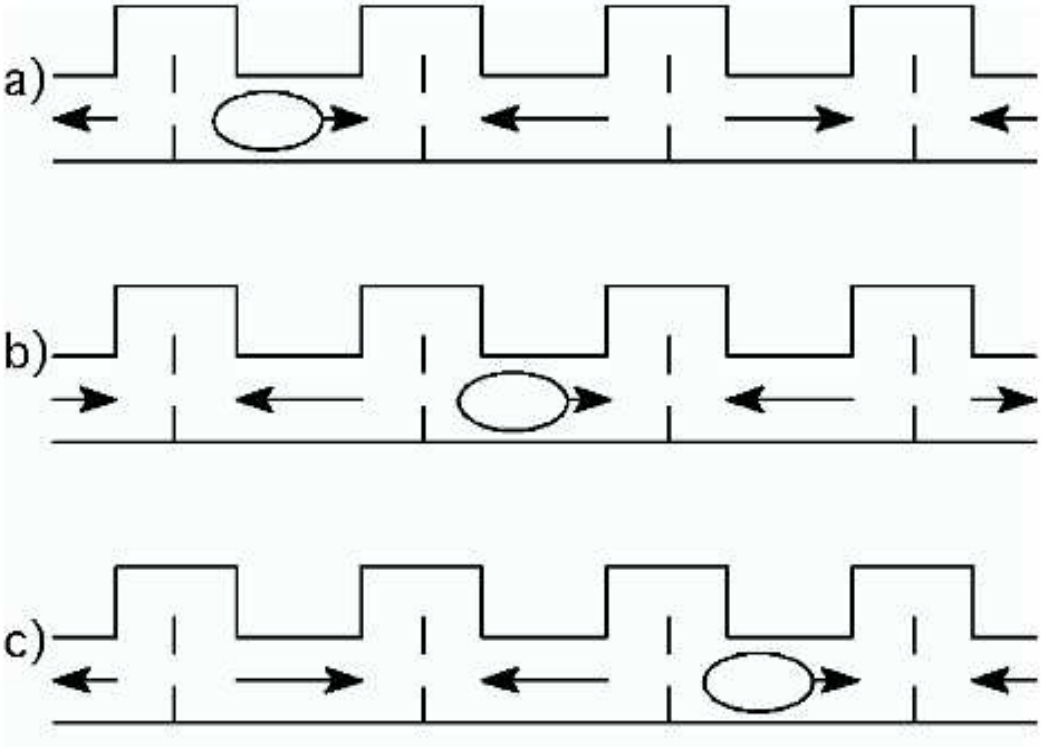

Figure 3.1: Mechanism of acceleration in a linear accelerator. The lengths of the accelerating RF cavities are tuned so that the particle bunch enters a cavity as the field is switching to the right direction. Taken from Ref. [5].

are calculated so that particles are exposed to accelerating field at the center of the cavity but have exit it as the field is switching to the "wrong" direction. As the particles enter the next cavity, the field in it has just switched to the "right" direction (Figure 3.1 illustrates this).

Another result of using sinusoidally oscillating fields is the accumulation of the beam into *bunches*. The variation of the accelerating field in time and space necessarily imparts structure to the beam: particles which are "out of time" with

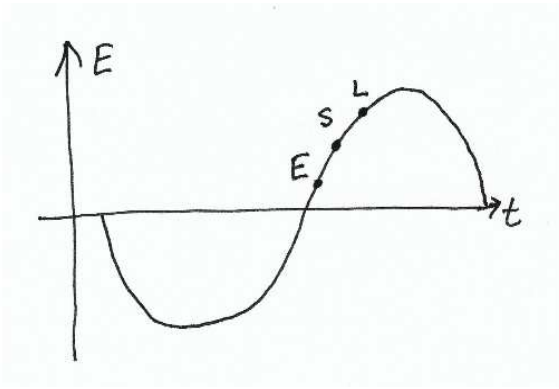

Figure 3.2: Particles entering the accelerating cavity are exposed to different accelerating potentials. Particles entering early are exposed to a weaker field, and those entering earlier are exposed to a stronger field. Those entering the cavity at dramatically different times are accelerated in the wrong direction or not at all. The net effect of this is to group a charged particle beam into bunches.

the accelerating field are lost. There is also a focusing effect on the particle momenta. A particle following the ideal trajectory encounters the field at a phase point where the strength of the accelerating field is increasing (See Figure 3.2). Particles with slightly too much momentum compared to the ideal arrive earlier, when the field is weaker, and are accelerated less. Conversely, particles with less arrive later than those on the ideal trajectory, and are accelerated more. Particles in a bunch oscillate about the ideal trajectory in phase space and are dealt with as a unit.

During transfer to the next stage of acceleration, the beam of H^- ions is sent through a foil to strip the electrons from the ions. The *Booster* accelerates

the resulting proton beam from 400 MeV to 8 GeV. The Booster is the first *synchrotron* in the accelerator sequence. In a linear accelerator, as the particles gain energy, they travel farther during each cycle of the accelerating voltage, so that each successive drift tube or accelerating cavity must be longer than the previous one. To achieve several-GeV proton beam energies, linear accelerators become prohibitively long, making it more cost-effective to bend the beam into a circle and reuse the same accelerating cavities. Dipole magnets with fields perpendicular to the plane of the accelerator ring constrain the beam to a circular path. Reusing the same RF cavities makes it impossible to adapt the length of the cavity to accommodate the increasing velocity of the protons. Instead, we increase the oscillation frequency of which the accelerating field. At the same time, the strength of the dipole magnetic field must be increased to keep the beam on the same path. The name *synchrotron* refers to the synchronization of the RF frequency and bending fields to the increasing beam momentum.

The next accelerator, also a synchrotron, is the *Main Injector*. In addition to accelerating protons to be sent into the Tevatron, it provides beam to fixed target experiments and sends protons to the antiproton source, discussed below. The Main Injector accelerates protons from the Booster to 150 GeV before transferring them into the Tevatron. As the first stage in antiproton production, it accelerates protons to 120 GeV and sends them to the antiproton source.

Antiproton production begins with the 120 GeV protons from the Main Injector striking a nickel target. The collision produces a variety of particles with a broad spectrum of momenta. Magnets separate out antiprotons with 8 GeV/c of momentum. Antiproton production is an inefficient process and the ability to accumulate antiprotons and cool them into a focused beam is the primary limit to the luminosity of the Tevatron. In a series of storage rings collectively named the *Antiproton Source*, the the antiprotons are accumulated and cooled as they are produced. Cooling is the process of reducing the spread of antiproton momenta relative to the central value of 8 GeV/c in the direction of the beam and reducing of any momentum transverse to the beam direction. Before being transferred into the Tevatron, the 8 GeV antiproton beam is transferred into the Main Injector and brought up to 150 GeV for injection.

In the *Tevatron*, the proton and antiproton beams are brought to their final energies of 980 GeV each. Once the final beam energy is attained, the RF structure changes so that the ideal trajectory carries particles through the RF cavities when the accelerating field is zero, but particles with too much or too little momentum are still given the required kick to keep the beam in bunches. The beams are held in the machine until the luminosity has dropped below a useful value or the beams are lost due to the malfunction of an accelerator component, in what is called a *store*. Thirty-six bunches of protons and thirty-six

of antiprotons circulate in in opposite directions, spiraling around each other in a helix, sharing the same magnet system and vacuum beam pipe. Electrostatic separators keep the beam orbits from decaying into each other. Each bunch of protons or antiprotons contains approximately 10^{11} particles. The bunches are grouped into three "trains" of 12 bunches each, with a 296 ns spacing between the bunches within a train, and a large gap between two of the trains to give enough time to ramp up "kicker" magnets to divert beam out of the ring either for an unplanned abort or at the end of a store.

At two interaction points in the ring, the beams are focused down into a point and cross, causing protons and antiprotons to collide. The measure of the quality of the beam, from the perspective of the experiments using the collider data, is the luminosity, which describes the rate at which interactions are produced. It is measured in units of $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ (note the relation to units of area, or cross-section). For colliding proton-antiproton beams, the luminosity is

$$\mathscr{L} = \frac{f \cdot n \cdot N_p \cdot N_{\bar{p}}}{2\pi \cdot (\sigma_p^2 \sigma_{\bar{p}}^2)} \cdot F\left(\frac{\sigma_\ell}{\beta^\star}\right)$$

[36]. The luminosity increases with the revolution frequency f for the beam, the number of bunches n in either beam, and the number of particles in the bunches (N_p and $\cdot N_{\bar{p}}$ for protons and antiprotons). At the Tevatron, f is about 47.6 kHz (the RF frequency is 53 MHz, and there are 1113 RF buckets, or

potential locations for particle bunches, around the ring). n is thirty-six, N_p is typically about 3×10^{11}, and $N_{\bar{p}}$ about 7.5×10^{10} at the beginning of a store. The luminosity decreases as the area of the interacting beams increases. Here, we express the area in terms of the Gaussian widths σ_p and $\sigma_{\bar{p}}$ of the proton and antiproton beams at the interaction points. The beam spot size is about 30 μm at the interaction point [6].

The form factor F is approximately 0.7 for the Tevatron and depends on the bunch length σ_ℓ (.37 m) and the value $\beta^\star$ (.35) of the amplitude function at the interaction point. The amplitude function has units of length and gives the wavelength of oscillations of particles in the beam about the ideal orbit, but the spatial extent of the beam also scales according to it. Typical instantaneous luminosities for the beginning of a store are currently in excess of 2×10^{32} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

Also relevant for the experimentalist is the distribution of interactions in time and space. At CDF, the luminous region is only tens of microns in radius, corresponding to the beam size, but extends well over a meter along the beam direction (z for the detector). The interactions are distributed in an approximately Gaussian distribution with a width of 40 cm. The length is in the z direction is a result of the length of the bunches, combined with the shallow crossing angle which is a consequence of the distance of the final focusing

magnets from the center of the detector. The extent of the interaction region motivates the extent of the tracking detectors in the beam direction. There are also typically multiple interactions per beam crossing, a number which increases with increasing instantaneous luminosity. Even though two interactions producing particles with enough p_T that they are both interesting is so rare as to be negligible (for now), this must be taken into account in analysis of data, since the additional energy in the calorimeter affects its energy resolution, and the $\not\!\!E_T$resolution in particular.

3.2 The CDF Detector

The CDF detector encloses one of the two interaction points at the Tevatron with layers of different types of detectors. Figure 3.3 shows the geometry of the whole detector; we describe the relevant subsystems in the following sections. The detector is approximately azimuthally symmetric, to reflect the symmetry in the interactions. It is also divided into "central" and "forward" regions, where the central part is the cylindrical portion encircling the interaction point, and the forward part, also referred to as the endcap or the plug, forms the ends of the cylinder. The central part of the detector captures event products emitted from the interaction point very close to transverse to the beam, with

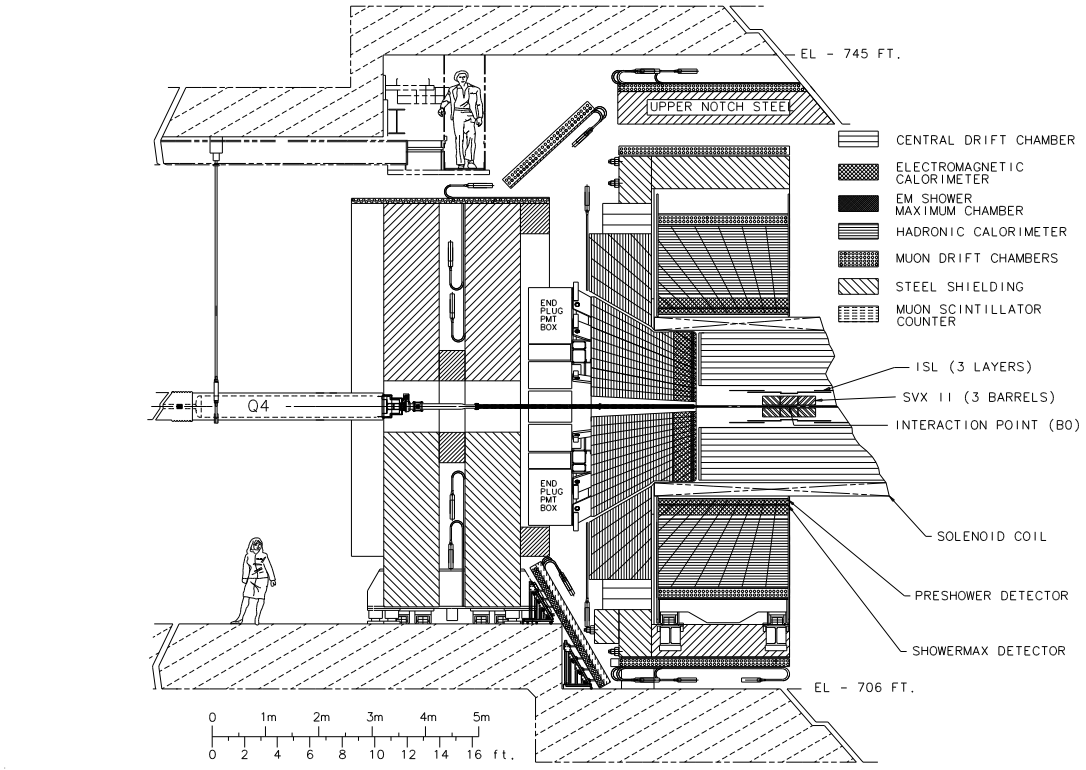

Figure 3.3: Geometry of the CDF detector.

pseudorapidity $|\eta| \lesssim 1.0$.[1] This includes a large fraction of the decay products of heavy particles such as the W, the Z, or the top quark.

[1]Pseudorapidity ($\eta = -\ln(\tan(\theta/2))$) is used as a convenient alternative to the standard polar angle, because it reflects the forward-backward symmetry of the detector, and because the distribution of particles in jets and the typical size of jets is approximately constant in this variable.

3.2.1 Tracking

Particles emerging with significant transverse momentum from the interaction point first cross the two tracking systems, the silicon semiconductor tracker and the Central Outer Tracker (COT), an open-cell wire drift chamber.

Silicon Tracking

The active part of the silicon tracker is a large number of silicon wafers. The bulk of each wafer is doped to be n-type, meaning that there is an excess of electrons in the semiconductor lattice. On one surface of the wafer are implanted lines of material with the opposite doping (p-type, an excess of holes). There are also stereo implants on the reverse side of some of the layers, which are of the same type as the bulk, but with a higher carrier concentration. These are oriented at an angle relative to the p-type implants, to provide z information when combined with information from the $r - \varphi$ side. For operation, a voltage (30-500V depending on the sensor) is placed across the wafer, generating an electric field perpendicular to the large flat surface and pointed from the p side to the n side. This "reverse bias" extracts the charge carriers from the silicon. Charged particles crossing a layer of the silicon tracker generate new electron-hole pairs, which drift apart in the electric field, inducing a signal in one or more of the implants on the surface. The implants are capacitively coupled to

metal strips through a silicon oxide insulating layer. The metal strips overlaid on the implants are wirebonded to a 128 channel readout chip, which amplifies and integrates the electrical signal.

The CDF silicon tracking detector in Run II consists of seven concentric measurement layers in the central region and eight layers in the forward. The detector is designed in three subsystems (Figure 3.4 shows the geometry of the silicon detectors in an r-z view).

The main part of the detector, SVX II, has five measurement layers, all of which are double-sided, with implants either at 90° or 1.2° degrees relative to the implants on the top, to provide information on the polar angle of particle tracks. Layer 00, single-sided wafers installed directly on the beam pipe, adds precision to the impact parameter measurement because of its proximity to the interactions. The Intermediate Silicon Layers (ISL) are at larger radii compared to the rest of the silicon. In the central region, a single layer provides information that can ease the connection of tracks between the silicon and the COT. In the forward region, where there are two layers, the ISL contributes to the momentum resolution and efficiency for silicon-only tracking, in the angular region where COT efficiency is falling off rapidly.

One advantage of silicon tracking, especially pertinent in the conditions of high instantaneous luminosity now present at the Tevatron, is that it can be

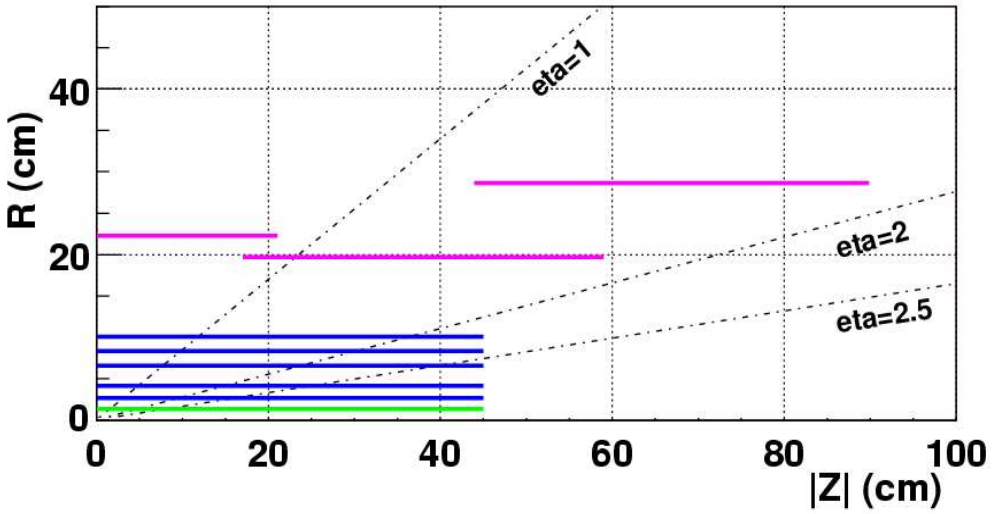

Figure 3.4: R-z view of the silicon tracking system geometry. The detector is symmetric around $z = 0$. Layer 00, the innermost detector, is in green. The SVX II, in blue, is the five layers above it. The ISL are the layers in magenta at large radius.

made relatively resistant to radiation damage. We can maintain detector performance in high-radiation environments by reading out the signal capacitively and, in the case of Layer 00, designing devices to tolerate high bias voltages without breakdown. Ionizing radiation damages the semiconductor lattice, creating an excess of potential available charge carriers relative to the original doping. This generates a "dark current" in the device which has both DC and AC components. Reading out the implants capacitively, through a signal induced across a layer of insulator, protects the signal from the DC component of the dark noise. As sensors are damaged, they become harder to deplete, so tolerance for a higher reverse bias voltage allows the device to be fully depleted after more exposure to radiation. Layer 00 is closest to the beam pipe, it will be exposed to the greatest radiation field of all the tracking detectors Using technology that will also be used at the Large Hadron Collider, the Layer 00 sensors can tolerate bias voltages up to 500 VDC, compared to a maximum of 120V for the other sensors in the detector.

The other main motivation for silicon tracking is the improved resolution on the position of the particle as it passes through the detector. Placed close to the interaction region, silicon detectors improve the resolution on the track origin, enough that the decay vertices from long lived particles can be distinguished from vertex of the primary interaction. Figure 3.5 compares the impact parame-

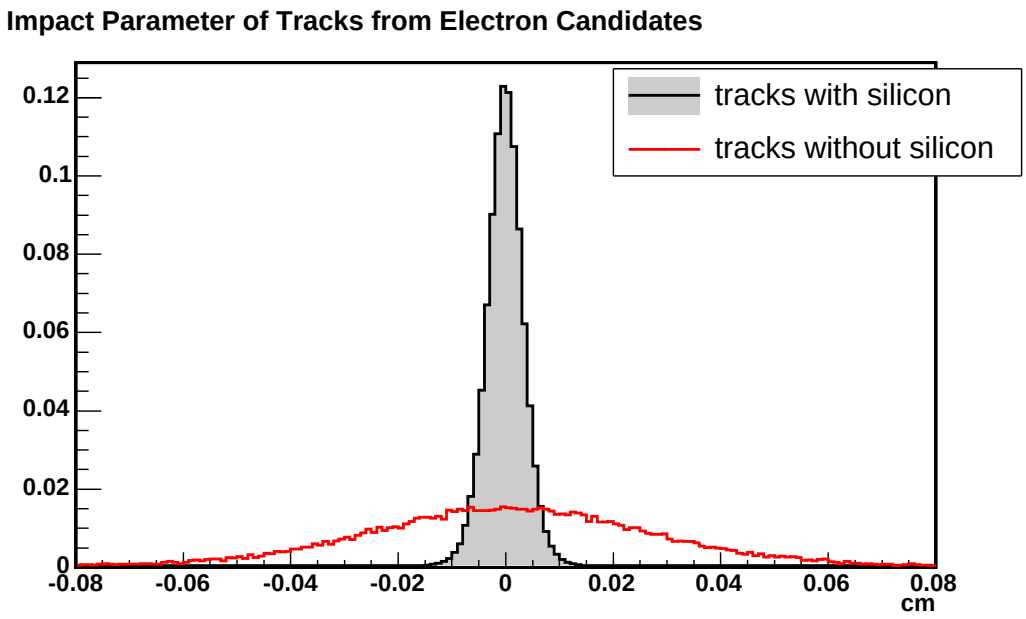

Figure 3.5: Impact parameter (distance of closest approach to the beamline) for tracks with and without silicon. Tracks are from fully-reconstructed central electron W candidates in data, which have an electron with $E_T > 20$ GeV and > 25 GeV of $\not\!\!E_T$.

ter distribution, or measured point of closest approach to the beamline, between tracks with and without silicon. The impact parameter resolution for this set of tracks is $\approx 30\mu$m, an order of magnitude smaller than for tracks without silicon. That makes it possible to reconstruct the displaced vertices of particles like b mesons from top quark decays, which typically travel 100 microns or more in the detector before decaying.

The Central Outer Tracker (COT)

The backbone of particle tracking at CDF is the Central Outer Tracker, an open-cell wire drift chamber which covers the central ($|\eta| \lesssim 1.0$) region and extends from a radius of 44 to 132 cm [37]. Figure 3.6 shows the geometric coverage of the COT in the context of the geometry of the rest of the detector, including the silicon tracker. The chamber is filled with an argon-ethane mixture, and charged particles passing through the chamber ionize the gas. A strong electric field causes them to drift toward wires strung through the chamber, inducing an electrical pulse on the wire. Both the position and timing of the pulse are recorded, as a "hit" on that wire. The COT consists of eight "superlayers", each of which has twelve such wires. Every other layer has its wires at a 2 degree angle relative to the wires in the surrounding layers, which are parallel to the beam direction. These angled layers, called the *stereo* layers, allow reconstruction of the polar as well as azimuthal angle (η and φ) of the particle track. We refer to the layers with wires parallel to the beam as the *axial* layers.

Tracks from the COT are a core requirement for the identification of electrons and muons, Quality requirements on the tracks can be made in terms of the minimum number of hits required for axial and stereo wires, or in terms of a

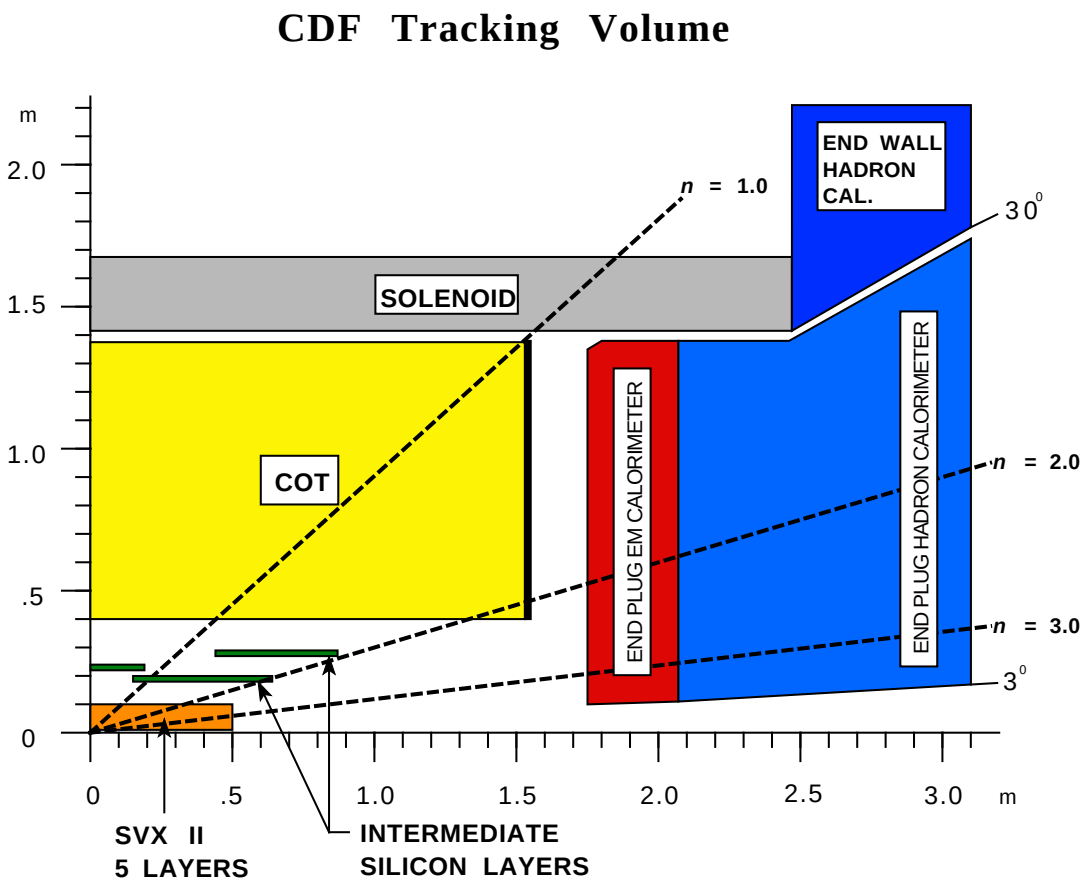

Figure 3.6: Geometry of the tracking system, including the COT and the silicon tracking detectors, with the solenoid and endplug calorimeter shown as well. The central calorimeter begins at a radius just beyond the solenoid. The detector is symmetric about $z = 0$ and is approximately azimuthally symmetric.

minimum number of segments, where a segment is a set of hits on a particular superlayer and is required to have some minimum number of hits.

The momentum of charged particles can be measured from their tracks because the entire tracking system is contained in the 1.4 Tesla magnetic field of a superconducting solenoid. The location of the solenoid is also shown in Figure 3.6. The solenoid generates an approximately constant magnetic field pointing in the z direction, coaxial with the beam. Particles with opposite charge bend in opposite directions in the plane transverse to the beam direction. The radius of curvature is directly proportional to their momentum in that plane: $\rho(\mathrm{m}) = \mathrm{p_T(GeV/c)}/(0.3 \cdot \mathrm{B(T)})$. This has the consequence that the resolution of the momentum measurement decreases with increasing momentum, as tracks approach a straight line. The resolution of COT tracks is $\sigma_{p_T}/p_t^2 \approx 1.7 \times 10^{-3}/GeVc^{-1}$, where the p_T^2 makes explicit the degradation of the resolution with increasing p_T[38].

3.2.2 Calorimetry

The calorimeter surrounds the tracking detector, down to $|\eta| = 3.6$, 3 degrees away from the beam, and contains enough material to stop essentially all particles coming from the interaction point, measuring their energy in the process. It is divided into central and plug regions, which have nearly identical

composition. The central calorimeter covers $|\eta| \lesssim 1.0$ and is cylindrical; the plug calorimeter covers $1.1 \lesssim |\eta| \lesssim 3.6$ and is composed of two identical pieces shaped like truncated cones, which fit into the ends of the central calorimeter (hence the name "plug"). Figure 3.3 shows the geometry of the calorimeters. The calorimeter is segmented in towers, in what is called a "projective geometry": the boundaries between towers are at a fixed angle relative to the center point of the detector, and the towers point at the center. The towers in the central calorimeter subtend 15 degrees in φ and 0.11 units of pseudorapidity. In the plug calorimeter, the tower size is more variable. For $|\eta| < 2.1$, the size in φ is 7.5 degrees, for the towers at higher rapidity it is 15 degrees. Because of the compression of the pseudorapidity relative to the physical polar angle θ as the angle approaches the beamline, the towers start out at a size 0.1 in η at $|\eta| = 1.1$, increase to about 0.2 at $|\eta| \approx 2.0$, and are over 0.5 at the end of the calorimeter at $|\eta| = 3.6$ [39].

Electromagnetic Calorimeter

High-energy electrons lose most of their energy through *bremsstrahlung*, or the emission of a photon carrying a fraction of the energy of the electron. High energy photons lose most of their energy through the production of positron-electron pairs. Both of these effects take place under then influence of the

nuclear electric fields of the atoms in material. The photons emitted by an energetic electron may in turn pair produce, and the electrons and positrons from pair production may then emit photons. The process continues until the electron energies have fallen below the *critical energy*, the energy at which they lose more energy by ionization than by radiation. The cascade of photons and electron-positron pairs originating from a single electron or photon is called a *shower*.

The *radiation length* (X_0) is defined as the mean distance over which an electron will lose all but $1/e$ of its energy to photon radiation, and is also the 7/9 of the mean free path of a photon before it pair produces. The size of a shower in the direction transverse to the direction of the original particle also scales with the radiation length. The *Molière Radius* R_M describes a circle which contains 90% of the particle energy. It depends linearly on X_0 and, up to some dependence on the material, is of the same order of magnitude ($R_M \approx 21 X_0/(E_c(\text{MeV}))$, where E_c is the critical energy, typically in the tens of MeV range) [6].

The radiation length scales with the inverse of the square of the atomic number, or nuclear charge, of a material [6]. This motivates the choice of lead, which has large atomic number and therefore a short radiation length (.56 cm), for the passive material in the CDF electromagnetic calorimeter. The

electromagnetic calorimeter is the innermost layer of the calorimeter and is intended to stop all electrons and photons. The central (plug) electromagnetic calorimeter is 19 (20) interaction lengths deep, to ensure that the majority of electromagnetic showers will be fully contained [39].

To measure particle energies, layers of scintillator are interleaved with the lead. Showering particles generate scintillation light, allowing us to record the evolution of the shower. The amount of scintillation light depends on the energy deposited, so that the energy of the particle can be reconstructed. The energy resolution of the central (plug) electromagnetic calorimeter is $14\%/\sqrt{E_T} \oplus 1\%$ ($16\%/\sqrt{E_T} \oplus 1\%$) [38].

Shower Maximum Detectors

The number of particles in an electromagnetic shower peaks at about 6 radiation lengths. At about this depth, a thin layer of more precise position detector is placed as an extra layer in the calorimeter, to measure the spatial location of the particles in the shower. This serves two purposes. First, the typical transverse size of an electromagnetic shower is smaller than the size of a tower ($R_M = 1.6$ cm for lead, towers are several cm on a side), so it allows measurement of the location where a particle struck the calorimeter. Second, the shape of a shower is an important discriminant between true electrons or

photons and hadronic particles, so the more detailed picture of the lateral shape of the shower is useful.

In the central calorimeter, the shower maximum detectors are gas trackers which work by the same principle as the COT. Two orthogonal layers, one of wires and one of metal strips, provide position measurements. In the plug calorimeter, strips of scintillator which cross at 45 degrees perform this function.

Hadronic Calorimeter

Outside of the electromagnetic calorimeter is the hadronic calorimeter, alternating layers of iron absorber and scintillator. The electromagnetic calorimeter is only one nuclear interaction length deep, so the showers of hadrons are only beginning to develop. Hadrons, both charged and neutral, lose energy through ionization and collisions with the nuclei of the material through which they pass. These particles in turn can interact, developing into a shower. Unlike the electromagnetically-driven interactions of electrons and photons in matter, these interactions do not depend on the nuclear charge. They depend instead the atomic number of the material, through the cross-sectional area of the nucleus. Typical nuclear interaction lengths are therefore much longer than electromagnetic interaction lengths, and depend more weakly on the material type. For this reason, there is more freedom in the choice of material. We use iron in the

the CDF calorimeter because it can also serve as a flux return for the solenoid.

The hadronic calorimeter is 4.5 interaction lengths deep in the central region and 7 deep in the plug. The energy resolution for the central hadronic calorimeter is $75\%/\sqrt{E_T} \oplus 5\%$; for the forward ones it is $80\%/\sqrt{E_T} \oplus 5\%$ [38].

3.2.3 Muon Detectors

Muons do not shower like electrons, due to their much greater mass, nor do they undergo nuclear interactions like hadrons. Rather, the dominant mechanism for energy loss by muons is ionization. Muons used for analysis of electroweak phenomena at CDF are typically in the energy range 20-200 GeV, near the minimum of the Bethe-Bloch curve for energy loss due to ionization. They exit the detector, having left only a small amount of energy in the calorimeter. To catch the unique signature of the only (to a good approximation) charged particle to exit the inner part of the detector, detectors are placed outside all of the calorimeter and, in most cases, outside additional shielding.

For the purposes of this analysis, there are three sections of muon detectors, distinguished by their geometry, but all built on the same model. The *Central MUon* (CMU) detectors lie directly outside the hadron calorimeter and cover the range $|\eta| \lesssim 0.6$. One difficulty in using the CMU is that the material in front of it, which consists mostly of the central hadron calorimeter,

is only 5.5 nuclear interaction lengths deep, so that charged particles that are not muons do occasionally escape the calorimeter to be detected in the CMU. This motivated the construction of the *Central Muon uPgrade* (CMP), which has approximately the same angular coverage as the CMU, except that there is additional steel shielding between it and the CMU. The *Central Muon eXtension* (CMX) detectors add coverage in the angular range $0.6 \lesssim |\eta| \lesssim 1.0$. Each of these detectors can be seen in Figure 3.3: the CMU is directly outside the central hadron calorimeter, the CMP is just outside of it, and the CMX is at an angle relative to the beamline.

The muon detectors all consist of four layers of rectangular cells. Each cell is a drift tube, with a single wire in the center. The wires run parallel, or approximately parallel, with the beam direction, so that they record the φ position of the particle crossing them. In the CMU, charge is read out from both ends of the tube to give local z (and therefore polar angle) information through charge division. In the CMX, polar angle information is available because some of the layers of drift tubes are at a small angle relative to the others.

One or more layers of scintillator are laid on one or both sides of the layers of drift tubes. These provide timing information, which allows us to associate muons with a particular interaction, since the drift time in the chambers is longer than the time between beam crossings. This is used to reject "out-of-

time" muons, which may be cosmic rays.

Hits from the layers of wires are connected into short tracks, or "stubs". The association of a stub to a track from the main tracking detectors are the basis of muon identification at CDF.

3.2.4 Luminosity Measurement

Measurement of the colliding beam luminosity is a key input to cross-section measurements: it quantifies how much data are in the sample. Luminosity is measured at CDF by a pair of Cerenkov detectors placed at large pseudorapidity (small angle to the beam) on each side of the interaction point [40, 41]. Each conical detector surrounds the beam pipe, and contains 48 smaller mylar cones which, filled with atmospheric-pressure isobutane, collect Cerenkov light from particles emerging from inelastic $p\bar{p}$ scattering.

The cross-section of inelastic $p\bar{p}$ scattering is well-known (to about 3%) [41]. Therefore by measuring the rate of this interaction the instantaneous luminosity can be inferred:

$$\mu f_{\mathrm{BC}} = \sigma_{\mathrm{in}} \mathscr{L}$$

In the above, σ_{in} is the inelastic $p\bar{p}$ cross-section, f_{BC} is the bunch-crossing frequency at the Tevatron (1.715 MHz, corresponding to 36 bunches and a

revolution frequency of 47.6 kHz), $\mathscr{L}$ is the instantaneous luminosity, and μ is the average number of $p\bar{p}$ interactions per bunch crossing. The luminosity detector measures μ to infer $\mathscr{L}$, by counting the number of beam crossings where *no* such interaction is observed. The number of $p\bar{p}$ interactions follows a Poisson distribution, so that by measuring the probability of no interaction occurring, the mean μ of the distribution can be inferred. Specifically, the probability of N interactions occurring is $\mathcal{P}(N,\mu) = e^{-\mu}\mu^N/N!$, so the probability of no interactions is $\mathcal{P}_0 = e^{-\mu}$. The uncertainties on the luminosity are from the understanding of the acceptance for the detectors for $p\bar{p}$ inelastic scattering as well as the 3% uncertainty on the value of σ_{in}, and combine to 6%.

3.3 Event Triggers and Data Acquisition

With a 2.5 MHz nominal bunch crossing-rate, corresponding to the 296 ns bunch spacing, even inelastic collisions occur at much higher rate than the rate at which events can possibly be recorded. A "trigger" system, consisting of three stages, performs fast event reconstruction to determine which events are of interest and should be recorded. A list of criteria defines what is considered interesting, and includes a variety of event classes: events with multiple high-p_T jets, events with one or more leptons, and so forth.

3.3.1 The CDF Trigger System

Level 1

The Level 1 trigger uses calorimeter, track, and muon detector information to decide, inside the 296 ns bunch crossing window, whether to keep or discard a given event [38]. Custom hardware quickly reconstructs simple event objects. Calorimeter information available at this stage is the number of towers with energy above a given threshold, the summed E_T over all towers, and the $\not\!\!E_T$. Tracking is done using only axial information from the COT, comparing hit information to a set of predefined patterns; the output information is the p_T and φ of the track. The tracks are also extrapolated to the calorimeter and muon systems for use in electron and muon triggers. Track segments from the drift tubes and scintillator information are both used for muon identification at this level. Level 1 accepts events at a rate of up to 50 kHz.

Level 2

The Level 2 trigger considers only events passing the Level 1 trigger. Also implemented at the hardware level, it accepts events at a maximum rate of 300 Hz. Decisions are made in 20 μs, so more reconstruction is possible. In particular, simple clustering of calorimeter towers is performed, for photon,

electron, and jet identification. Information from the shower maximum and silicon detectors is also available at this stage.

Level 3

At Level 3, full event reconstruction is done at the software level. Events passing the Level 2 trigger are sent to one of 300 commercial dual-processor Linux machines. Events can written to tape at about 75 Hz, so about 0.4 seconds are available to make Level 3 decisions. This time window allows full event reconstruction, including the application of calibration constants, using the same physics object definitions that are used offline for analysis. Because Level 3 reconstruction are nearly identical to final analysis reconstruction, more stringent cuts can be made for better background rejection while maintaining efficiency for signal. Selected events are written to tape for analysis.

3.3.2 Triggers Used in this Analysis

Trigger definitions are refined over time to reflect changing operating conditions, particularly the increasing instantaneous luminosity seen at the beginning of stores, but the high-p_T lepton triggers used in this analysis have, for the range of data used here, been relatively stable.

Central high-E_T Electrons

At Level 1 this trigger requires a tower in the central calorimeter to have at least 8 GeV of E_T, a ratio of hadronic to electromagnetic energy less than 0.125, and a matched track with $p_T > 8.3$ GeV/c. At Level 2 the calorimeter requirements are defined in terms of a cluster, which must be seeded by an > 8 GeV tower and have $E_T > 16$ GeV. At Level 3 the electron candidate is built using same electromagnetic clustering algorithm that is used offline, and the E_T required to be at least 18 GeV. The cluster must pass also some minimal shower shape requirements and be well-matched to a track with $p_T > 9$ GeV/c.

Forward high-E_T Electrons

Forward electrons trigger objects have a higher fake rate than central ones since they cannot be matched to Level 1 tracks. The main plug electron trigger also requires $\not\!\!E_T$ in the event to reduce background, preferentially choosing electrons from W decays, which includes W decays from top events. At Level 1, a cluster with $E_T > 8$ GeV and less than .125 of its energy in the hadronic calorimeter is required, analogous to the case for central electrons. But there is the additional criteria that the $\not\!\!E_T$ calculated using all towers with 1 GeV or more be greater than 15 GeV. As with the central electron trigger, Level 2 refines the requirement on the electron object, in this case by requiring a cluster

with $E_T > 20$ GeV. At Level 3 the electron E_T threshold is 20 GeV, and the reconstructed $\not\!\!E_T$is still required to be > 15 GeV.

Central high-p_T Muons

The CMUP trigger path starts at Level 1 with the requirement of stubs in both the CMU and CMP, which are matched to a track with $p_T > 4$ GeV/c. The only additional requirement at Level 2 is that the track have $p_T > 8$ GeV/c. At Level 3, the muon stub is required to be matched to a fully reconstructed track with $p_T > 18$ GeV/c. The CMX trigger requirements are identical except that the stub is in the CMX and the p_T thresholds at Level 1 and Level 2 are higher: 8 and 10 GeV/c, respectively.

3.4 Summary

The Tevatron accelerator produces proton-antiproton collisions at an energy of 1.96 TeV. The CDF detector has been assembled to study the products of these collisions. By surrounding a large solid angle around the interaction point, the detector catches the decay products from energetic inelastic scattering events. A combination of calorimeter and tracking detectors provide position measurements of the product particles. Used in the magnetic field of a super-

conducting solenoid, the inner tracking detectors, a combination of solid-state silicon and wire chamber trackers, provide momentum and charge measurement as well. Calorimeters absorb almost all outgoing particles and measure their energy. Muons are flagged by specialized tracking detectors that lie outside of the rest of the detector. Preliminary event reconstruction, called a trigger, is used to select the subset of events which are of interest for analysis. For this measurement, events which appear to have high-p_T leptons are selected and written to tape, so that they can be analyzed for the presence of top quark events.

Chapter 4

Event Selection

The aim of the event selection is to identify $t\bar{t}$ events while keeping contributions from other types of processes as small as possible. This measurement uses the dilepton subset of $t\bar{t}$ events, where both the Ws from the top quark decay to a charged lepton and a neutrino. Therefore the base of the event selection is a pair of opposite-sign leptons and missing transverse energy ($\not\!\!E_T$). The decay of each top quark also produces a b quark, which hadronizes and is detected as a jet, so top dilepton events typically have two or more jets. Because of the large mass of the top quark, each final state particle is typically produced with significant transverse momentum (p_T). We make additional event selection restricting the allowed relationships between these objects. in order to reduce background acceptance, particularly for Drell-Yan ($Z/\gamma^\star \to \ell\ell$) events. It is

worthwhile to try to reduce this background since estimating its contribution is difficult (see Chapter 6).

4.1 Lepton Definitions

The requirement of two charged leptons forms the base of the event selection. One lepton is identified clearly as an electron or a muon, using criteria chosen to minimize background from jets. These definitions make use of information in the tracking, calorimeter, and muon detectors. We cannot use tau leptons in the same way because they decay too quickly and into a variety of decay products, both leptons and hadrons (see Figure 4.1). Taus that have decayed into an electron or muon are reconstructed as such, and taus that decay hadronically are difficult to distinguish from jets, aside from their typically lower track multiplicity.

The second lepton, also referred to as the "track lepton", is identified only as an isolated track. This minimal selection increases acceptance for dilepton events. The track lepton definition does not refer to any detectors other than the tracker, so that the limited geometric acceptance of the calorimeter and muon detectors does not limit its efficiency. Also, the background-reducing criteria applied to the primary lepton types are only 80-90% efficient, so a minimal

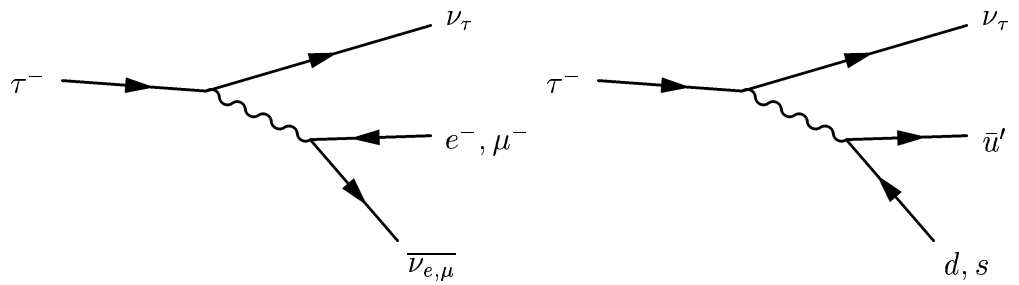

Figure 4.1: Kinematically allowed decay modes of the tau lepton. The τ is more massive than the estimated mass of the c quark, but no charmed meson is lighter than the τ, so $\bar{c}s$ and $\bar{c}d$ decays are not allowed.

selection recovers much of that efficiency as well.

Most significantly, this definition includes a large subset of events where one or both Ws have decayed to a τ and ν_τ. Because the W is equally likely to decay to any one of the three generations of $\ell + \nu$ pairs, 5/9 of top events where both Ws decay to leptons contain one or two tau leptons. Any dilepton selection will accept leptonic τ decay products, which are 35% of all decays, although there will be some inefficiency because some of the momentum of the original tau decay will be lost to the two neutrinos produced. Including both leptonic and hadronic decay modes, 85% of decays have just a single charged track in the final state [6]. Using an isolated track for the second lepton therefore significantly increases the acceptance. About 20% of the total acceptance with our selection is from events with one or two τs. Table 4.1 shows how the acceptance is divided

	ee	$e\mu$	$\mu\mu$	$e\tau$	$\mu\tau$	$\tau\tau$	total
electron	20.0	34.0	0.0	8.9	2.6	0.5	66.2
muon	0.0	7.3	19.3	0.6	6.3	0.5	33.8
total	20.0	41.3	19.3	9.5	8.9	1.0	100.0

Table 4.1: Fraction (in %) of the (opposite-sign, $\geq$ 2 jet) acceptance that is attributable to each possible generated lepton pair. The first two rows show the events divided according to whether they were reconstructed with an electron or muon as the primary lepton. The majority of events are accepted as electron plus track because there is more geometric acceptance for electrons, and because of way in which events are categorized when both leptons can be fully reconstructed.

between the different generated and accepted lepton types.

4.1.1 Electron Selection

Electrons leave a track in the detector and deposit essentially all of their energy in the electromagnetic calorimeter. Electron selection is therefore based on a track which points at a deposit in the electromagnetic calorimeter. This analysis uses two classes of electrons. Central electrons, typically in the range $|\eta| \lesssim 1.1$, have tracks in the central tracker (COT) and deposit their energy in the central electromagnetic calorimeter (CEM). Plug electrons are the range $1.2 \lesssim |\eta| \lesssim 2.0$, have tracks based on information from the silicon tracker, which extends farther forward than the central tracker, and deposit their energy in the plug electromagnetic calorimeter (PEM).

The two classes of electrons are identified using the same principles, applied

to the parts of the detector available in the two regions. The energy deposit in the calorimeter from an electron should be isolated from other energy in the event. Also, the amount of energy deposited in the electromagnetic calorimeter should be very large compared to the amount in the hadronic calorimeter, since the electromagnetic calorimeter is many interaction lengths deep. The transverse shape of the shower can be measured by the the shower maximum detectors. These characteristics distinguish the narrow, shallow showers of the single, light electrons from the broad, deep showers produced by the hadrons in a jet (See Chapter 3 Section 3.2.2 for more discussion of electromagnetic and hadronic shower development and calorimeter design).

Central Electrons

Kinematic and Geometric requirements The transverse energy of the cluster must be at least 20 GeV, and the momentum of the track at least 10 GeV/c. In addition, the ratio of the cluster E_T to the track p_T must be less than 2.0 unless the track p_T is greater than 50 GeV/c. The looser requirements on the track p_T are to account for the emission of photons (*bremsstrahlung*) by the electron as it passes through the material of the tracker. This reduces the measured track p_T, but with the exception of the rare case where the electron emits a photon carrying a significant fraction of its momentum, the photon

showers overlap with the electron shower and the energy of the cluster is still a good measure of the total energy of the electron. Also, the electron cluster must be within the fiducial volume of the CEM.

Track quality requirements The track is required to have at least 3 axial segments and 2 stereo segments, each consisting of at least 5 hits in the COT, and the track $|z_0|$ must be less than 60 cm. The track Z_0 requirement is designed to incorporate as much of the interaction region as possible (See Chapter 3 Section 3.1) while keeping the track fiducial in the tracking detectors. A track segment is a set of hits in the same superlayer of the COT (see Chapter 3 Section 3.2.1). The quality requirements reject fake tracks and ensure that there is enough information in the track to make a reliable p_T measurement.

Track-cluster matching Matching between the track and the electromagnetic cluster is done by extrapolating the track to the shower maximum detector. The distance between the point at which the extrapolated track crosses the plane of the shower maximum detector and the measured shower centroid should be small to consider the track and cluster a good match. The specific requirement is that the distance be less than 3 cm in local Z (along the direction of the beam) and that the track charge times the distance in local X (in the plane, transverse to the beam direction) satisfy -3.0 cm $\leq Q \cdot \Delta X \leq$ 1.5 cm.

The asymmetry in the $Q \cdot \Delta X$ requirement is due to the bending of the track in the magnetic field. The matching requirement both ensures that the electron object is well defined and rejects background from photons and $\pi^0 \to \gamma\gamma$ where a charged track happens to align with a photon cluster.

Shower Profile and Isolation Requirements on the shape of the shower are powerful tools for the rejection of jets. An isolated electron has very little energy in the area of the calorimeter immediately surrounding the electron cluster (Recall the size of the Molière radius in lead compared to the tower size from Chapter 3). Specifically, the total energy in the towers immediately surrounding the tower containing the electron shower is required to be less than 10% of the electron energy. The distribution of energy between the towers in the cluster is also required to be consistent with what is expected from an electron, and the shape of the shower in the shower maximum detector is required to be consistent with expectation. Finally, the amount of energy deposited in the hadronic part of the central calorimeter must be less than 5.5% of the amount deposited in the electromagnetic part, with a small correction allowing for the fact that the more energetic the electron, the more the shower is expected to extend into the hadronic calorimeter. The shower extends further because more energy is available for particle production before the all resulting particles cross below

the critical energy and the shower stops.

Conversion veto A source of electrons other than the decay of a W or Z is the conversion of a photon to a pair of opposite-sign electrons. If a second track can be found, close to the electron track in θ, with the opposite-sign, and which appears to originate from the same point, the electron is flagged as a conversion and not used.

Plug Electrons

Kinematic and Geometric requirements To be considered, an electromagnetic cluster must have $E_T > 20$ GeV and be within the fiducial region of the plug calorimeters and shower maximum detectors, $|\eta| > 1.2$. The upper $|\eta|$ bound, 2.0, is dictated not by the physical limits of the calorimeter but by analysis considerations. Because of the large mass of the $t\bar{t}$ system, its decay products are mostly central, and the distribution of leptons from top events falls off rapidly with pseudorapidity, but electron fake rates increase steadily as a function of η (see Figures 6.5 and 6.6 in Chapter 6). There is no additional p_T cut on the track since the track p_T is not independent from the E_T of the calorimeter cluster, because of the tracking algorithm used.

Silicon track reconstruction The algorithm used to reconstruct silicon tracks for the plug electrons is different from the COT-dependent tracking algorithms used for central tracks, because the COT acceptance is rapidly running out as a function of $|\eta|$. Instead, we reconstruct tracks using hits in the silicon and information from the electromagnetic calorimeter. The calorimeter cluster provides an energy measurement, specifying the curvature of the track, and the event vertex and shower maximum centroid provide two position measurements. This information specifies a helix, or "seed track", up to the choice of a sign. Silicon hits can be attached to the helix, and the track helix can be refit to using the hit information, which are far more precise position measurements than the original two points. Finally, one of the two track hypotheses is chosen based on which has more hits attached and which one is a better fit for the attached hits. This method finds tracks with good efficiency, extending as far in pseudorapidity as the silicon detector coverage, out to $|\eta| \simeq 2.5$. This tracking algorithm is referred to as "Phoenix" tracking and is the source of the abbreviation for this category of electrons: PHX.

Track quality requirements To reduce background for plug electrons, the most important requirement is the existence of the track itself. Tracks for plug electrons are found according to the algorithm described in Chapter 3

Section 3.2.1. The algorithm has a built-in quality requirement for tracks in that they must have at least three attached silicon hits (out of 6 or 7 possible hits depending on pseudorapidity). In addition, as with central tracks, the track $|z_0|$ must be less than 60 cm.

Shower profile and isolation Plug electrons meet the same calorimeter-based isolation requirement as central electrons. The total sum of the E_T in the towers surrounding the calorimeter cluster must be less than 10% of the E_T of the cluster. The energy deposited in the hadronic part of the plug calorimeter must be less than 5% of the amount of energy in the electromagnetic part. The shower profile requirements differ from those for central electrons because the calorimeter has different geometry and a different shower maximum detector. The cluster of calorimeter towers must be well-defined and have a $\chi^2 < 10$ compared to the expected cluster profile for electrons. The energy deposited in each of the two shower maximum layers must be concentrated in the middle 5 strips of the 9 used in the cluster, corresponding to a narrow shower (Compare the strip width of 5 mm to the 16 mm Molière radius in lead) or Also, the centroid of the clusters made from the calorimeter towers and in the shower maximum detectors must be close to each other.

4.1.2 Muon Selection

Muons are minimum ionizing particles, so they do not interact before exiting the detector, other than by ionizing the material through which they pass. Therefore an important criteria for identifying muons is the presence of ionization in detectors placed outside the calorimeter, and typically outside additional shielding, where no other charged particles should reach. A fully-reconstructed muon is a track which points at a track segment ("stub") in the muon detectors. This analysis uses two classes of fully-reconstructed muons, differentiated by which section of the muon detectors contains the track stub. CMUP muons have stubs in each of the two most central muon detectors, the CMU and CMP, which cover $|\eta| \lesssim 0.6$. CMX muons have a stub in the more forward muon detector, the CMX ($0.6 \lesssim |\eta| \lesssim 1.0$).

Kinematic and geometric requirements The track associated with the muon must have $p_T > 20$ GeV/c. Also, for CMX muons, the track is required to have crossed all of the measurement layers of the COT, to ensure that the trigger will be fully efficient. All muon stubs are required to be contained in the fiducial volume of the relevant muon detectors, meaning that they are not too close to the edge of any individual sub-detector.

Muon detector signature CMUP muons must have track stubs in both the CMU and CMP muon detectors. The requirement of both is an important veto of background from hadronic showers leaking out from the calorimeter. The CMU is directly outside the calorimeter, which is only 5.5 nuclear interaction lengths deep. Muons with stubs only in the CMU have large hadronic backgrounds. The CMP covers the same angular region as the CMU but lies outside a layer of additional shielding. CMX muons require only a stub in the CMX. For both CMUP and CMX muons, the quality of the match between the track and the stub(s) is checked. An extrapolated CMUP muon track must be less than 7 cm from the CMU stub and less than 5 cm from the CMP stub. For CMX, the maximum distance is 6 cm.

Track quality requirements Muons are subject to particularly stringent track requirements both because the track momentum is the only measure of the muon energy and to reject backgrounds particular to muons. The basic requirements are identical to those on central electron tracks: at least 3 axial and 2 stereo segments in the COT, and track $|z_0|$ less than 60 cm. We place additional restrictions on the impact parameter and the quality (χ^2) of the matching of the track to the COT hits, and check for cosmic rays. Cosmic rays are real muons passing through the detector. The rate for a cosmic muon to

be observed in coincidence with a collision event with multiple jets is extremely small, but nevertheless we reject events flagged by the cosmic veto. The veto flags tracks that are not consistent with the physical interaction vertex for the event and which cross the detector out of the time window of the beam crossing.

Another insidious but rare background to muons is kaons which decay *inside* the tracking chamber. The resulting particle track looks like a curved track with a kink in it: two circle segments joined at an angle. The track reconstruction assumes that every track can be described by a single helix and attempts to fit the kinked track in this manner. The resulting reconstructed track has a high momentum compared to the original particle but the quality of the fit is poor, since some of the hits lie far from the reconstructed track (see Figure 4.2). This class of muon background can be essentially eliminated by requiring the track to have a χ^2 such that the probability for the track to have a worse χ^2, given the number of degrees of freedom in the fit (the number of hits on the track minus the number of fit parameters), is small ($< 10^{-8}$). A χ^2 probability requirement is used rather than one on just the χ^2 per number of degrees of freedom, since the latter preferentially rejects tracks with a small number of hits, rather than selecting against badly measured tracks. Due to a problem with the simulation of tracking information, this requirement is made only in data. Timing information is used to reconstruct the precise position of hits on

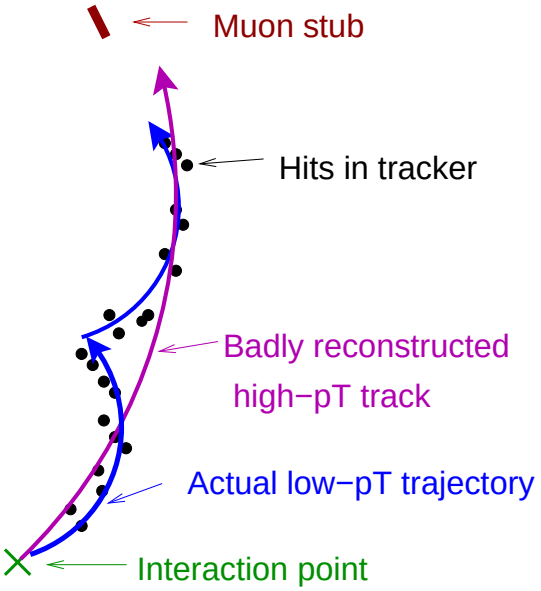

Figure 4.2: Cartoon of the mechanism by which the decay in flight of a medium- to low-p_T kaon can be reconstructed as a high-p_T muon.

COT wires, but the timing resolution was modeled incorrectly.

Calorimeter signature True muons are minimum ionizing particles and therefore typically deposit very little energy in the calorimeter. To reflect this, CMUP and CMX muons are required to deposit less than 2 GeV in the electromagnetic calorimeter and less than 6 GeV in the hadronic calorimeter. For muons with momentum greater than 100 GeV/c, minor corrections to both thresholds are made, as a function of the muon momentum, to account for the fact that more energetic muons deposit more energy in the calorimeter. We also make a calorimeter isolation requirement, analogous to the one for electrons.

This serves to rejects most muons produced by the semileptonic decay of a b hadron, in addition to most "punch-through" fake muons. The total sum E_T in the towers surrounding the ones at which the muon track points is required to be less than 10% of the muon p_T.

4.1.3 Track Lepton Selection

The use of track leptons as the second lepton is one of the unique features of this selection compared to that used in previous dilepton analyses [13, 14, 42, 43]. A track lepton simply a well-measured, isolated track, defined without reference to the calorimetry or muon detectors.

Kinematic and geometric requirements For the main result, track leptons are required to have $p_T > 20$ GeV/c, although we recalculate the cross-section with different p_T thresholds as a cross-check of the measurement. There are no fiducial requirements on the track except the implicit requirement, via the minimum COT hit requirement, that the track be contained in the COT. This effectively restricts track leptons to the range $|\eta| \lesssim 1.1$.

Track quality requirements As with muons, it is important for the track to be well-measured, both to reject background and because the track momentum is the only measure of the particle's energy. The track must have at least 24 hits

in the axial layers of the COT and at least 20 hits in the stereo layers. That this is not defined in terms of segments as with the primary leptons is incidental. As is the case for the well-reconstructed muons, there is an impact parameter threshold ($|d_0| < .025$ cm) for the track. The track must also satisfy the same χ^2 probability requirement. Because the track lepton will sometimes be a muon, the same backgrounds are a concern. In addition, the silicon information for the track is checked and required to be reasonable. Specifically, if the track passes through at least three active layers of silicon, it must have at least three silicon hits attached.

Track isolation To reject background from jets, the most important criteria is track isolation. Like the calorimeter isolation required for the primary leptons, it is defined in terms of a ratio of the candidate p_T and the p_T of objects in its immediate surroundings. The p_T of every track with $p_T > 0.5$ GeV/c in the angular region of the candidate track, including the candidate track itself, is summed. The angular region is defined in terms of the distance $\Delta R = \sqrt{(\Delta\eta^2 + \Delta\varphi^2)}$, and all tracks in the cone defined by $\Delta R < 0.4$ are included in the sum. The ratio of the candidate track p_T to the sum p_T in the cone is required to be at least 0.9. Tracks included in the sum pass minimal quality requirements similar to those for track leptons but looser.

Relationship to primary lepton To be used as the second lepton in the analysis, the isolated track must be distinct from the track belonging to the primary lepton, but to have a z_0 within 5 cm of the primary lepton track z_0, so that they are consistent with coming from the same interaction.

4.2 Jet Definition

Quarks and gluons with large transverse momentum produced in the interaction are not detected as single particles because they are strongly interacting. As they pull away from the interaction point, they generate additional colored particles (quarks and gluons), which combine into colorless hadrons. The direction the hadrons travel in the detector is correlated with the direction of the original parton, and, aside from energy spent in creating additional quarks and gluons, the hadrons carry the energy of the original parton. Therefore some information about the original parton can be recovered from the properties of this collection of hadrons, referred to as a jet.

Jets are reconstructed by grouping together energy in the calorimeter which are close together in $\eta - \varphi$ space. The collection of particles can then be treated as a single entity. The energy measured in the calorimeter is correlated with the energy of the original parton but typically does not fully reproduce it. How-

ever, jet energies can be corrected based on knowledge of jet formation and the detector response.

4.2.1 Jet Clustering

This analysis uses calorimeter-based jet reconstruction. Jet clustering begins with a single "seed" tower. Seed towers are taken from a list of all towers in the event with $E_T > 1$ GeV, starting with the highest-E_T tower. Then, all towers surrounding that tower, with $E_T > 1$ GeV and within a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} < R_0$, are included in the jet cluster. This process is repeated for every possible seed tower, in order of decreasing E_T, until all possible seed towers are used. Then jets sharing towers are merged if they overlap by more than 50%, or split into two jets if the overlap is less [44]. For this analysis, the jet cone is $\Delta R < 0.4$. The jet energy is the summed energy of the towers and the direction is the energy-weighted centroid of the calorimeter cluster.

4.2.2 Jet Energy Corrections

The energy in the towers of the jet is not by itself a good measure of the energy of the parton generating the jet. Extensive study has been done of the response of the CDF detector to jets, and correction factors measured that

bring the average measured jet energy to the energy of the parton [44]. These correct for the non-uniformity of the calorimeter response as a function of the jet angle, energy, time, and the fraction of energy deposited in the electromagnetic calorimeter. The overall absolute energy scale is also corrected.

We derive the corrections from a combination of real and simulated jet data. Events with two jets, one of which is in the best-understood region of the calorimeter ($0.2 < |\eta| < 0.6$) is used to determine the η-dependent non-uniformity corrections, since the two jets should have equal p_T. The absolute energy corrections rely on the accuracy of the simulation of the calorimeter. The calorimeter simulation is tuned on the response to single particles, and the correction factor measured from the difference between the reconstructed jet energy and the recorded true energy of the generated particles in the simulated jet.

4.2.3 Jets for Analysis

We correct the energies of all jets with more than 10% of their energy in the hadronic calorimeter, and count jets with corrected $E_T > 20$ GeV and $|\eta| < 2.0$. For the final jet count, used in checking that the number of jets is at least two, jets overlapping with either the primary or track lepton are excluded. In some situations the E_T threshold defining a jet is varied. In variations on the

measurement done as a check on the main result, 15 and 25 GeV are used as thresholds. For the $\not\!\!E_T$ calculation, we use jets down to a corrected E_T of 10 GeV.

4.3 Missing Transverse Energy Reconstruction

Missing transverse energy is a measurement of the transverse energy imbalance in the detector. When everything in the detector — jets, leptons, and unclustered energy — has been measured as well as possible, momentum conservation should be observed in the x and y directions. Therefore, assuming perfect measurement of all objects in the event, an imbalance in the transverse momentum in the event is interpreted as a neutrino escaping the detector. $\not\!\!E_T$ calculations are based on the summed energy in all the calorimeter towers in the detector, but in some cases, this does not represent the best measurement of the energy of all objects in the event. For muons in particular, the track associated with a particle is a better measure of the energy. Jets have known correction factors, as described above, so that it is better to use the corrected energy than the raw calorimeter energy originally used in the sum.

4.3.1 Basic $\not\!\!E_T$ Calculation

The x and y components of the $\not\!\!E_T$ are calculated from the E_x and E_y measured in each calorimeter tower, using the measured event vertex as $z = 0$:

$$\not\!\!E_{\{x,y\}} = -\sum_i (E_{\{x,y\}})_i$$

The sum is over all calorimeter towers in the detector.

4.3.2 Track E/P corrections

For muons and other minimum ionizing particles, or even for particles which land in a gap in the calorimeter, the momentum of the track, not the energy in the calorimeter, is the best measurement of the particle's energy. If the primary lepton in the event is a muon, the p_x and p_y of the muon are added to the sum of calorimeter tower energies. The energy in the calorimeter is small by the definition of the fully reconstructed muons.

After accounting for the primary muon if there is one, we make a more general correction for the E_T/p_T of tracks. For all reasonably well-measured tracks in the event with $p_T > 10$ GeV/c, excluding the track from the primary lepton, the E_T in the 3 by 3 block of calorimeter towers at which the track points is compared to the p_T of the track. For this purpose, "well-measured" is defined as passing all the track lepton selection except the isolation, track χ^2,

and silicon hit criteria. If any track has $E_T/p_T < 0.7$, the E_T is removed from the $\not\!\!E_T$ calculation and replaced by the track p_T.

4.3.3 Jet Corrections

For every jet in the event with $E_T > 10$ GeV, the corrected jet E_T is substituted for the uncorrected E_T in the above sum.

The motivation for taking such care in the $\not\!\!E_T$ correction is that one of the largest and most difficult backgrounds to calculate is $Z/\gamma^\star \to ee/\mu\mu$, which passes the analysis selection only if false $\not\!\!E_T$ is somehow generated in the event. By correcting for all objects known to be measured inaccurately, background from this source can be reduced.

4.4 Event Selection

The candidate event sample is defined in terms of the physics objects described above: leptons, jets, and $\not\!\!E_T$. The basic selection reflects the distinctive features of top dilepton decay events, and additional requirements on the relationships between these objects reduce background in the candidate sample while maintaining reasonable efficiency for top events.

4.4.1 Basic Selection

First, there must be at least one fully reconstructed lepton in the event. If there is more than one, we try the full event selection with each one in turn, in this order: central (CEM) electrons, then CMUP muons, then CMX muons, then plug (PHX) electrons. Within a particular lepton type, the leptons are tried in order of descending E_T. Each fully reconstructed lepton is tried until the event has passed all of the selection or has failed for every lepton type. In the data, the trigger corresponding to that lepton category must have fired for that event, and the relevant parts of the detector must be known to be fully functional at that time.

Once the primary lepton has been identified, the tracks are searched for a track lepton candidate. The highest p_T track isolated according to the definition above (Section 4.1.3) is taken as the track lepton for this primary lepton. If there is no isolated track, we try the next fully reconstructed lepton, if there is one.

Then we correct the $\not\!\!E_T$, and require the corrected $\not\!\!E_T$ to be greater than 25 GeV. This is an effective tool for reducing all backgrounds considered except for diboson events (Figure 4.3), but is quite efficient for $t\bar{t}$ events.

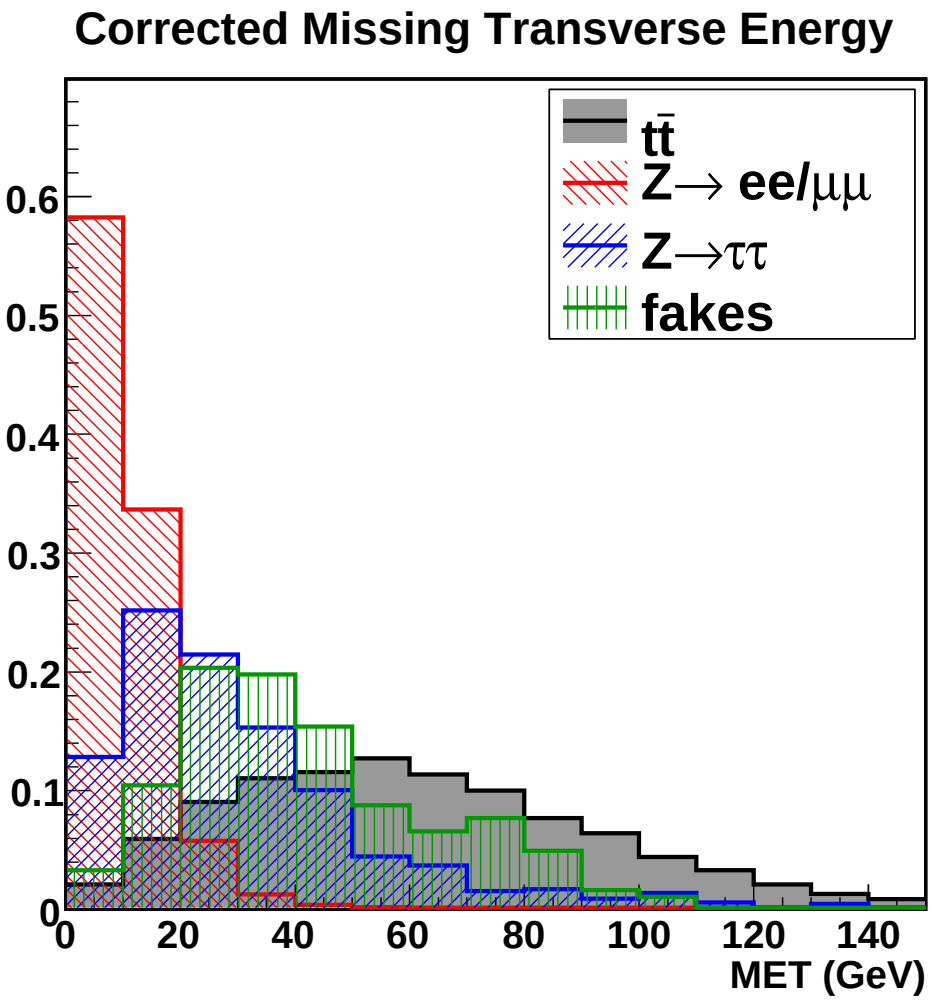

Figure 4.3: Missing transverse energy in $Z/\gamma^{\star} \rightarrow ee/\mu\mu$events with $\not\!\!E_T > 25$ GeV, compared to the distribution from dilepton $t\bar{t}$. Distributions taken from events simulated using the PYTHIA generator.

4.4.2 "Delta Phi" Selection

Missing transverse energy can only be produced in $Z/\gamma^\star \to ee/\mu\mu$ events through the mismeasurement of an object in the event. The probability of an object generating a significant amount of $\not\!\!E_T$ is small, so the direction of the $\not\!\!E_T$ in such events tends to be correlated with the direction of the mismeasured object. We therefore require that no object in the event be pointing right at the $\not\!\!E_T$ in φ. This reduces the rate for this background by effectively requiring that two objects be badly mismeasured, an even more improbable occurrence. Specifically:

1. The angle in φ between the primary lepton and the $\not\!\!E_T$ must be greater than 5°: $\varphi(\text{lep},\not\!\!E_T) > 5°$.

2. The track lepton must not be within 5° of pointing directly at or away from the $\not\!\!E_T$: $5° < \varphi(\text{trk},\not\!\!E_T) < 175°$.

3. No jet with $E_T > 10$ GeV may point within 25° of the $\not\!\!E_T$ unless $\not\!\!E_T > 50$ GeV.

Some comment is warranted on the different requirements for the different objects. The delta phi cut for the primary lepton is designed to reduce a very particular source of $\not\!\!E_T$. In a small fraction of $Z/\gamma^\star \to \mu\mu$ events, one muon

emits a bremsstrahlung photon carrying a significant fraction of its momentum. Such a muon may then reconstructed as an electron, because there is a good track pointing at a perfect electromagnetic cluster. If an electron is used as the primary lepton, false $\not\!\!E_T$ is generated since the "electron" energy will used in the $\not\!\!E_T$ calculation, not the muon momentum. It will not be corrected as a fully-reconstructed muon: this object will fail the muon selection because of the amount of energy in the calorimeter. It will also be missed by the generic track E/P correction because the electron, as the primary lepton, will be skipped. In this case the $\not\!\!E_T$ direction is very narrowly correlated with the primary lepton direction, so a 5° veto is sufficient.

For the track leptons, the lepton energy can be over- or under- measured, leading to over- or under-correction of the $\not\!\!E_T$, so $\not\!\!E_T$ both correlated and anti-correlated to the track lepton is vetoed.

The typical scale for a jet mismeasurement is only about 10 GeV, so mismeasured jets are not usually capable of generating large amounts of $\not\!\!E_T$. However, because jets cover a large solid angle in the detector compared to leptons, the angular correlation between a mismeasured jet and the $\not\!\!E_T$ is not as strong. Therefore the angular region vetoed is larger than for leptons, but the requirement can be released for $\not\!\!E_T > 50$ GeV to recover efficiency. The angle between the $\not\!\!E_T$ and the closest jet, in $t\bar{t}$ signal and $Z/\gamma^\star \to ee/\mu\mu$ events, is shown in

Figure 4.4.

If an event fails any of these requirements, the next fully reconstructed lepton candidate is tried, if there is one.

4.4.3 Drell-Yan Veto

Even accounting for the smearing of the reconstructed invariant mass in the case where a lepton energy is poorly measured, a large fraction of $Z/\gamma^{\star} \rightarrow ee/\mu\mu$ events with $\not\!\!E_T > 25$ GeV have a dilepton invariant mass in the region 76 GeV/$c^2 < M <$ 106 GeV/c^2 (Figure 4.5). Increasing the $\not\!\!E_T$ threshold to 40 GeV in this kinematic region reduces the contamination from Drell-Yan in the candidate sample. This requirement is referred to as the "Drell-Yan veto" or "Z veto".

If an event fails the Drell-Yan veto, the next fully reconstructed lepton candidate is tried, if there is one.

4.4.4 Candidate Event Classification

Events which have passed all of the event selection are classified according to the number of jets with $E_T > 20$ GeV and whether the primary and track lepton have opposite sign. We measure the cross-section using candidate events with two or more jets and opposite-sign leptons. Events with same-sign leptons

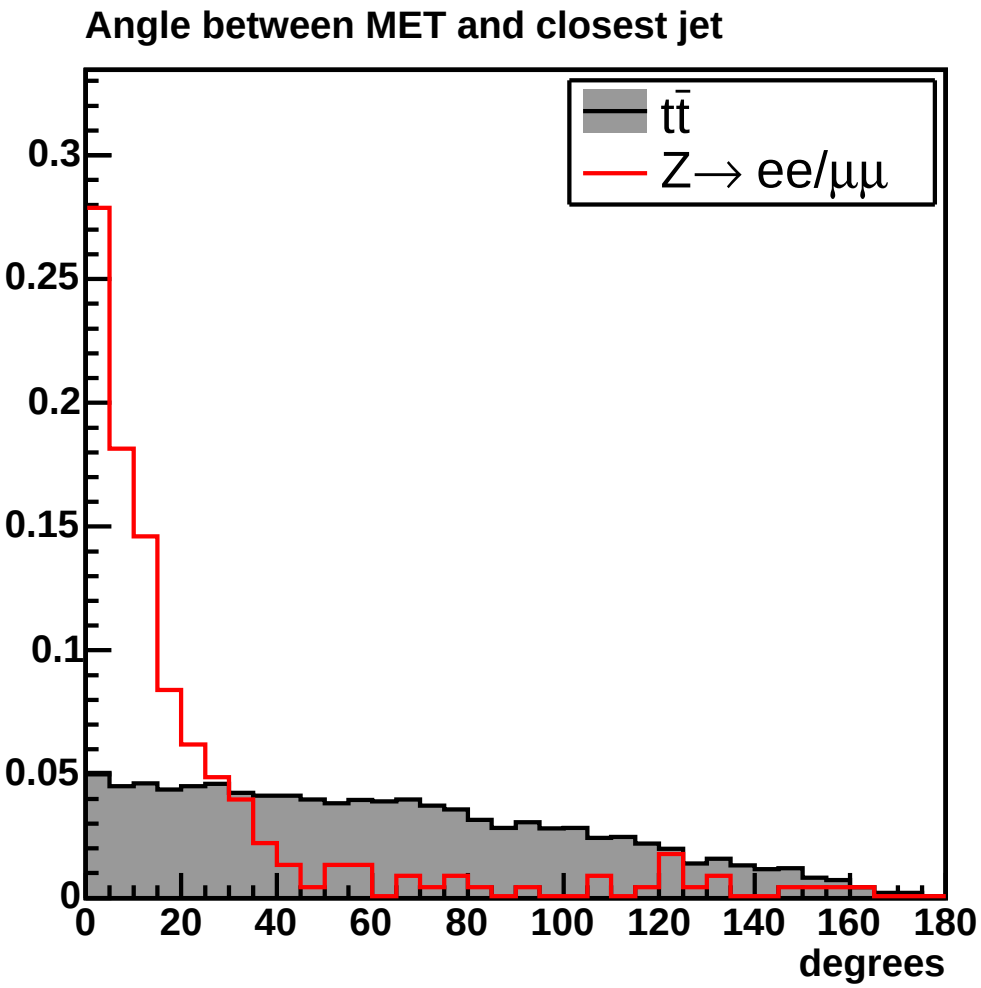

Figure 4.4: Angle between jets and $\not\!\!E_T$ in $Z/\gamma^\star \to ee/\mu\mu$ events with $\not\!\!E_T >$ 25 GeV, compared to the distribution from dilepton $t\bar{t}$. Distributions taken from events simulated using the PYTHIA generator.

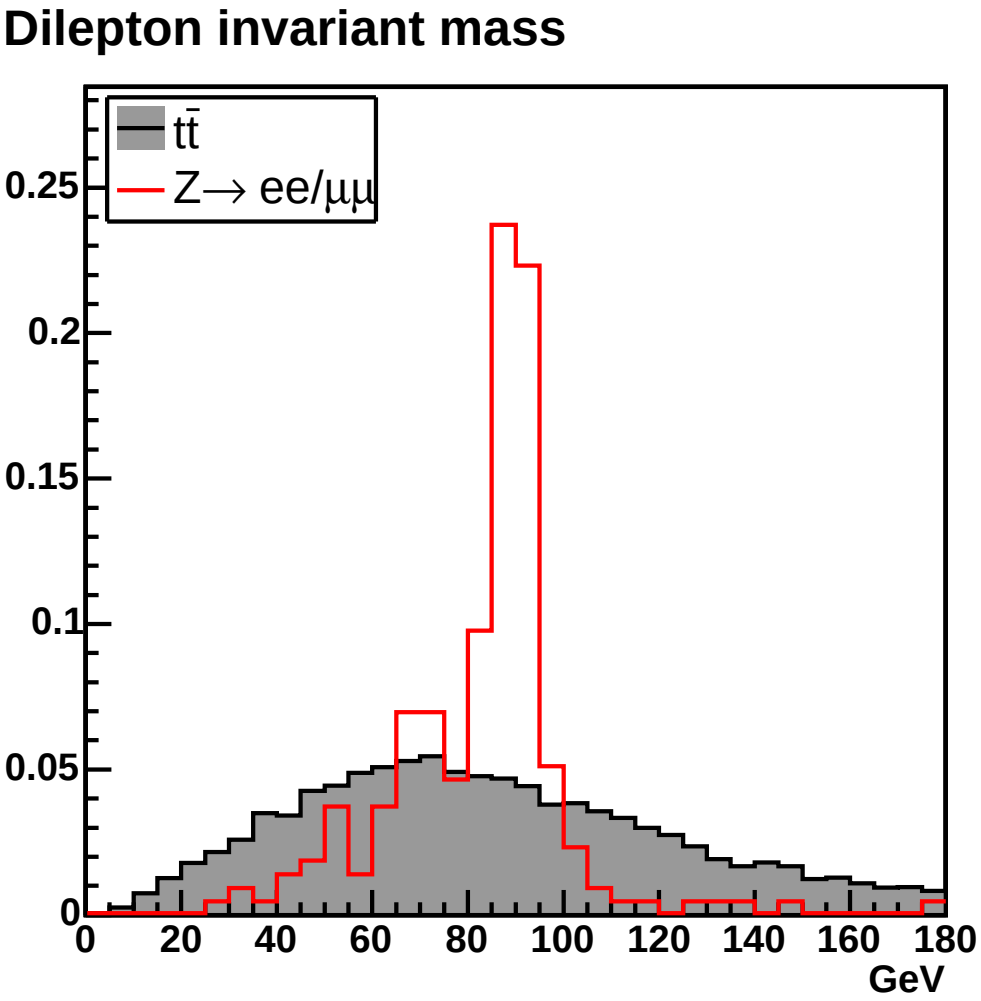

Figure 4.5: Dilepton invariant mass of $Z/\gamma^{\star} \to ee/\mu\mu$ events with $\not\!\!E_T > 25$ GeV, compared to the distribution from dilepton $t\bar{t}$. Distributions taken from events simulated using the PYTHIA generator.

and/or zero or one jets are used to validate the background estimates.

4.5 Data Sample

The data sample used for the cross-section measurement is determined by the amount of data taken for which all the relevant detectors are functional. We can define it now because it depends on the choice of event selection, through the physics object definitions. The integrated luminosity available is different for all four primary lepton types. To use the CEM, or central, electrons, we consider only those segments of the data where the calorimetry is in good operating condition. For the forward, or PHX, electrons, we make the additional requirement that the silicon tracker is functional, since the track reconstruction for those electrons depends on it. To use a muon for the primary lepton, we do not require that the silicon tracker be powered but do check that the relevant muon detectors, CMU, CMP, or CMX, are active. The integrated luminosities for all four primary lepton types are shown in Table 4.2.

4.6 Summary

The event selection defines the candidate sample, and sets the number of signal and background events expected in the sample. The basic event signa-

lepton type	luminosity (pb^{-1})
CEM	1097
CMUP	1047
CMX	1027
PHX	1015

Table 4.2: Integrated luminosity for each primary lepton type.

ture for top dilepton events is two opposite-sign leptons, two or more jets, and significant $\not\!\!E_T$. Restrictions on the relationship between those objects reduces the contribution to the candidate sample from backgrounds, particularly from $Z/\gamma^\star \to \ell\ell$. The use of an isolated track for the second lepton significantly increases the signal acceptance, by allowing more τ lepton decays and making the geometrical limitations of the muon detectors and calorimeter irrelevant.

Chapter 5

Top Dilepton Acceptance

With the event selection defined, we can determine the acceptance for dilepton $t\bar{t}$ events. The acceptance, defined as the expected fraction of signal events passing the selection, is one of the basic inputs to the cross-section measurement (see the formula in the Introduction). We use a simulated sample of $t\bar{t}$ events to measure it. The acceptance is corrected for known inaccuracies in the simulation, using information from other, more common and better-understood physics processes.

Limits in both experimental and theoretical knowledge introduce systematic uncertainty to the calculation of the acceptance. For each significant source of uncertainty, we quantify the size of the possible effect. For experimental limitations, we use studies comparing data and simulation, and for theoretical

ones, we use variations in the inputs for generating simulated $t\bar{t}$ samples.

5.1 Simulated $t\bar{t}$ Sample

The simulated $t\bar{t}$ event sample is generated using PYTHIA Monte Carlo [3]. Inputs are the top quark mass (175 GeV/c^2, in this case) and a recent measurement of the parton distribution functions. PYTHIA uses the leading-order matrix elements for top pair production to produce kinematic information for the final state quarks and leptons and discard some fraction of the events generated so that the kinematic distributions of particles follows the expected probability distributions. PYTHIA also decays unstable particles (such as the W from the top decay) and simulates the hadronization of quarks and gluons.

PYTHIA outputs a list of particles and their associated momentum four-vectors, which are input to a full simulation of the CDF detector. The ionization, showering, and energy deposition is modeled for each particle, and the expected signature in each detector produced. Event reconstruction identical to what is used on data is run on the result, and the output can be analyzed using the same techniques and software as the data.

Simulation has the advantage that the user knows the underlying physics process and can access information about all of the generated particles. For

example, we can check observed leptons to see if it matches a generated lepton, and match jets to generated quarks or gluons to infer the type of the originating parton. The weakness of simulation is that it does not reproduce the signature of particles in the detector with perfect accuracy.

5.2 Corrections to Simulation

The CDF detector simulation models the behavior of particles in the detector well but not perfectly. Corrections can be derived for known discrepancies. The efficiencies for identifying leptons and the trigger efficiencies for the high-p_T lepton triggers are both well-known and are used to bring the acceptance measured in simulation closer to the truth.

5.2.1 Primary Lepton Scale Factors

Lepton identification efficiencies are well known in both data and simulation and, in general, do not match. We measure efficiencies in Z data and simulation, since this process has two charged leptons in the final state and very pure Z samples can be selected in data. Also, very large samples of real and simulated exist for $Z/\gamma^\star \to \ell\ell$.

Efficiencies for fully reconstructed leptons, such as the ones used as primary

leptons in this analysis, are measured by first finding one fully reconstructed lepton in the data, and then looking for a second, opposite-sign lepton candidate of the same flavor, which meets minimal kinematic and identification criteria. The invariant mass of the leptons is required to be close to the Z resonance peak, to ensure that the dilepton signal far exceeds the background, even when the second lepton is only required to pass minimal selection requirements. Then, all of the second leptons are checked to see if they pass all of the selection requirements. The efficiency is the ratio of the number passing all selection to the total number of second leptons. It gives the probability that a real lepton will pass all of the requirements to be considered fully reconstructed.

The ratio of the efficiency in data to the efficiency in simulation is applied as a correction factor to the acceptance as calculated from simulation. That is, the acceptance for an event which passes all the selection with a particular lepton type will be multiplied with the data/simulation scale factor for that lepton. The trigger requirement is made in data but not simulation, so in this case we multiply the acceptance by the trigger efficiency rather than by a ratio. The measured efficiencies and scale factors for the four primary lepton types in this analysis are given in Table 5.1.

Lepton type	ϵ(data)	ϵ(simulation)	data/simulation
CEM electron	.793 $\pm$ .002	.808 $\pm$ .001	.981 $\pm$ .004
PHX electron	.715 $\pm$ .003	.764 $\pm$ .001	.935 $\pm$ .010
CMUP muon	.833 $\pm$ .004	.900 $\pm$ .001	.926 $\pm$ .004
CMX muon	.904 $\pm$ .004	.919 $\pm$ .001	.984 $\pm$ .005

Table 5.1: Efficiencies for identifying the different primary lepton types in data and simulation, and the ratio used to correct the acceptance.

5.2.2 Track Lepton Scale Factors

There are also discrepancies between data and simulation in the efficiency for identifying isolated track leptons. The most significant is the discrepancy in the isolation efficiency. The efficiency of the track isolation cut is measured analogously to the primary lepton efficiencies. We reconstruct a Z candidate from a well-reconstructed lepton and an opposite-sign track passing all of the track lepton cuts except isolation, with a pair invariant mass in the interval 76 GeV/c^2 $< M <$ 106 GeV/c^2. The efficiency of the isolation cut is the ratio of the number of tracks passing the isolation cut to the total number of tracks. This efficiency drops as a function of the number of additional jets, primarily because in events with more jets there is less solid angle in the detector for a track to be where it can be isolated. The scale factor also changes as a function of the number of jets, and is significantly different from one for events with two or more jets. Therefore we apply this scale factor, as a function of the number of jets considered. The efficiencies drop from $\approx 95\%$ for events with zero jets to $\approx$

90% for events with two or more jets. The resulting scale factors are 1.004±.001 for events with zero jets, 1.002±.003 for events with one jet, and .965±.011 for events with two or more jets.

The efficiency of the impact parameter requirement is another place where data and simulation could possibly differ. The total observed efficiency in data is .909±.003: .940 for runs with silicon and .53 for runs without (see Figure 3.5 in Chapter 3). The corresponding efficiency in .9185±.0007 in simulation, .947 for runs with silicon and .55 for runs without, leads to a scale factor of .989±.003.

5.2.3 Track χ^2 Probability Cut Efficiency

The χ^2 probability requirement, imposed on both fully reconstructed muons and on track leptons, is intended to reject kaon decays-in-flight that can be mistaken for prompt high-p_T muons. Because the requirement is made only in data, the acceptance is multiplied by the efficiency as measured in data rather than by a ratio.

We also measure this efficiency in the well-reconstructed lepton plus isolated track Z sample. One subtlety is that the probability cut efficiency is correlated between the two leptons through the hit timing information in the COT, so that the efficiency to apply it to both leptons is not equal to the product of

the efficiencies for each lepton. Therefore for electron+track events, where the requirement applies only to the track lepton, the efficiency is just the ratio of tracks that pass the requirement to all tracks. But for muon+track events, where it applies to both, the relevant efficiency is the ratio of muon+track events where both leptons pass the requirement, to all muon+track Z events. The measured efficiencies are $.962 \pm .001$ for electron+track events, $.944 \pm .001$ for CMUP+track events, and $.951 \pm .002$ for CMX+track events.

5.2.4 Event Trigger Efficiencies

Data used in dilepton analyses are collected on triggers which select events with at least one high-p_T ($p_T \gtrsim 18$ GeV/c) lepton candidate. The trigger thresholds are designed to be very efficient for fully-reconstructed leptons, but are not 100% efficient. Some characteristics of the fully-reconstructed leptons, especially the energy and the tracking information, are different for the "offline" reconstruction than in the fast reconstruction used in the trigger.

We measure single lepton trigger efficiencies with a combination of Z data and complementary trigger datasets. Z data is useful when the two lepton candidates land in sections of the detector corresponding to different triggers. Then, if lepton A fires the corresponding trigger A, the efficiency for the trigger for lepton B can be determined from how often lepton B fires its trigger. Comple-

mentary trigger datasets are often defined for important triggers for the express purpose of measuring trigger efficiencies. Such a trigger will share some but not all of the requirements of the trigger of interest, such as either its calorimeter or tracking requirements, or some stages of the trigger will be identical but others set to automatically accept every event.[1] Then the ratio of number of events passing the full requirements to the number passing the reduced requirements is the efficiency of the requirements omitted in the complementary trigger.

The probability for a dilepton event to fire one of the high-p_T lepton triggers is higher than the single lepton trigger efficiency since each event has two chances to fire one of the triggers. But the second lepton is not fully reconstructed, so the event trigger efficiency is not just a combination of the single-lepton trigger efficiencies. Instead the per-event trigger efficiency is calculated using simulation, with the single-lepton trigger efficiencies as inputs. For an event, we make a list of fully reconstructed leptons. For each lepton, a random number between zero and one is compared to the relevant single-lepton efficiency to decide if that lepton "passes" the trigger. If none of the leptons pass the trigger, the event is rejected; otherwise, it is accepted. Repeating this procedure for every event produces an average trigger efficiency for a sample and primary

[1] Such triggers are typically "prescaled" to actually accept only a fraction of events passing all the requirements, to avoid overwhelming the trigger system due to the high accept rate from increased, high-cross-section background.

	CEM	CMUP	CMX	PHX
single lepton	.971	.918	.910	.918
$t\bar{t}$	**.975**	**.916**	**.937**	**.918**
WW	.976	.917	.937	.935
WZ	.981	.921	.939	.925
ZZ	.983	.926	.941	.924
$Z/\gamma^\star \to \tau\tau$	.966	.911	.929	**.657**

Table 5.2: Single lepton and full per-event trigger efficiencies for the four primary lepton types. Numbers in bold are inputs to the cross-section measurement. Uncertainties are negligible compared to others uncertainties on the acceptance.

lepton type.

Single-lepton trigger efficiencies, as well as total per-event trigger efficiencies for each physics process for which an acceptance measured in simulation is used, are given in Table 5.2. For a given lepton type, the per-event efficiencies are very similar among the different physics processes, so the $t\bar{t}$ value is used. There is one exception, PHX+track $Z/\gamma^\star \to \tau\tau$ events. The leptons and $\not\!\!E_T$ from that process are less energetic than for top and diboson events, and the trigger efficiencies correspondingly lower since more events are near the threshold, where the trigger is not fully efficient. Therefore in that case we use the particular efficiency measured for that process and lepton type.

	0 jets	**1 jet**	**$\geq$ 2 jets**
CEM	0.009 $\pm$ 0.000	0.134 $\pm$ 0.002	0.472 $\pm$ 0.003
CMUP	0.004 $\pm$ 0.000	0.056 $\pm$ 0.001	0.196 $\pm$ 0.002
CMX	0.002 $\pm$ 0.000	0.023 $\pm$ 0.001	0.073 $\pm$ 0.001
PHX	0.002 $\pm$ 0.000	0.029 $\pm$ 0.001	0.104 $\pm$ 0.001

Table 5.3: Top dilepton signal acceptance, in percent, for $m_t = 175\text{GeV}/c^2$, measured in PYTHIA, corrected for all known discrepancies between data simulation. Uncertainty is statistical only.

5.3 Acceptance for Top Dilepton Events

The acceptance for the top dilepton signal is measured in a sample of about 4 million inclusive PYTHIA $t\bar{t}$ events generated at a top mass of $m_t = 175\text{GeV}/c^2$. Table 5.3 shows the signal acceptance as a function of the number of jets in the event, with all of the correction factors described above applied. The total acceptance is 0.845%. The 0.004% statistical uncertainty is negligible compared to the systematic uncertainties.

5.4 Systematic Uncertainties on the Acceptance

Systematic uncertainties reflect the limits on how accurately the top dilepton acceptance can be known. They arise from the limits of theoretical understanding of top quark pair production and the limits of experimental understanding of the final-state objects used to identify $t\bar{t}$ events. The first category includes uncertainties on QCD radiation, parton density functions, and the Monte Carlo

Source	uncertainty assigned
Primary lepton identification efficiency	1.1%
Track lepton identification efficiency	1.1%
Jet energy scale	1.3%
Initial-state QCD radiation	1.6%
Final-state QCD radiation	0.5%
Parton density functions	0.5%
Monte Carlo generator	1.5%

Table 5.4: Summary of systematic uncertainties on the signal acceptance. The total systematic uncertainty is 3%.

generator used to calculate the acceptance. The second includes uncertainties on lepton identification and jet energies.

The systematic uncertainties on the signal acceptance are discussed individually below and summarized in Table 5.4.

5.4.1 Primary Lepton Identification Efficiency

The dominant uncertainty on the identification efficiency for fully reconstructed leptons is its dependence on the level of additional activity, such as jets or unclustered low-p_T tracks in the event. The scale factors correcting the lepton identification efficiencies are derived from real and simulated Z data, in which most events have zero jets. Are they accurate for events, such as those in our signal sample, which have two or more jets?

Two methods exist for quantifying the possible disagreement between the

scale factors for zero jets and many jets. One is a direct comparison of the scale factor measured in $Z+ \geq 2$ jet events to the inclusive result. This has the advantage of accounting for the whole event environment, not just the reconstructed jets, but is severely limited by the lack of data with large numbers of jets. In the second method the scale factor is measured as a function of the distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2}$ between the lepton candidate and the nearest jet. Then a new scale factor is calculated, reweighted according to the distribution of distances between the primary lepton and the closest jet in the process of interest. This method accounts only for the direct influence of jets, but is much less statistically limited because all Z events with one or more jets are considered.

The two methods give consistent answers, in that the scale factor as a function of the distance to the nearest jet, reweighted using the distance distribution from $Z+ \geq 2$ jet events, reproduces the scale factor measured directly with $Z+ \geq 2$ jet events very well for all primary lepton types. Therefore the systematic uncertainty is taken as the larger of the discrepancy between the inclusive and reweighted scale factors and the statistical uncertainty on the reweighted scale factor. In most cases the statistical uncertainty is larger than the discrepancy.

The total systematic uncertainty is the weighted average of the uncertainties

on the individual lepton types, where the weights are the top dilepton acceptances for each lepton category. The resulting uncertainty is 1.1%.

5.4.2 Track Lepton Identification Efficiency

This uncertainty quantifies how well the simulation models the track isolation requirement in an environment with many jets, and is analogous to the uncertainty on well-reconstructed leptons. In this case, though, we base the uncertainty on the behavior of scale factor as a function of the number of jets rather than as a function of the ΔR between the lepton and the closest jet. Also, the scale factor measured in the higher jet multiplicity environment is taken as the central value. The scale factor is derived in Section 5.2.2. The systematic uncertainty used, 1.1%, is the statistical uncertainty on the scale factor measured for events with ≥ 2 jets.

5.4.3 Jet Energy Scale

Uncertainties on the jet energy scale are calculated along with the jet energy corrections [44]. The jet energy scale influences the top dilepton acceptance by making it easier or harder for events to pass the requirement of two or more jets. To estimate the uncertainty on the acceptance from the jet energy scale, we recalculate the signal acceptance in simulated $t\bar{t}$ events, with the jet energy

corrections varied within their uncertainties. the measured acceptance increases by 1.16% for the $+1\sigma$ variation and decreases by 1.4% for the -1σ variation. We take half the full difference between these, or 1.3%, as the systematic uncertainty.

5.4.4 Initial and Final State Radiation

Additional jets can be produced in association with the $t\bar{t}$ pair through QCD radiation of one or more partons from either the initial or final state. Additional jets can affect the acceptance through the minimum number of jets requirement. Extra jets can make it possible for an event to pass the selection even if one of the b jets would be too forward or soft to be counted Also, in the case of initial state radiation (ISR), an extra jet adds more energy to the event. The decay products of a $t\bar{t}$ pair recoiling against one or more jets have extra transverse momentum, making them likely to pass the kinematic selection. Final-state radiation (FSR) can work in the opposite direction: a jet radiated from one of the top quarks carries away some of the top quark's momentum, so that its decay products are less likely to pass the kinematic selection.

We test the dependence of the acceptance on QCD radiation using simulated $t\bar{t}$ samples where the rate of initial and final state radiation has been varied within the allowed range. The difference between the original and recalculated

acceptances quantifies this dependence. Numerically, the acceptances measured are:

- more ISR: 1.0067 ± .0103% (+1.3% from nominal)
- less ISR: .9743 ± .0100% (-2.0% from nominal)

This is to be compared to the nominal acceptance of .9938 ± .0050.[2] The behavior, more acceptance with more radiation, is as expected. The acceptances in the two samples are significantly different from each other statistically, so we take half the full difference, or 1.6%, as the systematic uncertainty.

The acceptance measured in the samples where the final state radiation was varied are:

- more FSR: .9977 ± .0107% (+0.4% from nominal)
- less FSR: .9992 ± .0118% (+0.5% from nominal)

These are not statistically different from the nominal acceptance, very possibly due to the competing effects discussed above. A conservative estimate of the systematic is the larger of the two observed differences, or 0.5%. It does not hurt to be conservative here since this contribution has a negligible impact on the total uncertainty.

[2]This is different from the 0.845% quoted in Section 5.3 because none of the lepton identification correction factors have been applied.

5.4.5 Parton Distribution Functions

Uncertainties on the parton distribution functions, or PDFs, which describe the fraction of the proton momentum carried by each type of parton, have a significant effect on the top cross-section (see Chapter 2 Section 2.1.2). They also have a smaller effect on the measured acceptance, through the kinematics of the $t\bar{t}$ decay products. The uncertainties on the PDFs are known. Twenty such uncertainties have been identified for the CTEQ PDF set [2]. We also consider differences between PDF sets and variations due to the uncertainty on the value of α_s as sources of uncertainty.

To quantify the dependence of the top dilepton acceptance on the PDFs used to generate the simulated $t\bar{t}$ samples, we recalculate the acceptance with events reweighted using PDFs which differ from the nominal in a specified way. In the truth information for a generated event, the incoming partons and their momenta are identified. The corresponding probabilities for those values are found in the nominal PDF and in the variation being studied. The event weight is the ratio of the product of the varied to nominal probabilities:

$$\text{weight} = \frac{\text{p}_1(\text{variation}) \cdot \text{p}_2(\text{variation})}{\text{p}_1(\text{nominal}) \cdot \text{p}_2(\text{nominal})}$$

We repeat this process for each variation in the PDFs record and the change in the acceptance.

For the CTEQ PDFs, the quadratic sums of the observed differences in the acceptance are +0.3% and -0.7% from the nominal. We also compare the acceptance as calculated using the nominal CTEQ PDFs to the acceptance calculated using PDFs determined independently by a second collaboration (MRST) [45]. The difference, .2%, is smaller than the uncertainties derived simply by consideration of the uncertainties on the CTEQ PDFs, and therefore ignored according to the current CDF convention. The uncertainty obtained by varying α_s is .1%, which is added in quadrature with the CTEQ uncertainties for a total uncertainty of +0.3% and -0.7%. Symmetrizing this gives a systematic uncertainty of 0.5%.

5.4.6 Monte Carlo Generator

To check for dependence of the measured acceptance on the Monte Carlo event generator, the $t\bar{t}$ acceptance is remeasured in $m_t = 175$ GeV HERWIG Monte Carlo and compared to the nominal value from PYTHIA. The two generators use different $W \to \ell\nu$ branching ratios. HERWIG uses $1/9 = .111$, motivated by the expected universal coupling between the W and all isospin doublets. PYTHIA uses .108, the current best measured value, from the experiments using data from LEP, an electron-positron collider at CERN. To exclude the effect of the branching fraction difference, we make the comparison between

acceptances calculated relative to the total number of generated *dilepton* events. The resulting difference, taken as a systematic uncertainty, is 1.5%.

5.5 Summary

We use the selection defined in Chapter 4 to measure the acceptance for top dilepton signal events using a sample of simulated $t\bar{t}$ events. The acceptance relates the number of signal events produced to the number observed in the cross-section calculation. The limits on how well the acceptance can be measured are quantified in a set of systematic uncertainties, which total to 3%. Contributing uncertainties are both theoretical and experimental in nature. The leading uncertainty at the present is the effect of additional jets from QCD radiation. With that uncertainty and all known corrections for discrepancies between data and simulation applied, the lepton + track acceptance for $t\bar{t}$ events is 0.845 ± 0.026 %.

Chapter 6

Dilepton Backgrounds

Backgrounds to the top dilepton sample all arise from the leptonic decay of one or more massive vector bosons. If enough jets are produced by extra energetic partons emerging from the interaction to substitute for the jets from the b quarks, these events can pass the selection if they have, or can mimic the existence of, two leptons and significant missing transverse energy ($\not\!\!E_T$).

Non-negligible backgrounds in the sample are from diboson production, W + jets where one of the jets is misidentified as a second lepton, and Drell-Yan ($Z/\gamma^\star \to \ell\ell$) events where $\not\!\!E_T$is produced by some combination of τ decays and the mismeasurement of objects in the event. Different techniques are used to estimate each category of background, using a combination of event simulation and control samples from the data. For the W + jets background, the estimate

can be derived almost entirely from data, but for the diboson contribution, the estimate must be based completely on simulated data.

The methods used to determine the backgrounds are limited in their accuracy, by the size of the samples used, the modeling of the relevant quantities in simulation, and experimental understanding of the physics objects used to reconstruct the events. For each class of backgrounds, we calculate statistical and uncertainties reflecting these limitations. By convention, all uncertainties on the background estimates, both statistical and systematic, are grouped with the systematic uncertainty on the cross-section.

6.1 Diboson

Diboson events (WW, WZ, and ZZ) events have small cross-sections, comparable to the $t\bar{t}$ cross-section, but when both bosons decay to leptons they can have two opposite-sign charged leptons and neutrinos in the final state. As such, they are near-perfect mimics of the top dilepton signature. Therefore it is not possible to isolate a large sample of such events in the data, and the acceptance calculated from simulated events is used to estimate the background. But because of the real charged leptons and neutrinos in the final state, the simulation is expected to be reliable, with one exception. PYTHIA is known to

not generate enough extra high-p_T jets, so a correction factor derived from a data-simulation comparison is applied to correct the acceptance for events with two or more jets up to the expected level.

6.1.1 Diboson Acceptance

We estimate the diboson background using the acceptance measured with simulated events, corrected for known discrepancies between simulation and data. The acceptance $\mathcal{A}$ is the fraction of generated events passing the top dilepton selection. Multiplied by the theoretical cross-section (σ_{VV}, given in Chapter 2 Section 2.3.3) and the sample integrated luminosity ($\int \mathscr{L} dt$), it determines the number of events in the candidate sample:

$$N_{VV}^{\text{bg}} = \mathcal{A} \cdot \sigma_{VV} \cdot \int \mathscr{L} dt$$

Diboson acceptances are measured in PYTHIA WW, WZ, and ZZ samples of ≈ 2.4 million inclusive generated events each. The correction factors applied to the top acceptance, described in Section 5.2, also apply to these acceptances. The integrated luminosities used are in Chapter 4 Section 4.5 and the corrected acceptances are shown as a function of the number of jets in Table 6.1.

WW

	0 jets	**1 jet**	**$\geq$ 2 jets**
CEM	33.70 $\pm$ 0.41	6.15 $\pm$ 0.17	1.35 $\pm$ 0.08
CMUP	13.56 $\pm$ 0.25	2.29 $\pm$ 0.10	0.64 $\pm$ 0.06
CMX	5.53 $\pm$ 0.17	0.93 $\pm$ 0.07	0.29 $\pm$ 0.04
PHX	12.15 $\pm$ 0.23	1.93 $\pm$ 0.09	0.49 $\pm$ 0.05

WZ

	0 jets	**1 jet**	**$\geq$ 2 jets**
CEM	10.16 $\pm$ 0.23	4.89 $\pm$ 0.15	1.57 $\pm$ 0.09
CMUP	6.24 $\pm$ 0.17	2.52 $\pm$ 0.10	0.82 $\pm$ 0.06
CMX	2.77 $\pm$ 0.12	1.17 $\pm$ 0.07	0.37 $\pm$ 0.04
PHX	4.65 $\pm$ 0.14	2.27 $\pm$ 0.10	0.66 $\pm$ 0.05

ZZ

	0 jets	**1 jet**	**$\geq$ 2 jets**
CEM	6.29 $\pm$ 0.17	1.83 $\pm$ 0.09	1.05 $\pm$ 0.07
CMUP	4.70 $\pm$ 0.14	1.16 $\pm$ 0.07	0.60 $\pm$ 0.05
CMX	1.90 $\pm$ 0.09	0.46 $\pm$ 0.04	0.24 $\pm$ 0.04
PHX	2.28 $\pm$ 0.10	0.63 $\pm$ 0.05	0.25 $\pm$ 0.04

Table 6.1: Diboson acceptances, in percent, as measured with PYTHIA simulated events, categorized by primary lepton type and number of jets, corrected for all known discrepancies between data and simulation. Uncertainty is statistical only.

6.1.2 Correcting for the Number of Jets

One of the largest corrections applied to the acceptance is for the insufficient number of extra jets produced by PYTHIA. To correct the one jet and two-or-more jet acceptances up to the expected level, we compare Z + jets data and simulation. A sample of Zs are selected with two opposite-sign, fully reconstructed electrons or muons, with an invariant mass is in the interval 76 GeV/c^2 $< M <$ 106 GeV/c^2. Both leptons are fully reconstructed, to reduce background in the data sample, particularly from top events, which can skew the distribution of the number of jets. The distribution of the number of jets per event is compared between data and simulation in Figure 6.1.

The fraction of events with zero, one, or two-or-more jets is calculated for data and simulation. The fraction of events with two or more jets is significantly lower in the simulation. Therefore a scale factor, the ratio of the fraction in data to the fraction in simulation, is calculated for each jet multiplicity. These ratios and scale factors are shown in Table 6.2 We multiply the acceptance for each jet multiplicity by the appropriate scale factor, and then rescale the acceptances so that the total acceptance, summed over the number of jets, is unchanged. This last step is necessary to keep the overall normalization correct.

This correction does not apply to the top dilepton acceptance since the two

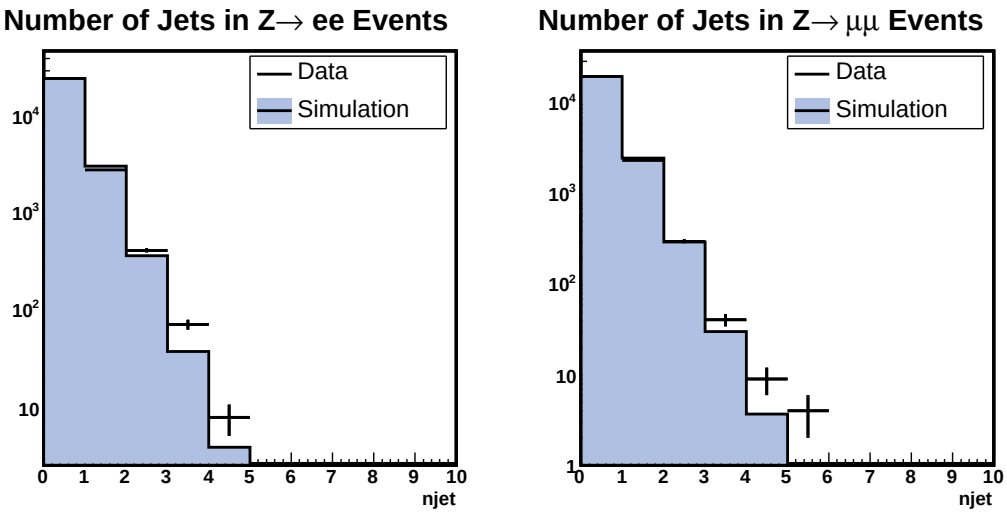

Figure 6.1: Comparison of the number of jets in Z candidate events between data and simulation.

	0 jets	1 jet	$\geq$ 2 jets
Data	.882 $\pm$ .002	.102 $\pm$ .002	.0153 $\pm$ .0008
PYTHIA	.8771 $\pm$ .0006	.1086 $\pm$ .0005	.0142 $\pm$ .0002
scale factor	1.006 $\pm$ .002	.940 $\pm$ .018	1.08 $\pm$ .06

Table 6.2: Fraction of events with a given jet multiplicity and scale factor correcting simulation to the level of data.

b jets are always produced by the top quark decays.

6.1.3 Systematic Uncertainties

Lepton Identification

As with the top dilepton acceptance, the diboson acceptance relies on lepton identification in events with multiple jets. Therefore the uncertainties on well-

reconstructed and track lepton identification that apply to the signal acceptance also apply to the diboson acceptance. Section 5.4 describes the derivation of these uncertainties in detail, and the result is a 1.1% uncertainty on each of the primary and track leptons.

Jet Energy Scale

Another systematic uncertainty common to the signal and the diboson backgrounds is a dependence on the jet energy scale. We repeat in the WW sample the procedure described in Chapter 5 Section 5.4, where the acceptance is recalculated with the jet energies varied within their uncertainties. The effect on the background acceptance is rather larger than on the signal acceptance. The E_T distribution of jets produced through QCD radiation falls off exponentially, but the b jets from top quark decay get a boost from the large top mass, so that the reconstructed jet E_T distribution for jets associated with top events falls off more slowly For this reason, the acceptance for these background events is more sensitive to small changes in the jet energy. Decreasing the jet energies by their uncertainty reduces the WW acceptance by 5.0%, and increasing them by their uncertainty increases the acceptance by 6.6%. Half the full difference, 5.8%, is taken as the systematic uncertainty, which is large compared the 1.3% uncertainty on the signal.

Jet Multiplicity Correction

The correction factor for the jet paucity in simulation, $1.08 \pm .06$, measured in Section 6.1.2, has a non-negligible statistical uncertainty due to the limited size of the $Z+ \geq 2$ jet sample. This uncertainty, 5.5%, is propagated through to the uncertainty on the diboson acceptance.

6.2 W + jets with a Fake Lepton

$W \to \ell\nu$ events with extra jets can pass the dilepton selection if one of the jets is misidentified as a second lepton. Such a fluctuation in the jet fragmentation is rare, but the W cross-section is so large compared to the $t\bar{t}$ cross-section that this is still a very significant source of background. The rate for a jet to be misidentified as a lepton, the "fake rate", is much higher for track leptons than for the fully reconstructed lepton, but there are contributions from both.

The fake rate is the probability for a jet to be misidentified as a lepton. It is derived from the jets in a high-energy photon sample, which are the closest usable analog to jets associated with Ws. Multiplied by number of the jets associated with single Ws, as a function of the jet E_T and $|\eta|$, it yields the expected number of identified "dilepton" events where one of the leptons is fake. To complete the estimate of the number of events with a fake lepton

passing all of the selection, it must be multiplied by efficiencies for the event selection criteria which refer to the existence of a track lepton: the opposite-sign requirement, the restriction on the angle between the track lepton and the $\not\!\!E_T$, and the Z veto.

6.2.1 Definition of the Fake Rate

The probability for a jet to be misidentified as a lepton, the fake rate, is defined as the number of observed fake leptons, divided by the total number of observed jets. For the fake rate denominator, the definition of jet is expanded to include all objects which might possibly be identified as a lepton, so that the numerator is a subset of the denominator. All jets used in the analysis are included, as are all tracks which are not associated with a jet. Further, jets with only a small energy deposition in the calorimeter but a sufficiently high p_T track to produce a fake lepton should not be excluded. For any jet containing a track with p_T greater than the corrected jet E_T, the kinematic information of the jet, including the E_T, is replaced with the corresponding information from the track. This set of regular jets, jets with kinematic information from tracks, and unaffiliated tracks is the complete denominator of the fake rate.

The fake rate must be measured in a data sample where the contamination from true leptons, which would inflate the fake rate, is expected to be negligible.

For this analysis, we measure fake rates in a sample of events triggered by a photon with $E_T > 25$ GeV. We further require the photon to have at least 80 GeV of energy (not E_T), for reasons to be explained more completely later in this section. Taking the jets for the lepton fake rate from a photon + jets sample is a departure from what has been done in past dilepton cross-section measurements. The motivation is that, after Z + jets production, which cannot be used for this yet because there is insufficient data, high-energy photon + jet production is the closest analog to W + jet production. Therefore the jets in that sample should have similar composition, in their quark and gluon content and the color environment, to the jets in the W sample. In previous analyses, the fake rate has been derived from data triggered by high-p_T jets, but consistency checks in simulation show that fake rates derived from multijet events do not predict the observed number of fake leptons in W + jets events well (see Figure 6.3).

6.2.2 Jets from Quarks, Jets from Gluons

Fake rates from multijet events do not adequately predict the number of fake leptons in W + jets events because jets in those events contain a different fraction of jets which begin as quarks rather than gluons, and quark jets have a much higher fake rate than gluon jets.

A large fraction of jets associated with W production are quark jets, but the majority of jets in plain multijet events are gluon jets. Quark jets are prominent in W + jets events because the contributing processes with an incoming gluon is favored over ones with an incoming quark, and the leading-order process with an incoming gluon has an outgoing quark (Chapter 2 Figure 2.7). In multijet events, processes with only gluons (see Figure 6.2) dominate because of the incoming gluons and because the matrix element is larger than for any of the ones involving quarks [15]. For higher-p_T jets, processes with incoming quarks become more important since the x region sampled is higher, but the E_T of jets associated with a W falls off exponentially, so most of them are just above threshold, where the gluon contributions dominate. The fraction of jets which are quark jets in W + jets and multijet events is shown in Figure 6.4.

It would not matter that jets associated with Ws and generic multijet production have different fractions of jets from quarks, except that jets which originate as quarks have a higher fake rate than jets originating as gluons. Because gluons must, in some sense, split into two quarks which fragment further, the average multiplicity of charged particles in a gluon jet is higher than in a quark jet (see Chapter 2 Section 2.3.2). The probability to find only one track in a jet, which essentially amounts to the probability of producing a fake track lepton is rather larger for quark jets than for gluon jets. Fake rates for quark and gluon

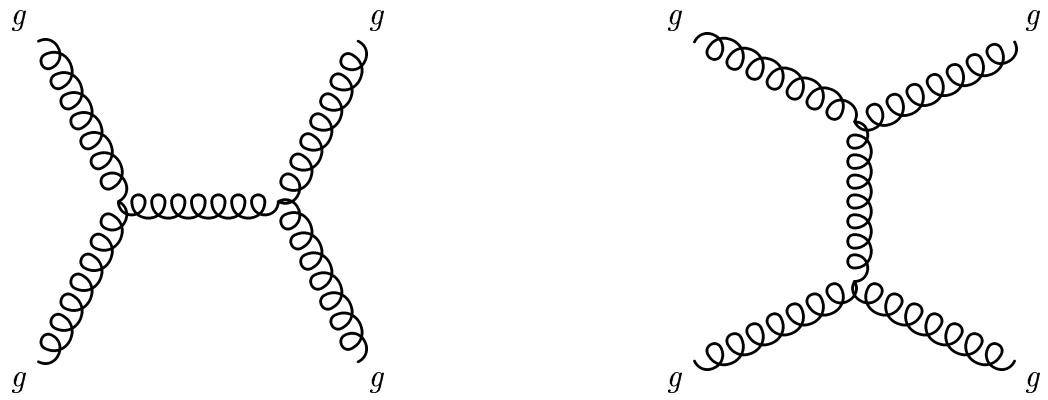

Figure 6.2: Leading diagrams for medium-p_T (20-50 GeV/c) multijet production at the Tevatron. These dominate because of the relative strength of the gluon PDF at low x and because the matrix elements are larger than those including quarks.

jets, as measured in simulated data, are shown in Figure 6.3. The quark jets fake rate is nearly an order of magnitude higher than the gluon jet fake rate.

The mechanisms for the relative frequencies of quark and gluon jets, and for why quark jets are expected to have higher fake rates than gluon jets, are discussed in more detail in Chapter 2, but the implications for fake lepton estimates are clear. To be reliable, background estimates for fake leptons must take into account the relative frequency of quark jets in the process that, with the addition of a fake lepton, will pass the dilepton selection. In this case, that process is predominantly W + jets, which is the motivation from taking the fake rates from a photon + jets sample. The $E_T > 25$ GeV photon sample actually has too many quark jets compared to the W + jets sample (see Figure 6.4),

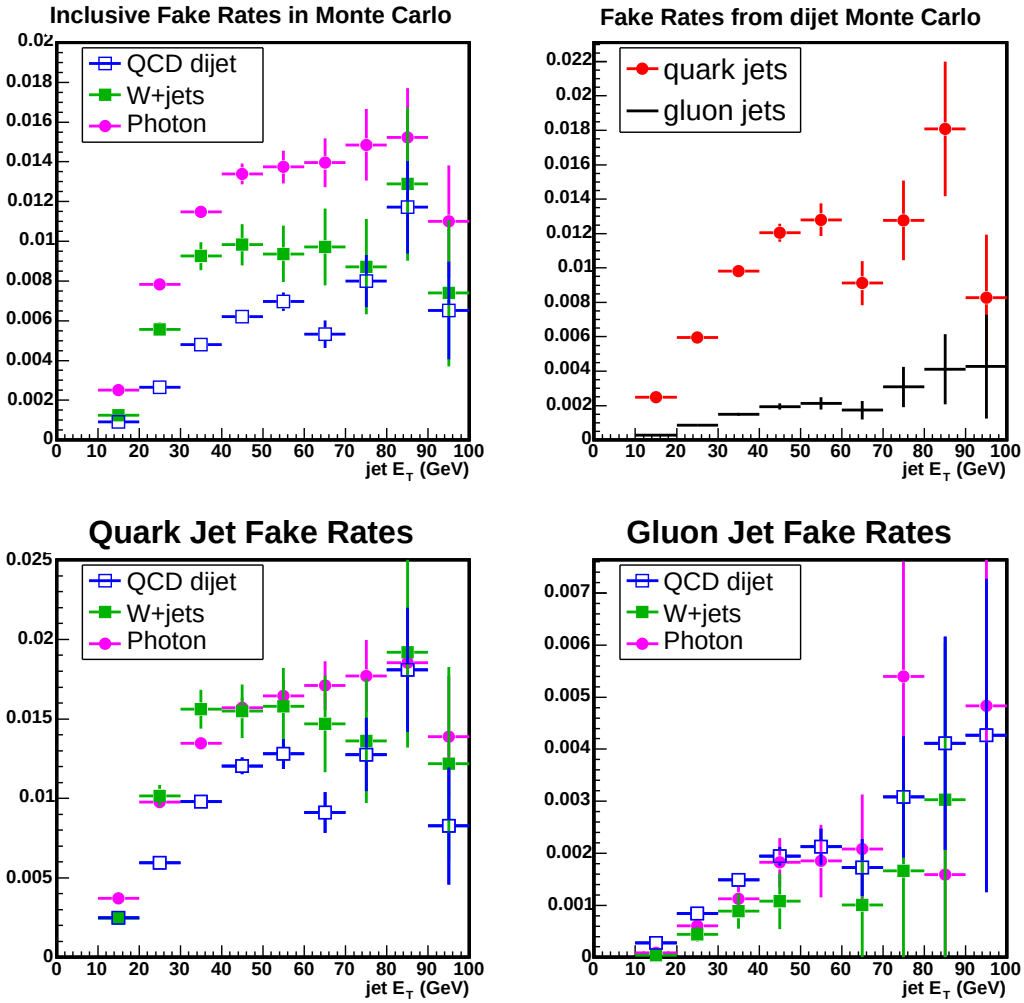

Figure 6.3: Track lepton fake rates in simulation, for different jet samples. as a function of jet E_T. In the upper left plot it is apparent that the fake rate from generic QCD multijet events (blue) disagrees with the fake rate for jets associated with a W. In the upper right, we show the difference between the fake rates for jets matched to a quark and jets matched to a gluon. The lower two plots check the matching of the quark and gluon jet fake rates between samples separately.

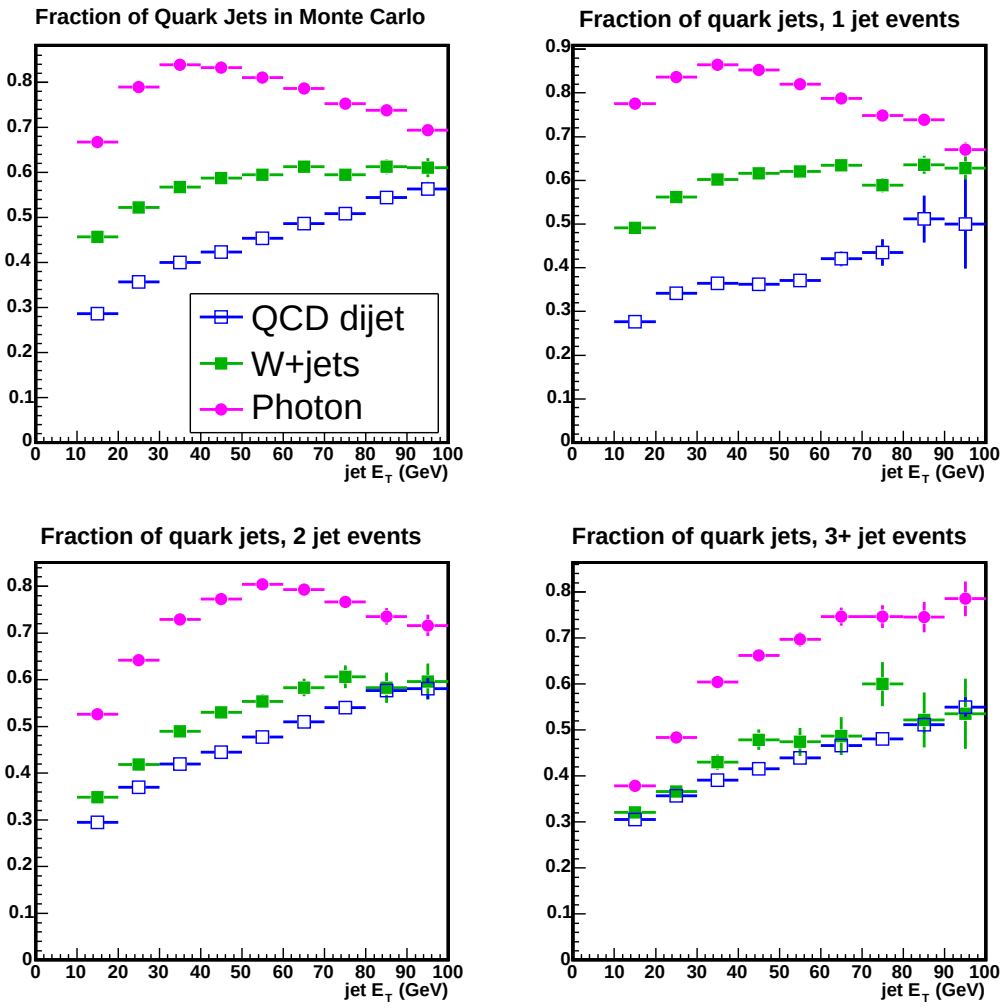

Figure 6.4: The fraction of jets matched to a quark in simulated W, photon, and multijet events. For any jet multiplicity, the fraction of quark jets is highest in the photon sample, followed by the W sample and the multijet sample. One can also see that in the W and photon samples, the fraction of quark jets falls off with increasing jet multiplicity.

but by requiring the photon to have $E > 80$ GeV, we require a larger fraction of the proton momentum, and the region of the PDFs sampled is similar to W production.

The fraction of jets which are quark jets also depends strongly on the number of jets in the event, as can also be seen in Figure 6.4. Additional jets can be produced as radiated gluons from either the process with an incoming quark or the one with an incoming gluon, so that the quark jet dominance slowly washes out with increasing jet multiplicity. For this reason, we measure the fake rate separately for events with one, two, or three or more jets.

Fake rates for track leptons, as measured in the 80 GeV photon sample, are shown in Figure 6.5 as a function of E_T, $|\eta|$, and the number of jets in the event. Fake rates are measured for all four primary lepton types as well as for track leptons. Because there are so few events in the numerator for the primary leptons, the fake rate is of necessity inclusive rather than measured separately for each jet multiplicity. Fake rates for fully-reconstructed leptons are shown in Figure 6.6 as a function of E_T and $|\eta|$. In these figures the smaller fake rates for fully reconstructed leptons, compared to track leptons, are apparent.

There is one additional subtlety in using the 80 GeV sample for the fake rates. In photon plus one jet events, conservation of momentum implies that it is hard to find such events where the E_T of the jet is significantly lower than

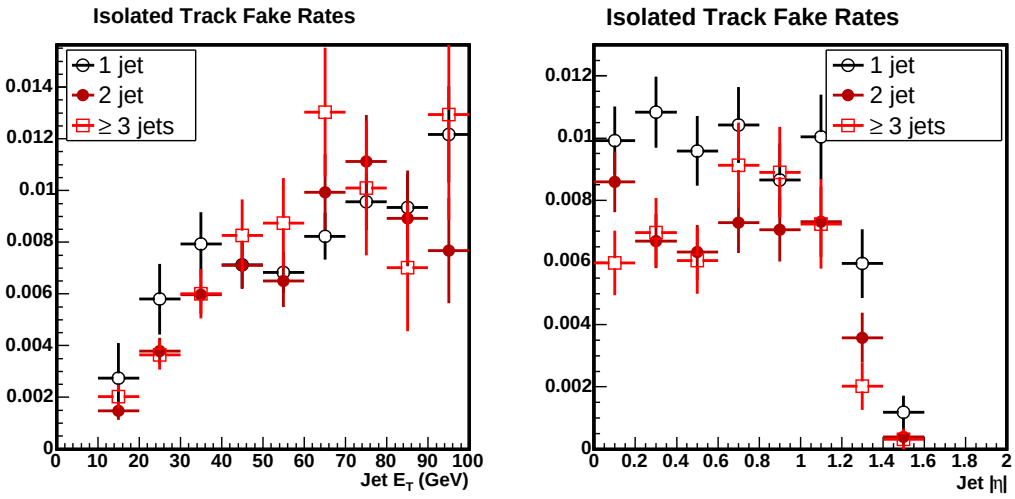

Figure 6.5: Track lepton fake rates from photon data selected using an 80 GeV photon, as a function of E_T and $|\eta|$.

the photon energy. As such, there are very few events in the lowest E_T part of the fake rate. This is not crucially important since that is an input to the zero jet event count prediction, which does not enter into the cross-section. But it is possible to shore up the one jet fake rate by including the Z plus one jet sample. The Z + jets sample is a near-perfect analog to the W + jets sample, up to the difference in their masses. There are not enough events for it to be useful in general for the fake rate, because most of the events with any jets have one only one jet, near the E_T threshold. But, this is exactly where the gap in the photon + jets fake rate is. Therefore the total fake rate used to predict the number of fake leptons in the zero jet dilepton sample is the combined rate from the photon and Z plus jets sample.

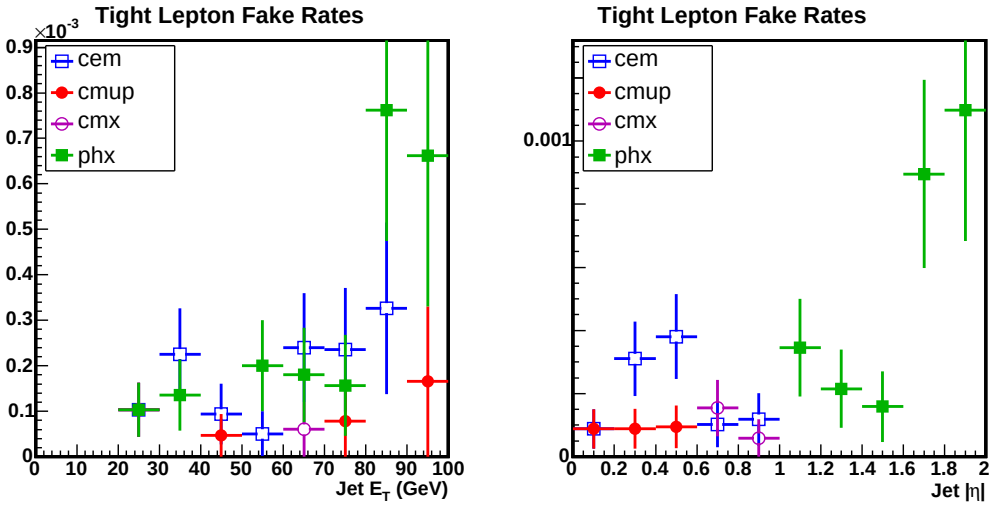

Figure 6.6: Fake rates for fully reconstructed leptons from photon data selected using an 80 GeV photon, as a function of E_T and $|\eta|$.

6.2.3 Testing the Fake Rate

The robustness of the track lepton fake rate is tested in both simulation and data. A lower kinematic threshold, 15 GeV for both the track leptons and jets, is used for these tests in order to give them more statistical power. This adds many jets to the sample because of the exponential falloff of the E_T distribution for jets. Also, the fake rate for 15 GeV track leptons is higher than for 20 GeV track leptons, because it requires a less severe fluctuation in the fragmentation.

Tests in Simulated Data

In the simulation it is possible to test the exact procedure used in the data, by using the fake rate from photon events to predict the number of fake leptons in jets in W events. Events with 0, 1, and ≥ 2 jets in addition to the fake lepton are considered, and in each jet multiplicity the predicted and observed number of fake leptons agree within their uncertainties. Figure 6.7 shows the predicted and observed number of fake leptons as a function of the E_T of the faking jet, for each jet multiplicity, and Table 6.3 shows the integral results numerically.

The fake rate from jets associated with an 80 GeV photon appears to systematically slightly overestimate the number of fake leptons observed in jets associated with a W. This effect is not statistically significant anywhere except in events with zero additional jets, and we will include that 18% discrepancy in the systematic uncertainty, as discussed later in this section.

Z + jets Data Test

In the data a direct test of the fake rate from photon data used in the analysis is possible. The number of fake leptons predicted in the Z + jets data is compared to the number observed, for events with 0, 1, and ≥ 2 jets in addition to the fake lepton. Note that only the fake rate from photon plus one jet is used to predict the number of fake leptons coming from the Z plus

Predict simulated W+jets with simulated γ and Z+jets

	predicted	observed
1 jets	5473 ± 147	4480
2 jet	1332 ± 200	1047
$\geq$ 3 jets	304 ± 67	226
$\geq$ 2 jets	1636 ± 211	1273

Predict Z+jets data with γ+jets data

	predicted	observed
1 jets	100 ± 13	101
2 jet	28 ± 2	26
$\geq$ 3 jets	12 ± 1	13
$\geq$ 2 jets	40 ± 2	39

Table 6.3: Predicted and observed number of fake leptons in the data and simulation cross-checks. The only statistically significant discrepancy observed is in the 1 jet category in the simulation cross-check, which is part of the basis of the systematic uncertainty on this background estimate.

one jet sample. Although the event sample is small, no statistically significant discrepancy is observed for any jet multiplicity in this test. Figure 6.8 shows the predicted and observed number of fake leptons as a function of the E_T of the faking jet, for each jet multiplicity, and Table 6.3 shows the integral results numerically.

6.2.4 Normalization from W + jets Data

The fake lepton estimate is normalized to the size of the W + jets sample with which the fake rate is multiplied. A sample of W + jets data is selected, where the W may be identified by any of the primary lepton types, and the event

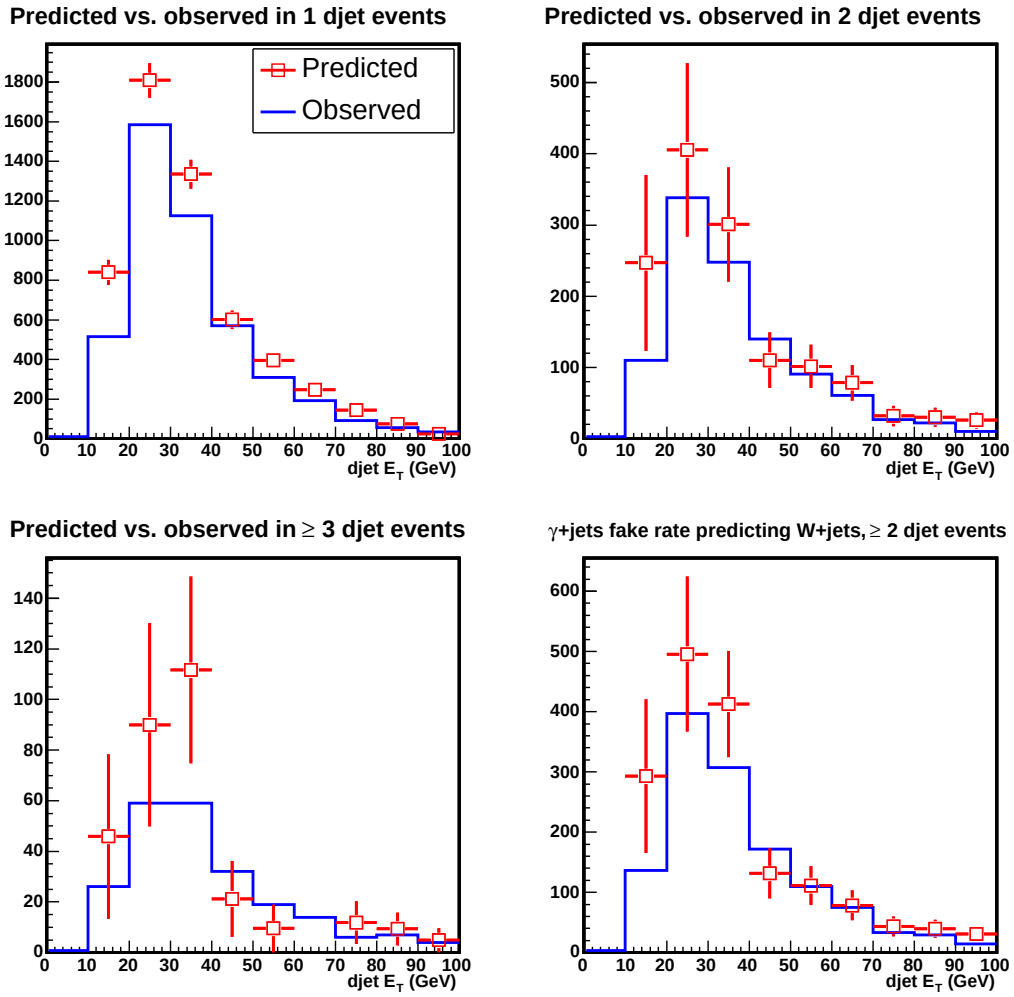

Figure 6.7: Predicted and observed number of isolated tracks in simulated W + jets events, as a function of the number of jets in the event. The fake rate used is derived from photon+jets and (for events reconstructed with zero jets) Z+jets simulation. The over-prediction for zero jet events is incorporated in the systematic uncertainty on the fake lepton background.

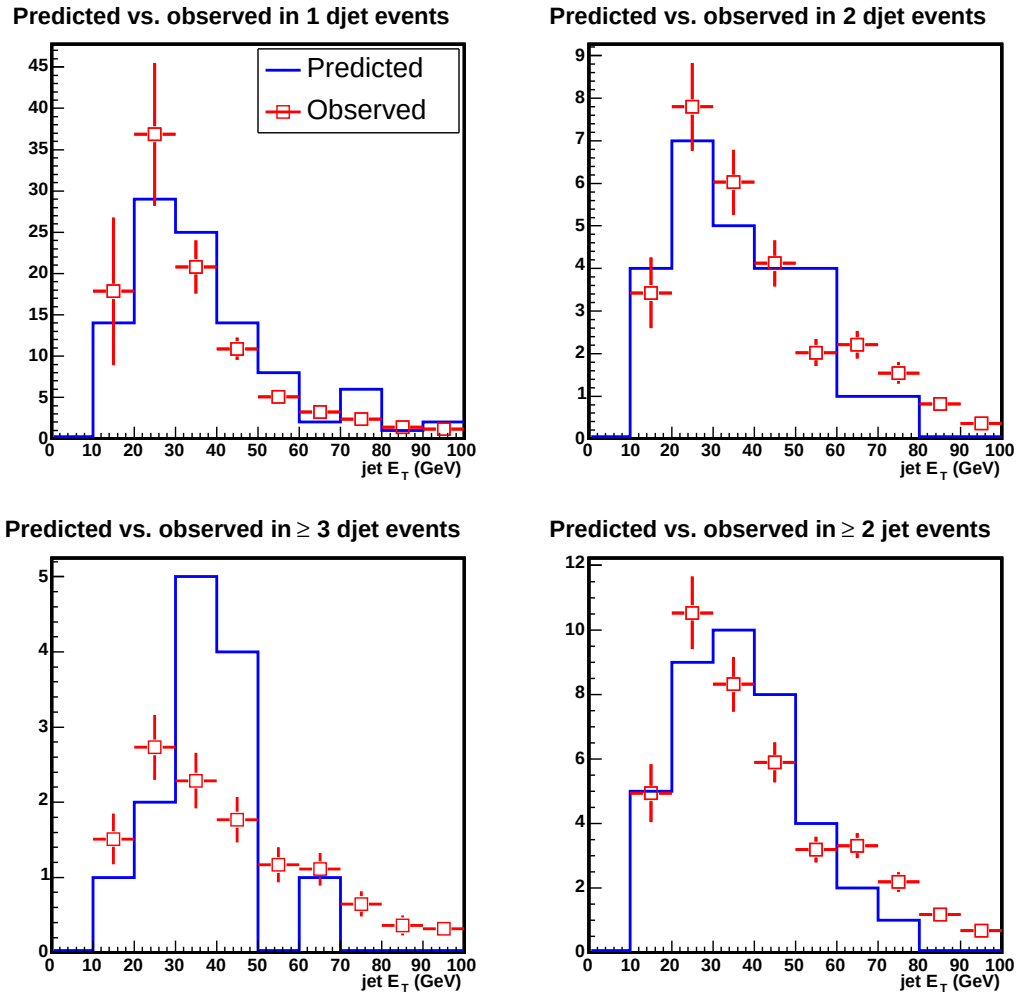

Figure 6.8: Predicted and observed number of isolated tracks in Z + jets data, as a function of the number of jets in the event. The fake rate used is derived from photon+jets data. The shape appears to be not well-predicted for events with three or more jets, but there are only a handful of events in this category, and it seems likely that the distribution fluctuated.

passes all of the event selection except for the criteria referring the existence of a track lepton. That is, events have a well-reconstructed lepton, corrected $\not\!\!E_T >$ 25 GeV, and pass the $\Delta\varphi$ requirements for the fully reconstructed lepton and jets relative to the $\not\!\!E_T$, but the Z veto, track lepton $\Delta\varphi$, and opposite-sign requirements are ignored.

For each primary lepton type, the E_T and $|\eta|$ of the jets in the event are recorded, separately for events with one, two, or three or more jets. The "jets" used are jets according to the expanded definition used in the fake rate, described above in Section 6.2.1. No lepton veto is applied to the jets. That is, jets which are actually leptons which have been included in the jet collection due to their track or calorimeter signature are included. They represent a very small fraction of the jets used.

The E_T and $|\eta|$ distributions of jets from these W samples are multiplied by the fake rate derived from events with the same number of jets, to give the basis for the fake lepton prediction for events reconstructed with one less jet than that number. That is, the fake rate from events with two jets is multiplied by the jet distributions for W plus one jet events to predict the number of dilepton events with one jet and a fake lepton.

Fake Leptons from Sources Other than W + jets

So far we have worked under the assumption that all events with a fake lepton passing our selection come from W + jets, and in fact, they constitute the majority of events chosen by the above selection. Other processes can also produce a lepton, $\not\!\!E_T$, and a large number of jets, one of which can then be misidentified, and these are included in the normalization sample. The inclusion of all potential sources of fake leptons is a benefit to using an inclusive W + jets data sample to normalize the estimate; other sources, such as $t\bar{t}$, $W\gamma$, and pure-QCD multijet do not, in general, have to be accounted for explicitly.

As determined by calculating an acceptance for the specialized W + selection, the only other process contributing significantly to the W + jets sample is $t\bar{t}$, which, at a cross-section of 6.7 pb. Lepton plus jets $t\bar{t}$ decays are 19% of the jets in the three or more jet sample. They are ignorable in the one and two jet samples, where they are less than 1% of jets, because of the inherent large jet multiplicity of $t\bar{t}$, especially when one or more of the Ws decays to quarks.

The lepton plus jets $t\bar{t}$ contribution, because of its size, deserves separate consideration. It need not be treated separately, as long as it behaves like the normal W + jets sample. Study of simulated $t\bar{t}$ events indicates that over 90% of the jets in lepton plus jets events come from quarks, a very different fraction

than for any other sample considered, although many of the jets are from heavy quarks, either the b from the t decay or a c from $W^+ \to c\bar{s}$. The isolated track fake rate for jets from a light quark is higher than the rate for jets from a heavy quark, because heavy quarks form hadrons which often decay through several stages, producing a large track multiplicity. As a result, the fake lepton comes from a jet from a quark from the W decay in about 90% of the simulated events with a fake lepton. If the lower fake rate from the heavy quarks compensates for the preponderance of quark jets, the fake rate designed for W + jets may be usable.

We can test the validity of the 80 GeV photon + jets fake rate for fakes from top in the same way as the validity can be checked for fakes from W + jets, in simulation. Figure 6.9 shows the comparison. The integral number predicted is 585±128, compared to 424 observed events. The agreement is good but the statistical uncertainty on the fake rate is large for events with three or more jets, where all the data from $t\bar{t}$ events is. There is a larger sample of jets available in the simulated W + jets sample, so, as an alternative, we can check whether the fake rate for jets associated with a W can predict the fake leptons observed in $t\bar{t}$ lepton plus jets events. This comparison is shown in Figure 6.10. The shape is not very well predicted, and 518±45 fake isolated tracks are predicted to the 424 observed. The level of disagreement in the normalization is not very

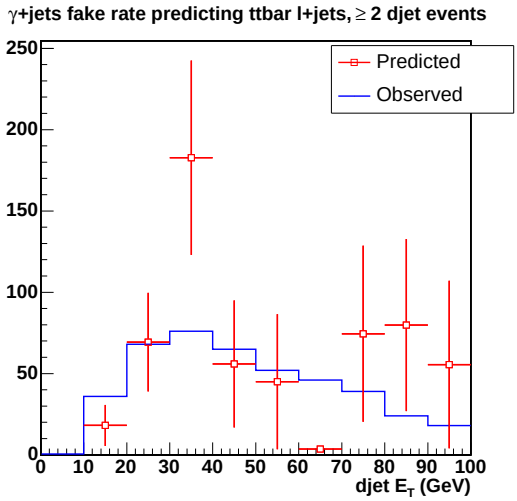

Figure 6.9: Testing the applicability of the 80 GeV photon + jets fake rate to fake leptons from top in the lepton + jets mode, in simulation. No fake leptons are observed from events with one jet.

significant and is the same size (18%) as the disagreement between the predicted and observed number of fake track leptons in the W + 1 jet simulation, so it will be covered by the systematic uncertainty. Therefore we will use the photon + jets fake rate to predict the number of fake leptons, regardless of the source of the events passing the W + jets selection.

6.2.5 Contribution from Fake Fully Reconstructed Leptons

The fake rates for fully reconstructed leptons are an order of magnitude or more smaller than for track leptons (see Figure 6.6), but they still contribute a

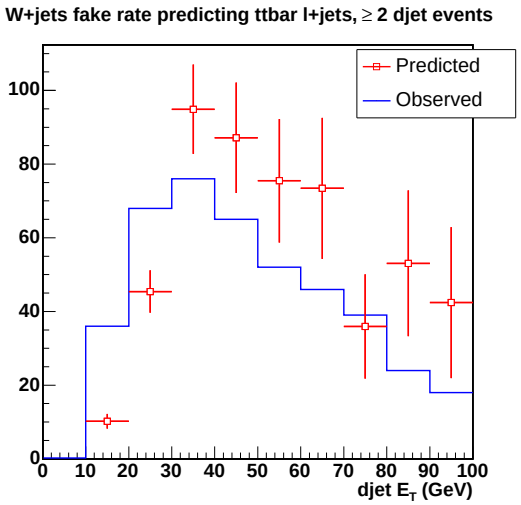

Figure 6.10: Testing the applicability of the using the same fake rate for from top in the lepton + jets mode as is used to predict fake leptons from W + jets events, by using a fake rate derived from the W + jets sample, in simulation. No fake leptons are observed from events with one jet.

non-negligible amount to the total number of events with fake leptons.

Primary lepton fake rates are measured in the same way as track lepton fake rates, except that a single fake rate is derived which includes events with any number of jets, to compensate for the far smaller number of fake leptons observed. With the current sample size, the fake rates for different jet multiplicities are not statistically distinguishable, so this is an acceptable strategy.

Ideally, the primary lepton fake rates would be multiplied by the jets from a W sample selected as described above except that the lepton from the W would be identified as a track lepton. The relatively large fake rate for track leptons, combined with the fact that the track lepton criteria are looser than

the lepton trigger criteria so that some events will be absent from the sample, makes this unfeasible. Therefore the fake rates are multiplied by the same W jet distributions as above, summed over the primary leptons used to generate the W jet distributions, and rescaled for the ratio of the W acceptance for track leptons to the total W acceptance for fully reconstructed leptons. That acceptance ratio, derived from simulated W samples with matrix elements generated by ALPGEN and using PYTHIA for parton showering, is about 1.1.

Total fake lepton predictions for 1.07 fb^{-1} at this stage of the calculation are shown in Table 6.4. The numbers are the result of multiplying the jet distributions with the fake rates. The effects of the opposite-sign, Z veto, and track lepton $\Delta\varphi$ criteria have not been included yet. Contributions from fully reconstructed and track leptons are shown separately. Events with a fake fully reconstructed lepton contribute only about 6% of the total for events with two or more jets.

6.2.6 Efficiency of the Z Veto and Track Lepton $\Delta\varphi$ Criteria for Events with a Fake Lepton

The efficiency of the kinematic selections involving a track lepton is measured, as a function of the number of jets in the event, using the same ALPGEN

fake track lepton contribution

	$n_j = 0$	$n_j = 1$	$n_j \geq 2$
CEM	182 ± 12	61.9 ± 3.9	23.2 ± 2.5
CMUP	96 ± 6	29.0 ± 1.9	11.6 ± 1.4
CMX	50 ± 3	14.8 ± 1.0	5.9 ± 0.7
PHX	82 ± 6	27.6 ± 1.8	11.0 ± 1.6

fake primary lepton contribution

	$n_j = 0$	$n_j = 1$	$n_j \geq 2$
CEM	14.4 ± 2.1	4.7 ± 0.6	1.5 ± 0.2
CMUP	5.3 ± 1.4	1.4 ± 0.3	0.6 ± 0.1
CMX	3.8 ± 1.1	0.9 ± 0.2	0.2 ± 0.1
PHX	10.3 ± 1.5	3.2 ± 0.4	1.1 ± 0.2

total

	$n_j = 0$	$n_j = 1$	$n_j \geq 2$
CEM	197 ± 12	66.6 ± 4.0	24.7 ± 2.5
CMUP	101 ± 7	30.4 ± 1.9	12.2 ± 1.4
CMX	54 ± 4	15.7 ± 1.0	6.2 ± 0.7
PHX	92 ± 6	30.8 ± 1.9	12.1 ± 1.6

Table 6.4: Fake predictions split by primary lepton type, jet bin, and fully reconstructed vs. track lepton contributions.

+ PYTHIA simulated W sample described in the previous subsection. The combined efficiencies are 0.82 ± 0.02 for events with zero jets, 0.85 ± 0.01 for events with one jets, and 0.89 ± 0.02 for events with two or more jets.

6.2.7 Opposite-Sign Fraction

A final piece of the W plus fake lepton estimate is the fraction of such events where the real and fake lepton have opposite sign. Because leading diagrams for W + jets production have the W recoiling against a quark which becomes a jet in the event, the charge of the tracks in that jet can be expected to be correlated to the charge of the W.

Ideally, one would like to measure the opposite-sign fraction in simulated W + jets events. The results from simulation, for PYTHIA alone or ALPGEN plus PYTHIA, are shown in Table 6.5. Both generators agree that the charge correlation is large, and falls off with increasing jet multiplicity. Agreement between the two is expected because both use PYTHIA to do parton showering. The uncertainties on the ALPGEN numbers are smaller because higher-order contributions (i.e. those with more high-p_T jets) are calculated explicitly, and the results recombined with the appropriate weights.

The difficulty with the numbers from the simulation is that the charge correlation, like the fake rate, depends on the details of the simulation of jet frag-

mentation. When the number of predicted events is compared to the number of observed candidate events in the zero jet data, the agreement is good within the statistical uncertainties, but when the predicted opposite-sign fraction fros simulation is applied, the number of predicted same-sign events is too small compared to the number observed, and the number of opposite-sign events correspondingly too large. This is suggestive that the simulation overestimates the fraction of events with a fake lepton where the lepton have opposite charge.

Because of the apparent problems with measuring the W + jets charge correlation in the simulation, we measure it in the data instead, using a subset of the candidate events with zero jets. Events with zero jets have a large contribution from events with a fake lepton, because of the large $W + 1$ jet cross-section, but very few events from $t\bar{t}$, because of its small cross-section and the two b quarks in the final state. We can further enhance the fraction of events with a fake lepton by requiring a significant amount of energy in the region of the hadronic calorimeter at which the track points. This about 70% efficient for events with a fake lepton and $Z/\gamma^{\star} \rightarrow \tau\tau$ events, but reduces all other contributions by a factor of ten.

Applying this leaves a sample dominated by events with a fake lepton, with a small contribution from $Z/\gamma^{\star} \rightarrow \tau\tau$ which is in the minority because of its smaller cross-section relative to $W + 1$ jet and smaller acceptance. The esti-

mated number of zero-jet events from sources with two real leptons (including $Z/\gamma^\star \to \tau\tau$) can be subtracted from the number of observed events, both for all events and for those with opposite sign. $67 \pm 3\%$ of remaining events are opposite-sign, which is taken as the opposite-sign fraction for all events with zero jets. As a check of this, the opposite-sign is also measured using *all* zero jet events. The fraction of events which have two real leptons is then larger, so there is a larger uncertainty on the measured fraction, based on the uncertainty on the number of events subtracted. In the full zero jet sample, an opposite-sign fraction of $69 \pm 5\%$ is measured. The derivation of these numbers is detailed in Table 6.6.

We cannot apply the same procedure to derive the opposite-sign fraction for higher jet multiplicities because the contribution from $t\bar{t}$ to the candidate sample becomes non-negligible. To obtain the charge correlation for events with more jets, we extrapolate the fraction from zero jet events using the jet multiplicity dependence of fractions measured in simulation. The reasoning behind the extrapolation is that the charge correlation for W+1 jet events is slowly washed with increasing jet multiplicity by the addition of fake leptons from jets which have a charge correlation of 50% (i.e., no correlation). Assuming that the fraction added is the same with each additional jet, the dependence of the opposite-sign fraction on the number of jets can be deduced. Solving for the

undiluted fraction x in the ALPGENsample (see Table 6.5 for input numbers):

$$\begin{aligned}
(.81 \pm .03)x + (.50)(1-x) &= .73 \pm .02 \\
((.81 \pm .03) - .50)x &= (.73 \pm .02) - .50 \\
(.31 \pm .03)x &= .23 \pm .02 \\
x &= .74 \pm .10
\end{aligned} \tag{6.6.1}$$

This can be checked by using the formula to predict the opposite-sign fraction for higher jet multiplicities in the PYTHIA sample. The opposite sign fraction for jet bin $(i+1)$ given opposite-sign fraction f_i in jet bin i is $f_{i+1} = (f_i)x + (.50)(1-x) = (f_i - .50)x + .50$. So, for PYTHIA, $f_1 = .74 \pm .03$, which compares favorably to the $.76 \pm .05$ from Table 6.5. Similarly, this method predicts $.68 \pm .04$ for events with two or more jets, where we measure $.61 \pm .10$. We can also predict the ≥ 2 jet fraction in ALPGEN, and find $.67 \pm .03$. This doesn't agree especially well with the measured $.75 \pm .05$, but the .75 is somewhat suspect anyways since it is higher than the 1 jet number.

Applying this technique to the $.67 \pm .03$ from the zero jet data yields opposite-sign fractions of $.63 \pm .03$ for events with one jet and $.59 \pm .03$ for events with two or more jets.

Again, we must account for the contribution from fake leptons from $t\bar{t}$ events

in the lepton plus jets channel. In the simulation, 78.9% of such events have opposite-sign leptons. The charge correlation here can be understood by recalling that the summed charge of the quark pair produced will be the opposite of the lepton charge, so that to produce a hadron of the same charge will require the production of a large number of additional quark-antiquark pairs and is unlikely to result in an isolated track. The studies above indicate that the simulation tends to overestimate charge correlations, so we dilute this number toward 50% using the same logic used to propagate the opposite-sign fraction to higher jet multiplicities. In this case, the dilution is by the amount required to scale .83, the PYTHIA number from Table 6.5, down to the .67 measured in data. This predicts a .66 opposite-sign fraction in data, for events with three or more jets.

To make the fake lepton background prediction for events with two or more jets, we must combine the two opposite-sign fractions, for W+jets and $t\bar{t}$ lepton+jets, in proportion to the number of fake leptons contributed by each process. Because the predicted fraction of $t\bar{t}$ in the W+jets data sample is based on an acceptance measured in simulation, this introduces a dependence on the input top cross-section. This dependence is iterated out by feeding the measured cross-section back in to the background calculation, a process that quickly converges (see the next section for more discussion on iterating out dependence on

W+jets, isolated track lepton

	0 jets	1 jet	$\geq$ 2 jets
PYTHIA	.83 $\pm$ .01	.76 $\pm$ .05	.61 $\pm$ .10
ALPGEN	.81 $\pm$ .02	.73 $\pm$ .02	.75 $\pm$ .04

Table 6.5: Fraction of lepton + isolated track events with opposite sign in simulated W + jets samples.

the top cross-section). The result is that 23% of jets in the normalizing W+jets sample are predicted to be from $t\bar{t}$, and the reweighted opposite-sign fraction is .61.

Having assembled all of the pieces of the W plus fake lepton background estimate, an important check can be done. Same-sign events in the data are expected to be predominantly events with a fake lepton, so that a comparison of the number of same-sign events predicted and observed is a check of the fake lepton estimate. These numbers are shown in Table 6.7. The contribution from events with a fake lepton and the contribution from all other sources are tabulated separately, although the uncertainties quoted are from the fake lepton estimate alone. In jet multiplicities other than zero, the simulation numbers seem to better reproduce the number of observed opposite-sign events. For the time being, this is covered by the systematic uncertainties on the fake lepton estimate, but it merits further study.

Zero jet events in data

	all events		opposite-sign events	
	all	hadronic	all	hadronic
WW	94.4	8.0	91.2	7.8
WZ	14.2	1.0	10.0	0.7
ZZ	2.6	0.	2.4	0.
$Z/\gamma^\star \to ee$	74.6	9.7	74.6	9.7
$Z/\gamma^\star \to \mu\mu$	19.3	1.5	19.3	1.5
$Z/\gamma^\star \to \tau\tau$	36.3	25.4	35.5	24.8
subtotal	241.4	45.6	233.5	44.5
data	547	245	443	179
data - subtotal = "fakes"	305.6	199.4	209.5	134.5

Table 6.6: Derivation of opposite-sign fraction for fakes from data. Attributing all data counts in excess of the diboson and Drell-Yan contributions to events with fake leptons, the opposite-sign fractions measured are 209.5/305.6 = .69 $\pm$ .05 in all zero jet data and 134.5/199.4 = .67 $\pm$ .03 in data where the track lepton is matched to a significant energy deposit in the hadronic calorimeter.

Same-sign Events

Same-sign fraction from simulation			
	0 jets	1 jet	$\geq$ 2 jets
fakes predicted	69.2 $\pm$ 2.4	32.9 $\pm$ 1.1	11.8 $\pm$ 0.7
total predicted	77.9 $\pm$ 2.4	36.3 $\pm$ 1.1	12.6 $\pm$ 0.7
candidates	104	28	12
Same-sign fraction from data			
fakes predicted	120.2 $\pm$ 23.2	45.1 $\pm$ 8.6	19.2 $\pm$ 3.8
total predicted	129.2 $\pm$ 23.2	48.6 $\pm$ 8.6	20.1 $\pm$ 3.8
candidates	104	28	12

Table 6.7: Predicted and observed same-sign candidates, where the opposite-sign fraction is derived from ALPGEN simulation or from zero jet data.

6.2.8 Systematic Uncertainties

There are three sources of uncertainty in the W + fake lepton estimate: the statistical uncertainty, which varies between 5-10% depending on the jet multiplicity, the systematic uncertainty on the overall normalization of the estimate, and the systematic uncertainty on the fraction of events with a fake lepton where the leptons have opposite sign.

The uncertainty on the overall normalization comes from the largest observed discrepancy in the data and simulation cross-checks described in Section 6.2.3. This is the 18% difference between the observed and predicted number of events in simulated events with zero reconstructed jets, which is also the right size to cover possible discrepancies between the fake rate for W + jets and the one for $t\bar{t}$ lepton plus jets.

The uncertainty on the opposite-sign fraction is the statistical uncertainty on the calculated opposite-sign fraction in data propagated to the ≥ 2 jet bin in Section 6.2.7 and is 5%.

Including both systematics and the 6% statistical uncertainty in the ≥ 2 jet bin, the total uncertainty on this background for the cross-section measurement is 20%.

6.3 Drell-Yan

Drell-Yan ($p\bar{p} \rightarrow Z/\gamma^\star \rightarrow \ell\ell$) events with $\not\!\!E_T$ are a significant source of background for $t\bar{t}$ since the inclusive cross-section is large and there are two real leptons in the final state. In the case of $Z/\gamma^\star \rightarrow \tau\tau$, the $\not\!\!E_T$ is mostly from the neutrinos from the τ leptons decaying, and the background calculation is based on simulation, similar to the case for dibosons. For $Z/\gamma^\star \rightarrow ee/\mu\mu$, there are no neutrinos in the final state and any $\not\!\!E_T$ reconstructed is a result of the mismeasurement of one or more objects, either lepton or jet, in the final state. As this process is not expected to be simulated adequately, and it is possible to select a sample of events enriched in $Z/\gamma^\star \rightarrow ee/\mu\mu$ content in the data, a hybrid method using both data and simulation is used. The drawback to this method is that it is limited by the sample sizes in data and simulation, and carries the largest uncertainty of any input to the cross-section.

6.3.1 $Z/\gamma^\star \rightarrow \tau\tau$

It is difficult to isolate $Z/\gamma^\star \rightarrow \tau\tau$ events in the data, but simulation is expected to do a reasonable job because of the real neutrinos responsible for the $\not\!\!E_T$ in the final state. Therefore the estimate of the $Z/\gamma^\star \rightarrow \tau\tau$ background is calculated in the same way as the diboson contributions, as detailed in Section

$Z/\gamma^\star \to \tau\tau$

	0 jets	**1 jet**	**$\geq$ 2 jets**
CEM	0.53 $\pm$ 0.03	0.40 $\pm$ 0.03	0.12 $\pm$ 0.01
CMUP	0.23 $\pm$ 0.02	0.19 $\pm$ 0.02	0.04 $\pm$ 0.01
CMX	0.14 $\pm$ 0.02	0.08 $\pm$ 0.01	0.03 $\pm$ 0.01
PHX	0.12 $\pm$ 0.01	0.09 $\pm$ 0.01	0.03 $\pm$ 0.01

Table 6.8: $Z/\gamma^\star \to \tau\tau$ acceptance, in percent, as measured in PYTHIA Monte Carlo simulation, by primary lepton type and number of jets, corrected for all known discrepancies between data and simulation. Uncertainty is statistical only.

6.1, including the rescaling to compensate for the deficit in generated extra jets. A sample of 7 million simulated $Z/\gamma^\star \to \tau\tau$ events generated using PYTHIA Monte Carlo is used to calculate the acceptance. Events are generated with $M(Z/\gamma^\star) > 30$ GeV; the corresponding cross-section is 327 $\pm$ 7 pb [4]. Table 6.8 shows the acceptance for this process.

6.3.2 $Z/\gamma^\star \to ee/\mu\mu$

Because the $\not\!\!E_T$ from mismeasured objects is not expected to be well-modeled in the simulation, the background estimate for $Z/\gamma^\star \to ee/\mu\mu$ is based on events observed in the data. We select a sample of events passing most of the standard analysis selection, with a few modifications to enhance the $Z/\gamma^\star \to ee/\mu\mu$ content. After subtracting the expected contributions to the sample from other sources, the number of events in this sample is the normalization

for the estimate. Then, the same selection is made in a sample of simulated $Z/\gamma^\star \to ee/\mu\mu$ events, and the properties of these events are used to transform the number of events passing the modified selection into the number of events passing the standard selection.

Normalization from Data

Events in the data sample that provides the basis for the $Z/\gamma^\star \to ee/\mu\mu$ background estimate pass all of the standard analysis requirements, with a few exceptions intended to enhance the Drell-Yan content of the sample. The dilepton invariant mass must be in the interval 76 GeV/c^2 $< M <$ 106 GeV/c^2. Also, events with any number of jets are allowed. The number of events selected is shown in Table 6.9. We count events with $\not\!\!E_T >$ 25 GeV and $\not\!\!E_T >$ 40 GeV for each primary lepton type (n_{25} and n_{40} in the formulas below). The two $\not\!\!E_T$thresholds are the ones used for candidate events with dilepton invariant mass inside and outside the Z resonance region. The number with $\not\!\!E_T >$ 25 GeV is used to calculate the number of events outside the 76 GeV/c^2 $< M <$ 106 GeV/c^2, and the number with $\not\!\!E_T >$ 40 GeV to calculate the number inside that window. Also, because there are very few $Z/\gamma^\star \to \mu\mu$ events with significant $\not\!\!E_T$, the muon numbers are collected into one category, and at the end, the estimate is divided between the CMUP and CMX categories using their relative geometrical

acceptances, 64:36.

Non-Drell-Yan Contributions to the Normalization Sample

The data sample is expected to contain many $Z/\gamma^\star \to ee/\mu\mu$ events with $\not\!\!E_T$, but it also contains events from many other sources, including $t\bar{t}$, WW, WZ, ZZ, $Z/\gamma^\star \to \tau\tau$, and W + jets events with a fake lepton, which are collectively referred to as "feeddown". These other contributions can be calculated and subtracted from the number observed to yield the number attributable to $Z/\gamma^\star \to ee/\mu\mu$. are included as sources. All of these "backgrounds" are calculated exactly as they are for the main analysis, except that the event selection is the altered one used for the data normalization counts.

The $t\bar{t}$ contribution is initially normalized to the theoretical cross-section, 6.7 pb, corresponding to the top quark mass at which the simulated $t\bar{t}$ is generated, 175 GeV/c^2. When the full cross-section measurement is completed, the resulting cross-section is put back into the feeddown calculation, and the cross-section recalculated. This process is repeated, until the change observed in the measured cross-section is less than 0.1%. Typically two iterations are required and the resulting change in the cross-section is 0.1 pb or less.

The total number of events expected from sources other than $Z/\gamma^\star \to ee/\mu\mu$ is shown, both before and after the top cross-section dependence is iterated out,

	Raw data counts		Initial feeddown		Final feeddown	
	n_{25}	n_{40}	n_{25}^{feeddown}	n_{40}^{feeddown}	n_{25}^{feeddown}	n_{40}^{feeddown}
CEM	148	54	74.	39.	77.	41.
CMUP+CMX	107	42	55.	28.	56.	30.
PHX	53	19	31.	15.	32.	15.

Table 6.9: Counts from the data used to normalize the $Z/\gamma^\star \rightarrow ee/\mu\mu$ background estimate, and the expected contribution from other processes ("feeddown").

in Table 6.9.

Ratios from Simulation

The same selection is applied to a simulated sample of $Z/\gamma^\star \rightarrow ee/\mu\mu$ events that have been generated using PYTHIA Monte Carlo. All of the standard event selection is used, except that the $\not\!\!E_T$requirement is not raised to 40 GeV for lepton + track pairs with invariant mass 76 GeV/c^2 $< M <$ 106 GeV/c^2. Then, for all events with invariant mass in that interval, the fraction with each jet multiplicity, zero, one, or two or more jets, is counted. This fraction, f_i for jet multiplicity i, is used to distribute the background estimate among the different jet multiplicities. Also, for each jet multiplicity individually, we calculate the ratio of the number of events outside of the invariant mass range 76-106 GeV/$c^2$to the number inside, R_i. These ratios are all collected in Table 6.10.

	f_i, the fraction with each jet multiplicity		
	0 jets	1 jet	≥ 2 jets
CEM	0.62 ± 0.02	0.30 ± 0.02	0.08 ± 0.01
CMUP+CMX	0.56 ± 0.02	0.33 ± 0.02	0.11 ± 0.01
PHX	0.68 ± 0.03	0.27 ± 0.02	0.05 ± 0.01
	R_i, (number inside 76-106 GeV/c^2)/(number outside)		
	0 jets	1 jet	≥ 2 jets
CEM	1.25 ± 0.09	0.96 ± 0.10	0.91 ± 0.20
CMUP+CMX	0.41 ± 0.04	0.31 ± 0.04	0.34 ± 0.08
PHX	0.50 ± 0.06	0.50 ± 0.09	0.89 ± 0.31

Table 6.10: Inputs to the Drell-Yan estimate from simulated $Z/\gamma^\star \to ee/\mu\mu$ events.

Assembling the Estimate

There are two more small correction factors in the Drell-Yan background calculation. One is the difference in the fraction of events inside the mass window between data and simulation. This is significantly different from one only for the CEM and PHX Z events, where the mass window cut in data only includes 98% and 94% as many events as it does in simulation. Those numbers, labeled C_{Zmass}, are a correction to R_i. The second correction is to the fraction of events f_i predicted to have each jet multiplicity. As with the acceptances measured for diboson and $Z/\gamma^\star \to \tau\tau$ events, these fractions depend on the ability of the PYTHIA generator to model the number of extra jets in the event, so the correction factors for the number of jets generated are applied, as described in Section 6.1.2. Then the fractions are rescaled so that they still

sum to one. These corrected fractions are labeled f_i'.

Combining all of the inputs described above, the estimated background from $Z/\gamma^\star \to ee/\mu\mu$ events is calculated for events with i jets as follows:

$$N^i_{\rm DY} = N^i_{\rm out} + N^i_{\rm in} \quad (6.6.2)$$

$$N^i_{\rm out} = ((n_{25} - n^{\rm feeddown}_{25}) \times (f_i' \cdot C_{\rm Zmass})) \times R_i \quad (6.6.3)$$

$$N^i_{\rm in} = ((n_{40} - n^{\rm feeddown}_{40}) \times (f_i' \cdot C_{\rm Zmass})) \quad (6.6.4)$$

In this formula, $n_{\{25,40\}}$ is the number of observed data events in the normalization sample and $n^{\rm feeddown}_{\{25,40\}}$ is the number of predicted events from other sources in the normalization sample.

6.3.3 Uncertainties on the $Z/\gamma^\star \to ee/\mu\mu$ Contribution

The largest uncertainty on the $Z/\gamma^\star \to ee/\mu\mu$ background estimate is the statistical uncertainty, which is determined by the size of the input data and simulation samples. It is determined by propagating the uncertainties on the data counts n_{25}, n_{40}, together with the uncertainties on R_i and f_i, the ratios from simulation. The total is 20%.

Since the scale factor correcting the number of extra jets produced by PYTHIA is applied as a correction to R_i, the fraction of events with jet multiplicity i, the statistical uncertainty on the correction factor (5.5%) also contributes to the

uncertainty on the $Z/\gamma^\star \to ee/\mu\mu$ background.

Finally, the reliability of the ratios measured in simulation depends on the ability of the simulation to model the $\not\!\!E_T$ from mismeasured objects. One way to make a quantitative comparison between the $\not\!\!E_T$ in the data and in the simulation is to compare the fraction of events above the 25 GeV threshold. Since many processes will contribute to the high-$\not\!\!E_T$ "Drell-Yan" sample, as described above, some of the $\Delta\varphi$ selection requirements are inverted to ensure that the comparison is mostly between real and simulated $Z/\gamma^\star \to ee/\mu\mu$ events. That is, the $\not\!\!E_T$ is required to either point at a jet or a lepton. Then the fraction of events with $\not\!\!E_T >$ 25 GeV is measured in the data, and the $\not\!\!E_T$ distribution from the simulation is integrated to find the threshold that would give the same fraction of events above threshold. The analogous threshold in the simulation is 24 GeV. All of the ratios from the simulation are re-derived with the $\not\!\!E_T$ threshold at 24 GeV instead of 25 GeV, and the background recalculated. The recalculated background estimate is 13.5% lower than the default estimate, and the full difference is taken as a systematic uncertainty.

Combined in quadrature, the three uncertainties give a total of 25% on this background estimate.

The systematic uncertainties on all of the backgrounds, and the corresponding uncertainty contributed to the cross-section measurement, are collected in

Source	Uncertainty on background	Uncertainty on cross-section
Primary/track lepton	1.1%/1.1%	0.2%/0.2%
Jet Energy scale	5.8%	0.9%
Jet multiplicity correction	5.5%	0.9%
$Z/\gamma^\star \to ee/\mu\mu$	25%	3.5%
W + fake lepton	20%	7.8%

Table 6.11: Summary table of systematic uncertainties on the background estimate.

Table 6.11.

6.4 Summary

Major backgrounds to the lepton + track $t\bar{t}$ dilepton sample come from diboson, Drell-Yan, and W events. Where possible, background estimates include information from control samples in the data to make the estimates more robust. In the case of W + jets with a fake lepton, the background estimate is based almost entirely on data. For the $Z/\gamma^\star \to ee/\mu\mu$ background, data sets the normalization but simulation is used to fill in the details. Diboson and $Z/\gamma^\star \to \tau\tau$ contributions are estimated using simulation alone, with correction factors from data/simulation comparisons where appropriate.

Systematic uncertainties quantify the level of confidence in a particular background estimation technique. For backgrounds like Drell-Yan $ee/\mu\mu$, and W +

jets with a fake lepton, the estimates rely on understanding of rare fluctuations such as a significantly mismeasured jet and/or lepton or the hadronization of a parton producing only one charged particle with significant transverse momentum. The larger uncertainties on these processes, the limiting systematic uncertainties on the cross-section, reflect the experimental challenges of understanding these backgrounds. These uncertainties are a significant improvement on the previous iteration of this analysis [46]. This is attributable both to the much increased amount of data available as well as to improved techniques for estimating the fake lepton background and the uncertainty on the $Z/\gamma^{\star} \to ee/\mu\mu$ background.

Chapter 7

Results

Using the criteria described in Chapter 4, we select a sample of events in the data expected to contain a significant contribution from $t\bar{t}$ events. From the number of observed events $N_{\rm obs}$, the number of expected background events $N_{\rm bkgnd}$ (calculated in Chapter 6), the corrected acceptance $\mathcal{A} \cdot \epsilon$ (Chapter 5), and the size of the data sample $\int \mathscr{L} dt$, the cross-section can be derived:

$$\sigma = \frac{N_{\rm obs} - N_{\rm bkgnd}}{\mathcal{A} \cdot \epsilon \cdot \int \mathscr{L} dt}$$

We also measure the cross-section as a function of the input top mass, on which it depends strongly, and show some kinematic features of the candidate event sample.

To test the robustness of the analysis, we present results for candidate sam-

ples selected with slightly different criteria. Variations considered are a 15 GeV/c and 25 GeV/c minimum p_T threshold on the track lepton and jets, and the case where the track lepton is also a fully-reconstructed lepton. Changing these requirements alters both the size of the background relative to the expected signal and the relative amounts of background from different sources. Consistency of the results with the main result tests the reliability of the background estimates.

7.1 Main Result and Cross-section Calculation

Prior to calculating the cross-section, the background and signal estimates for the data sample are compared to the total number of observed events, for events with zero, one, and two or more jets. The zero and one jet event comparisons test the background predictions, because the contribution from $t\bar{t}$ in those events is very small. The zero jet events are not as strong of a check, since a subset of these is used to derive the opposite-sign fraction for W + jets events with a fake lepton (see Chapter 6 Section 6.2.7). The results are shown in Table 7.1, and are split by the primary lepton type in Tables 7.2 and 7.3. The agreement is good overall in the zero and one jet bins.

We calculate the cross-section using the formula in the first paragraph. The

number of observed events is 129, the background is calculated to be 54.0 ± 6.6 events, the summed acceptance times efficiency is $0.85 \pm 0.03\%$, and the total acceptance-weighted integrated luminosity is 1070 ± 64 pb^{-1} (see Chapter 4 Section 4.5). Assuming $m_t = 175$ GeV/c^2 and $BR(W \to \ell\nu) = 10.8$ %,

$$\sigma_{t\bar{t}} = 8.3 \pm 1.3(\text{stat.}) \pm 0.7(\text{sys.}) \pm 0.5(\text{lum.})\text{pb} ,$$

consistent with the Standard Model prediction of $6.7^{+0.7}_{-0.9}$ pb [7].

The cross-section calculated at other values of m_t is shown in Table 7.4. The measured cross-section decreases with increasing input m_t because the number of observed events and the background predictions are unchanged, but the acceptance increases with input top mass because the $t\bar{t}$ decay products are more energetic and therefore more likely to pass the kinematic selection.

7.2 Results with Other Kinematic Thresholds

We repeat the cross-section analysis with the p_T threshold on both the track lepton on the jets moved either up or down by 5 GeV/c. This varies the expected signal to background ratio (for a top mass of 175 GeV/c^2) by over a factor of 2, but the cross-section result should stay consistent, since the number of observed events should increase or decrease in accordance with the predicted acceptance and background, if the backgrounds have been estimated correctly.

	0 jets	1 jet	≥ 2 jets
WW	91.7 ± 7.5	16.0 ± 1.3	3.9 ± 0.4
WZ	10.0 ± 0.8	4.6 ± 0.4	1.4 ± 0.1
ZZ	2.4 ± 0.0	0.7 ± 0.0	0.3 ± 0.0
$Z/\gamma^\star \to ee$	72.4 ± 15.8	25.9 ± 6.1	7.8 ± 2.2
$Z/\gamma^\star \to \mu\mu$	18.9 ± 5.3	8.9 ± 2.7	3.4 ± 1.2
$Z/\gamma^\star \to \tau\tau$	35.5 ± 3.2	26.5 ± 2.5	7.3 ± 0.9
Fakes	244.1 ± 46.4	76.8 ± 14.6	29.9 ± 5.9
All Backgrounds	475.0 ± 51.6	159.2 ± 16.9	54.0 ± 6.6
$t\bar{t}, \sigma = 6.7$ pb	1.2 ± 0.1	17.3 ± 0.6	60.5 ± 1.9
Predicted	476.2 ± 51.6	176.5 ± 17.0	114.5 ± 7.0
Observed	443	187	129

Table 7.1: Predicted and observed events in 1.1 fb^{-1}, with details of the background contributions. Systematic uncertainties are included in the prediction numbers.

Signal acceptances and background estimates are all updated with the new selection but correction factors for discrepancies between data and simulation are unchanged from the default analysis. The change in the correction factors for different track and jet thresholds is small in the cases checked.

The comparison of predicted and observed events for the 1.1 fb^{-1} data sample for a track and jet threshold of 15 GeV/c is shown in Table 7.5. The corrected signal acceptance increases to 1.04 % and the predicted ratio of signal events to background events is .68, compared to 0.84% and 1.1 for the main analysis. 189 candidate events are observed and the cross-section (using the same systematic uncertainties as the main analysis) is:

$$\sigma_{t\bar{t}} = 7.2 \pm 1.2(\text{stat.}) \pm 1.2(\text{sys.}) \pm 0.4(\text{lumi.})\text{pb}.$$

CEM

	0 jets	1 jet	≥ 2 jets
Diboson	54.4 ± 4.5	11.4 ± 1.0	2.8 ± 0.3
Drell-Yan	81.9 ± 14.6	36.8 ± 5.7	10.7 ± 2.2
Fakes	108.0 ± 21.3	35.7 ± 7.0	13.4 ± 2.8
All Backgrounds	244.4 ± 26.5	83.8 ± 9.2	26.9 ± 3.6
$t\bar{t}, \sigma = 6.7$ pb	0.7 ± 0.0	9.9 ± 0.3	34.7 ± 1.1
Predicted	245.1 ± 26.5	93.7 ± 9.3	61.6 ± 3.8
Observed	220	81	64

PHX

	0 jets	1 jet	≥ 2 jets
Diboson	18.6 ± 1.6	3.6 ± 0.3	1.0 ± 0.1
Drell-Yan	13.6 ± 4.2	6.6 ± 1.7	2.1 ± 0.7
Fakes	50.8 ± 10.1	16.5 ± 3.2	6.5 ± 1.5
All Backgrounds	82.9 ± 11.0	26.7 ± 3.7	9.6 ± 1.6
$t\bar{t}, \sigma = 6.7$ pb	0.1 ± 0.0	2.0 ± 0.1	7.1 ± 0.2
Predicted	83.0 ± 11.0	28.6 ± 3.7	16.7 ± 1.7
Observed	97	31	12

Table 7.2: Numbers of events expected from signal and background compared to the number of observed events in 1.1 fb^{-1} of data, for events where the primary lepton is a fully reconstructed electron. Recall that events are selected first as CEM + track, then as CMUP + track, then CMX + track, then PHX + track. Systematic uncertainties are included.

CMUP

	0 jets	1 jet	≥ 2 jets
Diboson	22.1 ± 1.9	4.4 ± 0.4	1.3 ± 0.1
Drell-Yan	19.8 ± 4.7	12.0 ± 2.5	3.6 ± 1.1
Fakes	55.8 ± 11.0	16.3 ± 3.2	6.6 ± 1.4
All Backgrounds	97.7 ± 12.2	32.6 ± 4.2	11.5 ± 1.8
$t\bar{t}, \sigma = 6.7$ pb	0.3 ± 0.0	3.9 ± 0.1	13.8 ± 0.5
Predicted	98.0 ± 12.2	36.6 ± 4.2	25.3 ± 1.9
Observed	92	54	40

CMX

	0 jets	1 jet	≥ 2 jets
Diboson	8.9 ± 0.8	1.8 ± 0.2	0.6 ± 0.1
Drell-Yan	11.6 ± 2.0	5.9 ± 1.1	2.1 ± 0.5
Fakes	29.5 ± 5.9	8.4 ± 1.7	3.3 ± 0.7
All Backgrounds	50.1 ± 6.3	16.1 ± 2.0	6.0 ± 0.9
$t\bar{t}, \sigma = 6.7$ pb	0.1 ± 0.0	1.6 ± 0.1	5.0 ± 0.2
Predicted	50.2 ± 6.3	17.7 ± 2.0	11.0 ± 0.9
Observed	34	21	13

Table 7.3: Numbers of events expected from signal and background compared to the number of observed events in 1.1 fb^{-1} of data, for events where the primary lepton is a fully reconstructed muon. Recall that events are selected first as CEM + track, then as CMUP + track, then CMX + track, then PHX + track. Systematic uncertainties are included.

Input m_t (GeV/c^2)	Theoretical σ (pb)	$t\bar{t}$ Acceptance	Measured σ (pb)
170	$7.8^{+0.9}_{-1.0}$	$0.69 \pm .02$ %	10.2 ± 1.5 (stat.) ± 0.9 (sys.)
175	$6.7^{+0.7}_{-0.9}$	$0.85 \pm .03$ %	8.3 ± 1.3 (stat.) ± 0.7 (sys.)
180	$5.8^{+0.6}_{-0.8}$	$1.01 \pm .03$ %	6.9 ± 1.1 (stat.) ± 0.6 (sys.)

Table 7.4: The cross-section as calculated at several input top masses. Theoretical prediction from Ref. [7]. A common uncertainty of 6%, due to the uncertainty on the integrated luminosity, is omitted.

	0 jets	1 jet	$\geq$ 2 jets
WW	100.9 $\pm$ 8.3	23.0 $\pm$ 1.9	7.4 $\pm$ 0.7
WZ	9.6 $\pm$ 0.8	5.7 $\pm$ 0.5	2.4 $\pm$ 0.2
ZZ	2.5 $\pm$ 0.0	0.8 $\pm$ 0.0	0.6 $\pm$ 0.0
$Z/\gamma^\star \to ee$	56.0 $\pm$ 15.2	26.0 $\pm$ 7.4	11.1 $\pm$ 3.4
$Z/\gamma^\star \to \mu\mu$	12.8 $\pm$ 4.4	7.9 $\pm$ 3.2	4.6 $\pm$ 1.8
$Z/\gamma^\star \to \tau\tau$	42.0 $\pm$ 3.8	43.4 $\pm$ 3.9	18.4 $\pm$ 1.8
fakes	410.1 $\pm$ 78.1	130.9 $\pm$ 24.8	64.2 $\pm$ 12.2
All backgrounds	633.9 $\pm$ 81.0	237.7 $\pm$ 27.0	108.7 $\pm$ 13.2
$t\bar{t}(\sigma = 6.7)$ pb	0.7 $\pm$ 0.0	14.3 $\pm$ 0.46	74.4 $\pm$ 2.3
Total pred.	634.6 $\pm$ 81.0	252.0 $\pm$ 27.0	183.0 $\pm$ 13.6
Observed	648	295	189

Table 7.5: Predicted and observed events for the 1.1 fb^{-1} lepton + track sample, where the track lepton and jet p_T thresholds are both at 15 GeV/c instead of the default 20 GeV/c.

For the case where the track lepton and jet thresholds are both 25 GeV/c, the corrected signal acceptance is 0.86% but the expected signal to background ratio is 1.5. The predicted and observed number of events is shown in Table 7.6 and the cross-section result is

$$\sigma_{t\bar{t}} = 7.6 \pm 1.3(\text{stat}) \pm 0.6(\text{sys}) \pm 0.5(\text{lumi})\text{pb}.$$

These cross-sections agree well with the nominal value of 8.3 pb, which implies that the backgrounds in the lepton plus track sample are well-understood.

	0 jets	1 jet	$\geq$ 2 jets
WW	83.3 $\pm$ 6.9	11.0 $\pm$ 0.9	2.3 $\pm$ 0.2
WZ	10.0 $\pm$ 0.8	3.5 $\pm$ 0.3	0.9 $\pm$ 0.1
ZZ	2.3 $\pm$ 0.0	0.5 $\pm$ 0.0	0.2 $\pm$ 0.0
$Z/\gamma^\star \to ee$	56.2 $\pm$ 12.5	18.4 $\pm$ 4.4	3.9 $\pm$ 1.3
$Z/\gamma^\star \to \mu\mu$	21.4 $\pm$ 5.7	7.9 $\pm$ 2.3	2.7 $\pm$ 0.9
$Z/\gamma^\star \to \tau\tau$	29.2 $\pm$ 2.7	16.4 $\pm$ 1.6	4.1 $\pm$ 0.6
fakes	177.3 $\pm$ 33.2	52.0 $\pm$ 9.8	17.5 $\pm$ 3.3
All backgrounds	379.6 $\pm$ 38.3	109.8 $\pm$ 11.6	31.7 $\pm$ 3.9
$t\bar{t}(\sigma = 6.7)$ pb	2.0 $\pm$ 0.1	20.6 $\pm$ 0.7	48.8 $\pm$ 1.5
Total pred.	381.6 $\pm$ 38.3	130.3 $\pm$ 11.7	80.4 $\pm$ 4.2
Observed	325	145	87

Table 7.6: Predicted and observed events for the 1.1 fb^{-1} lepton + track sample, where the track lepton and jet p_T thresholds are both at 25 GeV/c instead of the default 20 GeV/c.

7.3 Cross-section with Two Fully Reconstructed Leptons

Requiring that both leptons are fully reconstructed dramatically reduces the background in the dilepton sample. For simplicity and consistency with the main analysis, we select events, after all other requirements are applied, where the track lepton is closely matched in η and φ to a lepton meeting the requirements for one of the four primary lepton types. Note that the two fully reconstructed leptons need not be of the same type, and that the track lepton matching requirement makes this candidate sample is a subset of the lepton + track candidate sample.

The signal acceptance for this selection is calculated identically to the acceptance for the main analysis, except that data/simulation correction factors for the second identified lepton type are included. The corrected acceptance is 0.38%, less than half of what it is for the main analysis, illustrating the acceptance gained from taus and non-fiducial leptons by using an isolated track for the second lepton. The background contributions are all derived analogously to the lepton + track case. The expected backgrounds decrease significantly, particularly the contribution from events with a fake lepton, which is over an order of magnitude smaller than it is for the main analysis. The total expected signal to background ratio is 2.3. The numbers of predicted and observed events in 1.1 fb^{-1} are compared in Table 7.7, and the cross-section measured is

$$6.1 \pm 1.5(\text{stat.}) \pm 0.5(\text{sys.})\text{pb}.$$

This result is in very good agreement with the Standard Model prediction, and is consistent with the results using the track lepton selection.

	$n_j = 0$	$n_j = 1$	$n_j \geq 2$
diboson	45.9 ± 2.8	9.4 ± 0.6	2.6 ± 0.2
Drell-Yan	41.4 ± 9.6	21.4 ± 4.2	6.9 ± 1.6
fakes	15.8 ± 3.3	4.8 ± 1.0	1.6 ± 0.3
background	103.1 ± 10.5	35.6 ± 4.3	11.2 ± 1.7
$t\bar{t}(\sigma = 6.7)$ pb	0.5 ± 0.0	7.2 ± 0.2	26.2 ± 0.8
Predicted	103.6 ± 10.5	42.7 ± 4.3	37.4 ± 1.9
Observed	79	45	35

Table 7.7: Predicted and observed events in the subsample where the track lepton is matched to a fully reconstructed lepton. Uncertainties are statistical only.

7.4 Kinematic Features of the Candidate Event Sample

In the process of measuring the cross-section we have isolated a data sample whose main component is $t\bar{t}$ events. We can then examine the main kinematic features of the sample, and compare them to what is expected for the $t\bar{t}$ signal and the various backgrounds.

First, we show the number of jets in the event in Figure 7.1, which is a graphical representation of Table 7.1. Here all the normalizations are absolute, taken from the background predictions and the theoretical cross-section for $t\bar{t}$ at $m_t = 175$ GeV/c^2.

The remainder of the kinematic features are considered separately for events with 0, 1, or ≥ 2 jets. For these figures (7.2 through 7.4), the top contribution is

set to the measured cross-section and the background is rescaled so that the total number of predicted events is equal to the number observed. For events with ≥ 2 jets, the cross-section is set by the overall normalization, so the background numbers are not scaled. Also, the systematic uncertainties on the background normalizations are not shown.

Figure 7.2 shows the $\not\!\!E_T$ in candidate events. $\not\!\!E_T$ is one of the distinguishing features of top quark leptonic decays, and the $\not\!\!E_T$ spectrum for the backgrounds, except for a small contribution from diboson events, is rather softer than for the signal, and the distribution of events in the data reflects that.

The dilepton invariant mass, shown in Figure 7.3, also has a distinctly harder shape for the signal than for the backgrounds. There is a distinct spike in the distribution from Drell-Yan events between 50-100 GeV/c^2, and the fact that the candidate events follow the predicted distribution is evidence for the accuracy of that background estimate.

The transverse momentum of the leptons is shown in Figure 7.4. Both leptons are included, so there are two entries in the histogram per event. Leptons from $t\bar{t}$ events are more energetic on average but there is not as large a difference in the shapes for signal and background as there is for other quantities considered.

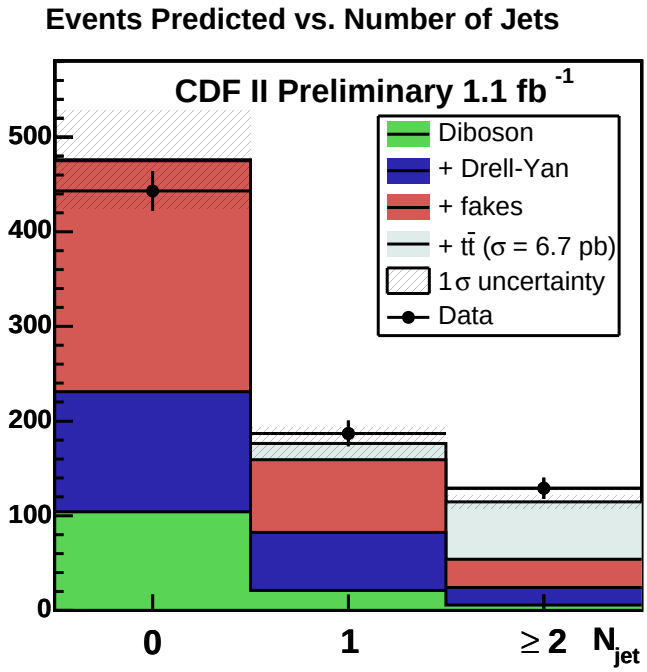

Figure 7.1: Number of predicted events compared to the number observed in the data. The cross-hatched areas show the (1 σ) uncertainties on the predicted numbers (statistical and systematic).

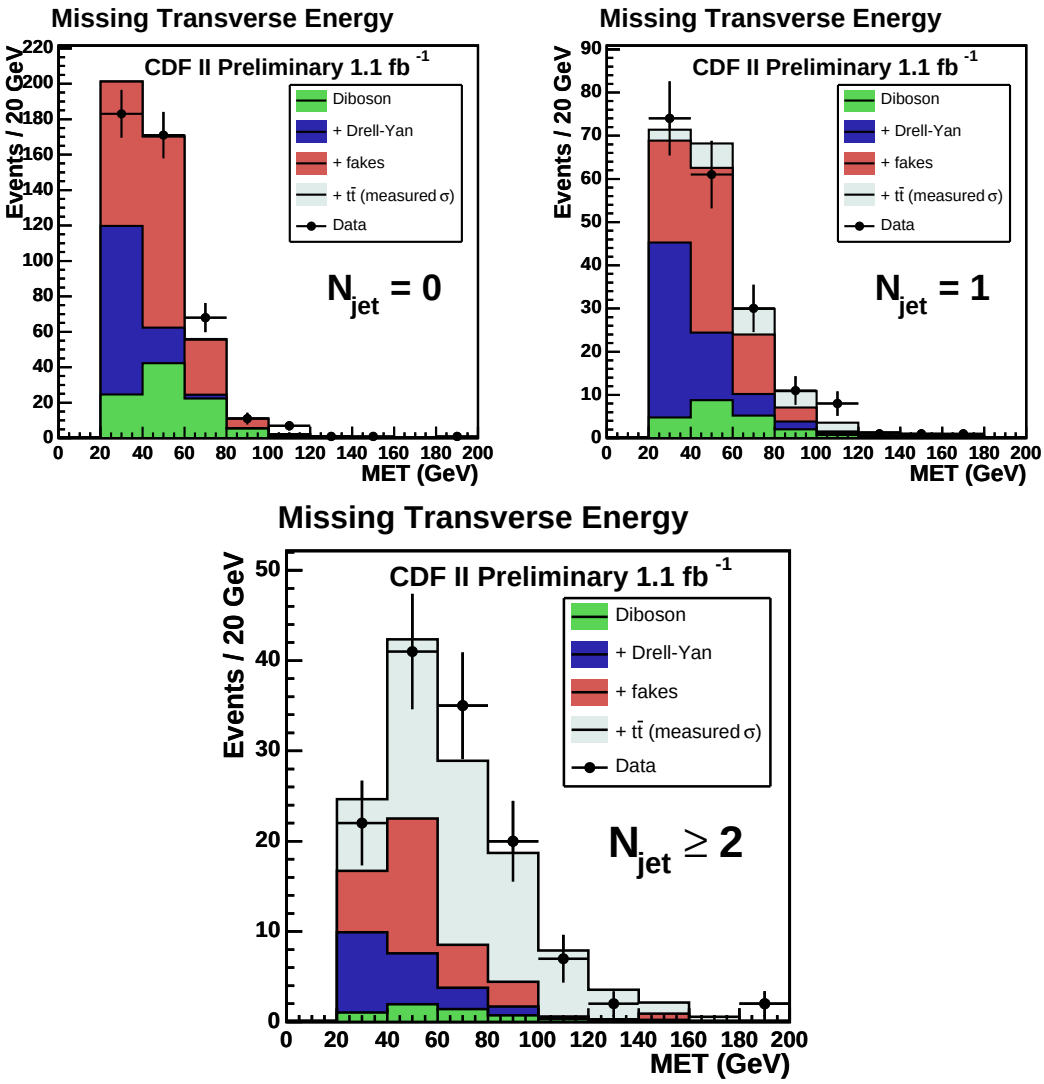

Figure 7.2: Missing transverse energy of predicted and candidate events. The measured cross-section is used to normalize the $t\bar{t}$ distribution and the prediction is scaled so that the predicted number of events is equal to the observed.

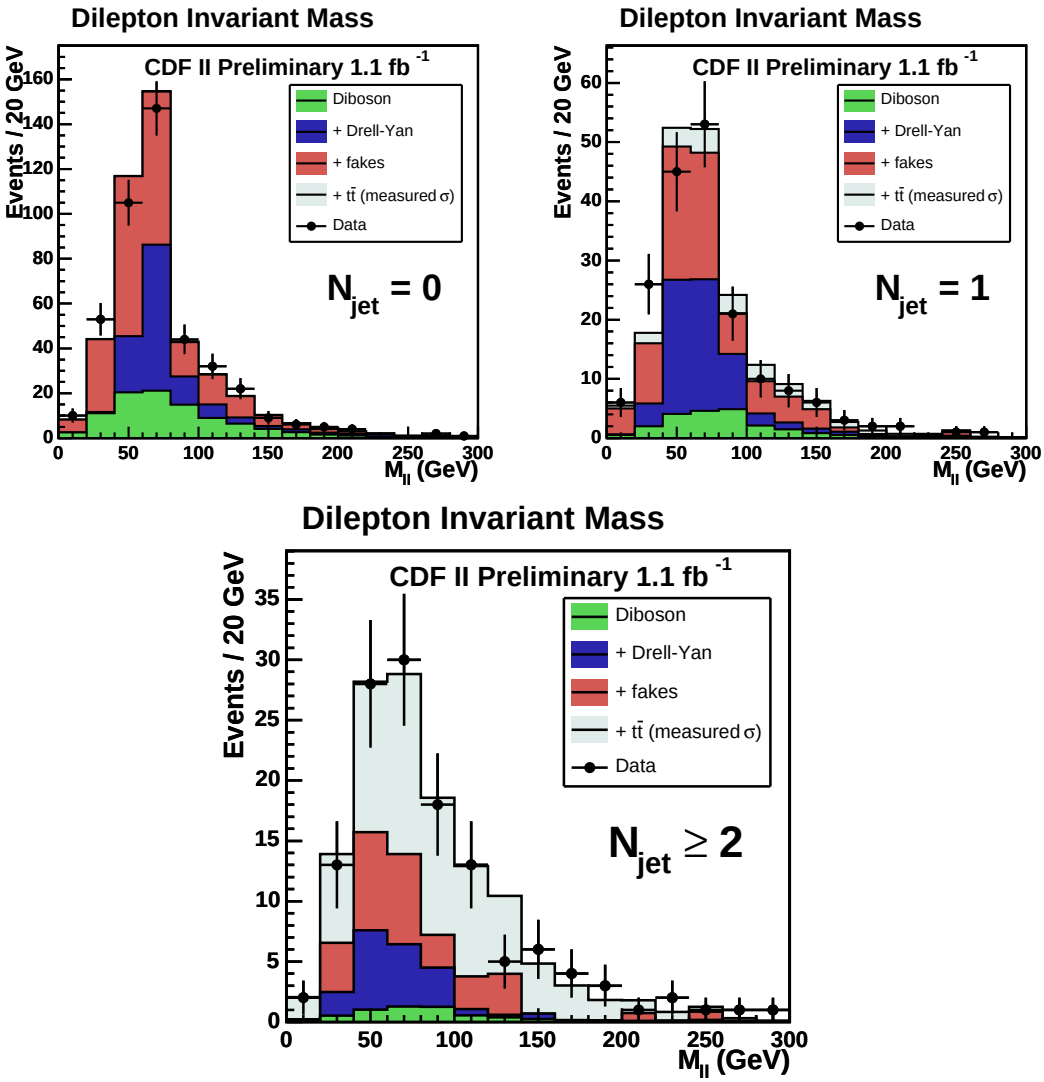

Figure 7.3: Invariant mass of the lepton pair in predicted and candidate events. The measured cross-section is used to normalize the $t\bar{t}$ distribution and the prediction is scaled so that the predicted number of events is equal to the observed.

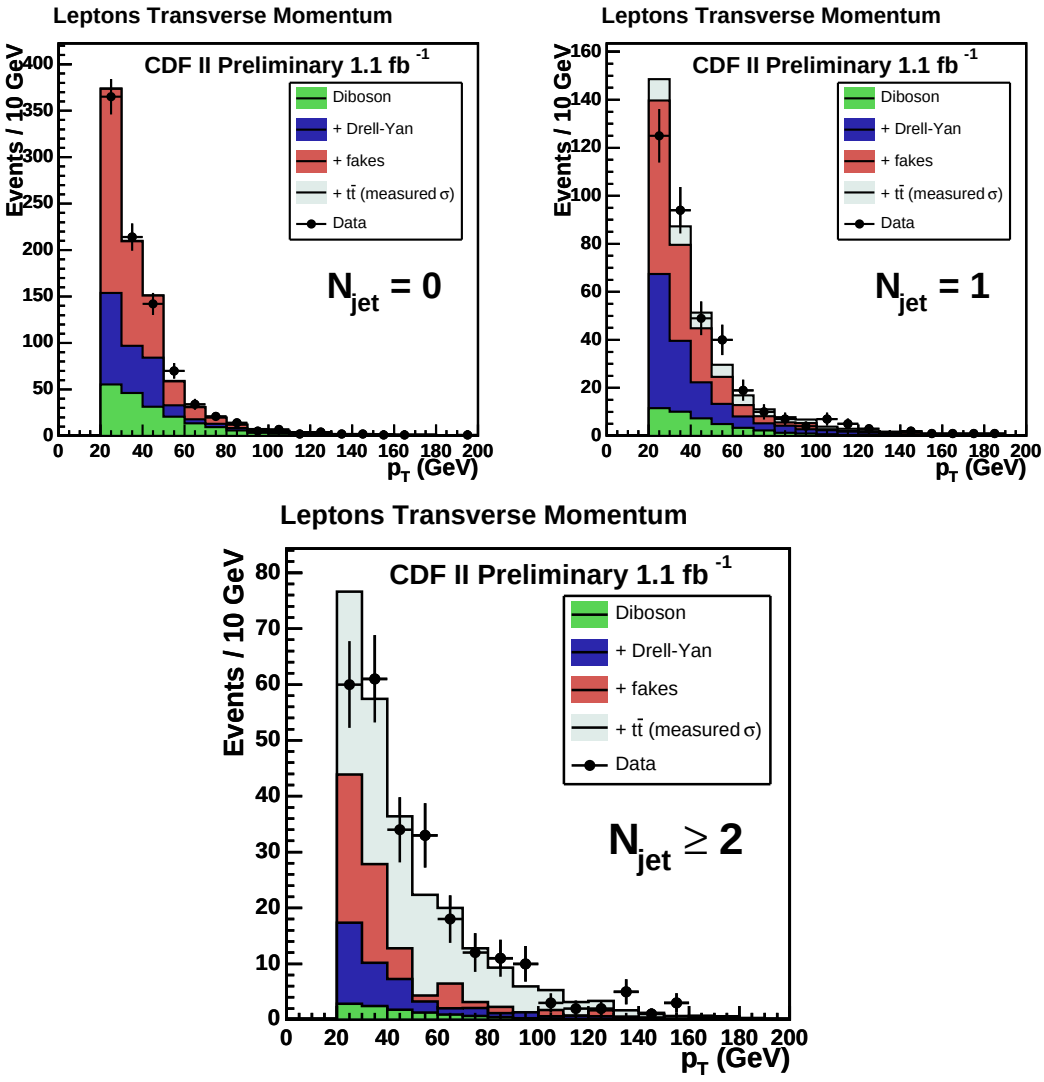

Figure 7.4: Transverse momentum of both lepton candidates in predicted and candidate events. The measured cross-section is used to normalize the $t\bar{t}$ distribution and the prediction is scaled so that the predicted number of events is equal to the observed.

Chapter 8

Conclusion

We have measured the $t\bar{t}$ production cross-section in the dilepton channel using events identified with one fully reconstructed lepton and one isolated track. The result is

$$\sigma_{t\bar{t}} = 8.3 \pm 1.3(\text{stat.}) \pm 0.7(\text{sys.}) \pm 0.5(\text{lum.})\text{pb} \ ,$$

consistent both with the Standard Model prediction and with variations on the cross-section where the p_T threshold for the track leptons and jets has been moved. We also perform a check in the low-background sample where the track lepton is matched to a second fully-reconstructed lepton. The results from all variations on the cross-section are consistent, validating that the background estimates are robust. We also show that the kinematic features of the candi-

date sample are qualitatively consistent with the expectations for signal and background.

This result is consistent with previous measurements in the dilepton channel at both CDF and D0 [42, 43]. Since the dataset is approximately a factor of five larger and the background systematic uncertainties have been reduced, this is a much more precise measurement, at least compared to what has been published. Right now the amount of data available is doubling approximately every year, and cross-section results are regularly updated as data becomes available.

Using lepton + track selection yields a larger candidate sample than using two fully reconstructed leptons, at the expense of increased background. As the amount of data available increases, statistical uncertainties shrink and the precision of top cross-section measurements will eventually become limited by systematic uncertainties. The largest systematic uncertainties for dilepton cross-section measurements are on the background estimates. Therefore as the systematic uncertainties become a larger fraction of the total uncertainty, it is beneficial to reduce the level of background in the candidate sample. This suggests that it will be necessary to re-examine the benefits of lepton + track selection.

One way to maintain the benefits of a more open selection is to reduce the systematic uncertainty on the backgrounds by improving the background

estimation techniques. Such considerations have motivated the study of the fake lepton background. We have introduced the use of a sample triggered on high-energy photons to generate a lepton fake rate as close as possible to that for jets associated with Ws, and used both data and simulation to reconstruct the fraction of those events where the fake lepton will have the opposite sign from the true one.

Another approach is to alter the event selection in a way that reduces the background while maintaining some appealing aspect of the selection. In the lepton + track case, a likely goal would be to reduce background without tightening the requirements on the second lepton, so that acceptance for τ leptons is maintained. One tactic is to require that one of the jets in the event be "tagged" as a likely the product of a b quark, due to the apparent long lifetime or semileptonic decay of a particle in the jet. Making this tagging requirement reduces the background by over a factor of four while removing less than half of the $t\bar{t}$ signal, depending on the tagging criteria used. Such an analysis using this selection is currently in progress at CDF.

Aside from tagging, it is possible to tune the selection to specifically reduce the expected contribution from backgrounds with larger uncertainties. With the large data and simulation samples now available, the background from events with a fake lepton is slowly proving tractable, but a complete understanding

of the pathological set of $Z/\gamma^\star \to ee/\mu\mu$ events with large $\not\!\!E_T$ has been more elusive. Therefore a possible refinement of the selection could benefit from more stringent requirements on the $\not\!\!E_T$ in addition to further study of $\not\!\!E_T$ production in the absence of neutrinos.

Even if the $t\bar{t}$ sample from the Tevatron does not itself hold evidence of physics beyond the Standard Model, careful study of the physics of the top quark sector will be important for the searches for new physics at future hadron colliders. The anticipated signatures for new physics, particularly supersymmetry, may have multilepton, multijet final states, possibly with $\not\!\!E_T$ from the escaping dark matter candidate. This means that all of the backgrounds to top, as well as top production itself, will very likely be backgrounds to at least some of the indications for new physics. Even more generally, improved understanding of the mechanisms that generate $\not\!\!E_T$ in the absence of neutrinos and fake leptons from jets will be necessary to accurately estimate the contributions from known physics processes in the data collected at new hadron colliders. Measurements of the top quark cross-section, particularly in the dilepton channel, are the best testing ground currently available for our understanding of these backgrounds, in addition to providing the normalization for tomorrow's backgrounds. Even as top quark measurements become staples of hadron collider physics, they will remain crucial to our understanding of the fundamental par-

ticles and their interactions, and discovering what, if anything, will supersede the Standard Model.

Appendix A

Glossary of Acronyms

For the abbreviations used frequently in the text, we provide definitions and references to more complete descriptions in the text where applicable.

CEM *Central ElectroMagnetic Calorimeter* Part of the calorimeter (See Chapter 3 Section 3.2.2). Also a shorthand name for fully reconstructed electrons in the central part of the detector (See Chapter 4 Section 4.1.1).

CDF *The Collider Detector at Fermilab* The name of both one of the colliding-beam detectors at Fermilab and of the collaboration of scientists that built the detector and analyze the data from it. See Chapter 3 Section 3.2.

CMUP *Central MUon and Central Muon UPgrade* A muon type identified using track segments in the two more central muon detectors. The muon

detectors are described in Chapter 3 Section 3.2.3, and the selection is described in Chapter4 Section 4.1.2

CMX *Central Muon eXtension* A muon type identified using a track segment in the more forward muon detectors. The muon detectors are described in Chapter 3 Section 3.2.3, and the selection is described in Chapter4 Section 4.1.2

COT *Central Outer Tracker* A gas-filled open cell drift chamber used to reconstruct particle trajectories. See Chapter 3 Section 3.2.1.

FSR *Final State Radiation* The emission of a parton (quark or gluon) with significant p_T from the products of a high-energy scattering event. See Chapter 5 Section 5.4.4.

ISR *Initial State Radiation* The emission of a parton (quark or gluon) with significant p_T by the incoming partons before they participate in a high-energy scattering event. See Chapter 5 Section 5.4.4.

LHC *Large Hadron Collider* A high-energy ($\sqrt{s} = 14$ TeV), high-luminosity proton-proton collider currently under construction at CERN, a laboratory on the border of France and Geneva. First collisions currently scheduled for 2008.

PDF *Parton Distribution Function* Probability distributions for the fraction of the proton momentum carried by a particular constituent parton (quarks and gluons). See Chapter 2 Section 2.1.2.

PEM *Plug ElectroMagnetic* The forward part of the electromagnetic calorimeter. See Chapter 3 Section 3.2.2.

PHX *abbreviation for Phoenix* Fully reconstructed forward ($1.2 \lesssim |\eta| \lesssim 2.0$) electrons, named after the tracking algorithm. See Chapter 4 Section 4.1.1

QCD *Quantum ChromoDynamics* The theory describing the strong interaction, the main force by which colored particles (quarks and gluons) interact. See Chapter 2.

Bibliography

[1] CDF, Top Physics Group Blessed Plots (Run I), http://www-cdf.fnal.gov/internal/physics/top/RunITop/blessed_plots.html.

[2] CTEQ, J. Pumplin *et al.*, JHEP **07**, 012 (2002), hep-ph/0201195.

[3] T. Sjostrand *et al.*, Comput. Phys. Commun. **135**, 238 (2001), hep-ph/0010017.

[4] CDF II, D. Acosta *et al.*, Phys. Rev. Lett. **94**, 091803 (2005), hep-ex/0406078.

[5] C. Gattuso, D. Patterson, and B. Worthel, Linac Rookie Book, Fermilab Accelerator Division Documentation, http://www-bdnew.fnal.gov/operations/rookie_books/LINAC_v2.pdf, 2004.

[6] Particle Data Group, W. M. Yao *et al.*, J. Phys. **G33**, 1 (2006).

[7] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi, JHEP **04**, 068 (2004), hep-ph/0303085.

[8] CDF, F. Abe *et al.*, Phys. Rev. Lett. **74**, 2626 (1995), hep-ex/9503002.

[9] D0, S. Abachi *et al.*, Phys. Rev. Lett. **74**, 2632 (1995), hep-ex/9503003.

[10] DONUT, K. Kodama *et al.*, Phys. Lett. **B504**, 218 (2001), hep-ex/0012035.

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

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

[13] CDF, F. Abe *et al.*, Phys. Rev. Lett. **80**, 2779 (1998), hep-ex/9802017.

[14] D0, V. M. Abazov *et al.*, Phys. Rev. **D67**, 012004 (2003), hep-ex/0205019.

[15] R. K. Ellis, W. J. Stirling, and B. Webber, *QCD and Collider Physics* (Cambridge University Press, 1996).

[16] R. Bonciani, S. Catani, M. L. Mangano, and P. Nason, Nucl. Phys. **B529**, 424 (1998), hep-ph/9801375.

[17] P. Nason, S. Dawson, and R. K. Ellis, Nucl. Phys. **B303**, 607 (1988).

[18] W. Beenakker, H. Kuijf, W. L. van Neerven, and J. Smith, Phys. Rev. **D40**, 54 (1989).

[19] S. Catani, M. L. Mangano, P. Nason, and L. Trentadue, Phys. Lett. **B378**, 329 (1996), hep-ph/9602208.

[20] N. Kidonakis and R. Vogt, Phys. Rev. **D68**, 114014 (2003), hep-ph/0308222.

[21] D0, V. M. Abazov *et al.*, Phys. Lett. **B639**, 616 (2006), hep-ex/0603002.

[22] CDF, D. Acosta *et al.*, Phys. Rev. Lett. **95**, 102002 (2005), hep-ex/0505091.

[23] D. Albert, W. J. Marciano, D. Wyler, and Z. Parsa, Nucl. Phys. **B166**, 460 (1980).

[24] P. J. Sutton, A. D. Martin, R. G. Roberts, and W. J. Stirling, Phys. Rev. **D45**, 2349 (1992).

[25] P. J. Rijken and W. L. van Neerven, Phys. Rev. **D51**, 44 (1995), hep-ph/9408366.

[26] R. Hamberg, W. L. van Neerven, and T. Matsuura, Nucl. Phys. **B359**, 343 (1991).

[27] R. V. Harlander and W. B. Kilgore, Phys. Rev. Lett. **88**, 201801 (2002), hep-ph/0201206.

[28] CDF, A. A. Affolder *et al.*, Phys. Rev. Lett. **87**, 211804 (2001).

[29] J. M. Campbell and R. K. Ellis, Phys. Rev. **D60**, 113006 (1999), hep-ph/9905386.

[30] Tevatron Electroweak Working Group, A Combination of CDF and D0 Results on the Mass of the Top Quark, hep-ex/0703034, 2007.

[31] H. P. Nilles, Phys. Rept. **110**, 1 (1984).

[32] C. T. Hill and S. J. Parke, Phys. Rev. **D49**, 4454 (1994), hep-ph/9312324.

[33] H. T. Edwards, Ann. Rev. Nucl. Part. Sci. **35**, 605 (1985).

[34] C. Gattuso *et al.*, Accelerator Concepts, Fermilab Accelerator Division Documentation, http://www-bdnew.fnal.gov/operations/rookie_books/Concepts_v3.1.pdf, 2002.

[35] C. W. Schmidt and C. D. Curtis, AIP Conf. Proc. **158**, 425 (1987).

[36] Tevatron Rookie Book, Fermilab Accelerator Division Documentation, http://www-bdnew.fnal.gov/operations/rookie_books/Tevatron_v1.pdf.

[37] CDF, A. A. Affolder *et al.*, Nucl. Instrum. Meth. **A526**, 249 (2004).

[38] CDF, A. Abulencia *et al.*, Measurements of Inclusive w and z Cross Sections in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ tev, Submitted to Phys. Rev. D, 2005.

[39] CDF, CDFII Collaboration Technical Design Report, FERMILAB-PUB-96/390-E, http://www-cdf.fnal.gov/upgrades/tdr/tdr.html.

[40] CDF, D. Acosta *et al.*, Nucl. Instrum. Meth. **A461**, 540 (2001).

[41] D. Acosta *et al.*, Nucl. Instrum. Meth. **A494**, 57 (2002).

[42] CDF, D. Acosta *et al.*, Phys. Rev. Lett. **93**, 142001 (2004), hep-ex/0404036.

[43] D0, V. M. Abazov *et al.*, Phys. Lett. **B626**, 55 (2005), hep-ex/0505082.

[44] A. Bhatti *et al.*, Nucl. Instrum. Meth. **A566**, 375 (2006), hep-ex/0510047.

[45] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Eur. Phys. J. **C35**, 325 (2004), hep-ph/0308087.

[46] C. Collaboration, Measurement of $\sigma(p\bar{p} \to t\bar{t} \to b\bar{b}\ell^+\nu\ell^-\nu)$ at $\sqrt{s} = 1.96$ tev, Public CDF Analysis Document, http://www-cdf.fnal.gov/physics/new/top/confNotes/cdf7942_top_dilepton_ltrk.ps, 2005.