

**Measurement of the Top Quark Pair Production
Cross Section with 1.12 fb^{-1} of $p\bar{p}$ Collisions at \sqrt{s}
 $= 1.96 \text{ TeV}$**

A dissertation presented by

Daniel Joseph Sherman

to

The Department of Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachusetts

August 2007

©2007 - Daniel Joseph Sherman

All rights reserved.

Measurement of the Top Quark Pair Production Cross Section with 1.12 fb^{-1} of $p\bar{p}$ Collisions at $\sqrt{s} = 1.96 \text{ TeV}$

Abstract

We present a measurement of the top quark pair production cross section in e +jets and μ +jets events recorded with the CDF II detector. The data correspond to 1.12 fb^{-1} of $p\bar{p}$ collisions with center-of-mass energy 1.96 TeV taken during Run 2 of the Fermilab Tevatron. We identify $t\bar{t}$ candidate events by requiring at least one secondary vertex b -tag, and we measure a production cross section of 8.2 ± 0.5 (stat) ± 0.8 (syst) ± 0.5 (lum) pb, the most precise single measurement of $\sigma_{t\bar{t}}$ to date. As a check, we also measure $\sigma_{t\bar{t}}$ in events with multiple b -tagged jets using a higher-efficiency tagging algorithm, and there we measure $8.8^{+0.8}_{-0.7}$ (stat) ± 1.2 (syst) ± 0.5 (lum) pb, also the most precise single measurement with this signature. Both results assume a top quark mass of $175 \text{ GeV}/c^2$, for which the theoretical expectation is $6.7^{+0.7}_{-0.9}$ pb; the measured excess is of the order 1σ , and becomes less significant at the current world average top mass of $171 \text{ GeV}/c^2$.

Contents

Title Page	i
Abstract	iii
Table of Contents	iv
Figures	vii
Tables	xii
Acknowledgments	xiv
1 Introduction	1
2 Theoretical and Experimental Top Results	3
2.1 The Top Quark	4
2.1.1 Top Production	4
2.1.2 Top Decay	15
2.1.3 Non-Standard Model Top Phenomenology	18
2.2 Experimental Top Quark Results	20
3 Experimental Apparatus	25
3.1 The Accelerator Complex	25
3.1.1 Luminosity	28
3.2 The CDF Detector	32
3.2.1 Tracking Systems	33
3.2.2 Calorimetry	44
3.2.3 Muon Detectors	48
3.2.4 Luminosity Counters	54
3.2.5 Trigger and Data Acquisition	54
4 Event Reconstruction	59
4.1 Tracking	59
4.1.1 Primary Vertex Identification	64
4.2 Lepton Identification	66
4.2.1 Electrons	66

4.2.2	Muons	73
4.3	Jet Clustering	78
4.4	Missing Energy	80
4.5	Monte Carlo	81
5	<i>b</i>-Tagging	83
5.1	The SecVtx Algorithm	85
5.2	<i>b</i> -Tagging Efficiency	92
5.2.1	Electron Method	93
5.2.2	Muon Method	107
5.2.3	Combined Result	116
5.3	Light Flavor Tagging	118
5.3.1	Mistag Matrix Performance	126
5.3.2	Systematic Uncertainties	132
5.3.3	Mistag Asymmetry	134
6	Composition of the <i>W</i>+Jets Sample	146
6.1	Method 2 Overview	146
6.2	Event Selection	149
6.2.1	Triggers and Data Event Selection	149
6.2.2	<i>b</i> -Tagging	150
6.2.3	Monte Carlo	151
6.3	Signal Expectation	151
6.3.1	Acceptance	151
6.3.2	<i>b</i> -Tagging Efficiency	155
6.4	Backgrounds	157
6.4.1	Electroweak & Single Top (MC Backgrounds)	157
6.4.2	Non- <i>W</i>	159
6.4.3	<i>W</i> + Heavy Flavor	167
6.4.4	Mistags	168
6.5	Summary Tables	172
7	Results	176
7.1	Optimization	177
7.2	Top Signal	182
7.3	Backgrounds	185
7.4	Results	190
7.5	Uncertainties	197
7.5.1	Statistical Errors	202
8	Conclusion and Outlook	205
A	Data Selection Details	207

B Calibration of Heavy Flavor in W+Jets	211
B.1 W +jets Monte Carlo	212
B.2 Raw Heavy Flavor Fractions	215
B.2.1 Heavy Flavor Overlap Removal	216
B.2.2 Initial Results	224
B.3 K -Factor Determination	224
B.4 Systematic Uncertainties	231
B.5 Results	232
C Kinematics of the W+Jets Sample	235
Bibliography	244

Figures

2.1	Feynman diagrams for $t\bar{t}$ production	5
2.2	Parton distribution functions	10
2.3	Leading-order $t\bar{t}$ cross section vs. $\sqrt{\hat{s}}$	11
2.4	$t\bar{t}$ invariant mass and p_T distributions	11
2.5	Top quark p_T and η distributions	12
2.6	W p_T and η distributions	12
2.7	b p_T and η distributions	13
2.8	Theoretical $t\bar{t}$ cross section vs. top mass	13
2.9	Feynman diagrams for single top	14
2.10	$t\bar{t}$ branching fractions	17
2.11	Example Feynman diagrams for non-Standard Model $t\bar{t}$ production	18
2.12	Top and charged Higgs branching fractions	19
2.13	Top mass results from the Tevatron	22
2.14	$t\bar{t}$ cross section results from CDF in Run 2	23
2.15	Measured $t\bar{t}$ cross section vs. mass, comparison with theory	24
2.16	Dependence of the $t\bar{t}$ cross section on the center-of-mass energy	24
3.1	Schematic view of Fermilab	26
3.2	Peak instantaneous luminosities in Run 2	30
3.3	Integrated luminosity in Run 2	31
3.4	Cross-sectional view of the CDF II detector	33
3.5	Illustration of the CDF II detector	34
3.6	Cross-sectional view of the CDF II tracking volume	35
3.7	Cross section of the silicon system	36
3.8	Cross section of L00	37
3.9	Cross section of the SVX-II	38
3.10	Signal-to-noise in the SVX-II	39
3.11	Residual distribution of the SVX-II	40
3.12	Layout of the ISL	40
3.13	Schematic of the COT superlayer structure	42

3.14	Wire pattern in COT cells	43
3.15	Layout of the forward calorimeters	46
3.16	Coverage of the muon system	50
3.17	Position of the muon chambers relative to the calorimeter	51
3.18	Close-up of a block of CMU cells	52
3.19	Trajectory of an ideal CMUP muon	52
3.20	Schematic of the CMX	53
3.21	Schematic of the CDF trigger system	57
3.22	Block diagram of the CDF trigger system	58
4.1	Illustration of the signatures for various types of particle	60
4.2	Impact of L00 on tracking resolution	64
4.3	$\eta - \varphi$ distribution for high- p_T electrons	67
4.4	E_T, p_T of high- p_T electron candidates	70
4.5	$\frac{E}{p}, \eta$ of high- p_T electron candidates	70
4.6	$\frac{E^{had}}{E^{EM}}, q\Delta(r\varphi)$ of high- p_T electron candidates	71
4.7	$\Delta z, \chi^2$ of high- p_T electron candidates	71
4.8	L_{shr} , isolation of high- p_T electron candidates	72
4.9	$\eta - \varphi$ distribution for high- p_T muons	73
4.10	p_T, η of high- p_T muon candidates	75
4.11	Electromagnetic, Hadronic energy of high- p_T muon candidates	76
4.12	Track χ^2 , CMU $\Delta(r\varphi)$ of high- p_T muon candidates	76
4.13	CMP $\Delta(r\varphi)$, CMX $\Delta(r\varphi)$ of high- p_T muon candidates	77
4.14	L_{shr} , isolation of high- p_T muon candidates	77
4.15	Systematic uncertainties on jet energy corrections	79
5.1	E_T and η distributions for b -jets in $t\bar{t}$	84
5.2	Illustration of a SECVTX vertex	87
5.3	d_0 and σ_{d_0} distribution for tight SECVTX tracks	88
5.4	d_0 significance and z_0 distribution for tight SECVTX tracks	89
5.5	p_T and silicon χ^2 distribution for tight SECVTX tracks	89
5.6	Numbers of pass 1 and pass 2 tracks with loose SECVTX	90
5.7	L_{2d} and $\sigma_{L_{2d}}$ of loose SECVTX-tagged jets	90
5.8	Invariant mass and p_T of loose SECVTX-tagged jets	91
5.9	Conversion position from 8-GeV electron sample	97
5.10	Jet E_T, η distributions in the low- p_T electron sample	98
5.11	Electron E_T, η distributions in the low- p_T electron sample	98
5.12	Tag L_{2d}, L_{2d} significance distributions in the low- p_T electron sample	99
5.13	Vertex mass fits in the 8-GeV electron sample	101
5.14	Dependence of electron-method SF on jet E_T	104
5.15	Dependence of electron-method SF on jet η	104
5.16	Dependence of electron-method SF on jet ϕ	105

5.17	Dependence of electron-method SF on jet E_T	105
5.18	Dependence of electron-method SF on primary vertex position	106
5.19	Dependence of electron-method SF on number of vertices	106
5.20	Data and Monte Carlo kinematics of the low- p_T muon sample	108
5.21	Templates used in the p_{Trel} fits	109
5.22	Muon p_{Trel} fits for loose SECVTX	110
5.23	Muon p_{Trel} fits for tight SECVTX	110
5.24	Tagged and untagged b templates in the muon p_{Trel} fits	112
5.25	Dependence of muon-method SF on muon jet E_T	114
5.26	Dependence of muon-method SF on muon jet η	114
5.27	Dependence of muon-method SF on the primary vertex position	115
5.28	Dependence of muon-method SF on number of vertices	115
5.29	E_T and η dependence of the b -tagging efficiency	117
5.30	Kinematics of the Jet20 dataset	122
5.31	Kinematics of the Jet50 dataset	123
5.32	Kinematics of the Jet70 dataset	124
5.33	Kinematics of the Jet100 dataset	125
5.34	Performance of the mistag matrix vs. jet E_T	127
5.35	Performance of the mistag matrix vs. jet η	127
5.36	Performance of the mistag matrix vs. track multiplicity	128
5.37	Performance of the mistag matrix vs. ΣE_T	128
5.38	Performance of the mistag matrix vs. the number of z vertices	129
5.39	Performance of the mistag matrix vs. primary vertex z	129
5.40	Performance of the mistag matrix vs. instantaneous luminosity	130
5.41	Performance of the mistag matrix vs. number of jets in the event	130
5.42	Performance of the mistag matrix vs. number of tracks in the event	131
5.43	Mistag asymmetry fits in Jet20	137
5.44	Mistag asymmetry fits in Jet50	138
5.45	Mistag asymmetry fits in Jet70	139
5.46	Mistag asymmetry fits in Jet100	140
5.47	Dependence of α on jet E_T for loose SECVTX	142
5.48	Dependence of α on jet E_T for tight SECVTX	143
6.1	Feynman diagrams with gluon splitting in all-jet and W events	148
6.2	Missing E_T fits used to evaluate the non- W background in 1-jet events	162
6.3	Missing E_T fits used to evaluate the non- W background in 2-jet events	162
6.4	Missing E_T fits used to evaluate the non- W background in ≥ 3 -jet events	163
6.5	Missing E_T -Isolation scatter plots for pretag electron and muon events	164
6.6	Missing E_T -Isolation scatter plots for b -tagged electron and muon events	165
6.7	Event mistag probabilities, loose tagger	171
6.8	Event mistag probabilities, tight tagger	172
6.9	Composition of W +jets sample with tight tags for base selection	173

6.10	Composition of W +jets sample with loose tags for base selection . . .	175
7.1	Optimization curves for ≥ 1 tight tag	179
7.2	Optimization curves for ≥ 1 loose tag	180
7.3	Optimization curves for ≥ 2 tight tags	180
7.4	Optimization curves for ≥ 2 loose tags	181
7.5	Composition of W +jets sample with tight tags for the optimized selection	186
7.6	Composition of W +jets sample with loose tags for the optimized selection	189
7.7	Composition of W +jets sample with ≥ 1 tight tag, optimized selection, measured $\sigma_{t\bar{t}}$	193
7.8	Composition of W +jets sample with ≥ 2 tight tags, optimized selec- tion, measured $\sigma_{t\bar{t}}$	194
7.9	Composition of W +jets sample with ≥ 1 loose tag, optimized selection, measured $\sigma_{t\bar{t}}$	195
7.10	Composition of W +jets sample with ≥ 2 loose tags, optimized selec- tion, measured $\sigma_{t\bar{t}}$	196
7.11	Pull distribution for single-tag pseudo-experiments	204
B.1	Feynman diagrams for W +light flavor	213
B.2	Feynman diagrams for W +heavy flavor	213
B.3	Comparison of $c\bar{c}$ p_T 's from the PYTHIA shower and from ALPGEN . .	218
B.4	Comparison of $c\bar{c}$ ΔR 's from the PYTHIA shower and from ALPGEN .	219
B.5	Comparison of $c\bar{c}$ p_T 's using alternate overlap-removal schemes	220
B.6	Comparison of $c\bar{c}$ ΔR 's using alternate overlap-removal schemes . . .	221
B.7	Comparison of $b\bar{b}$ p_T 's using alternate overlap-removal schemes	222
B.8	Comparison of $b\bar{b}$ ΔR 's using alternate overlap-removal schemes . . .	223
B.9	Feynman diagrams with gluon splitting in all-jet and W events	225
B.10	ΔR between jets in the all-jet samples	226
B.11	ΔR between b -tagged jets in the all-jet samples	227
B.12	Dependence of K on jet energy	229
B.13	Template fits to extract the K factor	230
B.14	Comparison of simulated and generated jets used to evaluate W +jets systematic uncertainties	232
C.1	H_T distribution in $W+\geq 3$ jets	236
C.2	Missing E_T distribution in $W+\geq 3$ jets	236
C.3	W transverse mass distribution in $W+\geq 3$ jets	237
C.4	Lepton p_T distribution in $W+\geq 3$ jets	237
C.5	Leading jet E_T distribution in $W+\geq 3$ jets	238
C.6	Leading jet η distribution in $W+\geq 3$ jets	238
C.7	Second jet E_T distribution in $W+\geq 3$ jets	239
C.8	Second jet η distribution in $W+\geq 3$ jets	239
C.9	Third jet E_T distribution in $W+\geq 3$ jets	240

C.10 Third jet η distribution in $W+\geq 3$ jets	240
C.11 SECVTX tag mass distribution in $W+\geq 3$ jets	241
C.12 SECVTX tag p_T distribution in $W+\geq 3$ jets	241
C.13 SECVTX tag L_{2d} distribution in $W+\geq 3$ jets	242
C.14 SECVTX tag pseudo- $c\tau$ distribution in $W+\geq 3$ jets	242
C.15 SECVTX tag L_{2d} significance distribution in $W+\geq 3$ jets	243

Tables

5.1	SECVTX parameters	86
5.2	Event yields used to calculate SF	96
5.3	Electron-method scale factor systematics	102
5.4	Systematic errors on the muon method scale factor	113
5.5	Tag rates in the mistag matrix samples	120
5.6	Binning of the mistag matrix	121
5.7	Tag rate studies for mistag matrix systematic uncertainties	133
5.8	Mistag matrix systematic uncertainties	133
5.9	Fit heavy flavor content of the tagged and pretag jet samples	144
5.10	Final results for α and $\alpha\beta$ for both taggers.	145
6.1	Pretag and b -tagged event yields	150
6.2	Event yields at each stage of event selection in $t\bar{t}$ Monte Carlo	153
6.3	Summary of pretag efficiencies and scale factors	154
6.4	$t\bar{t}$ acceptance in PYTHIA Monte Carlo	155
6.5	Expected $t\bar{t}$ acceptance in data	156
6.6	b -Tagging efficiency in $t\bar{t}$ events	157
6.7	Expected $t\bar{t}$ event yields for a cross section of 6.7pb	157
6.8	Cross sections used for electro-weak backgrounds	159
6.9	Expected b -tagged diboson, single top, and $Z \rightarrow \tau\tau$ backgrounds	159
6.10	Expected double-tagged diboson, single top, and $Z \rightarrow \tau\tau$ backgrounds	160
6.11	Fit results for the non- W background	164
6.12	Comparison of non- W background methods	166
6.13	W +jets Monte Carlo samples	168
6.14	Heavy flavor fractions for W +jets	169
6.15	Expected b -tag efficiencies in W +jets	170
6.16	Mistag matrix prediction in W +jets data	172
6.17	Mistag background summary	173
6.18	Composition of W +jets sample with ≥ 1 tight tag for base selection	173
6.19	Composition of W +jets sample with ≥ 2 tight tags for base selection	174

6.20	Composition of W +jets sample with ≥ 1 loose tag for base selection .	174
6.21	Composition of W +jets sample with ≥ 2 loose tags for base selection	175
7.1	Jet energy scale uncertainty impact on acceptance	178
7.2	Event yields for optimized selection	181
7.3	Corrected acceptances for the optimized single-tag selection	183
7.4	Corrected acceptances for the optimized double-tag selection	183
7.5	Signal event tagging efficiencies, optimized selection	184
7.6	Expected $t\bar{t}$ event yields, optimized selection	184
7.7	W +Heavy Flavor fractions, optimized selection	186
7.8	Composition of W +jets sample with ≥ 2 loose tags, optimized selection	187
7.9	Composition of W +jets sample with ≥ 2 loose tags, optimized selection	187
7.10	Composition of W +jets sample with ≥ 2 loose tags, optimized selection	188
7.11	Composition of W +jets sample with ≥ 2 loose tags, optimized selection	188
7.12	Composition of W +jets sample with ≥ 2 loose tags, optimized selection, measured $\sigma_{t\bar{t}}$	191
7.13	Composition of W +jets sample with ≥ 2 loose tags, optimized selection, measured $\sigma_{t\bar{t}}$	191
7.14	Composition of W +jets sample with ≥ 2 loose tags, optimized selection, measured $\sigma_{t\bar{t}}$	192
7.15	Composition of W +jets sample with ≥ 2 loose tags, optimized selection, measured $\sigma_{t\bar{t}}$	192
7.16	Acceptance systematics	198
7.17	Parton distribution function systematics	200
7.18	Background systematic summary	201
7.19	Summary of all systematic uncertainties	202
A.1	Event yields at each stage of event selection for CEM electrons	208
A.2	Event yields at each stage of event selection for CMUP muons	209
A.3	Event yields at each stage of event selection for CMX muons	210
B.1	W Monte Carlo samples used to derive the heavy flavor fractions	214
B.2	ALPGEN settings used in this analysis	215
B.3	Acceptances for different ALPGEN samples	220
B.4	Heavy flavor fractions taken directly from ALPGEN W Monte Carlo	224
B.5	Fit results in the all-jet QCD sample	228
B.6	Dependence of K on the number of jets	228
B.7	Heavy flavor fractions used to extract systematic uncertainties	233
B.8	Summary of systematic uncertainties on the W +heavy flavor fractions	233
B.9	Corrected heavy flavor fractions	234

Acknowledgments

Melissa Franklin deserves the most thanks for her patience, intelligence, and kindness. She has been a wonderful teacher and a great friend. I was very lucky to have her as an advisor, and I owe whatever success I have in physics to her. Thanks, Boss.

João Guimarães da Costa and Sebastian Grinstein have contributed significantly to this analysis, but I am more grateful for their mentoring and friendship. João was particularly helpful when I was starting out at CDF, and deserves special thanks for not throttling me when I probably deserved it. Sebastian was also fantastic to work with over the past few years; I'm not sure living at Fermilab would have been tolerable without him there.

Other members of the Harvard CDF group have provided considerable encouragement and support over the last few years. I am especially grateful to Andy Foland, my first advisor, and Lester Miller and Kevin Burkett, who supervised me during my first few years. My fellow graduate students – Ayana, Sal, Ingyin, and John Paul – were all brilliant people and easy to work with, and I thank them for making the last several years more fun.

Chapter 1

Introduction

This dissertation presents a measurement of the top quark pair production cross section ($\sigma_{t\bar{t}}$) in the lepton+jets decay topology with displaced vertex b -tags. The analysis was completed in April 2007 using 1.12 fb^{-1} of $p\bar{p}$ collision data from Run 2 at the Fermilab Tevatron with a center-of-mass energy 1.96 TeV. The data were acquired with the CDF II detector between the summers of 2000 and 2006.

This document includes all information immediately relevant to the measurement of the $t\bar{t}$ cross section; more details can always be found in the references. Chapter 2 describes the phenomenology of top quark production, motivates the measurement of $\sigma_{t\bar{t}}$, and summarizes recent experimental results related to the top quark. Chapter 3 introduces both accelerator chain at Fermilab and CDF II detector used in this analysis. Chapter 4 discusses the tools and techniques used to process the data, including tracking, jet clustering, and lepton identification. Chapter 5 covers in some detail the methods used to identify heavy flavor (b -tagging), including measurements of the b -tagging efficiency and purity; we concentrate on the SECVTX b -tagging algorithm

central to this analysis. Studies of the composition of the b -tagged W +jets sample are presented in Chapter 6, and the final measurements of $\sigma_{t\bar{t}}$ and its uncertainty are shown in Chapter 7. Chapter 8 provides a brief summary of the results and an outlook for future updates of this analysis.

Three appendices are also included. Appendix A shows the data event yield in the $t\bar{t}$ signal samples at each stage of event selection. Appendix B presents a detailed study of the ALPGEN generator, validation of a W +jets Monte Carlo simulation, and the measurement of the heavy flavor content of the W +jets data sample. In Appendix C, we show kinematic distributions for $t\bar{t}$ candidate events, and we compare them with those predicted from Monte Carlo.

Chapter 2

Theoretical and Experimental Top Results

Fundamental particles and their interactions are currently understood as the constituents of the *Standard Model*, an effective field theory that has been very successful in explaining observed particle physics phenomena. Among the particles predicted by the Standard Model, the top quark was among the most recent to be experimentally confirmed, discovered at the Tevatron in 1995 [1].¹ This chapter provides an introduction to the Standard Model top quark, including the phenomenology of top quark production and decay, and we review past experimental results from the Tevatron.

¹The tau neutrino ν_τ was the last to be observed, in 2000.

2.1 The Top Quark

The masses of the quarks and leptons are free parameters in the Standard Model, though the phenomenology is well-determined once the mass spectrum is known. The most striking feature of the top quark, and what makes it unique among the quarks, is its mass, roughly $175 \text{ GeV}/c^2$ [2].² This is nearly 40 times heavier than the next heaviest quark, bottom, and twice as massive as the next heaviest observed particle, the Z boson. Because of its large mass, theoretical predictions for its production and decay are on much firmer ground than they are for lighter quarks [3]. In the next three sections, we summarize the theoretical expectations for the top quark, and we discuss its role in possible extensions to the Standard Model.

2.1.1 Top Production

This dissertation presents a measurement of the *cross section* for $t\bar{t}$ production ($\sigma_{t\bar{t}}$), the total rate for top quark pairs to be produced in proton-antiproton collisions. The minimum center-of-mass energy required to produce a top quark pair is $2m_t$, roughly $350 \text{ GeV}/c^2$, an energy only presently achievable at the Tevatron (see Section 3.1). Here, $t\bar{t}$ pairs are produced strongly through a combination of $q\bar{q}$ annihilation and gluon fusion, the leading-order Feynman diagrams for which are shown in Figure 2.1.

At hadron colliders, not all of the beam energy is used in the collision. Protons are composed of quarks and gluons (collectively called *partons*), each of which carries

²As of this writing, the current world average was $170.9 \text{ GeV}/c^2$. However, we performed the analysis assuming a mass of $175 \text{ GeV}/c^2$, an older result. Most of the theoretical literature and all $t\bar{t}$ simulations used in this dissertation make the same assumption.

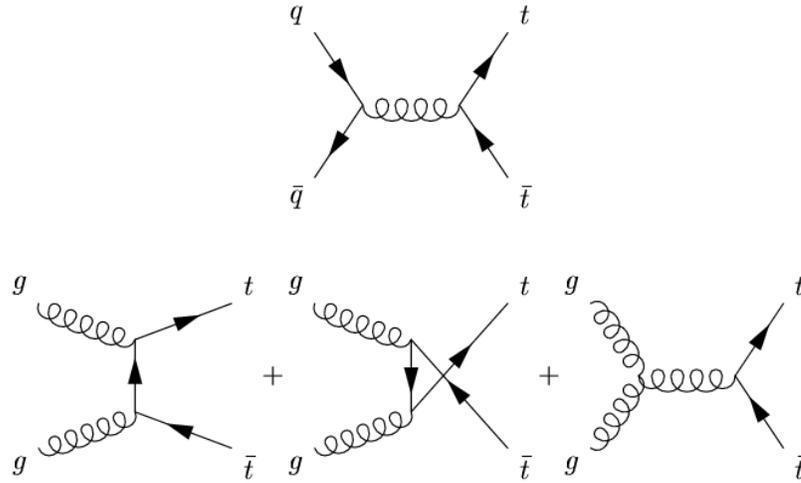


Figure 2.1: Feynman diagrams for $t\bar{t}$ production. The $q\bar{q}$ diagram dominates over gg at the Tevatron.

a fraction x of the hadron momentum. The distribution of x for a given type of quark or gluon is called its *parton distribution function* (PDF), some examples of which are shown in Figure 2.2 [4]. The energy in an inelastic collision is not the center-of-mass energy of the two beams (\sqrt{s}), but rather the energy of the two interacting partons, $\sqrt{\hat{s}}$:

$$\sqrt{\hat{s}} = \sqrt{x_a x_b s}, \quad (2.1)$$

where a and b are the two partons involved in the interaction.

In Run 2 at the Tevatron, the total $p\bar{p}$ center-of-mass energy $\sqrt{s} = 1.96$ TeV. To produce top quark pairs, the collision must take advantage of the high end of the PDF's. Assuming both partons have comparable momenta, the average momentum fraction of the partons x_{min} is large:

$$x_{min} \simeq \frac{2m_t}{\sqrt{s}} = 0.18. \quad (2.2)$$

As shown in Figure 2.2, quarks are more likely to have this large a momentum fraction, so the $q\bar{q}$ annihilation process dominates over gluon fusion at the Tevatron, though the gluon contribution is not negligible.

$t\bar{t}$ Cross Section Calculation

The total cross section for $p\bar{p} \rightarrow t\bar{t}$ can be written as an integral over x_a and x_b :

$$\sigma_{t\bar{t}} = \sum_{a,b} \int_0^1 \int_0^1 dx_a dx_b f_a^p(x_a) f_b^{\bar{p}}(x_b) \hat{\sigma}_{ab \rightarrow t\bar{t}}(\hat{s}, m_t). \quad (2.3)$$

Here, the sum runs over all all different types of partons for a and b , and $f^{p(\bar{p})}$ is the PDF for that parton in a proton (antiproton). The quantity $\hat{\sigma}$ is the point cross section, calculated for given a , b , and \hat{s} ; this term is zero when the center-of-mass energy is well below the kinematic threshold or when ab is not gg or $q\bar{q}$.

Ignoring the gluon fusion contribution, we can write a simple expression for the leading-order point cross section [5]:

$$\hat{\sigma} = \frac{8\pi\alpha_s}{27\hat{s}} \sqrt{1 - \frac{4m_t^2}{\hat{s}}} \left(1 + \frac{2m_t^2}{\hat{s}}\right). \quad (2.4)$$

Figure 2.3 shows this cross section before the PDF effects are included. The central line is given for $m_t = 175 \text{ GeV}/c^2$, and the two bounding curves are for $m_t = 170$ and $180 \text{ GeV}/c^2$. The cross section peaks very near the threshold at $2m_t$, and falls as the inverse of \hat{s} when the center-of-mass energy is large.

To evaluate Equation 2.3 completely, the PDF's are evaluated at a predetermined factorization scale, μ_f , and the point cross section depends on α_s evaluated at a second scale, μ_r . Both of these are commonly assumed to be m_t , and the dependence of the result on this assumption is factored into the uncertainties [6]. The cross section itself is not sensitive to the choice of scales, but fixed-order calculations will have some residual dependence. As higher orders are included, the uncertainty in the theoretical cross section due to these choices gets smaller [3]. At leading order (LO), for instance, the relative uncertainty on $\sigma_{t\bar{t}}$ can be as large as 50%, while at next-to-leading order (NLO) this uncertainty is reduced to 12% [7]. At present, the $t\bar{t}$ cross section has been calculated to next-to-next-to-leading-order (NNLO) and includes soft gluon corrections to next-to-next-to-next-to-leading logarithmic (NNNLL) terms. The result is approximately 30% higher than the leading order calculation. The logarithmic terms incorporate the behavior of the calculation near threshold when a soft gluon is emitted in the initial state; the total impact of these terms is roughly 3% for $\mu_f = \mu_r = m_t$ and increases for larger μ [3].

For a top quark mass of $175 \text{ GeV}/c^2$, the cross section is calculated to be 6.7 pb for $\sqrt{s} = 1.96 \text{ TeV}$ [3][6]. The maximum of $\frac{d\sigma_{t\bar{t}}}{d\sqrt{s}}$ occurs when the center-of-mass energy is approximately 50% greater than the threshold value. At this energy, top quarks will be produced with semi-relativistic energies, $\beta \approx 0.5$. As expected, the $q\bar{q}$ annihilation dominates, contributing 85% of the total cross section [7]. In Run 1, with $\sqrt{s} = 1.80 \text{ TeV}$, the calculation yields 5.2 pb, with 90% of the cross section coming from $q\bar{q} \rightarrow t\bar{t}$ processes [3][7]. The kinematics of $t\bar{t}$ events in Run 2, taken from a Monte Carlo simulation, are shown in Figures 2.4 through 2.7. The $t\bar{t}$ invariant mass is equivalent

to $\sqrt{\hat{s}}$. Distributions of p_T and η for top decay daughters, W bosons and b quarks, are also shown.

The uncertainty in the calculation is derived from two main sources: the scale dependence mentioned above and the uncertainty on the parton distribution functions [6]. To evaluate the scale dependence, the value of $\mu(= \mu_f = \mu_r)$ is varied between $\frac{m_t}{2}$ and $2m_t$ [8]. (Varying μ_f and μ_r independently has been shown to provide a negligibly different uncertainty.) The PDF uncertainty is evaluated by repeating the calculation with different PDF parameterizations, taking the RMS deviation in the result as the uncertainty. The total cross section for a top quark mass of 175 GeV/c^2 is $6.7_{-0.9}^{+0.7}$ pb, where the largest contribution to the uncertainty is from the high- x gluon PDF [6].

Because top quarks are produced so close to their kinematic threshold (see Figure 2.3), the cross section is also a strong function of the mass of the quark. Figure 2.8 shows the expected cross section as a function of the top mass from two different calculations, with systematic uncertainties on the curves illustrated as well. For a top quark mass of 170 GeV/c^2 , for instance, $\sigma_{t\bar{t}}$ is 7.8 ± 0.9 pb [6].

Single Top

In addition to $t\bar{t}$ production, single top quarks can be produced through the weak interaction. The two dominant diagrams are shown in Figure 2.9. The NLO theoretical cross section for s -channel top production is 0.88 ± 0.05 pb, and the cross section for t -channel production is 1.98 ± 0.08 pb [9]. Since the process is electroweak rather than QCD, the calculation is much more precise than for $t\bar{t}$. Single top has not yet

been observed experimentally, but is expected to constitute a small background to $t\bar{t}$.

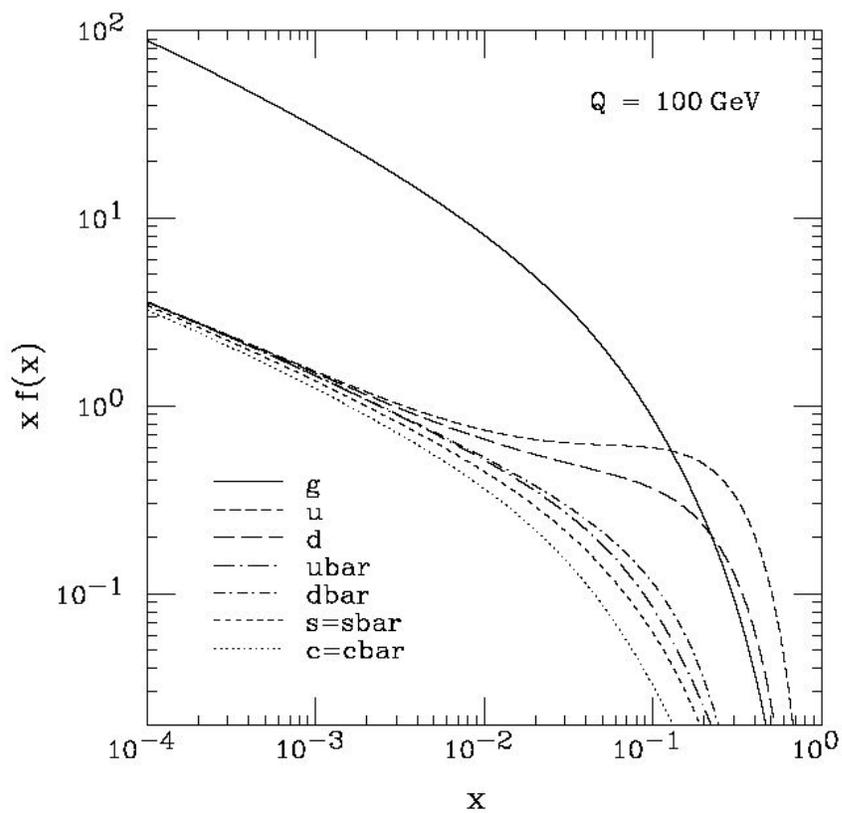


Figure 2.2: Example parton distribution functions inside protons, taken from the CTEQ5L parameterization and evaluated at a momentum scale $Q = 100$ GeV. The curves are multiplied by x to illustrate better the behavior at both ends [4].

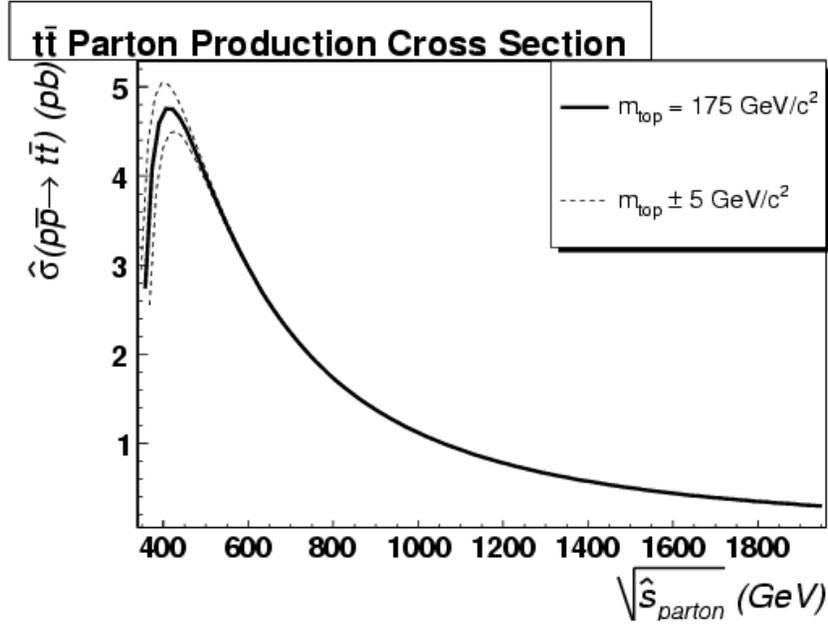


Figure 2.3: Leading-order cross section for $q\bar{q} \rightarrow t\bar{t}$ as a function of the center-of-mass energy of the quarks. The central value is for a top mass of $175 \text{ GeV}/c^2$, and the others are for masses of 170 and $180 \text{ GeV}/c^2$. The effect of the PDFs is not incorporated.

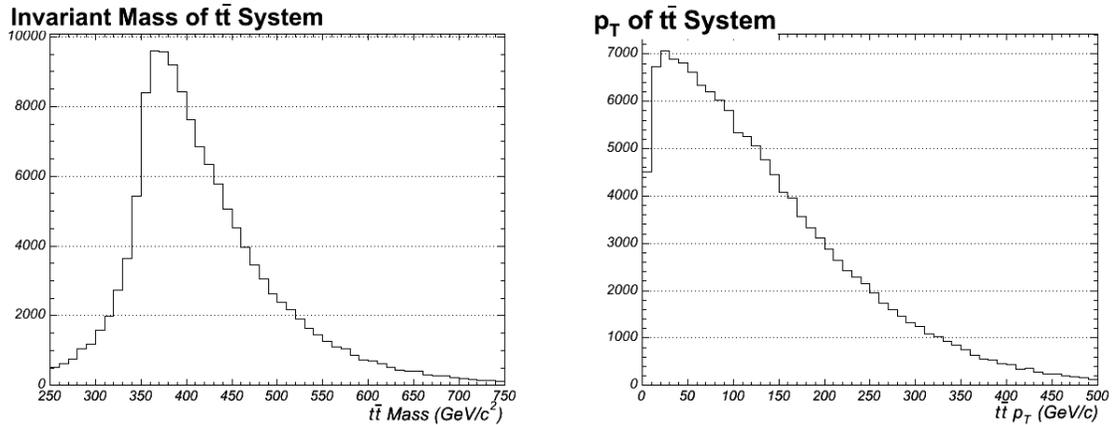


Figure 2.4: Distributions of $t\bar{t}$ mass (left) and p_T (right) taken from a $t\bar{t}$ simulation. We assume a top quark mass of $175 \text{ GeV}/c^2$. The invariant mass distribution shows the same dependence on $\sqrt{\hat{s}}$ as the LO cross section shown in Figure 2.3.

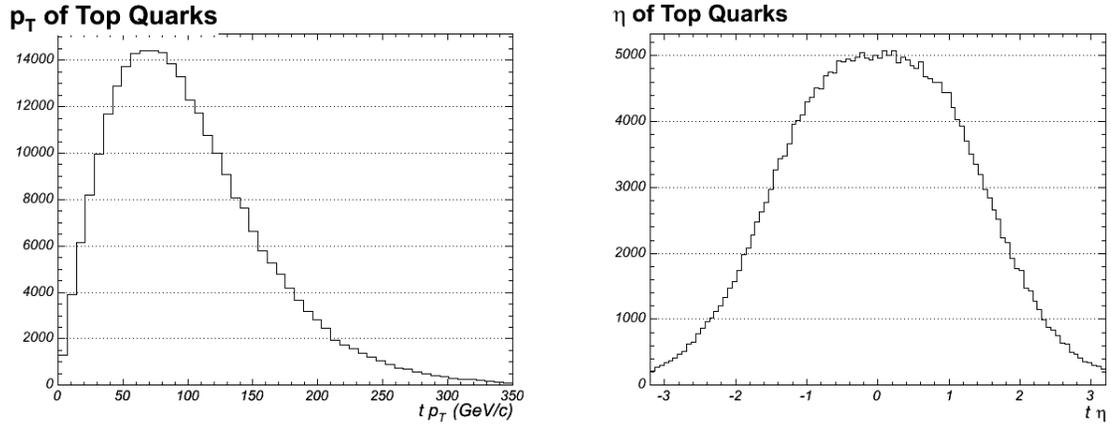


Figure 2.5: Distributions of top quark p_T (left) and η (right) taken from a $t\bar{t}$ simulation. We assume a top quark mass of $175 \text{ GeV}/c^2$.

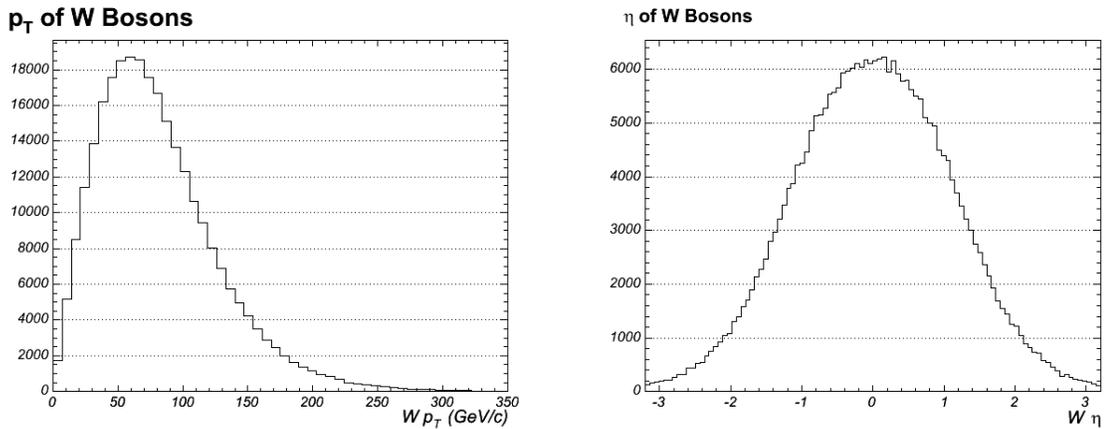


Figure 2.6: Distributions of p_T (left) and η (right) for W bosons from top quark decays, taken from a $t\bar{t}$ simulation. We assume a top quark mass of $175 \text{ GeV}/c^2$.

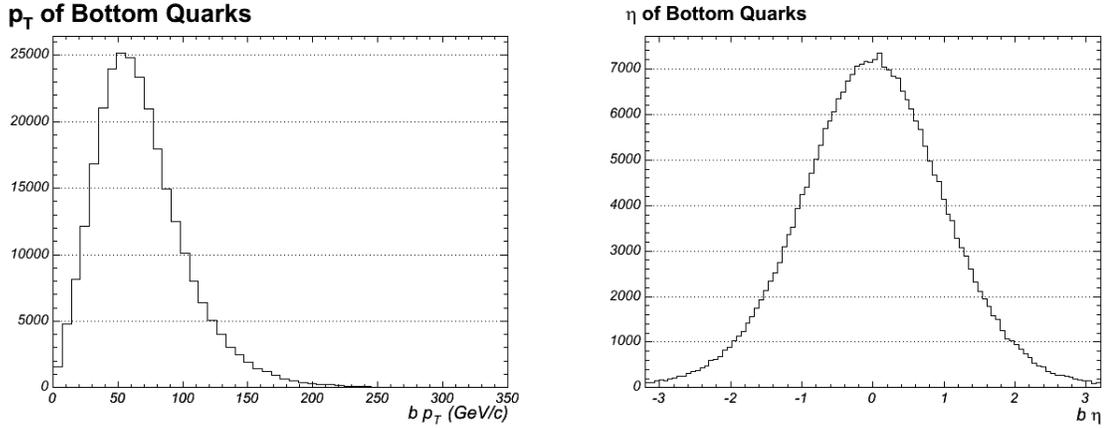


Figure 2.7: Distributions of p_T (left) and η (right) for b quarks from top quark decays, taken from a $t\bar{t}$ simulation. We assume a top quark mass of $175 \text{ GeV}/c^2$.

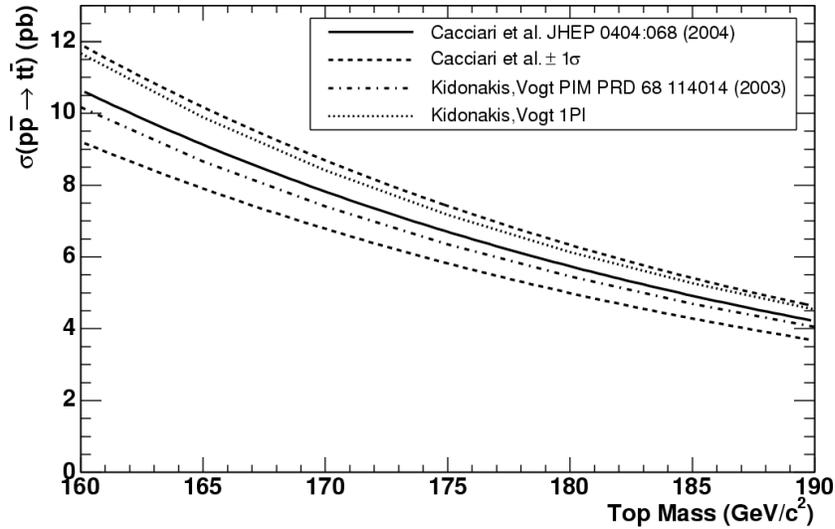


Figure 2.8: Expected $t\bar{t}$ cross section as a function of the top mass. The Cacciari *et al.* curve and its uncertainties are evaluated at NLO with next-to-leading-logarithmic corrections; The two results from Kidonakis *et al.* are evaluated at NNLO with NNLL corrections, with two different choices of reference frame.

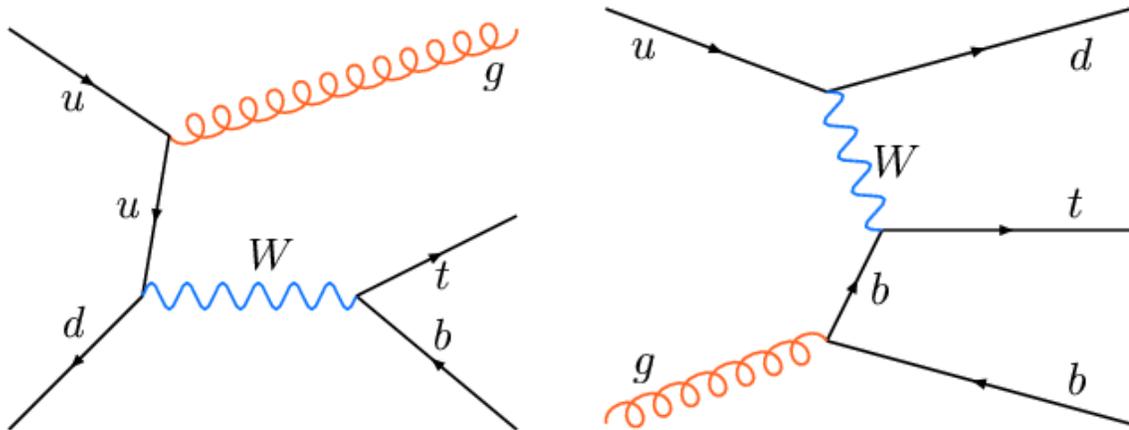


Figure 2.9: Feynman diagrams contributing to single top production, s -channel (left) and t -channel (right).

2.1.2 Top Decay

Because the top quark's mass is so large, its decays happen on a time scale shorter than that required for hadronization. Its width can be approximated with the expression [10]:

$$\Gamma_t \approx \Gamma(t \rightarrow Wb) \approx \frac{G_F}{8\pi\sqrt{2}} m_t^3 |V_{tb}|^2 \approx 1.6 \text{ GeV}, \quad (2.5)$$

which corresponds to a lifetime on the order of 10^{-25} s. Top is the only quark capable of being studied as a free particle.

The Standard Model top quark is expected to decay to a W boson and a bottom (b) quark nearly 100% of the time; other possible weak decays, $t \rightarrow Ws$ and $t \rightarrow Wd$, are suppressed by squares of the CKM matrix elements V_{ts} and V_{td} (4% and 0.8% in a global Standard Model fit [2]) and contribute less than one percent to top decays. More exotic flavor-changing neutral current decays are suppressed by many more orders of magnitude.

Because the top decays nearly universally to Wb , $t\bar{t}$ final states will all be composed of 2 W 's and two bottom quarks. We therefore classify top final states according to the decays of the two W bosons, which are predicted by the Standard Model and have been confirmed in experiment. A W^+ can decay to a $q\bar{q}'$ or to a lepton-neutrino pair. For the former, there are six possibilities: $u\bar{d}$ or $c\bar{s}$, each in three color combinations. We ignore CKM-suppressed combinations, and note that $t\bar{b}$ pairs are kinematically forbidden. For the latter, $e^+\nu_e$, $\mu^+\nu_\mu$, and $\tau^+\nu_\tau$ are equally likely. To first approximation, each of these has a branching fraction of $\frac{1}{9}$, and the rates are equivalent for the charge conjugates.

The relative sizes of $t\bar{t}$ final states is shown in Figure 2.10. The most common final state is the *all-hadronic* mode, in which both W 's decay to jets, with 44% of all $t\bar{t}$ decays. The *dilepton* channel includes all final states where both W 's decayed leptonically, with a total branching fraction of 11%. Finally, the *lepton+jets* signature, including one leptonic and one hadronic W decay, occupies the remaining 44%; however, τ particles are short-lived and decay 64% of the time to hadrons, which changes the fraction of events with leptons in the final state. For this analysis, we focus on the lepton+jets channel, where the identified lepton is an electron or a muon; when τ decays to leptons are included, the total branching fraction for this channel is roughly 35%.

Top Pair Decay Channels

$\bar{c}s$	electron+jets	muon+jets	tau+jets	all-hadronic		
$\bar{u}d$						
τ^-	$e\tau$	$\mu\tau$	$\tau\tau$	tau+jets		
μ^-	$e\mu$	$\mu\mu$	$\mu\tau$	muon+jets		
e^-	$e\bar{e}$	$e\mu$	$e\tau$	electron+jets		
W decay	e^+	μ^+	τ^+	$u\bar{d}$	$c\bar{s}$	

Figure 2.10: Relative sizes of $t\bar{t}$ final states. In this analysis, we focus on the lepton+jets signature, where the lepton is either a muon or an electron.

2.1.3 Non-Standard Model Top Phenomenology

Within the Standard Model, the $t\bar{t}$ cross section is theoretically determined to a precision of 10%, once the top quark mass is known [6]. The measurement of $\sigma_{t\bar{t}}$, however, is more than a test of QCD. In many extensions to the Standard Model, top quarks can be pair-produced through non-QCD processes, or may decay to something other than Wb . Precise knowledge of the $t\bar{t}$ cross section provides direct constraints on these models.

Examples of non-Standard Model production modes include resonant production and heavy particles decaying to top quarks. Feynman diagrams for these processes are shown in Figure 2.11. Heavy neutral particles, such as a Z' boson or a massive gluon, could be produced in $q\bar{q}$ annihilation, and can decay to a $t\bar{t}$ pair [11]. Similarly, pair-produced particles, such as the T proposed in Little Higgs theories, may preferentially decay to top quarks and gauge bosons (A) which escape undetected [12]. In both cases, the experimental signature would be consistent with QCD $t\bar{t}$, and would augment the measured cross section.

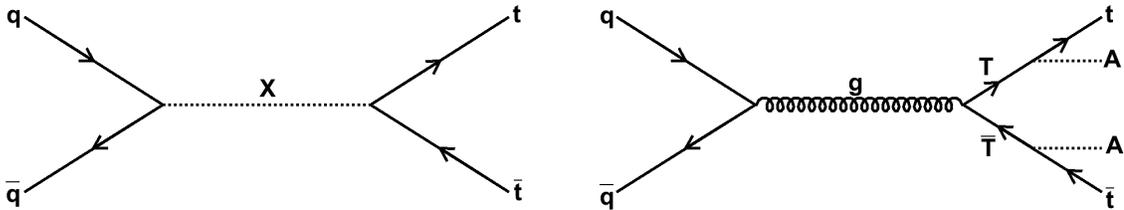


Figure 2.11: Example Feynman diagrams for non-Standard Model production of $t\bar{t}$. On the left, a new neutral boson, such as a Z' , decays to $t\bar{t}$. On the right, pair-produced particles, such as the heavy top in Little Higgs models, each decay to a top quark and a neutral boson (A).

A final possibility is for new physics to alter the decays of top quarks. We expect top quarks to decay nearly 100% of the time to Wb , and we determine the relative sizes of $t\bar{t}$ final states using the known W branching ratios. If an alternate top quark decay channel, such as $t \rightarrow H^+b$, is significant, the measured cross sections in different final states will disagree with one another [13]. As shown in Figure 2.12, the charged Higgs decays differently from the W , favoring τ final states to those with lighter decay products [14]. Comparisons of $\sigma_{t\bar{t}}$ measurements in the electron, muon, and τ channels provide constraints on non-Standard Model top decays.

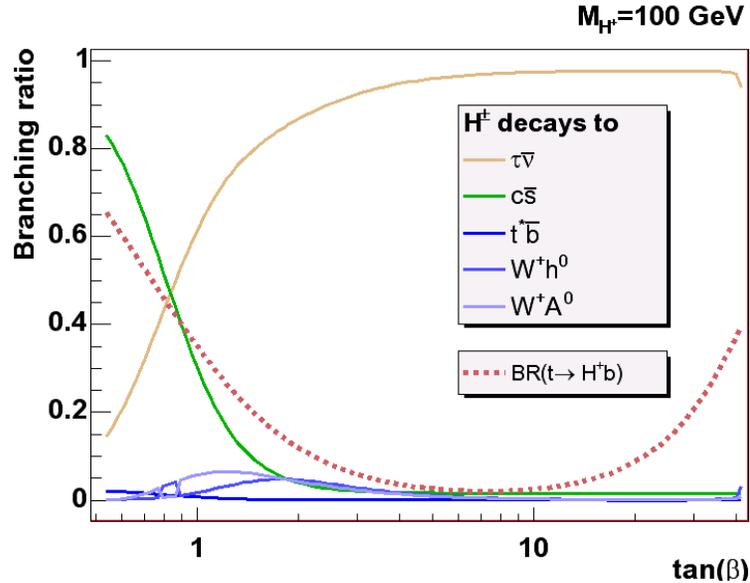


Figure 2.12: Branching fractions for the top quark and charged Higgs as a function of $\tan \beta$, a free parameter in the Minimal Supersymmetric Standard Model corresponding to the ratio of expectation values for the two Higgs doublets [14]. We assume a Higgs mass of $100 \text{ GeV}/c^2$.

2.2 Experimental Top Quark Results

To date, all experimental results on the top quark have come from the Fermilab Tevatron, described in detail in Chapter 3. Measurements have been made with two detectors experiments, CDF and DØ, and at two center-of-mass energies, 1.8 TeV in Run 1 and 1.96 TeV in Run 2.

Most top quark properties are independent of \sqrt{s} , and results from Run 1 and Run 2 can be combined. As mentioned above, the quark masses are free parameters in the Standard Model, but the production and decay rates are well-defined once the mass is known. If we use the best results from both experiments in all decay channels, the average top quark mass is $170.9 \pm 1.8 \text{ GeV}/c^2$, with excellent consistency between each contributing measurement [15]. A summary of the individual results is shown in Figure 2.13.

At this mass, the theoretical $t\bar{t}$ cross section is $7.6 \pm 0.9 \text{ pb}$ in Run 2 [3]. A summary of the measured cross sections, which were calculated assuming $m_t = 175 \text{ GeV}/c^2$, is shown in Figure 2.14 [16]. Because the kinematic acceptance for $t\bar{t}$ events has some dependence on the mass as well, measurements of the cross section will vary inversely with the top quark mass. Using the most precise $\sigma_{t\bar{t}}$ results from CDF in Run 2 in all $t\bar{t}$ final states, and adjusting the results to the average top quark mass, we see excellent agreement with the theoretical predictions, as illustrated in Figure 2.15.

Because top production is so near the kinematic threshold, the theoretical cross section calculation in Run 2 is 25% larger than it was in Run 1. Figure 2.16 shows the combined cross section results from CDF in Run 1 and Run 2 as a function of \sqrt{s} ; both points are consistent with the Standard Model prediction, although the

uncertainties are too large to provide a stringent test of the slope.

Additional top quark analyses have provided measurements of the top quark's spin (via the W helicity), charge, production mechanism ($q\bar{q}$ vs. gg , limit on resonant production), branching fractions (V_{tb} , flavor-changing neutral currents, charged Higgs), and lifetime [16]. To date, there is no strong evidence for non-Standard Model physics in the top sector. Finally, the production of single top quarks through the weak interaction has not yet been conclusively observed, though recent evidence suggests that the cross section for this process is at or near its Standard Model prediction [17].

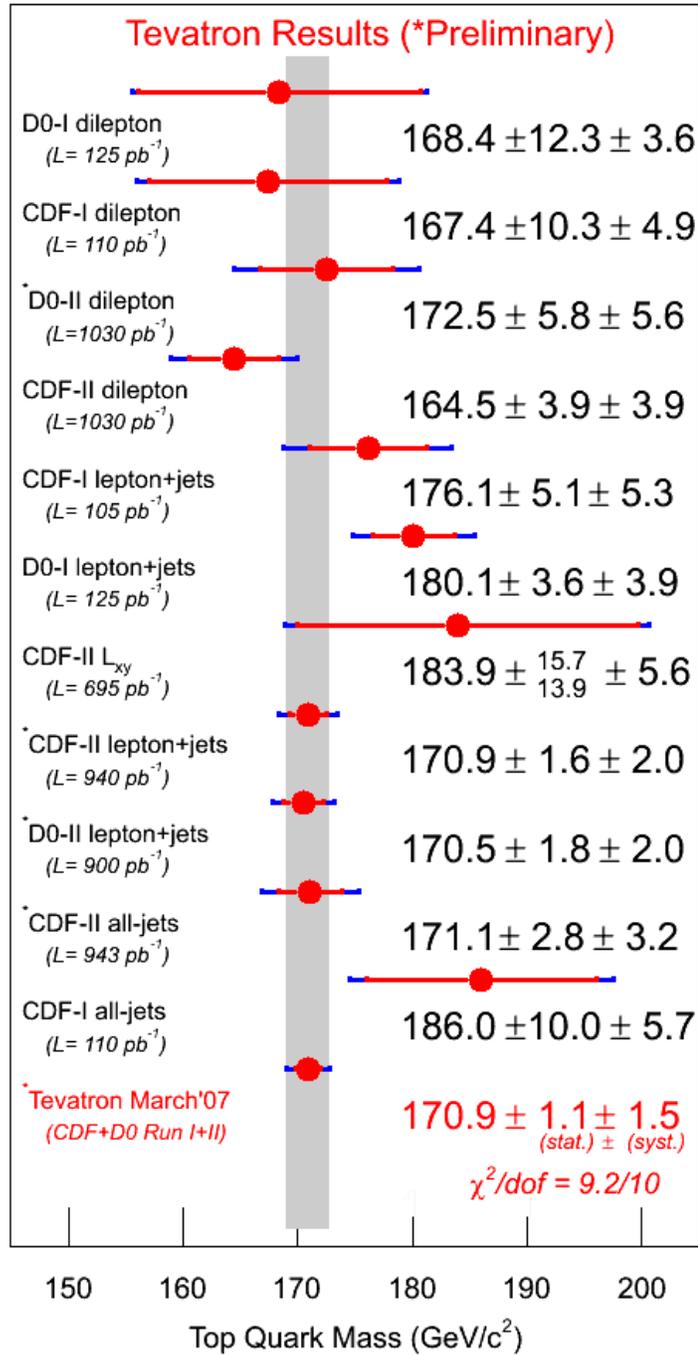


Figure 2.13: Summary of top mass results from the Tevatron. The combined value includes the best results from CDF and DØ in both Run 1 and Run 2 in all decay modes.

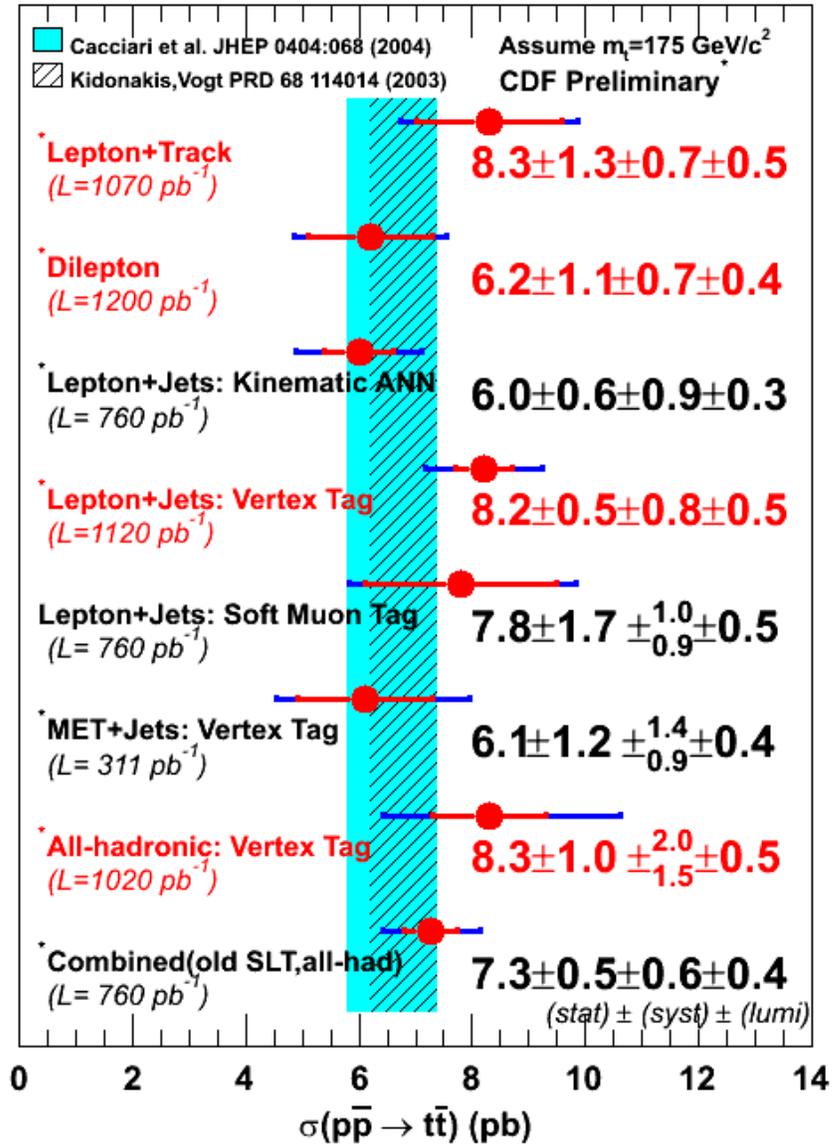


Figure 2.14: Summary of measurements of the $t\bar{t}$ cross section at CDF in Run 2, including multiple measurements in the same decay channel. Results are quoted assuming a top mass of $175 \text{ GeV}/c^2$. The “Combined” value includes previous results (with $\leq 330 \text{ pb}^{-1}$) from the dilepton, all-hadronic, and three lepton+jets (soft muon tag, kinematic ANN, and vertex tag) analyses.

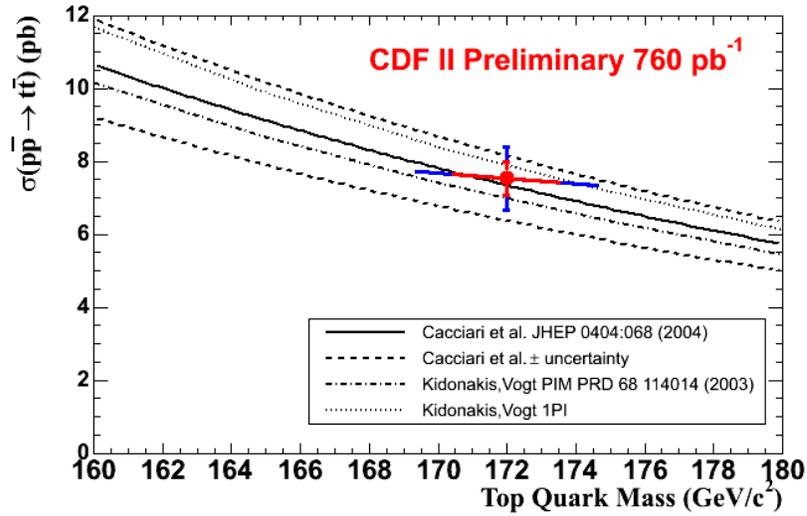


Figure 2.15: Combination of the best measurements of the $t\bar{t}$ cross section in CDF Run 2 quoted at the CDF-average top mass. The residual slope shown in the error bars illustrates the dependence of the measured cross section on the top mass.

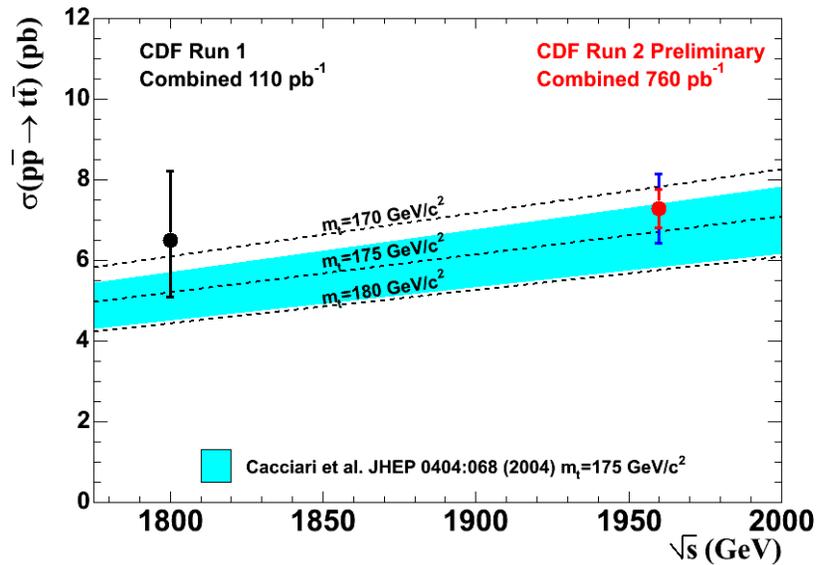


Figure 2.16: The dependence of the $t\bar{t}$ cross section on the center-of-mass energy, \sqrt{s} . The two points are combined results from CDF Run 1 and Run 2.

Chapter 3

Experimental Apparatus

In this Chapter, we introduce the physical systems used in this experiment. We describe the accelerator chain at Fermilab in Section 3.1 and the CDF II detector in Section 3.2. In both cases, more technical details can be found in the references.

3.1 The Accelerator Complex

The Fermi National Accelerator Laboratory (Fermilab) is a Department of Energy-operated facility in Illinois dedicated to high-energy physics research [18]. Founded in 1967, it has hosted the experiments that discovered the bottom quark (1977), top quark (1995), and τ neutrino (2000). The current physics program includes neutrino, astrophysics, and collider experiments. Fermilab is also the site of the Tevatron, the world's highest-energy accelerator, which boosts beams of protons and antiprotons to energies of 980 GeV. The beams travel in opposite directions through a four-mile circular tunnel, crossing at five fixed locations. At two of these, called B \emptyset and D \emptyset ,

the protons and antiprotons are allowed to interact, and a multi-purpose detector records the energy and momenta of outgoing particles from collisions at each. The CDF II Detector (CDF) sits at $B\emptyset$, and the $D\emptyset$ Detector is located at $D\emptyset$. The analysis presented in this dissertation uses data from CDF only.

The Tevatron comprises multiple subsystems: the pre-accelerator, the linear accelerator, the Booster, the Main Injector, an antiproton source, and the main Tevatron ring, all of which are described in great detail in Ref. [19]. The layout of the accelerator is shown schematically in Figure 3.1.

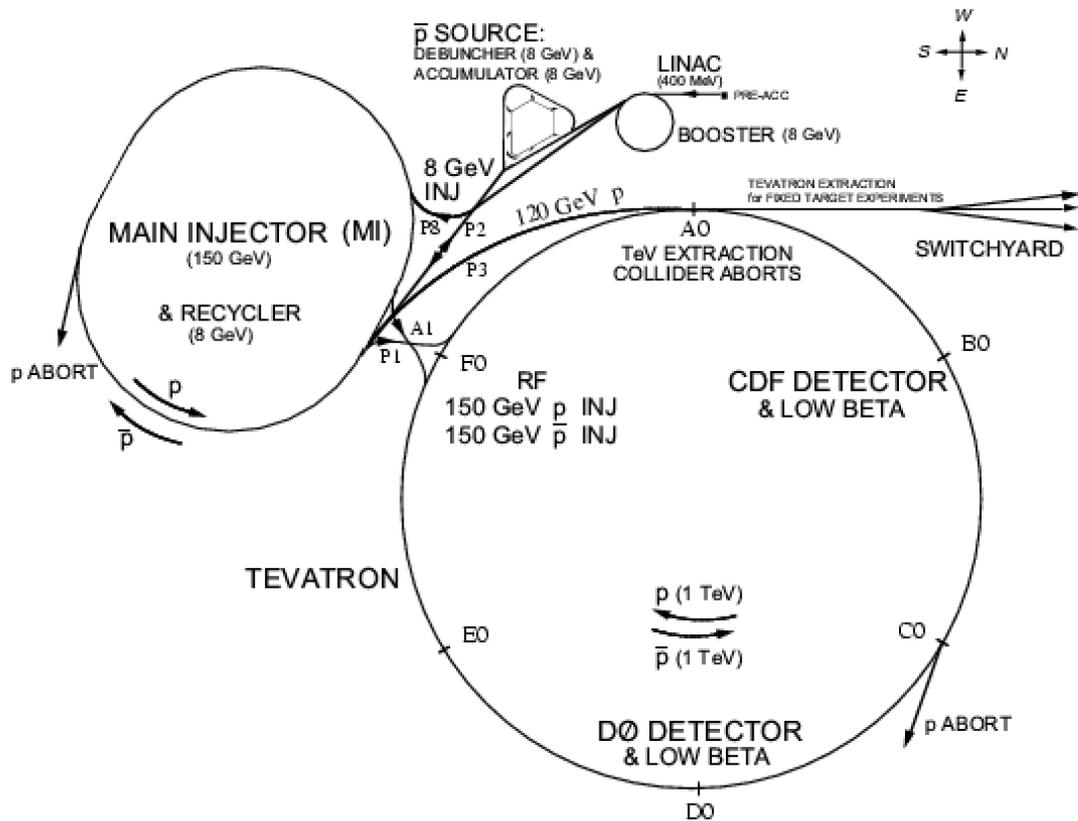


Figure 3.1: Layout of the accelerator chain at Fermilab.

The Pre-accelerator initiates the chain by producing negatively-charged hydrogen ions at an electric potential of -750 kV. Every 66 milliseconds, the H^- ions drift through an accelerating column to a grounded wall, picking up a kinetic energy of 750 keV. A transfer line then transports the ions to the linear accelerator (Linac). The Linac consists of 150 meters of RF cavities, which both accelerate the ions to a kinetic energy of 400 MeV and separate the continuous stream particles from the Pre-accelerator into discrete clusters, or *bunches*.

At the end of the Linac, the ions pass through a thin copper foil, which strips the electrons. The beam of protons next enters the Booster, the first circular accelerator. The Booster comprises 75 magnets arranged in a circle with radius 75 meters, and it accelerates the protons to a kinetic energy of 8 GeV before passing them to the Main Injector, a 530-meter-radius synchrotron. The Main Injector has two functions, preparing 150-GeV protons for transmission into the Tevatron and producing a 120-GeV beam of protons for the production of \bar{p} 's in the Antiproton Source.

At the entrance to the Antiproton Source, 120-GeV protons strike a large nickel target, producing a spray of particles containing only 20 antiprotons for every million protons. The outgoing particles pass through a "lithium lens," which produces a solenoidal magnetic field to collimate and focus them, and are then filtered for 8-GeV antiprotons using magnetic spectroscopy. Since the antiprotons will be non-uniform in energy and direction, they are sent to the Debuncher, a rounded triangular synchrotron with a mean radius of 90 meters. The Debuncher's main function is to convert the broad, diffuse antiproton clumps into a stable, uniform beam. This is accomplished through a technique called "stochastic cooling," in which the beam's

properties (horizontal and vertical profile, momentum) are recorded at one point in its path, and the spread is corrected at another point. After about 100 milliseconds, a steady beam of 8-GeV antiprotons is achieved.

Antiprotons from the Debuncher are moved to the Accumulator, which shares the same structure and *accumulates* all the antiprotons into a single *stack*. The antiproton beam is continuously cooled, so that its physical size in the transverse plane and the width of the \bar{p} momentum distribution are reduced as much as possible until it is sent to the Main Injector to be prepared for collisions. Antiprotons are accelerated to 150 GeV in Main Injector, the same energy as the protons.

The act of extracting protons and antiprotons for collisions in the Tevatron, called a *shot*, begins with the passage of 36 proton bunches from the Main Injector. The beam structure comprises three trains of 12 bunches, with a 396 ns gap between bunches and 2.6 μ s between trains for beam aborts. The antiproton beam has the same structure as the protons, but travel in a direction opposite the proton beam. With both protons and antiprotons loaded, the Tevatron provides the final acceleration to 980 GeV, at which point collisions at $D\bar{O}$ and $B\bar{O}$ begin.

3.1.1 Luminosity

The rate at which interactions occur in these collisions is determined by a product of the interaction *cross section*, σ , and the *instantaneous luminosity* of the beams, \mathcal{L} , with units of $(\text{area time})^{-1}$. In terms of the beam properties, the luminosity can be expressed as:

$$\mathcal{L} = \frac{f N_p N_{\bar{p}} B}{4\pi\rho^2} \approx 10^{31-32} \text{cm}^{-2} \text{s}^{-1} \quad (3.1)$$

- f - revolution frequency
- N_p ($N_{\bar{p}}$) - # protons (antiprotons) per bunch
- B - # bunches
- ρ - Gaussian width of the beam, typically 25 μm

Since the instantaneous luminosity is dependent on the number of protons and antiprotons in the beam and the beam width (which increases over time), it continually decreases after collisions begin. Shots are typically performed every 24 to 36 hours, and instantaneous luminosities vary from as high as $250 \times 10^{30} \text{cm}^{-2} \text{s}^{-1}$ immediately after a shot to $20 \times 10^{30} \text{cm}^{-2} \text{s}^{-1}$ just before dropping the beam. The peak luminosity after each shot is shown in Figure 3.2. During normal operations, there will be a short downtime of about four hours before the next shot, with the accelerator operating year-round, 24 hours a day.

For physics analyses, it is common to refer to the amount of data collected as an *integrated luminosity*, the integral of the instantaneous luminosity with respect to time over the data collection period. The total integrated luminosity since the beginning of Run 2 is shown in Figure 3.3 [20]. All results described in this document correspond to 1.12fb^{-1} of data collected between the summers of 2000 and 2006.¹

¹Due to some inefficiencies in data collection and occasions when the full detector was not functioning, the integrated luminosity used here is smaller than the total delivered over the same time.

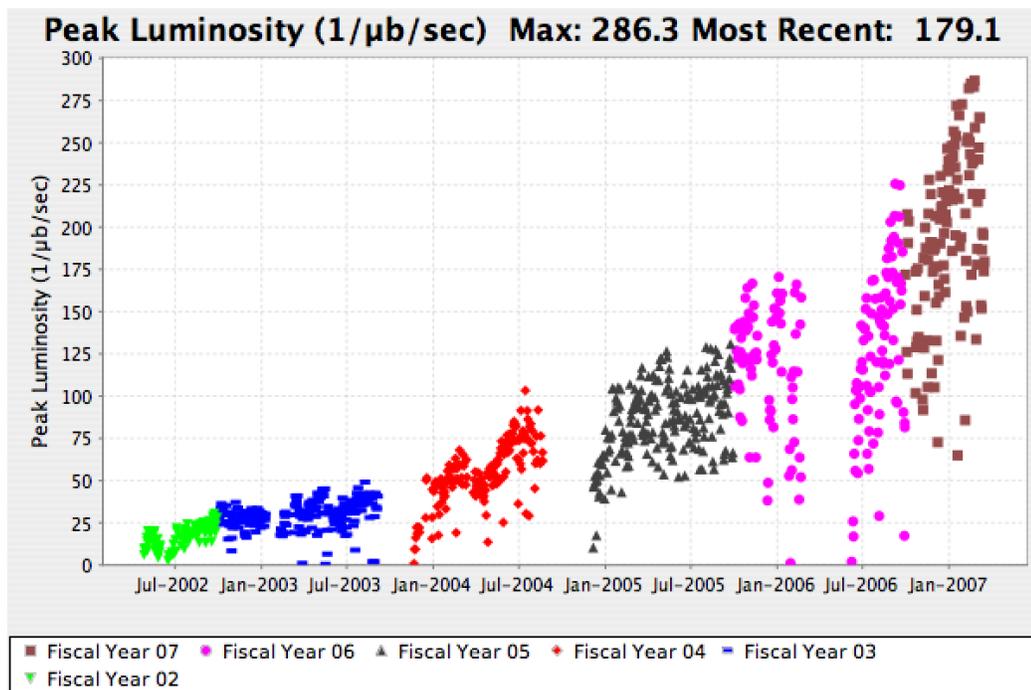


Figure 3.2: The peak instantaneous luminosity for every shot in Run 2. The points are an average of the luminosity measured at CDF and DØ. This analysis uses data though September of 2006.

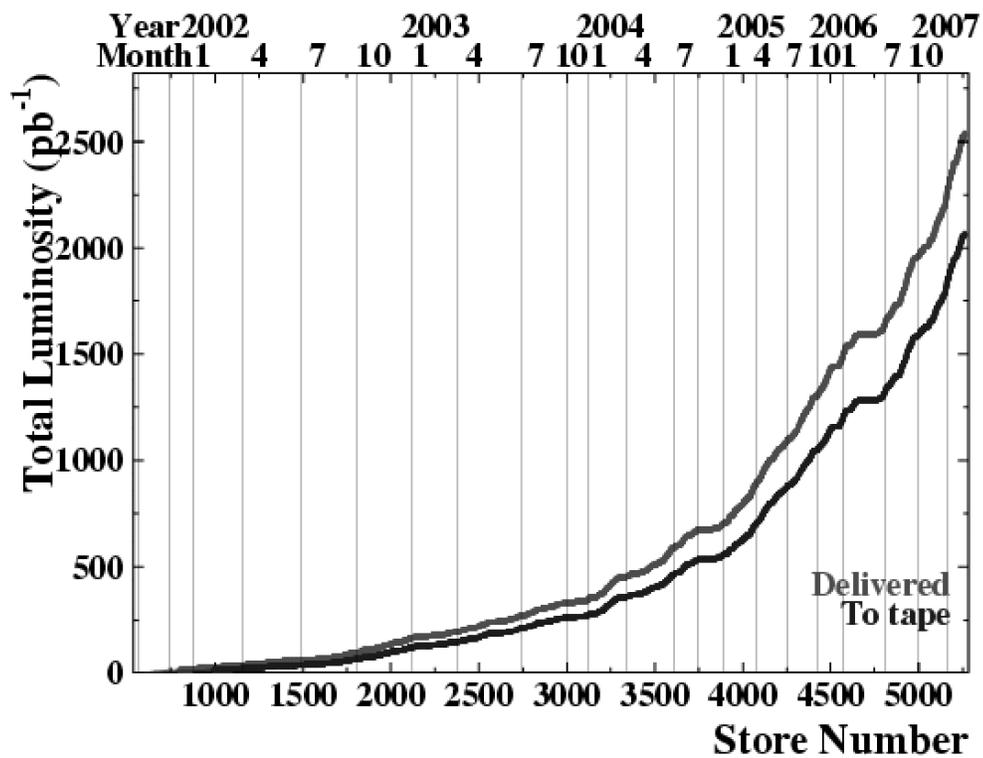


Figure 3.3: The total integrated luminosity recorded at CDF in Run 2. This analysis uses data roughly through store 4500, comprising roughly 1.12 fb^{-1} .

3.2 The CDF Detector

CDF II is a general-purpose detector designed to record the energies and momenta of particles produced in proton-antiproton collisions [21]. The detector is roughly cylindrically symmetric, with the beamline constituting the polar axis in its coordinate system. The $+\hat{z}$ axis is defined to be the direction of the proton beam at the collision point (east), and the $+\hat{x}$ axis is oriented north.

The particles produced in collisions will be highly relativistic, motivating a set of Lorentz-invariant coordinates. In particular, in place of the polar angle θ , we use the pseudo-rapidity η :

$$\eta \equiv -\ln\left(\tan\frac{\theta}{2}\right) \approx \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) \quad (3.2)$$

At hadron colliders, particle production is constant as a function of rapidity (the last term in the equation above); for massless particles, or when $E \gg m$, the rapidity and pseudo-rapidity are equivalent, and η has the advantage of only depending on the polar angle θ . With the azimuthal angle φ , the angular separation between two boosted tracks $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2}$ is also Lorentz invariant.

The CDF II system comprises several subsystems, each of which is responsible for recording a different aspect of the collisions. The tracking system, a drift chamber and a silicon tracker inside an axial magnetic field, records the positions of charged particles as they pass through, allowing for reconstruction of their paths and precise measurement of their momenta. Outside the magnetic field, electromagnetic and hadronic calorimeters absorb and record the energies of electrons, photons, and hadrons. At the largest radial distance from the interaction region, a muon detec-

tion system identifies tracks that pass through the calorimeter and surrounding steel. Each of these systems will be discussed in turn. We will also describe the CDF luminosity monitor and data acquisition system, both critical to physics analyses carried out with the detector. A cross-sectional view and a three-dimensional cartoon of the detector are shown in Figures 3.4 and 3.5.

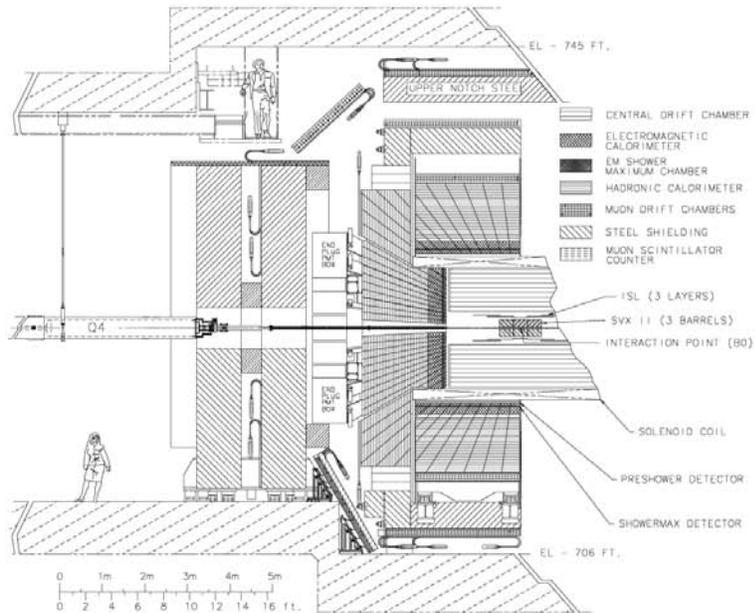


Figure 3.4: Cross-sectional view of the CDF II detector.

3.2.1 Tracking Systems

The tracking system is the most critical set of subdetectors for the analysis presented in this thesis. The system has two key components: an eight-layer silicon system covering radii from roughly 1 cm to 30 cm, and a large open-cell drift chamber (COT) between radii of 40 cm and 1.3 m, both sitting in a 1.4 T solenoidal

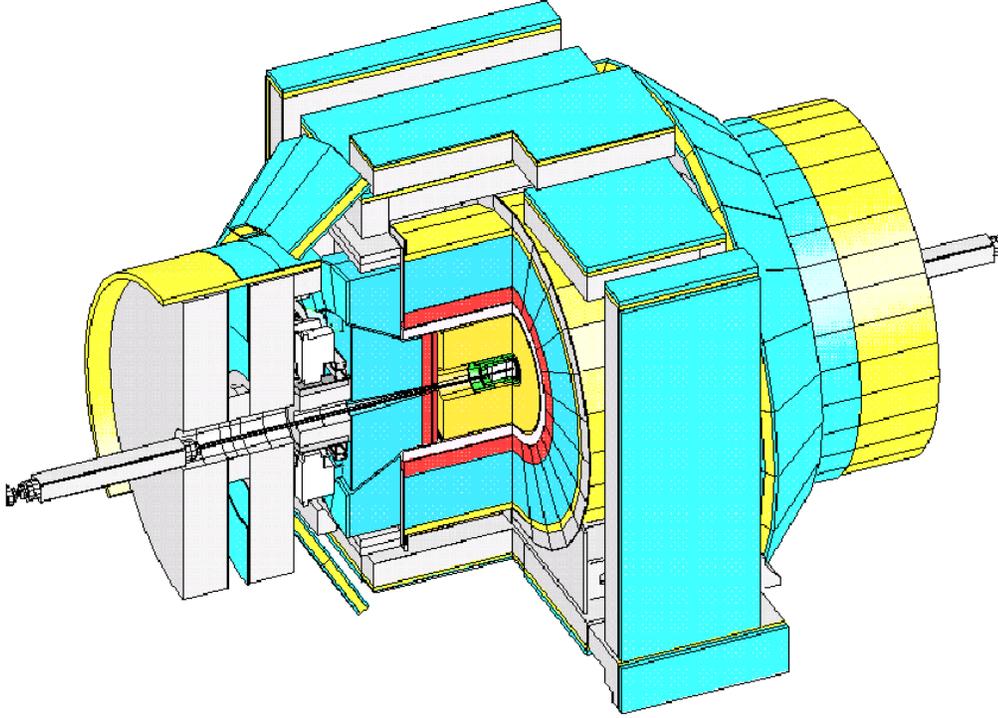


Figure 3.5: Cartoon illustrating the shape of the CDF II detector. The various subdetectors are shown to scale.

magnetic field pointed in the $+\hat{z}$ direction. The layout of these detectors is shown in Figure 3.6. Helical tracks are reconstructed from hit information recorded in both detectors, allowing for precise transverse momentum (p_T) measurements and excellent impact parameter resolution for displaced vertex bottom quark tagging (b -tagging). We discuss the technical specifications for each detector in this section; details on tracking and b -tagging algorithms and their performance will follow in Chapters 4 and 5.

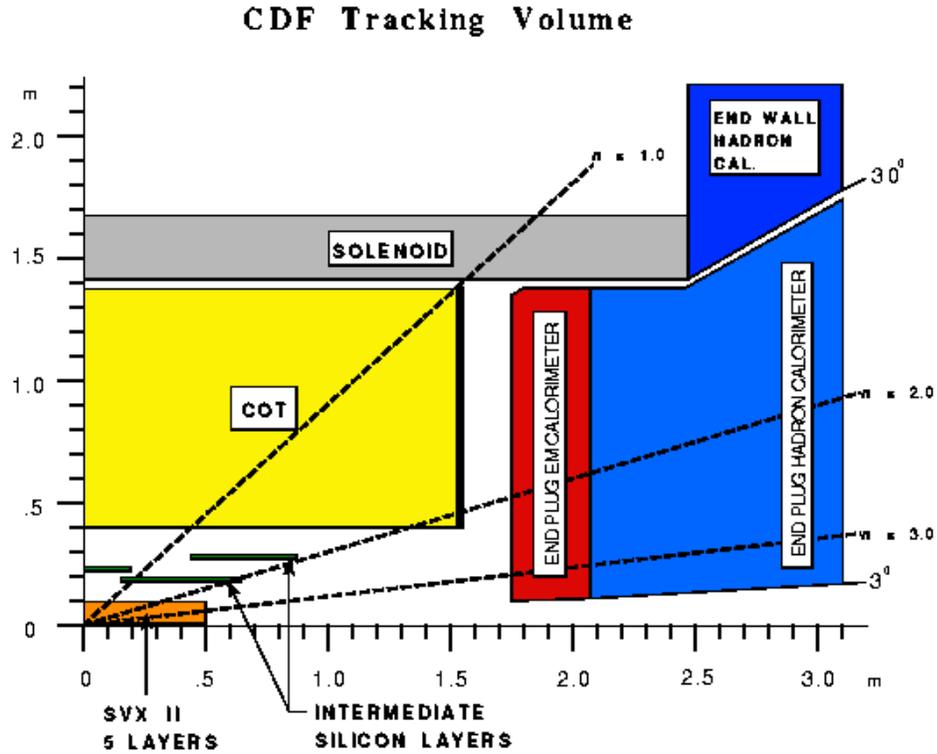


Figure 3.6: Schematic of the CDF II tracking volume.

Inner Tracker

The silicon detector comprises eight layers of microstrip sensors arranged with cylindrical symmetry between radii of 1.3 cm and 28 cm. It is composed of three distinct subsystems: Layer 00 (L00), SVX-II, and the Intermediate Silicon Layer (ISL), each of which is new in Run 2. The full system contains over six square meters of silicon and more than 700,000 readout channels, and was the largest silicon detector ever built at the time it was commissioned. The layout of the full system in the transverse plane is shown in Figure 3.7.

The innermost component of the system, L00, is a single-sided layer spanning

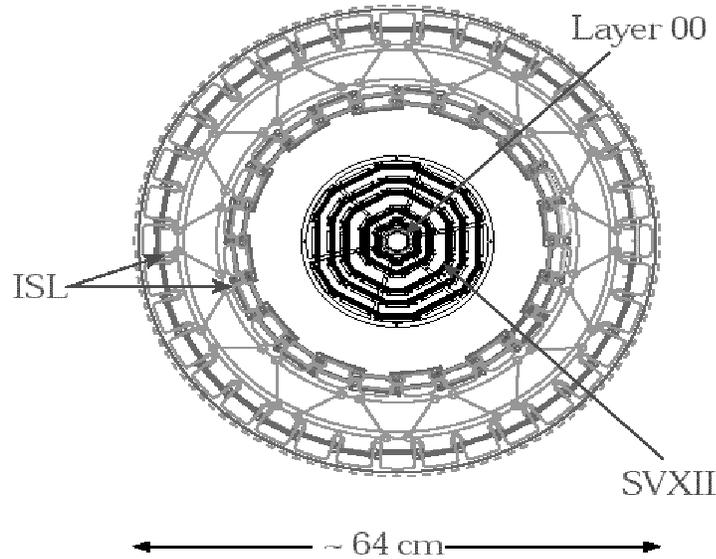


Figure 3.7: Schematic view of the three-component silicon system in the x - y plane. Each subdetector is drawn to scale.

80 cm in z and mounted directly on the beampipe, a distance of 1.3 cm from the $p\bar{p}$ interaction region [22]. Because L00 provides a position measurement so close to the collision, it substantially improves the impact parameter resolution, especially for low- p_T tracks. The sensors are spaced $25\ \mu\text{m}$ apart, but only alternate strips are read out. The detector geometry is illustrated in Figure 3.8.

The SVX-II is the most important component of the silicon system, comprising five layers of double-sided strips spanning radii from 2.45 cm to 10.6 cm [23]. SVX-II, like L00, is arranged with sixfold rotational symmetry, containing 12 castellated wedges as shown in Figure 3.9. It extends 96 cm in z , divided into three identical barrels with bulkheads at each end. The *port cards*, mounted on the outer radii of the

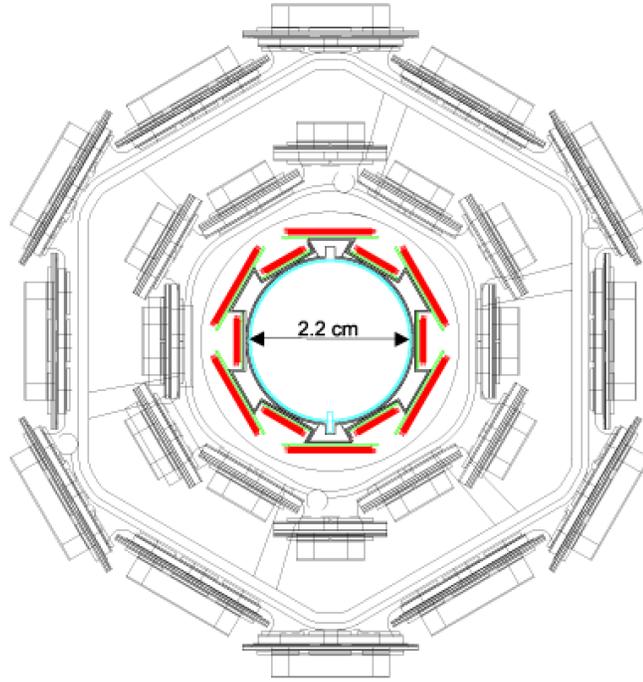


Figure 3.8: Cross-sectional view of the L00 detector in x - y plane. The detector is shown nested in the inner two layers of the SVX-II.

bulkheads, communicate with the front-end chips and convert electrical into optical signals, which are transmitted outside the tracking volume.

One side of each strip in the SVX-II is oriented parallel to the beam to determine the azimuthal angle of particles traversing the detector. These strips are 15-20 μm wide, 300 μm thick, and spaced by 60 μm in $r\phi$. On the opposite side, sensors are oriented differently, 1.2° (in opposite directions) for the third and fifth innermost layers, 90° for the others; sensors with small-angle stereo orientation are 15 μm wide and separated by 60-65 μm , other are 20 μm wide and separated by 125-141 μm , depending on the layer.

More than 85% of the silicon channels in the detector are operational, and the

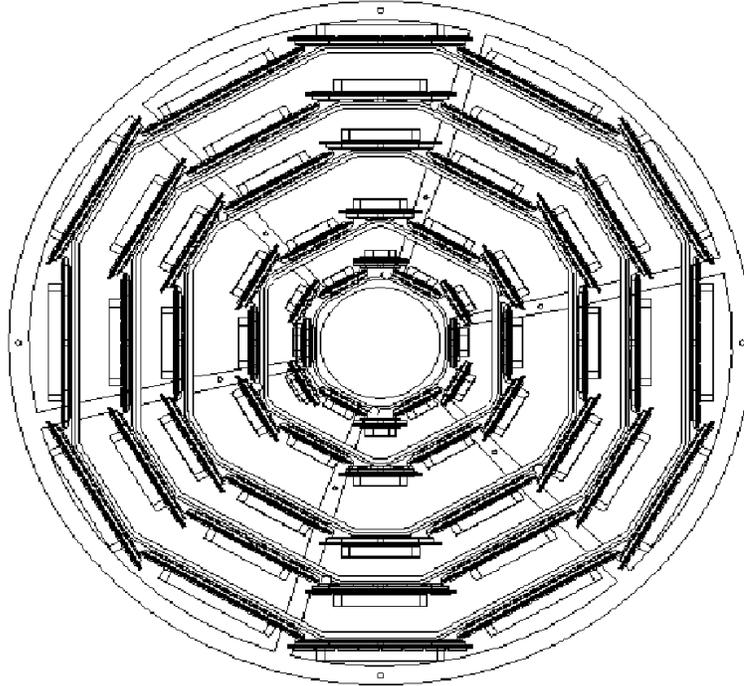


Figure 3.9: Cross-sectional view of the SVX-II detector in x - y plane. The segmentation and rotational symmetry are clearly visible.

signal-to-noise ratio is larger than 10 in functioning parts of the detector, as shown in Figure 3.10. Combining hits in neighboring strips, we can identify the position of a track on a given layer of the SVX with a precision of roughly $9 \mu\text{m}$, as illustrated in Figure 3.11.

Finally, the ISL consists of two symmetric silicon layers for $|\eta| \geq 1.1$ and one in the central region [24]. The high- η layers are at radii of 20 and 28 cm, and the central layer is at a radius of 23 cm. The detector is shown in Figure 3.12. In the central region, the ISL layer improves the linking of COT tracks (with radius > 40 cm) to the SVX (radius < 11 cm). The forward layers help provide sufficient track information to form a standalone tracker with the SVX-II where the outer tracker is less effective,

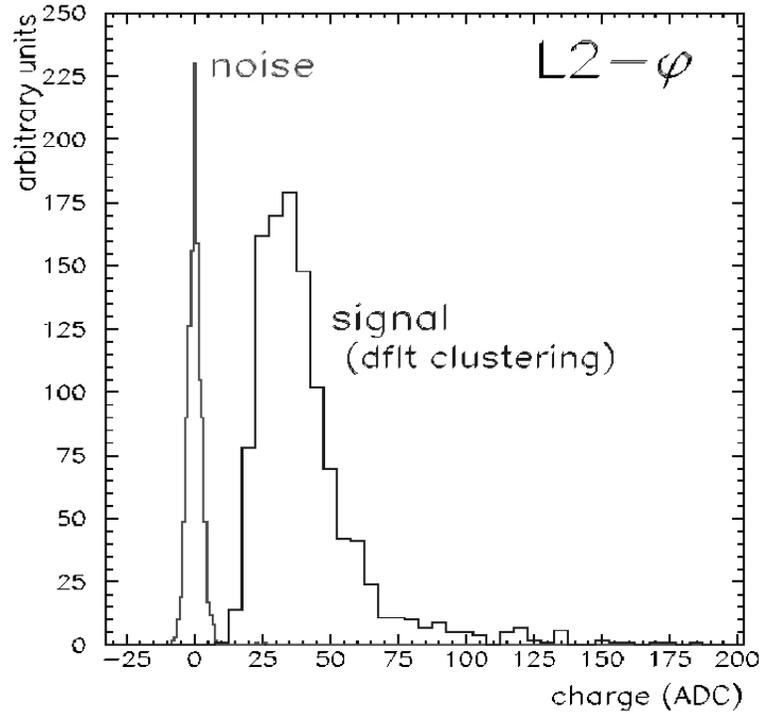


Figure 3.10: Distribution of collected charge for signal and noise in Layer 2 of the SVX-II. The ratio $\frac{S}{N}$ is larger than 10 nearly everywhere in the detector.

out to $|\eta| \leq 2.0$. Like the SVX-II, the ISL uses double-sided sensors with a small 1.2° stereo angle between the sides. The pitch is $112 \mu\text{m}$ on both sides and in all layers.

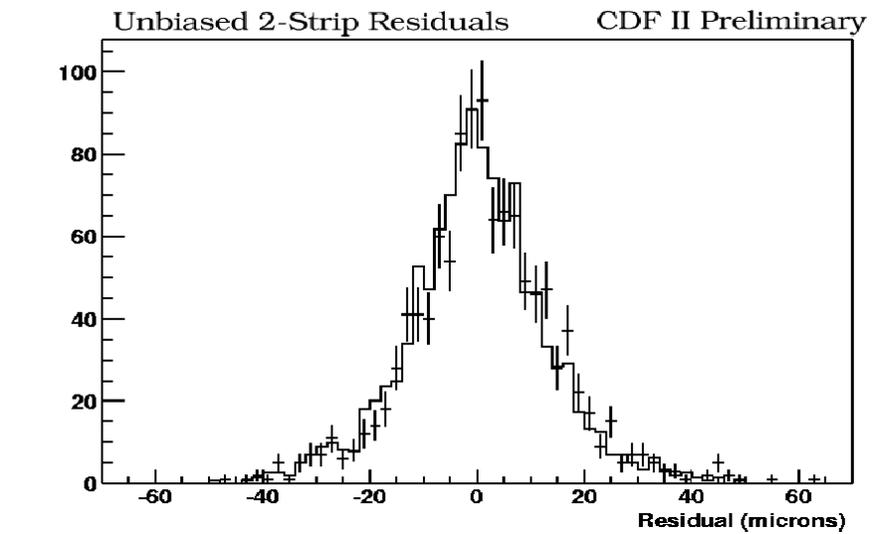


Figure 3.11: Residuals for two-strip clusters in Layer 2 of the SVX-II. The resolution of this layer, determined by the width of a Gaussian fit to this distribution, is $9\mu\text{m}$.

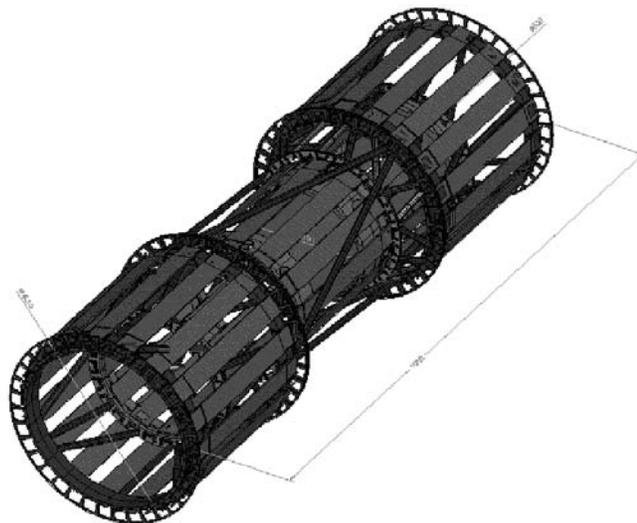


Figure 3.12: Three-dimensional sketch of the ISL detector. The full detector is nearly 2m long and 60 cm across. The barrel structure is visible in the image.

Outer Tracker

The Central Outer Tracker (COT) is positioned immediately outside the silicon system, still contained within the 1.4 T solenoidal field [25]. It is designed as a standalone tracker, since charged particles with $|\eta| < 1$ and transverse momentum above 400 MeV/c will traverse all eight superlayers. Ultimately, tracks found in the COT are linked to hits in the silicon detector for improved resolution.

The COT extends 3.1 m in z , covering radii from 44 to 132 cm. Between the two endplates, 30,240 sense wires are strung in a pattern of eight *superlayers*, alternating between axial alignment and a stereo angle of 2° . The information contributed by the stereo layers allows for a measurement of the track pseudo-rapidity using only the COT. The superlayers are rotationally symmetric, with a repeating *supercell* pattern of 12 sense wires. The number of supercells increases at large radius, so the physical spacing between wires (and therefore the typical drift time) is the same in all superlayers. The layout of the COT is illustrated in Figure 3.13.

The supercell structure is shown in Figure 3.14. The cell comprises 13 equally spaced potential wires, with a sense wire located directly in between each pair. The wire plane is rotated 35° relative to the radial direction to compensate for the Lorentz angle inside the solenoidal field. High voltage is applied to the potential and shaper wires to maintain a 1.9 kV/cm drift electric field. Additionally, the supercells are isolated from one another by gold field sheets, which are kept grounded.

To avoid event pileup, the drift time in the COT is required to be less than the beam crossing time of 396 ns. The chamber is therefore filled with a 50-50 mix of Argon and Ethane, chosen for its relatively fast drift velocity, 5×10^4 m/s. The

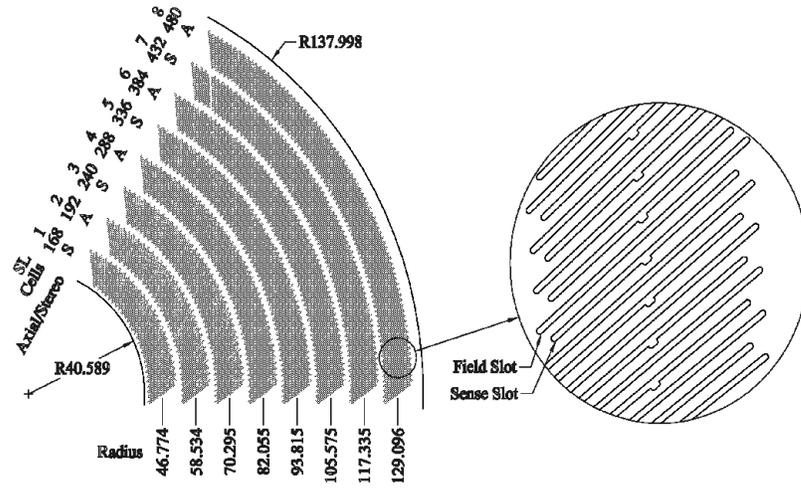


Figure 3.13: Illustration of the COT superlayer structure. The repeating cell structure, its skewness from the radial direction, and the increase in cell multiplicity at larger radius are all visible.

maximum drift time is 180 ns.

Hit information from the wires is recorded by an amplifier-shaper-discriminator (ASD) board mounted directly on the COT endcap. The board is capable of detecting multiple hits on a single wire, especially important at high instantaneous luminosity and in the innermost superlayers where the occupancy is highest.

The COT can also be used for particle identification through measurements of $\frac{dE}{dx}$, the ionization energy loss of each track in the detector. An additional subdetector at the outer radius of the COT measures each particle's time-of-flight through the tracking volume, which provides a measurement of the particle mass if the path length and momentum are known. Neither of these methods are used in this analysis.

The resolution and performance of the full tracking system will be discussed in Section 4.1.

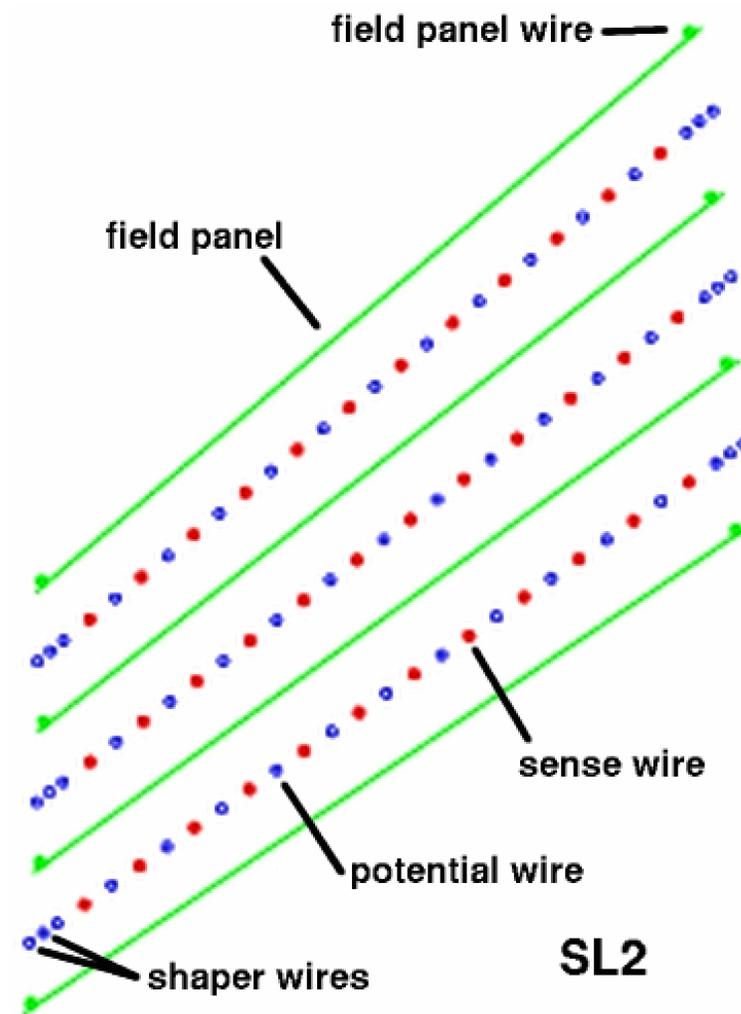


Figure 3.14: Schematic view of three COT cells. Each cell alternates between 13 potential and 12 sense wires, and cells are separated by field sheets.

3.2.2 Calorimetry

Immediately outside the solenoid, a set of electromagnetic and hadronic calorimeters record the energies of particles with sufficient energy to leave the tracking volume. (Particles with $p_T < 300 \text{ MeV}/c$ curl up inside the magnetic field.) Each of these systems is further divided into the central calorimeter ($|\eta| < 1.1$) and the plug calorimeter ($1.1 < |\eta| < 3.4$); both systems are segmented in a projective tower geometry, with the electromagnetic calorimeter closer to the interaction region.

The central calorimeter is approximately 5 m in length, is centered at $z = 0 \text{ cm}$, and has an inner radius of 173 cm [26][27]. The repeated structure, called a *tower*, covers an azimuthal angle $\Delta\varphi$ of 15° and a pseudo-rapidity range $\Delta\eta$ of 0.11; the detector is azimuthally symmetric, but the towers become increasingly narrow at high $|\eta|$. The segmentations of the electromagnetic and hadronic components are identical.

The towers are constructed from alternating layers of high- Z absorber material and active readout material. As particles travel in the absorber, they lose energy and produce *showers*, cascades of secondary particles. The active material samples the showers as they pass through several layers, and the total amount of light collected there corresponds to the initial energy of the particle. The radial segmentation is illustrated in Figure 3.4.

Lead and iron are used as the absorbers in the electromagnetic and hadronic subdetectors, respectively, with polystyrene scintillator as the active material. In the Central Electromagnetic Calorimeter (CEM), the absorber and scintillator thicknesses are 5 mm and 3 mm, respectively. In the central hadronic calorimeter (CHA), they

are 1.0 and 2.5 cm.

The total depth of the CEM is $18X_0$, where X_0 is the radiation length, defined as the characteristic length over which an electron will be reduced to $1/e$ of its initial energy due to *bremsstrahlung*. The energy resolution of the detector is:

$$\frac{\sigma_E}{E} = \frac{13.5\%}{\sqrt{E_T}} \oplus 2\%, \quad (3.3)$$

where E_T is always measured in GeV and the two terms are added in quadrature [28]. The first term is the *stochastic* term, due to sampling fluctuations and photon statistics in photo-multipliers. The second term is due to intrinsic non-uniformity in the detector response.

The CHA has a thickness of $4.5\lambda_t$, where λ_t is the interaction length, defined as the mean free path of particles between inelastic collisions. With E_T measured in GeV, the single pion resolution is given by:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E_T}} \oplus 3\%. \quad (3.4)$$

To further improve the spatial precision of the energy measurement, the Central Electromagnetic Shower Counter (CES), a proportional strip and wire chamber, is placed at the average position of the maximum shower deposition. The CES is embedded directly in the CEM, at a position $5.9X_0$ from the solenoid, and it consists of cathode strips running in the azimuthal direction and anode wires running parallel to the beamline; for 50-GeV electron tracks, its position resolution is 2 mm. These subdetectors are critical for electron identification, as will be discussed in Section 4.2.

The basic structure of the plug calorimeter is similar to the central calorimeter;

the inner component of the detector measures electromagnetic energy, and the outer measures hadronic energy [29]. The towers again comprise alternating absorber-scintillator layers that each cover a fixed width in η , again with lead and iron as the absorber materials, as shown in Figure 3.15. The detector is segmented azimuthally into 24 15° towers for the four lowest- η rings, up to $|\eta| < 2.1$, and 12 30° towers for the eight most forward rings, covering up to $|\eta| < 3.6$.

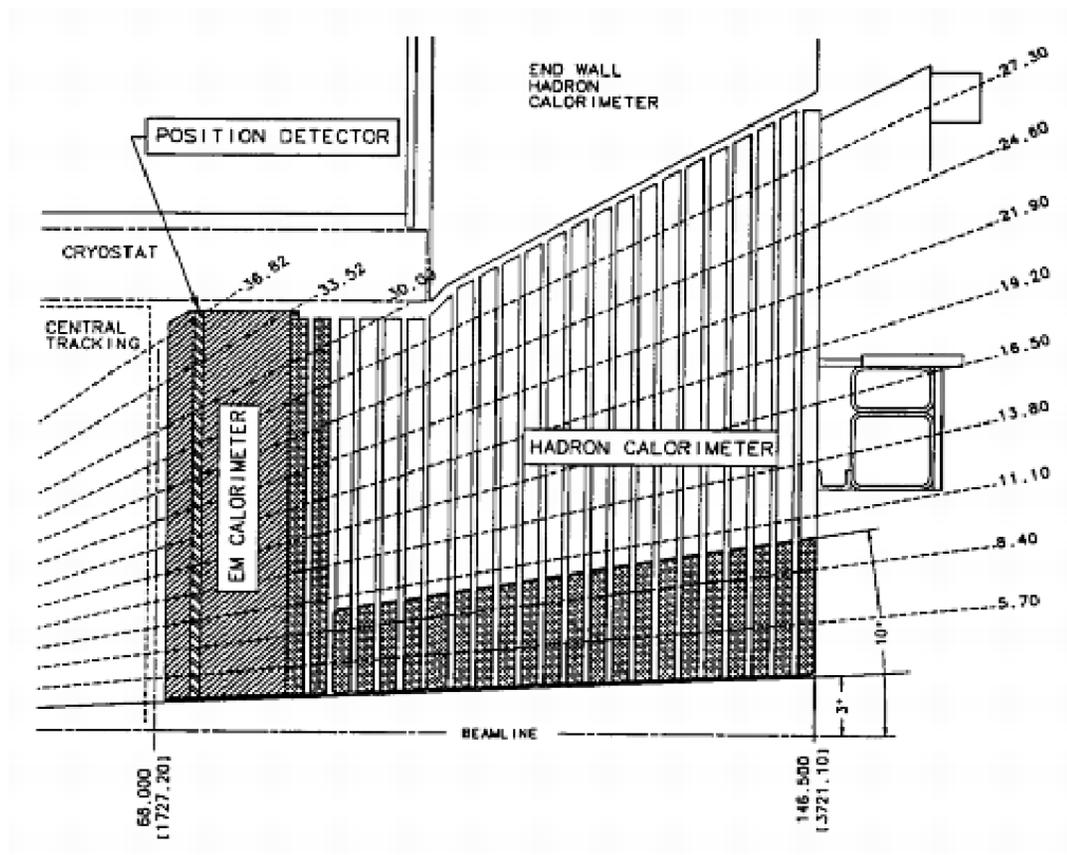


Figure 3.15: Illustration of the PEM and PHA detectors. The segmentation of the calorimeter and the location of the PES are shown as well.

The Plug Electromagnetic Calorimeter (PEM) has a total thickness of 21 radiation lengths (X_0) and an energy resolution of:

$$\frac{\sigma_E}{E} = \frac{14.4\%}{\sqrt{E_T}} \oplus 0.7\%. \quad (3.5)$$

The Plug Hadronic Calorimeter (PHA) has a total thickness of 7 interaction lengths (λ_I) and a single pion energy resolution of:

$$\frac{\sigma_E}{E} = \frac{80\%}{\sqrt{E_T}} \oplus 5\%. \quad (3.6)$$

As in the central region, the PEM is equipped with a position-detector at shower maximum (PES) and a preshower scintillator detector (PPR), making electron identification possible even at very high pseudo-rapidity. These detectors are not used in this analysis.

3.2.3 Muon Detectors

The outermost component of the CDF II detector identifies muons, which traverse the tracking chambers and calorimetry with minimum ionizing traces. The muon system comprises three main subdetectors, the Central Muon Chambers (CMU), the Central Muon Upgrade (CMP), and the Central Muon Extension (CMX) [30]. Additional muon detectors, the Intermediate Muon chambers (IMU), extend the coverage of the muon system beyond the central tracker, but are not used in this analysis. The position of each of the subdetectors in η - φ space is shown in Figure 3.16.

The CMU structure is mounted on the outside of the CHA, each wedge contained inside the tower in azimuth but extending across multiple η towers, as shown in Figure 3.17. Wedges are segmented sixfold in the z direction, covering the region $|\eta| < 0.6$, and subtend an angle of 12.6° in φ , leaving gaps of 2.4° between adjacent wedges. The wedges are further subdivided into sixteen cells in the $r - \varphi$ plane, four radial layers of four cells each. Each cell contains a single sense wire stretched in the z direction, the third and fourth layers offset from the inner two by 2 mm to eliminate an ambiguity in the φ -direction of muon tracks. A CMU cell is illustrated in Figure 3.18. Tracks with hits on all four radial layers can be measured with a typical $r\varphi$ resolution of $250 \mu\text{m}$. By comparing the pulse heights on opposite ends of the sense wires, the z position of the track can be measured with a resolution of 1.2 mm. A minimum of three hits out of four is necessary to constrain a muon track; a group of three or more hits in the muon chambers is called a *muon stub*.

The CMP is a second set of drift chambers located beyond an additional 60 cm of steel, which both covers the gaps between CMU wedges and helps reject hadrons

(pions and kaons) that reach the CMU but not the CMP. Unlike nearly all other components of CDF II, the CMP is not azimuthally symmetric, but rather consists of flat stacks on the top, bottom, north, and south sides of the detector. Unlike the CMU, the CMP does not record a z position, but the $r\varphi$ resolution is comparable in the two components. Commonly, as in this analysis, stubs from the CMU and CMP are used together (called CMUP muons) to achieve the highest resolution and purity for muons in the central region. The path of a CMUP muon is illustrated in Figure 3.19.

A third muon system, the CMX, consists of a pair of truncated cones located at either end of the CMU and CMP. The chambers are oriented with large circular bases at $|\eta| = 0.6$, tapering to a smaller radius at $|\eta| = 1.0$. The basic cell geometry in the CMX is similar to the other detectors as well, with coincidence required in three of four layers to form a stub. The top 30° of the east CMX arch is not instrumented, instead used for cooling lines for the solenoid; the equivalent piece on the west end, the *keystone*, was added in Run 2. The lower 90° on both sides, the *miniskirts*, were also commissioned in Run 2. The miniskirt and keystone did not become operational until after data-taking in Run 2 began. The first 330 pb^{-1} of data in the analysis (out of 1.12 fb^{-1}) were recorded with these components turned off.

All muon subdetector are additionally outfitted with scintillator tiles, which provide timing information and help reject muon signatures from secondary interactions, cosmic rays, and other non-collision backgrounds.

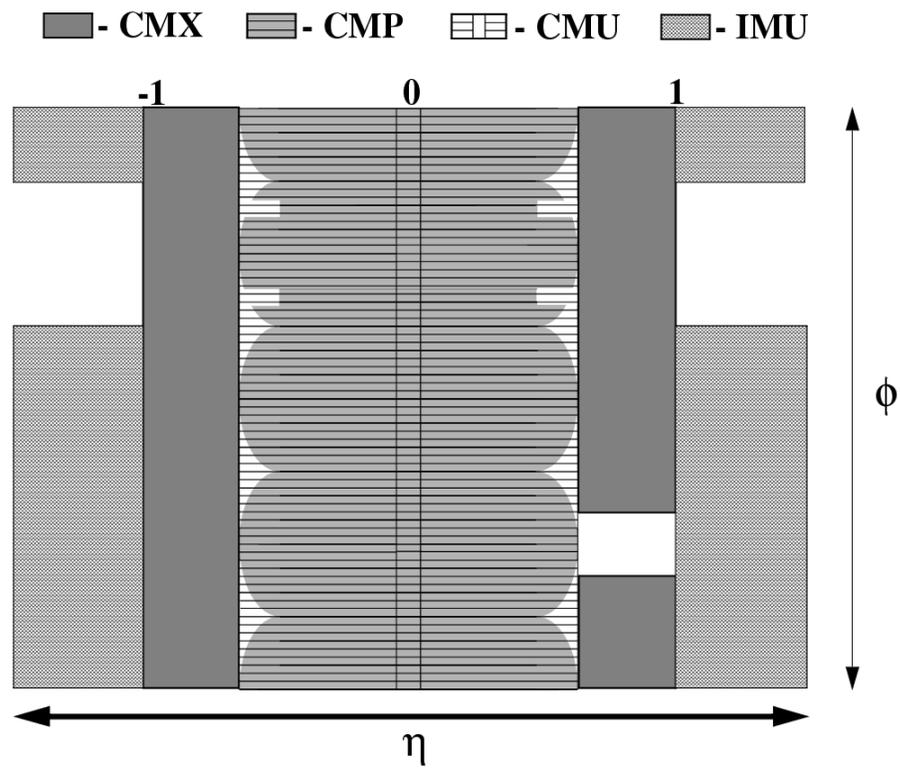


Figure 3.16: Coverage in η and ϕ of the full muon detector. The IMU is not used in this analysis.

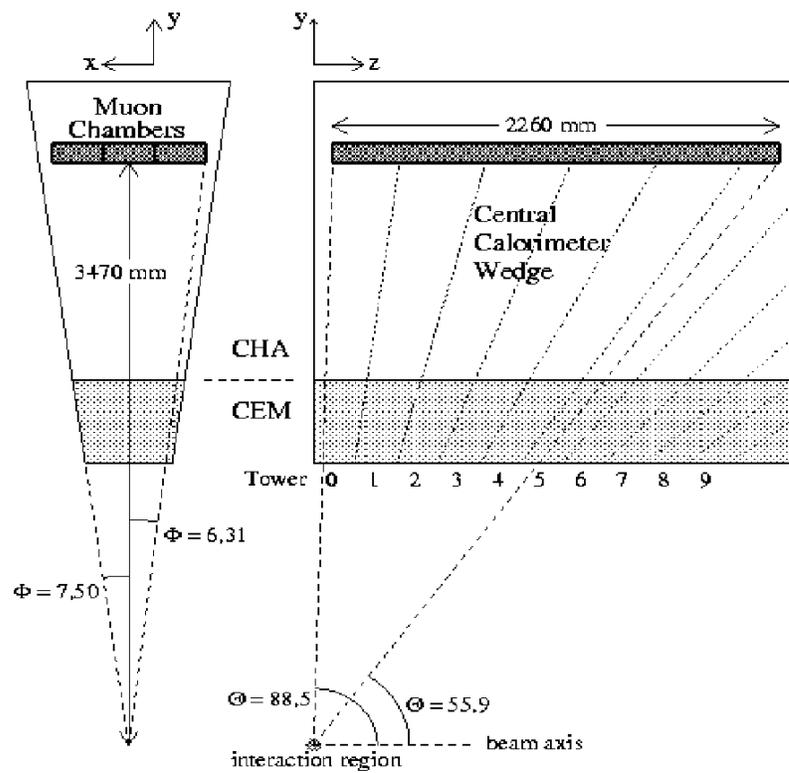


Figure 3.17: Cross-sectional view of the muon chambers mounted directly on **CHA** wedges.

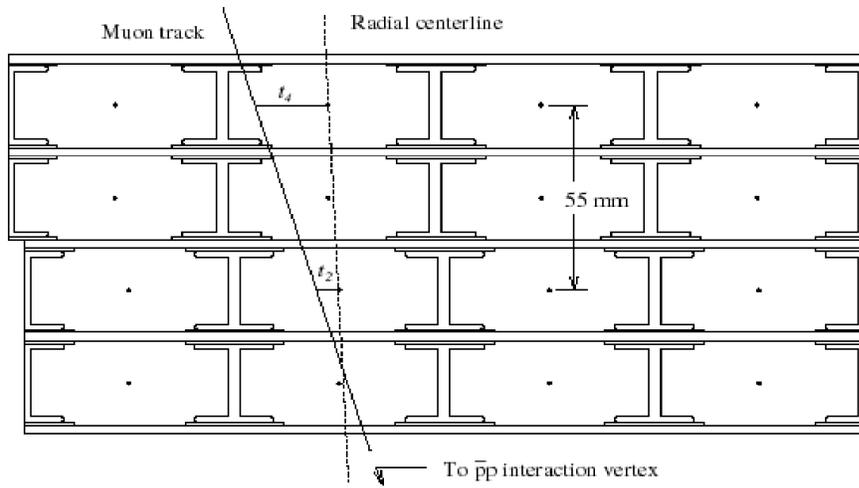


Figure 3.18: Schematic of a track passing through the CMU. The offset between the second and third layers allows for unambiguous stub reconstruction given only a set of drift times. A radial line pointing to the center of the detector is also shown.

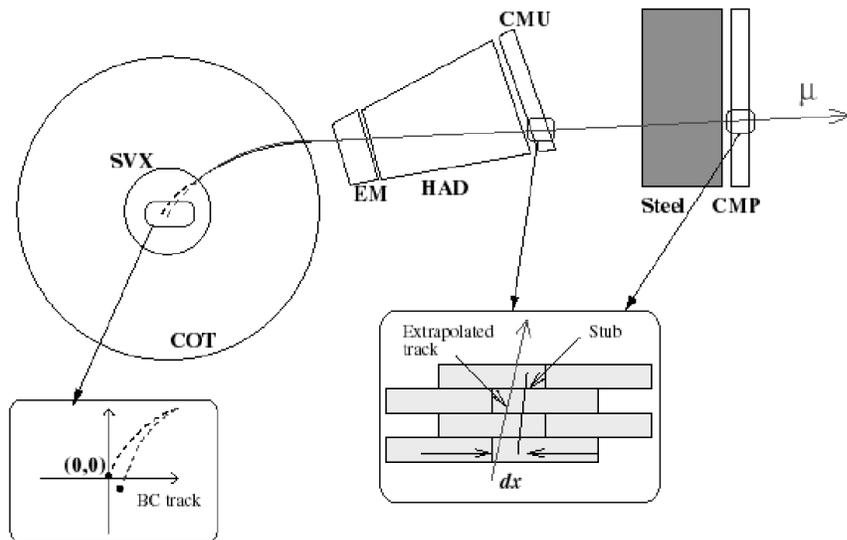


Figure 3.19: Transverse path of a CMUP muon. The track becomes straight outside the solenoid and passes through the calorimeter and both the CMU and CMP detectors.

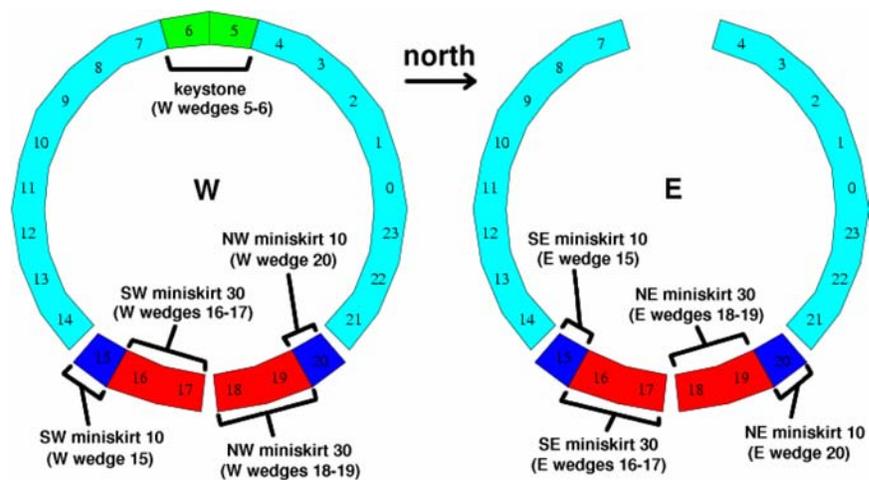


Figure 3.20: Segmentation of the CMX detector. The miniskirt and keystone were commissioned after the beginning of Run 2, and so are not used in the earliest data.

3.2.4 Luminosity Counters

The Cherenkov Luminosity Counter (CLC) is an essential component of the CDF II system, though it operates independently of the other subdetectors. The CLC records the instantaneous luminosity during collisions by measuring the rate of inelastic proton-antiproton interactions [31]. It records very forward outgoing particles, between $|\eta|$ of 3.7 and 4.7 (a polar angle $\theta < 3^\circ$). In the plug at each end of the detector, 48 conical counters, each between 1.1 and 1.8 m long, are arranged three-deep in η around the beampipe, and each is connected to its own photomultiplier tube.

When particles travel through the counters, Cherenkov light will be generated almost collinear with them. The number of generated photoelectrons is proportional to the path length through the cell, minimizing contributions from other sources (beam halo, secondaries from the detector volume) not parallel to the counter. Each particle from a $p\bar{p}$ collision typically contributes a signal of 100 electrons, with the total signal from multiple particles simply the sum of the signals from each individual particle. We measure the total number of inelastic $p\bar{p}$ collisions with a precision of 4.4%; combined with a 4% relative uncertainty on the inelastic $p\bar{p}$ cross section, we measure the luminosity with a total precision of 5.9% [32].

3.2.5 Trigger and Data Acquisition

The final critical component of the CDF II online system is the trigger and data acquisition (DAQ) system. During collisions, a beam crossing occurs every 396 ns, or at a rate of 2.5 MHz. The full detector cannot be read out in that time, nor can every event be recorded to tape, given bandwidth and storage limitations. In-

stead, we use a three-tier trigger system to determine whether an event contains a potentially-interesting physics signature, with each trigger level using progressively more advanced reconstruction techniques on progressively fewer events. A schematic and a logic diagram of the trigger system are shown in Figures 3.21 and 3.22. Each CDF *dataset* consists of events which pass a specific trigger path, or set of selection criteria at each of the three levels.

Level 1, the most primitive stage of triggering, makes a decision to pass or ignore an event by simply counting the number of physics objects present. The calorimeter, muon chamber, and tracking system each has its own data pipeline, and a decision is made within $5.5 \mu\text{s}$, the total time an event may be stored. The trigger is based solely on hardware, with custom electronics capable of performing fast, simple scans for energy deposits in the calorimeter, stubs found in the muon chambers, and some tracks found with the eXtremely Fast Tracker (XFT) in the COT [33]. An additional module, the XTRP, can extrapolate tracks from the COT to the calorimeter or muon chambers, critical for lepton-based triggering [34]. Energy and momentum thresholds are adjustable online, and each trigger path can be prescaled to control rates and deadtime. At typical instantaneous luminosities, the Level 1 accept rate is between 12 and 15 kHz.

Level 2, a second hardware-based trigger system, performs more advanced event reconstruction, improving the signal purity in each trigger path. The system comprises four asynchronous buffers which store event information passed from Level 1. When the buffers are full, any further events accepted in Level 1 are lost, resulting in trigger *deadtime*; prescales and thresholds are tuned to keep deadtime below a few

percent. The two main strengths of Level 2 are its cluster finder (L2CAL), which allows triggering on reconstructed calorimeter jets rather than single towers, and its displaced vertex tagger (SVT), which identifies well-measured tracks in the silicon detector not originating from the interaction point [35]. The latter is critical for B physics analyses, but it is not used in this analysis. Altogether, the total Level 2 acceptance rate is 300 Hz.

The final stage of triggering, Level 3, is based in software and is run on a farm of more than 200 dual-processor computers. Events are reconstructed using the full detector, more advanced tracking and jet clustering algorithms, and up-to-date detector calibration information. If an event is accepted in Level 3, it is passed to the Consumer-Server Logger (CSL), which sends the event to be recorded to tape. The trigger selection at Level 3 is limited only by the data transfer rate, around 20 MB/s, which corresponds to an acceptance rate of 75 Hz.

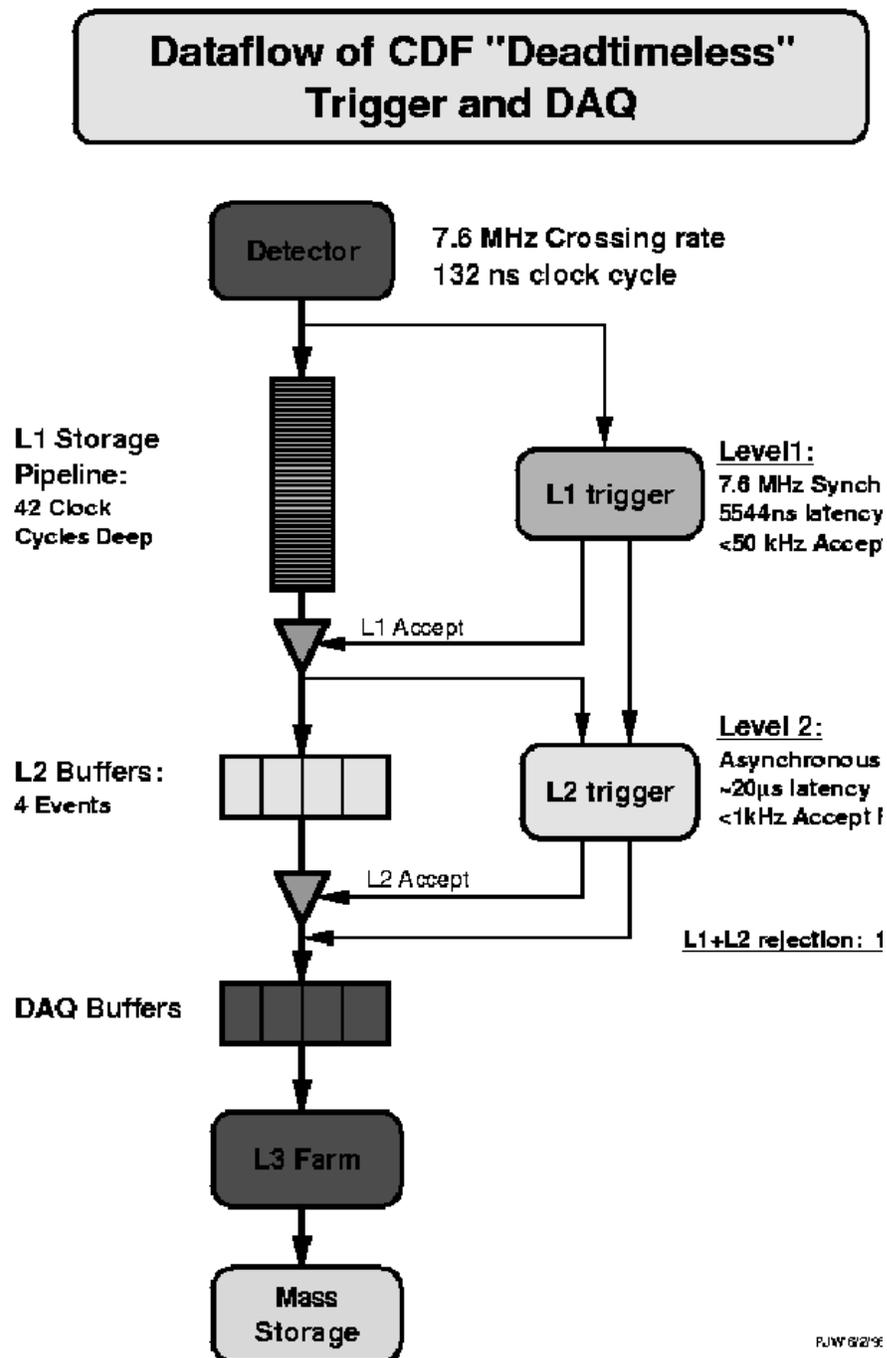
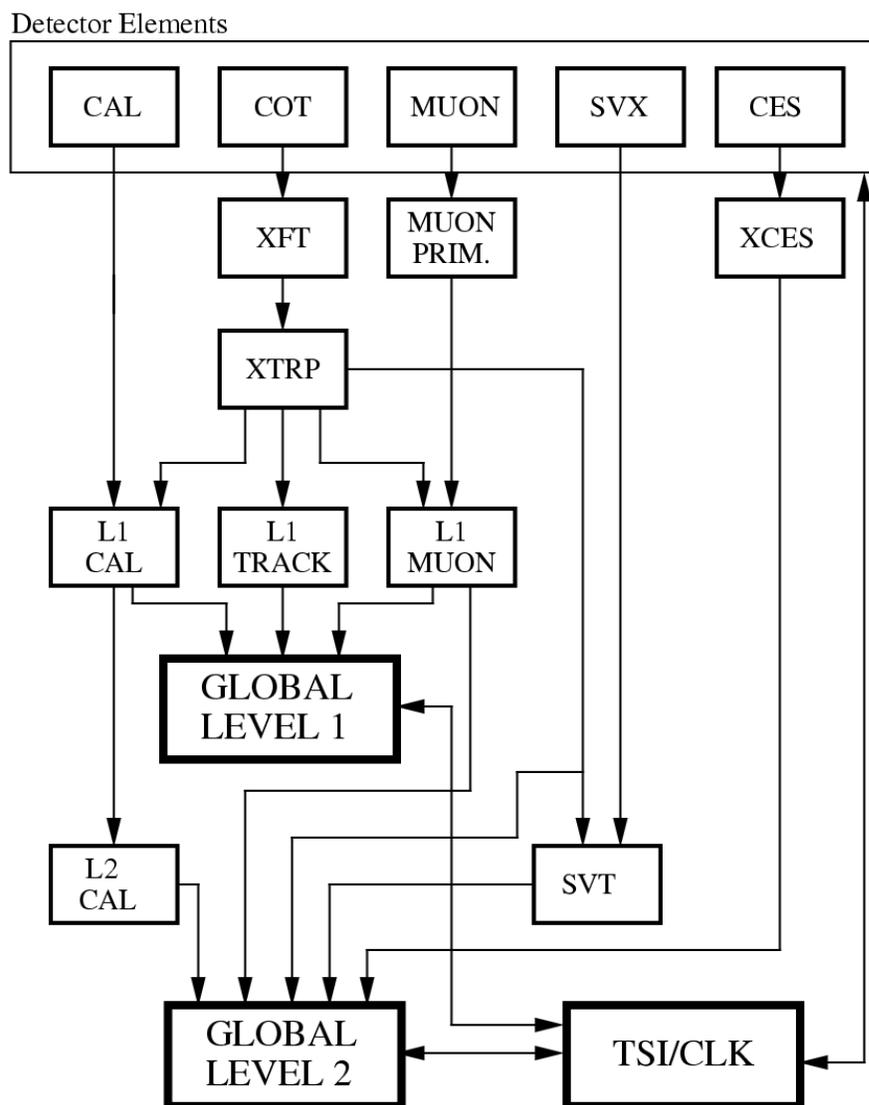


Figure 3.21: Schematic view of the CDF trigger system.

RUN II TRIGGER SYSTEM



PJW 9/23/96

Figure 3.22: Block diagram of the CDF trigger system, illustrating when information from each of the main subdetectors is used in the trigger decision.

Chapter 4

Event Reconstruction

Both at Level 3 of the trigger and again offline, we use event reconstruction techniques to determine the physics content of each event; COT tracking, lepton identification, and jet clustering are all performed online, but with algorithms that are optimized both for speed and maximum acceptance. Recorded events are reconstructed offline with slower, more sophisticated algorithms, which combine the information from all CDF subdetectors to distinguish different types of particles (leptons, hadrons, neutrinos, etc., as shown in Figure 4.1). In this chapter, we describe these reconstruction techniques, in particular the tracking, lepton identification, and jet clustering algorithms.

4.1 Tracking

The most basic reconstructed objects in collision are tracks, the paths of charged particles through the silicon detectors and COT described in Section 3.2.1. Since this

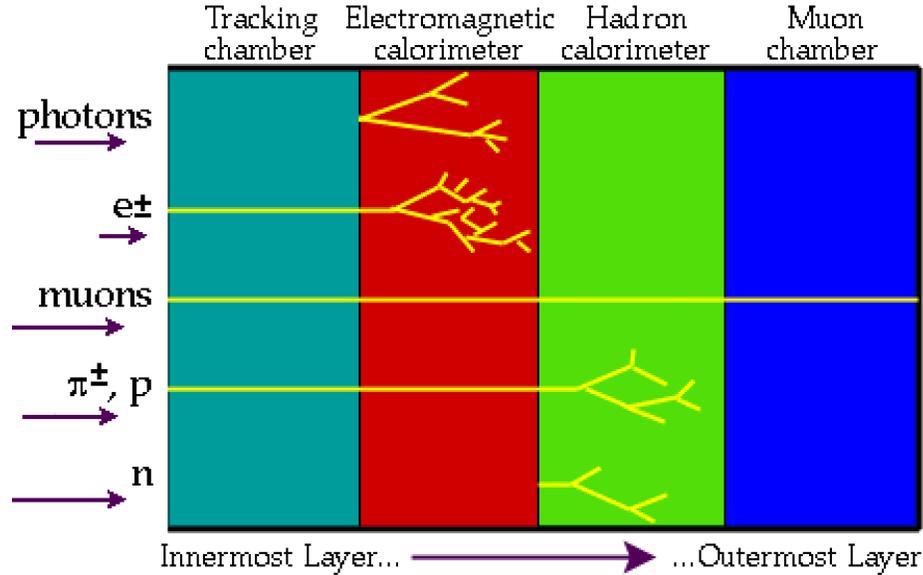


Figure 4.1: Illustration of the different detector signatures left by particles commonly produced in collisions.

volume is contained in a uniform axial magnetic field, a charged particle will travel along a *helical* path, which we parameterize in terms of five quantities:

- d_0 : the impact parameter, defined as the distance of closest approach to a fixed reference point in the transverse plane. The reference point is chosen to be the measured location of the $p\bar{p}$ collision, as described in Section 4.1.1;
- φ_0 : the azimuthal direction of the track at the point of closest approach, *i.e.* at the same location where d_0 is determined;
- C : the half curvature, $\frac{1}{2R}$, where R is the radius of curvature in the transverse plane. C is signed according to the charge of the particle, and is proportional to its transverse momentum p_T ;
- Λ : the cotangent of the polar angle θ ;

- z_0 : the z -coordinate at the point of closest approach.

The offline CDF tracking algorithm maps the set of position measurements from the detector (*hits*) to a set of tracks with these five parameters. We reconstruct tracks in two phases: first, we reconstruct tracks using only the information in the COT; second, we extrapolate these tracks through the silicon detectors and associate nearby hits to the COT tracks [36].

The event record contains timing information for each hit on each wire in the drift chamber. The hit time is corrected for wire-by-wire gain differences, the collision time relative to the CDF clock, and the propagation time of the particles through the detector; what remains is the drift time, which corresponds to a distance of closest approach between the track and the wire.

Starting from the innermost wires in each COT superlayer, track seeds are constructed using triplets of hits in adjacent wire planes. Since the direction of the drift relative to the wire is ambiguous, each triplet actually defines a pair of tracks, but only the one more consistent with originating at the interaction point is considered. Seeds are extrapolated through the superlayer with straight line fits, and additional hits in each superlayer are added when they fall within 1 mm (20 ns in drift time) of the seed track. A *segment* is defined as a set of five or more hits in a superlayer which are part of the same track.

Next, we produce tracks by linking together segments from the four axial superlayers. Each segment in the outermost superlayer (SL8) is extrapolated inward toward the origin, assuming a circular path and no impact parameter. Track segments found in SL6 are checked for consistency with these extrapolated tracks, pairing segments

only if the angular difference is less than 50 mrad. When two segments are linked, we calculate the two-dimensional track parameters (C , d_0 , and φ_0) and use those to extrapolate in to the next axial superlayer. Each time a new segment is attached, the track is recalculated.

We use a second tracking algorithm in parallel to improve the track-finding efficiency [37]. Extrapolating from the origin through each segment with 8 or more hits, we calculate the residual for every hit in the detector within 1 cm of the track candidate. These residuals are histogrammed in 200 μm bins, and when the bin at the peak of the histogram includes 10 or more hits, a track is made. Any additional hits within 750 μm of this initial track are also included. Tracks duplicated using the two methods are removed at this stage.

The stereo superlayers are included using the segment-linking technique, again working from the outermost to innermost superlayers. Stereo segments are included in the fit when adjacent to a reconstructed axial segment and with an angular difference of less than 10 mrad from the axial track. Each time a stereo segment is added to the track, the three-dimensional track parameters are refit. The z information in COT tracks provides a rough measurement of the longitudinal position of the collision, which is approximately Gaussian with a width of 25 cm.

The track-finding efficiency in the central region of the COT is nearly 99% in low-occupancy environments, but can drop by 2-3% in very high-occupancy events, especially those with multiple $p\bar{p}$ collisions. Tracks found in the COT typically have a d_0 resolution of 250 μm and a z_0 resolution of 0.3 cm [38]. The transverse momentum resolution for COT-only tracks, which depends on the curvature of the track, is given

by the expression:

$$\frac{\sigma_{p_T}}{p_T} = 0.0015 p_T \text{ (GeV}/c \text{)}. \quad (4.1)$$

We next attempt to attach silicon hits to the tracks found in the COT. Though the long lever arm of the COT is most important for measuring the curvature of the tracks, the d_0 and z_0 resolutions are much improved with the addition of silicon information very near the interaction point. The silicon detector has a position resolution of $9 \mu\text{m}$ in $r\varphi$ for two-strip clusters in the SVX-II, compared with $140 \mu\text{m}$ in the COT [38].

We include hits in the silicon layers within 4σ of the COT track in the fit, and iteratively refit the track for each additional hit added. The search starts in the outermost silicon layers and proceeds inward, and the σ used is recalculated at each step, effectively narrowing the search window, or *road*, over subsequent passes. When multiple hits are found within a road on a single silicon layer, all possibilities are considered, but only the track with the best fit χ^2 is ultimately retained.

The efficiency to attach silicon hits to COT-based tracks with $p_T > 2 \text{ GeV}/c$ is 94% [39], and improves the impact parameter resolution from $250 \mu\text{m}$ to $20 \mu\text{m}$, similar to the intrinsic width of the luminous region [40]. The addition of a hit from L00 is especially critical for low- p_T tracks, as demonstrated in Figure 4.2. The p_T resolution for tracks with COT, SVX, and ISL hits is:

$$\frac{\sigma_{p_T}}{p_T} = 0.0007 p_T \text{ (GeV}/c \text{)}. \quad (4.2)$$

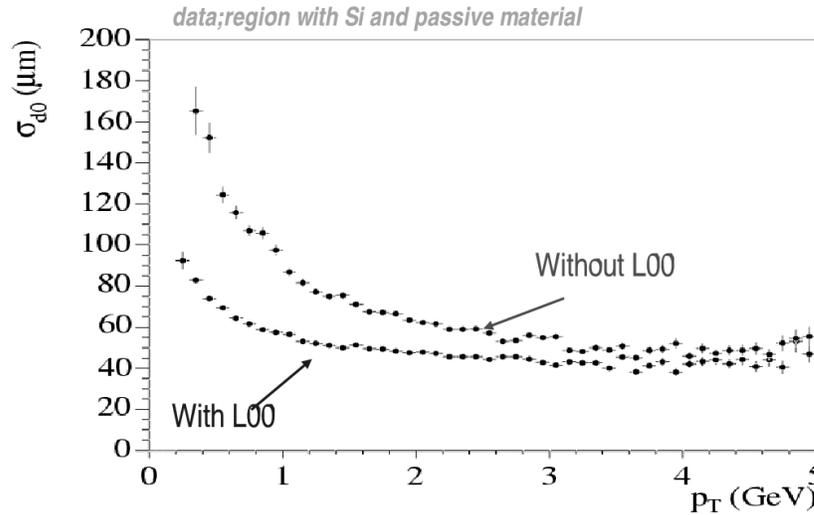


Figure 4.2: Impact parameter resolution (σ_{d_0}) as a function of track momentum for tracks with COT and SVX information, with and without hit information from L00. The improvement is most striking at low- p_T , where multiple scattering effects become large. The resolution includes a quadrature contribution of $25 \mu\text{m}$ from the intrinsic width of the beam.

4.1.1 Primary Vertex Identification

We identify the position of $p\bar{p}$ collisions on an event-by-event basis by first reconstructing the average beamline over a large number of events. With a large sample of tracks, the beamspot position (x_0, y_0) can be derived online (as a function of longitudinal position) through the correlation between impact parameter and azimuthal angle:

$$d_0 = y_0 \cos \varphi_0 + x_0 \sin \varphi_0, \quad (4.3)$$

where d_0 and φ_0 are initially calculated relative to the origin of the detector. A two-dimensional distribution of d_0 and φ_0 will have a sinusoidal shape, where the size of

the modulation and the phase determine x_0 and y_0 . Independent calculations of the beam position using COT-only and silicon-only tracks are used to monitor the relative alignment of the two subdetectors [41]. Offline, the average beamline is recalculated using an iterative method, in which only well-measured tracks contribute to the fit.

The three-dimensional coordinate of the interaction point is calculated for each event using `PrimeVtx`, a seeded vertexing algorithm [42]. `PrimeVtx` iteratively defines the most likely common origin for a set of tracks, where the tracks are required to be consistent with a predetermined “seed” position. The transverse position of the seed is taken from the run-averaged beamline. We determine its longitudinal position from the maximum in the z_0 distribution for all tracks in the event. In general, when there are multiple seeds, we use the one corresponding to the set of tracks with the largest total p_T .¹

Track parameters are recalculated relative to the seed position, and those tracks which are consistent with the beamline and seed ($|z_0| < 5$ cm, $|d_0| < 1$ cm, $|\frac{d_0}{\sigma_{d_0}}| < 3$) are fit together in a χ^2 minimization; we use the CTVMFT constrained vertexing algorithm to identify the most likely three-dimensional position of a common vertex [43]. The `PrimeVtx` algorithm iteratively removes the track with the largest contribution to the vertex χ^2 and refits, until the worst track $\chi^2 < 10$. The vertex is fit with a Gaussian constraint to the beamline, so the fit uncertainty is at worst the width of the beam, about 28 μm . With `PrimeVtx`, we reconstruct primary vertices with a resolution of $\tilde{15}$ μm , depending on the occupancy of the event.

¹The analysis presented here requires the presence of a lepton, so we only consider seeds such that the z_0 of the lepton track relative to this vertex is smaller than 5 cm.

4.2 Lepton Identification

To identify electron and muon candidates, we first impose strict requirements on the quality of the candidate track, requiring ≥ 3 axial and ≥ 3 stereo segments from the COT and $|z_0| < 60$ cm. The next two sections describe the lepton-specific requirements for high- p_T electrons and muons; when an alternate selection is used, as in Section 5.2, the requirements will be listed there.

4.2.1 Electrons

In this analysis, we use electrons found in the CEM only, with $|\eta| < 1.1$.² The signature of a central electron consists of a track pointing to an energy deposit in the electromagnetic calorimeter, coupled with little or no activity in the hadronic calorimeter. The position of electrons in the detector is shown in Figure 4.3. The selection criteria are [44]:

- Transverse electromagnetic energy ($E \sin \theta$) $E_T > 20$ GeV: A two-tower electromagnetic calorimeter cluster is constructed around the largest energy deposit, and the result is multiplied by the sine of the polar angle of the COT track associated with it. This energy is corrected for non-linearity and tower-to-tower variations.
- Transverse track momentum $p_T > 10$ GeV/ c : The cluster must be associated with a COT track.

²The analysis is limited by systematic uncertainties rather than statistics, and background to electrons in the PEM is large.

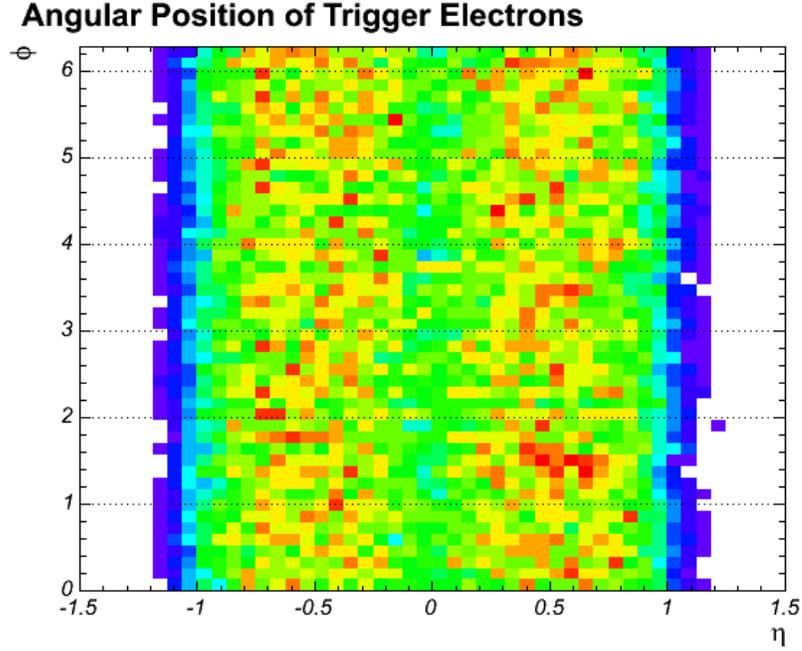


Figure 4.3: Scatter plot of η vs. φ for high- p_T electron candidates in the CEM.

- Ratio of electron energy and track momentum $\frac{E}{p} < 2$: The deposited electromagnetic energy and the momentum of the electron track are similar, but not necessarily identical; the ratio of the two will peak at or near unity, but will be broadened by measurement resolution and have a high tail due to *bremsstrahlung*. Radiated photons will contribute to the energy accumulated in the calorimeter, but the track momentum will be reduced. We impose this selection to reject excess electromagnetic activity (such as that from π^0 's) being mistakenly paired with an unrelated track, which would otherwise be indistinguishable from an electron signature.
- Ratio of calorimeter deposits in the hadronic and electromagnetic calorimeters

$\frac{E^{had}}{E^{EM}} < 0.055 + 0.00045(E^{had} + E^{EM})$ (GeV): The hadronic activity near an electron should be minimal, but the second term compensates for some leakage when the electron is especially energetic.

- Track extrapolation matches identified CES cluster centroids, $-3 \text{ cm} < Q\Delta(r\varphi) < 1.5 \text{ cm}$ and $|\Delta z| < 3 \text{ cm}$: The electron track candidate should point to the location of the cluster found at shower maximum. The cut in $r\varphi$ is asymmetric to account for photon *bremsstrahlung*, which is correlated with the direction of curvature and is therefore signed.
- Consistency of the CES shower profile with results from a test beam, $\chi^2 < 10$: The strip component of the CES has an expectation for the shower shape from electron test beams, and the χ^2 cut enforces consistency with the electron hypothesis.
- The lateral tower shower profile matches with test beam expectations, $L_{shr} < 0.2$: L_{shr} is a sum over towers of the difference between observed and expected deposits divided by the root-mean-square uncertainty on the energy measurement and on the test beam expectation. This requirement ensures that the sharing of energy between towers is consistent with electron test beam results. Excess energy in nearby towers drives L_{shr} high.
- The electron is required to be isolated from other activity in the calorimeter, $I < 0.1$: We define the *isolation energy* of an electron to be the total hadronic and electromagnetic energy within a ΔR cone of 0.4 from the electron, less the electron cluster energy. The *isolation* I of an electron is the ratio of the

transverse isolation energy to the cluster energy. Electron candidates which are separated from any hadronic activity in the event will tend to have small or zero I .

- The electron has no partner track consistent with being a photon conversion pair: We reject electrons which come from photon conversions. Since they include real electrons, conversions constitute a large background to leptonic W and Z decays that is irreducible through tightened electron selection. Instead, we reject electrons found to have oppositely-charge partner tracks consistent with forming a conversion pair. We use two variables to discriminate conversions from random track pairings: sep , the perpendicular distance between the tracks when they are parallel (they must curve in opposite directions); and $\Delta \cot \theta$, the difference between the cotangents of the polar angles of the two tracks. When a second track exists such that $sep < 0.02$ cm and $\Delta \cot \theta < 0.04$, the electron is considered a conversion. The efficiency of the conversion-finding algorithm is 90% for electrons with $E_T > 20$ GeV [45].

The efficiency of these cuts is measured in a sample of Z boson decays to electron pairs; averaged over all runs, the identification efficiency to identify electrons (excluding the trigger efficiency) is 80% [44]. This value has some run dependence, which will be described in Section 6.3.1. Distributions of high- p_T electron properties are shown in Figures 4.4 to 4.8.

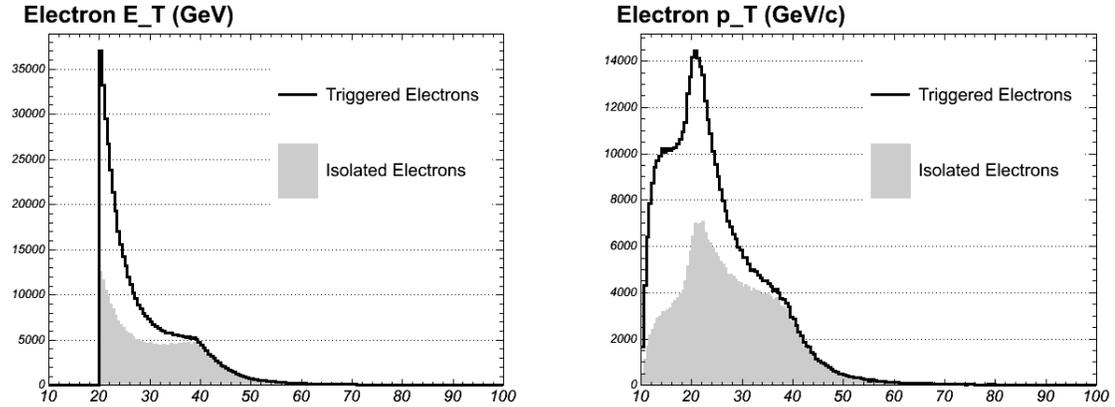


Figure 4.4: Distributions of E_T (left) and track p_T (right) for high- p_T electron candidates in data from the 18 GeV electron trigger. Distributions are shown before and after the isolation requirement.

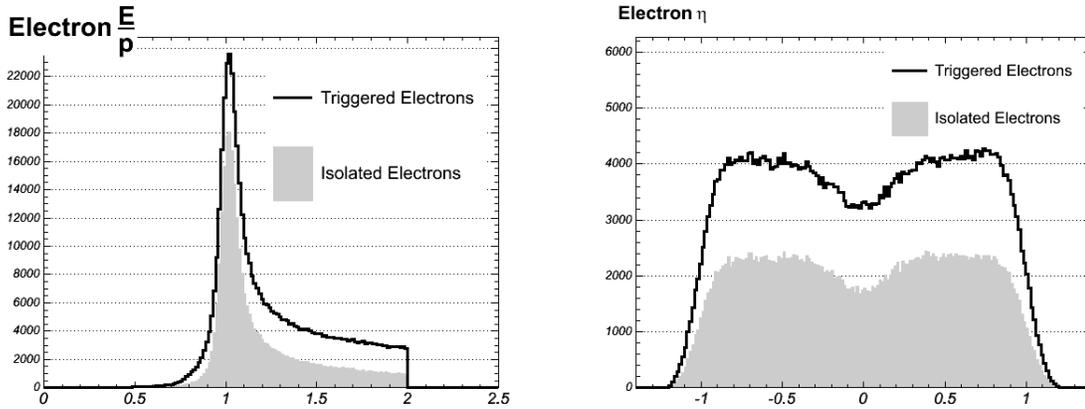


Figure 4.5: Distributions of $\frac{E}{p}$ (left) and η (right) for high- p_T electron candidates in data from the 18 GeV electron trigger. Distributions are shown before and after the isolation requirement.

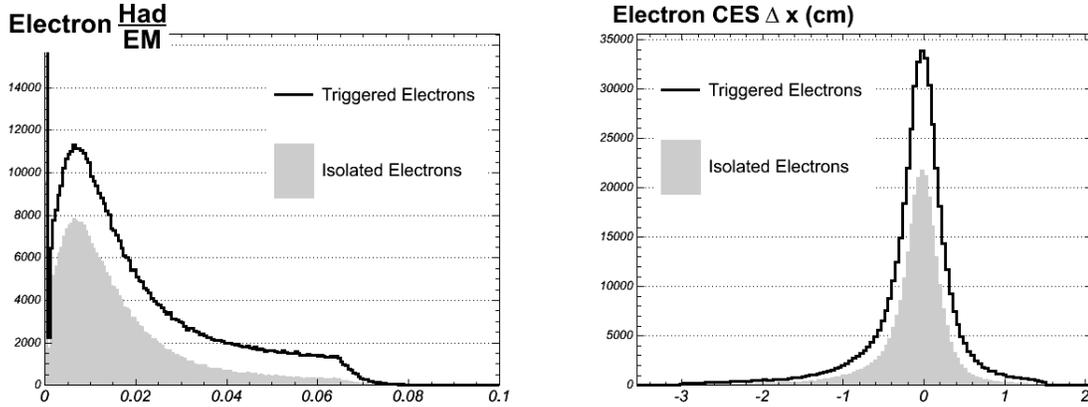


Figure 4.6: Distributions of $\frac{E^{had}}{E^{EM}}$ (left) and $q\Delta(r\varphi)$ (right) for high- p_T electron candidates in data from the 18 GeV electron trigger. Distributions are shown before and after the isolation requirement.

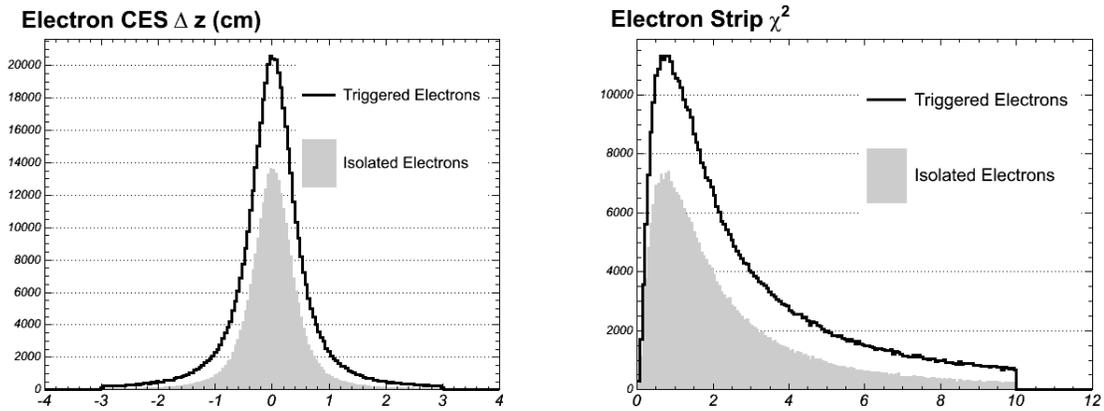


Figure 4.7: Distributions of Δz (left) and CES strip χ^2 (right) for high- p_T electron candidates in data from the 18 GeV electron trigger. Distributions are shown before and after the isolation requirement.

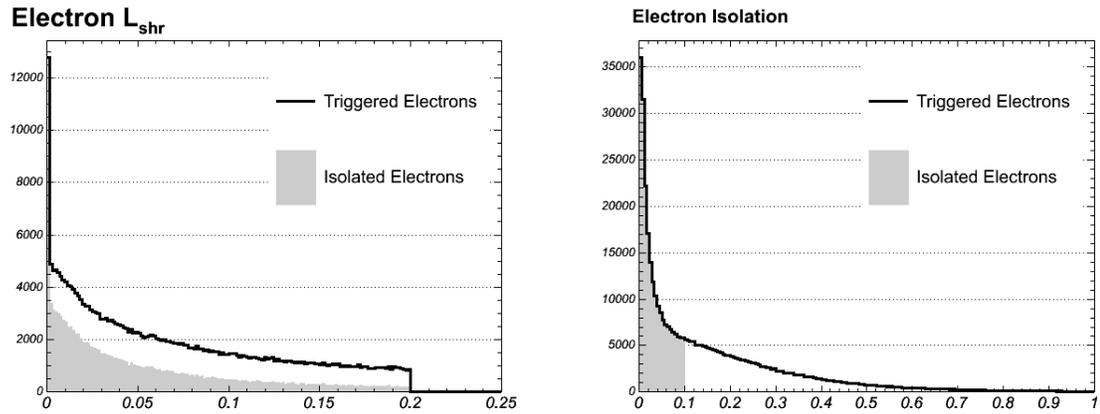


Figure 4.8: Distributions of L_{shr} (left) and isolation I (right) for high- p_T electron candidates in data from the 18 GeV electron trigger. Distributions are shown before and after the isolation requirement.

4.2.2 Muons

Muons are minimum-ionizing, and the signature in the CDF II detector therefore consists of a track with little or no calorimeter activity and a stub in the muon chambers. In this analysis, we accept muons only in the central regions, where the stub appears either in the CMX or in both the CMU and CMP, called CMUP. CMX muons include the Miniskirt and Keystone regions when they were active, as described in Section 3.2.3. The muon $\eta - \varphi$ distribution is shown in Figure 4.9, and can be compared to the muon detector diagram in Figure 3.16. As with electrons, the offline muon selection criteria are more stringent than those applied at trigger level [46]. They are:

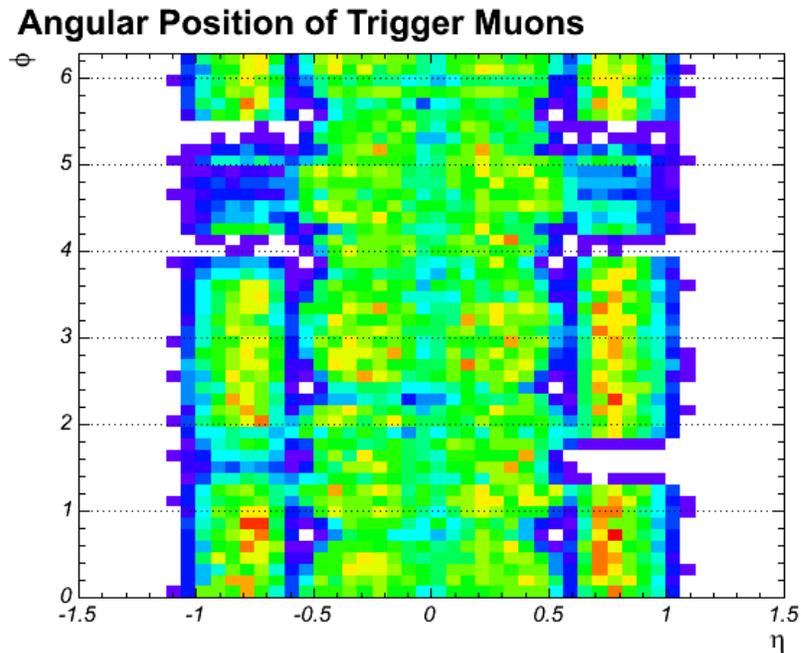


Figure 4.9: Scatter plot of η vs. φ for high- p_T muon candidates in the CMU, CMP, and CMX.

- Transverse track momentum $p_T > 20 \text{ GeV}/c$.
- Calorimeter energy deposits $E^{EM} < 2 \text{ GeV}$ and $E^{had} < 6 \text{ GeV}$: Minimum-ionizing particles will leave very small amounts of energy in the calorimeters. When the muon momentum p exceeds $100 \text{ GeV}/c$, the requirements are $E^{EM} < 2 \text{ GeV} + 0.0115 c(p - 100 \text{ GeV}/c)$ and $E^{had} < 6 \text{ GeV} + 0.0280 c(p - 100 \text{ GeV}/c)$. Some tracks miss the calorimeter altogether, as shown in Figure 4.11.
- Track extrapolation to muon chambers should match stub position in $r\varphi$: Similar to the CES matching for electrons, this requirement removes random pairings of tracks and stubs which can imitate the muon signature. Since we expect multiple scattering to affect the muon's path through the detector material, the tightness of the requirement is different for each of the three subdetectors. For the CMU, CMP, and CMX, $|\Delta r\varphi|$ is required to be less than 7 cm, 5 cm, and 6 cm, respectively.
- The overall quality of the COT track fit $\chi^2 < 2.3$ (2.75): To limit contamination from decays-in-flight ($K \rightarrow \mu\nu$, $\pi \rightarrow \mu\nu$), the track is required to be well-measured and consistent with a single helix. The tightness of this requirement was selected to be 99% efficient for real muons and varies over time, as described in Section 6.3.1.
- Impact parameter $|d_0| < 0.02$ (0.2) cm when the track uses (does not use) silicon information: To reject cosmic ray muons, which will pass through the detector but only rarely cross the beamline, we require the muon have a small impact parameter.

- Muons are required to be isolated from calorimeter activity, $I < 0.1$: As we do for electrons, we require that the energy in the calorimeter within a cone of $\Delta R \leq 0.4$ (excluding the muon tower energy) does not exceed 10% of the muon p_T .

The efficiency of these requirements for muons with p_T above 20 GeV/ c (measured in $Z \rightarrow \mu^+\mu^-$ events, again excluding the trigger efficiency) is 90%; differences between CMUP and CMX and the time-dependence of this figure will be reported in Section 6.3.1. Distributions of high- p_T muon properties are shown in Figures 4.10 to 4.14.

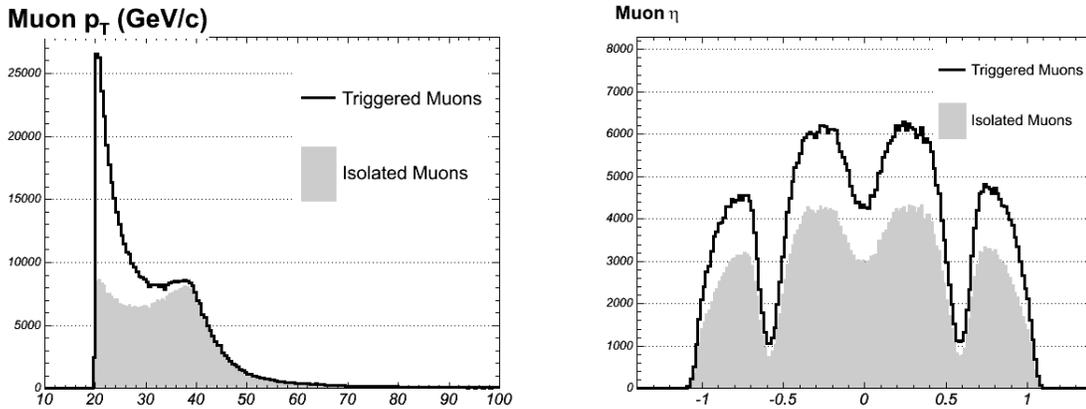


Figure 4.10: Distributions of p_T (left) and η (right) for high- p_T muon candidates in data from the 18 GeV/ c muon trigger. Distributions are shown before and after the isolation requirement.

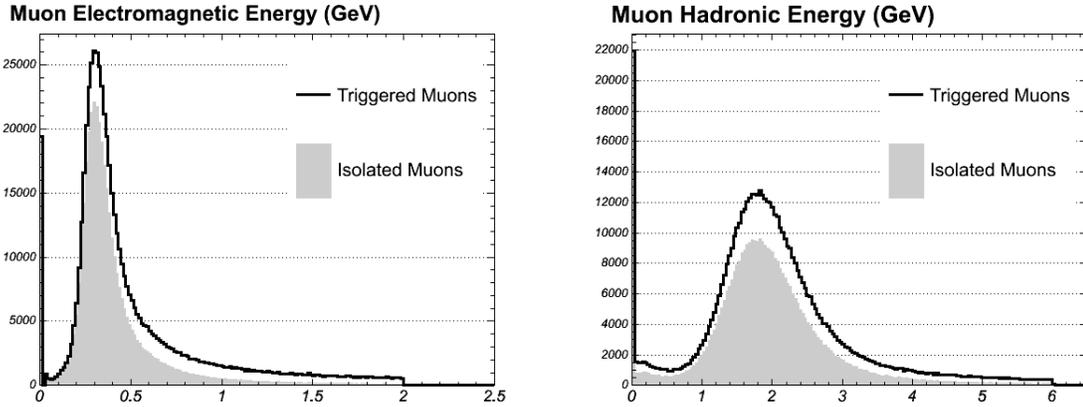


Figure 4.11: Distributions of electromagnetic (left) and hadronic (right) energy deposited for high- p_T muon candidates in data from the 18 GeV/ c muon trigger. Distributions are shown before and after the isolation requirement.

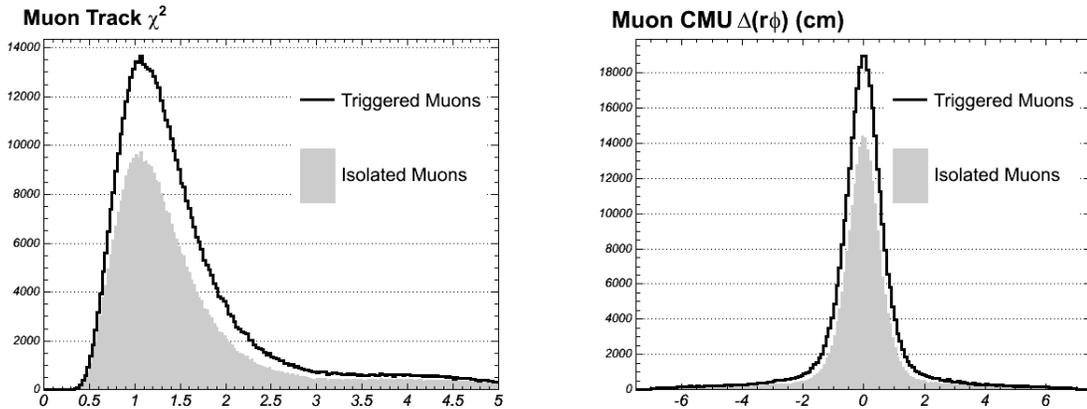


Figure 4.12: Distributions of track χ^2 (left) and CMU $\Delta(r\phi)$ (right) for high- p_T muon candidates in data from the 18 GeV/ c muon trigger. Distributions are shown before and after the isolation requirement.

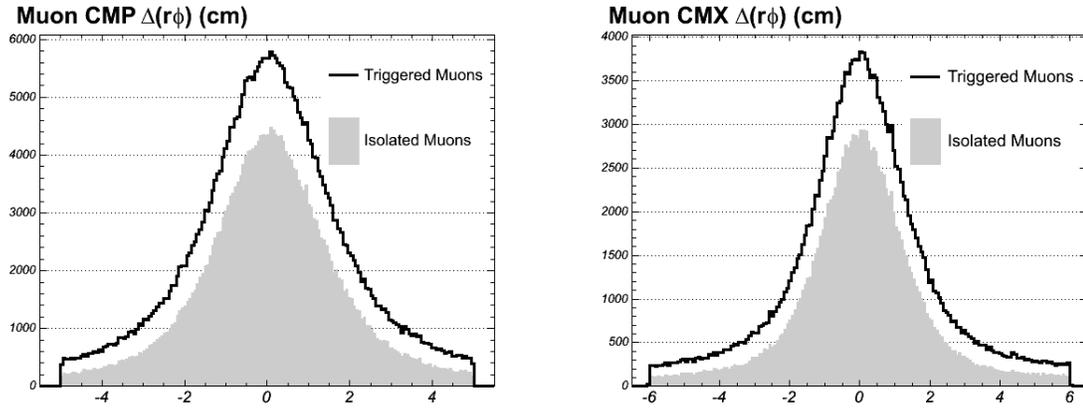


Figure 4.13: Distributions of $\Delta(r\phi)$ for high- p_T muon candidates in the CMP (left) and CMX (right). Distributions are shown before and after the isolation requirement.

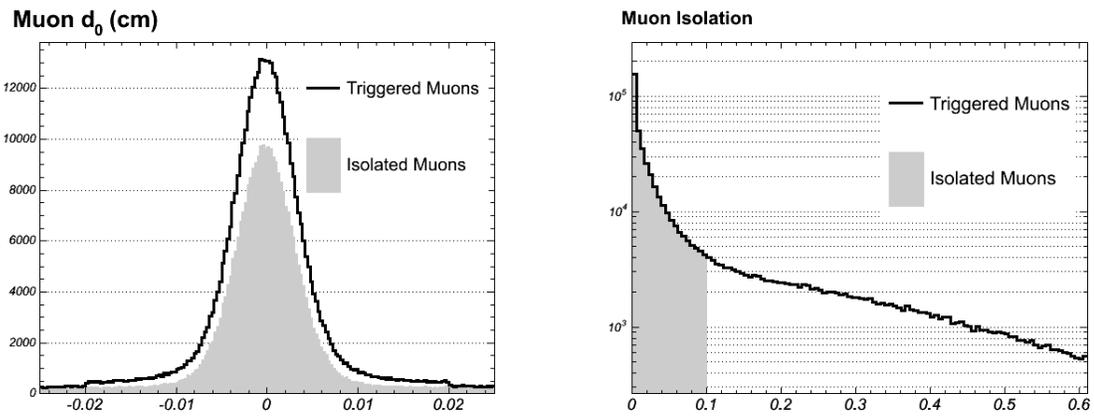


Figure 4.14: Distributions of L_{shr} (left) and isolation I (right) for high- p_T muon candidates in data from the 18 GeV/ c muon trigger. Distributions are shown before and after the isolation requirement.

4.3 Jet Clustering

When quarks and gluons are produced in $p\bar{p}$ collisions, they hadronize before traveling through the detector, resulting in a stream of color-neutral particles roughly collinear to the initial parton. When they pass through the detector, these clusters of particles will leave a broad energy deposit in the calorimeters, often accompanied by multiple tracks in the COT and silicon detector. These objects are called *jets*.

Since jets are commonly wider than a single calorimeter tower, the total transverse energy and direction of a jet are determined through a clustering algorithm called `JetClu` [47]. Starting from seed towers with deposited energy >1 GeV, the algorithm considers the energy deposits in all towers within a cone $\Delta R \leq 0.4$. The transverse energy-weighted centroid of this cluster is taken as the new center, and a new *jet cone* is defined, iteratively moving the center until the jet energy and its centroid remain stable.

Since the algorithm is iterative and uses multiple seeds, two jets can be found with overlapping cones. In this case, we determine f_{merge} , the fraction of a jet's energy that falls inside the cone of another jet. When f_{merge} is greater than 75%, the two jets are merged, and the E_T -weighted centroid is recalculated from all towers in both initial jets. If $f_{merge} < 0.75$, the two jets are separated, with each tower individually associated to the jet with the nearer center; again, the centers of the two jets are recalculated.

The raw E_T of the clustered jet is defined as the vector sum of transverse energies deposited in each tower. We correct the raw energy for various detector and physics effects, such as the intrinsic tower-by-tower non-uniformity, degradation in detector

performance over time, non-linearity in detector response, and excess energy from additional $p\bar{p}$ collisions in a particular crossing [48]. We derive an additional p_T -dependent correction in Monte Carlo to account for the difference between simulated jet E_T 's and initial parton transverse momenta. On average, the corrected jet E_T is equal to the parton p_T . Each level of jet energy corrections carries a p_T -dependent systematic error shown in Figure 4.15.

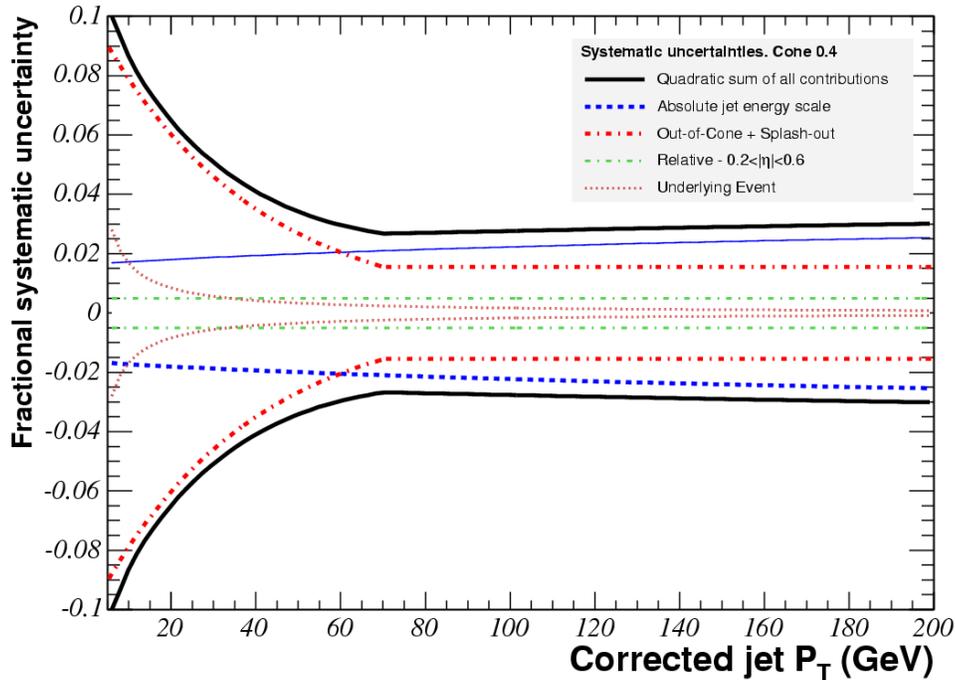


Figure 4.15: Fractional systematic error on the correction to the jet energy as a function of jet transverse momentum. Several levels of correction are applied, and the systematics are taken in quadrature. This analysis does not make use of the out-of-cone or underlying event corrections, so those uncertainties are ignored.

We use the central and plug calorimeters together for jet clustering, so we reconstruct jets out to a pseudo-rapidity of 2.4. In this analysis, we require jet E_T 's to exceed 20 GeV.

4.4 Missing Energy

Some particles produced in $p\bar{p}$ collisions will not interact with the detector at all. Neutrinos are the only example in the Standard Model, but many hypothetical particles in extensions to the Standard Model exhibit the same behavior. There will be no direct evidence of such particles in the tracker or calorimeter, but we can infer their presence indirectly, through an imbalance in the total measured E_T in the event. The total vector E_T can be written as:

$$E_T^{\vec{r}aw} \equiv \sum E_i^{raw} \sin \theta_i \hat{n}_i, \quad (4.4)$$

where the sum runs over all calorimeter towers and \hat{n}_i is the unit vector in the $x - y$ plane pointing from the beam axis to the tower. Since momentum is conserved, and the initial p and \bar{p} transverse momenta are small, this quantity will be zero when the calorimeter is hermetic and infinitely precise.

The total transverse energy recorded in the event is corrected for the non-neutrino contribution to the imbalance. First, the difference between the calorimeter deposit and the track momentum for identified muons is added back into this sum, since muon tracks will typically leave less than 2 GeV of energy in the calorimeter independent of its momentum. Second, the discrepancy between the raw energy in jets and their *corrected* transverse energies (as was mentioned in Section 4.3) is added back in the vector sum, and a similar correction is made for the unclustered energy in the event. Since the corrected vector sum of transverse energies, \vec{E}_T^{corr} , will point opposite the direction of a real particle that goes undetected, it is more convenient to define the *missing energy*, \vec{E}_T , simply as $-\vec{E}_T^{corr}$.

4.5 Monte Carlo

The performance of the CDF II detector and event reconstruction techniques can be studied in a few simple data samples (such as $Z \rightarrow \mu^+\mu^-$ events), but more complicated event signatures require the use of Monte Carlo simulations. The simulations are structured to have the same format as real collision data, but with a record of the underlying physics process. Monte Carlo is critical for understanding and interpreting data, and can be tuned to account for some inherent limitations and imperfections in the detector and in our analysis methods.

Monte Carlo simulation proceeds in three steps: *generation*, *detector simulation*, and *production*. The *generation* of a Monte Carlo event is the determination of the four-momenta of all particles in the initial and final states for the physics process being modeled. Unless otherwise noted, we use PYTHIA [49] (version 6.216) to generate our signal ($t\bar{t}$) and some background samples, and ALPGEN [50] (version 2.10) is used to generate W +jets background samples; the latter of these will be given special attention in Appendix B. In both cases, the EvtGen [51] software package is used specifically to model the decays of generated heavy flavor (bottom and charm) hadrons.

The generated event record is subjected to a *detector simulation*, such that the Monte Carlo events contain the same information as real data. This stage involves converting track trajectories into hits in the tracking chambers and simulating the shower of particles as they reach the calorimeters. The CDF simulation uses GEANT 3, a software package that models the energy loss of particles as they pass through the detector volume [52]. GEANT is given a three-dimensional map of the CDF II detector

material, and will incorporate multiple scattering, photon conversions, *bremsstrahlung*, and nuclear interactions into the generated event record. Additional tools fill in for GEANT for specific tasks: GFLASH [53] employs a fast shower simulator to speed up calorimeter simulations, and a tuned parametric model accurately simulates charge deposition in the silicon tracker.

The production stage, in which physics quantities are reconstructed in simulated events, is identical to the treatment of data samples. Because the detector and beam change over time (silicon wedges being off or on, movement of the beam, alignment of the detector, instantaneous luminosity effects, etc.), the Monte Carlo is interfaced with the CDF database to incorporate the run dependence and make the simulation more similar to data. For this analysis, the most critical component of this run-by-run simulation is the inclusion of additional $p\bar{p}$ interactions at high luminosity, which are simulated with PYTHIA and superimposed on the event record.

Chapter 5

b-Tagging

The salient feature of $t\bar{t}$ final states is the presence of heavy flavor, two bottom quarks from the two top quark decays and a charm quark from every other hadronic W decay. Identification of heavy flavor jets, called *b*-tagging, is critical to suppress backgrounds to $t\bar{t}$, most of which include only light flavor jets in the final state.

At CDF, we have developed *b*-tagging algorithms that take advantage of two key properties of bottom hadrons, large semi-leptonic branching fractions and long lifetime. *Soft lepton taggers* search for low- p_T muons or electrons inside jets, which are likely to be the result of a heavy flavor hadron decay. Though the *efficiency* of soft lepton taggers, the fraction of electrons or muons correctly identified, is quite high (>90%), the total semi-leptonic branching fraction for B hadrons is only 20%, which puts an upper limit on the fraction of *b*-jets which can be *b*-tagged. The fake rate, the fraction of non-leptons which are incorrectly tagged, can also be as high as 0.5%, resulting in a poor signal-to-background ratio in a $t\bar{t}$ analysis [54].

Lifetime tagging, on the other hand, is much less restricted, since bottom hadrons

have a nonzero lifetime, $450 \mu\text{m}$. In $t\bar{t}$ events, bottom quark jets (*b*-jets) are produced with transverse momentum around 50 GeV, as shown in Figure 5.1, corresponding to a typical flight distance of 4 to 5 mm; on the scale of the SVX tracking resolution ($\sigma_{d_0} \approx 30\mu\text{m}$), decays with this displacement are often distinguishable from the primary interaction. The fake *b*-tag rate for lifetime taggers is typically 0.5% per jet [55].

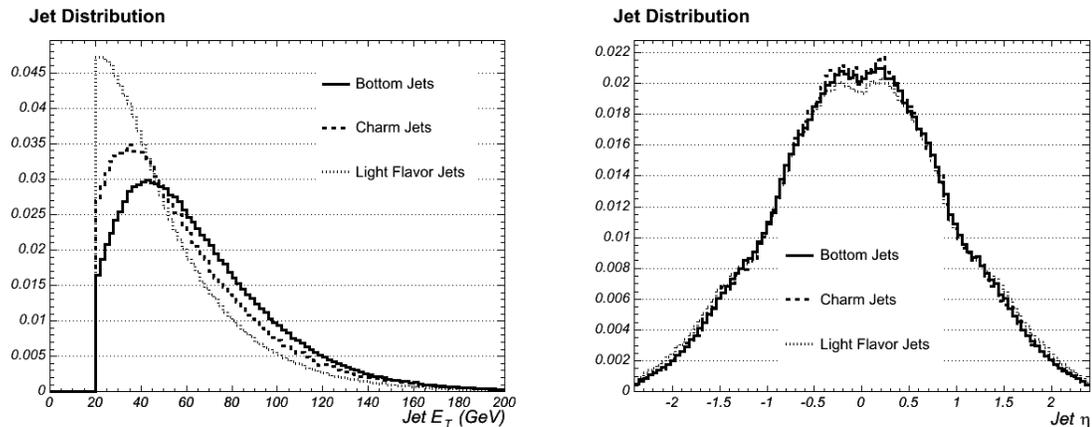


Figure 5.1: Distributions of jet E_T (left) and η (right) for bottom, charm, and light flavor jets in $t\bar{t}$ Monte Carlo. Central *b*-jets have an average E_T of 50 GeV, and therefore typically travel 5 mm before decaying.

In this analysis, we use an algorithm (SECVTX, described in detail in the following section) that selects tracks inside a jet with large impact parameter and attempts to reconstruct the best single decay vertex from them [56]. The simplicity of this approach allows us to identify heavy flavor with high efficiency, but we can not distinguish between actual *B* decay vertices and those from charm or light flavor (*mistags*) on a jet-by-jet basis. Instead, we study large samples of *b*-tagged jets and determine on average the tag rates for each type of jet.¹ The methods used to quantify the

¹The terms *mistag rate* and *tag rate* are misnomer, but are commonly used at CDF. We define the tag (mistag) rate simply to be the fraction of jets which are *b*-tagged (mistagged). Mistags will be described in more depth in Section 5.3.

algorithm's performance in terms of its bottom and charm tagging efficiencies and mistag rate are discussed in Sections 5.2 and 5.3. The application of these results to the lepton+jets dataset is described in Chapter 6.

5.1 The SecVtx Algorithm

We use two tunings of the algorithm in this analysis, *tight* and *loose*, which have been optimized separately for high purity (*tight*) and high *b*-tagging efficiency (*loose*). The loose tagger is 20% more efficient for identifying *b*-jets in $t\bar{t}$ events, but at the expense of a 200% increase in light flavor tags [57]. The loose tuning is especially useful in analyses requiring multiple *b*-tags, since the background contribution from multiple fake tags is small.

The SECVTX algorithm runs on individual jets, reconstructing heavy flavor decay vertices from tracks within a cone $\Delta R \leq 0.4$ from the axis of the jet. Tracks are required to include multiple hits in the silicon detector, and to be displaced significantly from the event primary vertex determined by `PrimeVtx` (as described in Section 4.1.1). All tracks are required to have $p_T > 0.5$ GeV/ c , at least two axial and at least two stereo COT segments, $|d_0| < 1.5$ mm (to reduce the contribution from conversions and tracks from material interactions), and a z_0 within 1 cm of the primary vertex [55].

SECVTX is a two-stage algorithm, first attempting to form a vertex with at least three tracks (*Pass 1*), then making two-track vertices with tighter track selection (*Pass 2*). The algorithm only attempts *Pass 2* when no acceptable vertices are found in *Pass 1*. *Pass 1* and *Pass 2* track selections (silicon requirements and impact parameter

significance²) are different for the tight and loose tunings, and are shown in Table 5.1.

		Loose		Tight	
		Pass 1	Pass 2	Pass 1	Pass 2
SVX layers	>	2	3	3	
Track- χ^2	<	8.0			
$ z_0 $ relative to seed (cm)	<	2.0			
Track- d_0 (cm)	<	0.15			
Track- p_T (GeV)	>	0.5	1.0	0.5	1.0
Track- d_0 Sign.	>	2.0	3.0	2.0	3.5
Seed Vertex χ^2	<	50			
At Least One Track- p_T (GeV/ c)	>	1.0	1.5	1.0	1.5
Track prune χ^2	>	90	1000	45	30
Vertex fit χ^2	<	120	2000	50	
L_{2d} Significance	>	6.0		7.5	

Table 5.1: Comparison of pass 1 and pass 2 track selection and vertex requirements for the loose and tight SECVTX taggers.

We use the CTVMFT algorithm [43] to find displaced vertices, which are characterized in terms of their two-dimensional displacement from the primary vertex projected along the jet direction, denoted L_{2d} or L_{xy} and illustrated in Figure 5.2. The *significance* of L_{2d} is the final indicator of whether or not a jet is *b*-tagged, where the denominator includes a contribution from the primary vertex uncertainty. In detail, the tight algorithm proceeds as follows [55]:

1. **Track Selection:** Loop over tracks
 - (a) Define *good tracks* as tracks passing the following criteria:
 - $p_T > 0.5$ GeV/ c
 - $|d_0| < 0.15$ cm relative to the primary vertex
 - $|z_0| < 2.0$ cm relative to the primary vertex
 - ≥ 2 axial and ≥ 2 stereo segments in the COT
 - Number of SVX $r - \phi$ hits ≥ 3

²The *significance* of a quantity is defined as its value divided by its uncertainty.

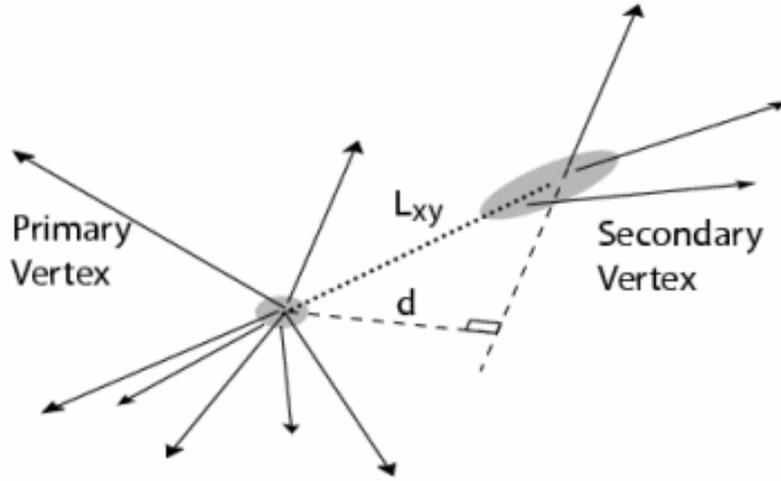


Figure 5.2: Illustration of a SECVTX vertex. The significance of the two dimensional displacement, L_{2d} determines whether or not a jet is *b*-tagged.

- Track fit $\chi^2/dof < 8.0$.
- (b) Remove tracks with an oppositely charged partner track that have invariant mass between 0.4876 and 0.5076 GeV/c^2 , consistent with $K_0^s \rightarrow \pi^+\pi^-$, or between 1.10963 and 1.12163 GeV/c^2 , consistent with $\Lambda \rightarrow p\pi$
- (c) Associate tracks to jets if they are within a $\Delta R < 0.4$ cone

2. Obtain a tag candidate: Loop over jets

- (a) **Pass 1:** Select tracks with $\frac{d_0}{\sigma_{d_0}} > 2.0$ and $p_T > 0.5 \text{ GeV}/c$, loop over pairs of tracks
 - i. Form a seed vertex between the pair of tracks
 - ii. Add tracks to the vertex with $\frac{d_0}{\sigma_{d_0}} < 3.0$ (relative to seed vertex)
 - iii. Remove tracks that contribute $\chi^2 > 50$
 - iv. If no 3-track vertex is found, select next track pair
 - v. If a 3-track vertex is found, require vertex to satisfy:
 - Highest track p_T , $p_T^{max-track} > 1.0 \text{ GeV}/c$
 - Vertex fit $\chi^2 < 50$
 - $\frac{L_{2d}}{\sigma_{L_{2d}}} > 7.5$
 - Pseudo- $c\tau < 1.0 \text{ cm}$
 - $|L_{2d}| < 5.0 \text{ cm}$

- $R_0 < 2.5$ cm, where R_0 is the transverse distance of the vertex from the center of the SVX
- vi. If no Pass 1 vertex is found, examine next pair of seed tracks
- (b) **Pass 2:** If no 3-track vertex is found, select tracks with $d_0/\sigma_{d_0} > 3.5$ and $p_T > 1.0$, form vertices between pairs of tracks
- Highest track p_T , $p_T^{max-track} > 1.5$ GeV/ c
 - Overall vertex $\chi^2 < 50$
 - $L_{2D}/\sigma_{L_{2D}} > 7.5$
 - $|L_{2D}| < 5.0$ cm
 - Distance of the vertex from the center of the SVX $R_0 < 1.2$ cm or 1.45 cm $< R_0 < 2.5$ cm, unless vertex has more than 2 pass 1 tracks. This is to mitigate the effects of conversions and nuclear interactions.

The loose algorithm is equivalent, but with no cut on R_0 and different requirements on the tracks and vertices, as listed in Table 5.1. The properties of good, pass 1, pass 2, and used tracks from tight SECVTX in $t\bar{t}$ Monte Carlo are shown in Figures 5.3 to 5.5.

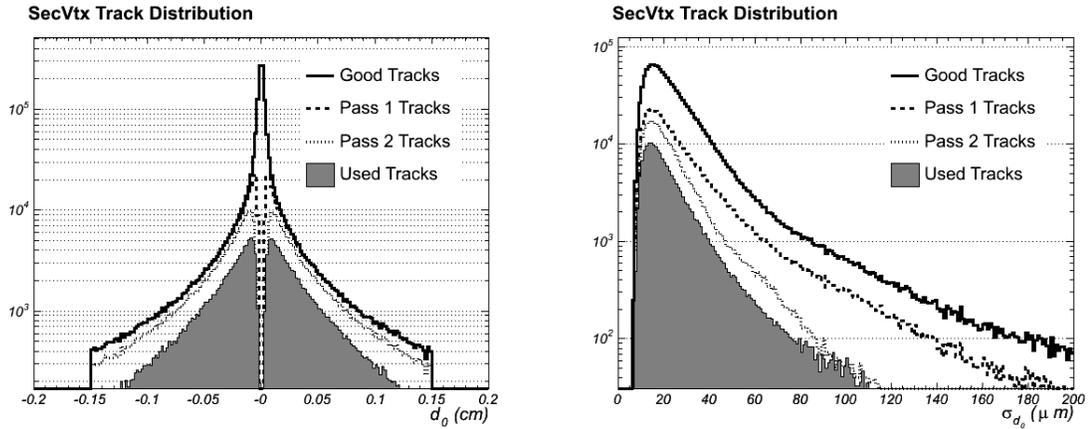


Figure 5.3: Distributions of d_0 (left) and σ_{d_0} (right) for tracks considered by tight SECVTX in $t\bar{t}$ Monte Carlo. “Used” tracks are those included in a vertex fit.

If a final vertex passes the selection listed above and has L_{2d} significance greater than 6.0 (7.5) for the loose (tight) tuning, the jet is *b*-tagged. The vertex properties

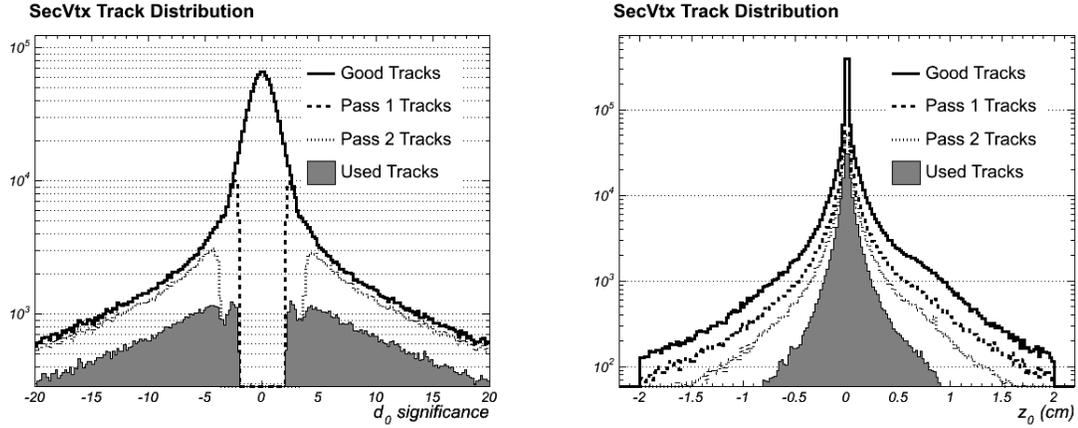


Figure 5.4: Distributions of d_0 significance (left) and z_0 (right) for tracks considered by tight SECVTX in $t\bar{t}$ Monte Carlo. “Used” tracks are those included in a vertex fit.

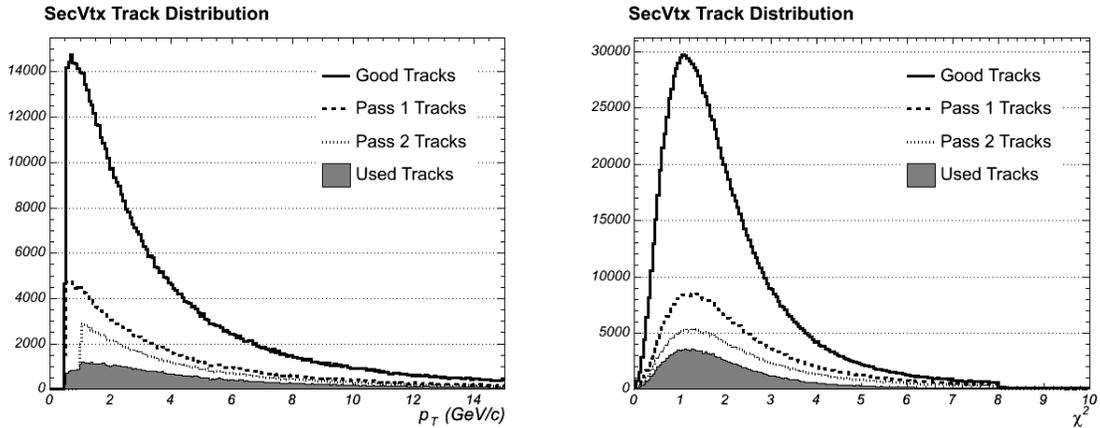


Figure 5.5: Distributions of p_T (left) and silicon χ^2/dof (right) for tracks considered by tight SECVTX in $t\bar{t}$ Monte Carlo. “Used” tracks are those included in a vertex fit.

are also recorded, as the invariant mass of the vertex tracks (m^{vtx} , assuming the pion mass for each track), the vector sum of the track p_T 's (p_T^{vtx}), and the effective lifetime of the tag (pseudo- $c\tau \equiv L_{2d} \frac{m^{vtx}}{p_T^{vtx}}$) provide some discrimination between bottom, charm, and light flavor decays, and can therefore be used to estimate the composition of a set of b -tagged jets. This technique is explored in Section 5.3.3 and Appendix B.

Some comparisons of SECVTX distributions for bottom, charm, and light flavor jets are shown in Figures 5.6 to 5.8.

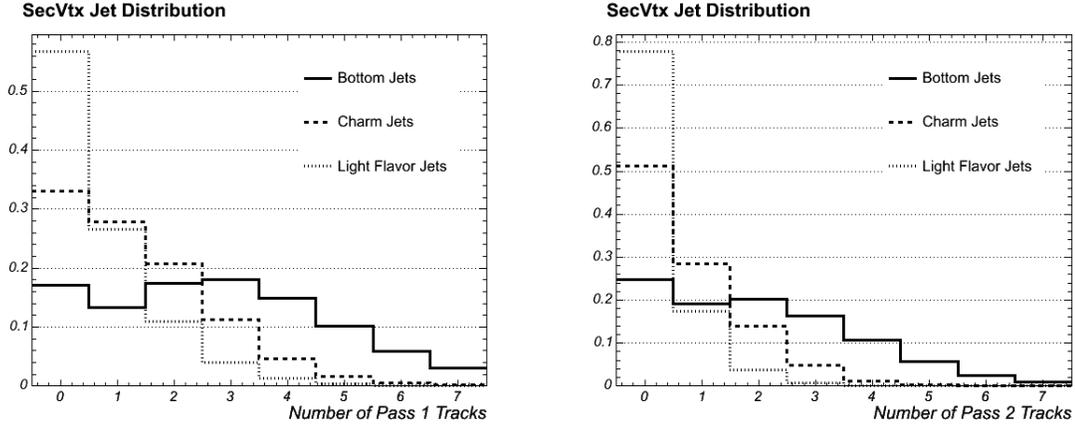


Figure 5.6: Numbers of pass 1 (left) and pass 2 (right) tracks with loose SECVTX for bottom, charm, and light flavor jets. Distributions are taken from $t\bar{t}$ Monte Carlo.

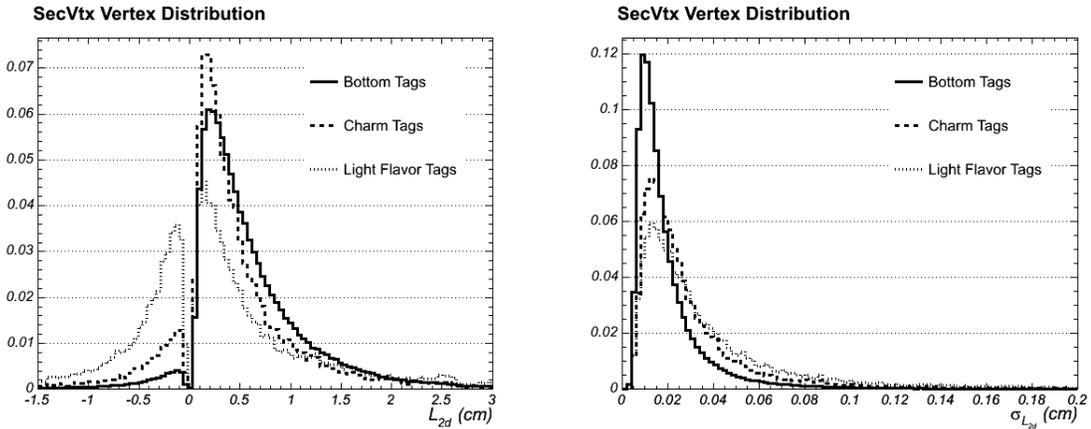


Figure 5.7: L_{2d} (left) and $\sigma_{L_{2d}}$ of loose SECVTX-tagged jets. Distributions are taken from $t\bar{t}$ Monte Carlo.

A critical feature of the SECVTX algorithm is its symmetry with respect to the primary vertex, in that tracks with either positive or negative d_0 can be used in a vertex fit. Secondary vertices may be reconstructed *behind* the primary, with negative

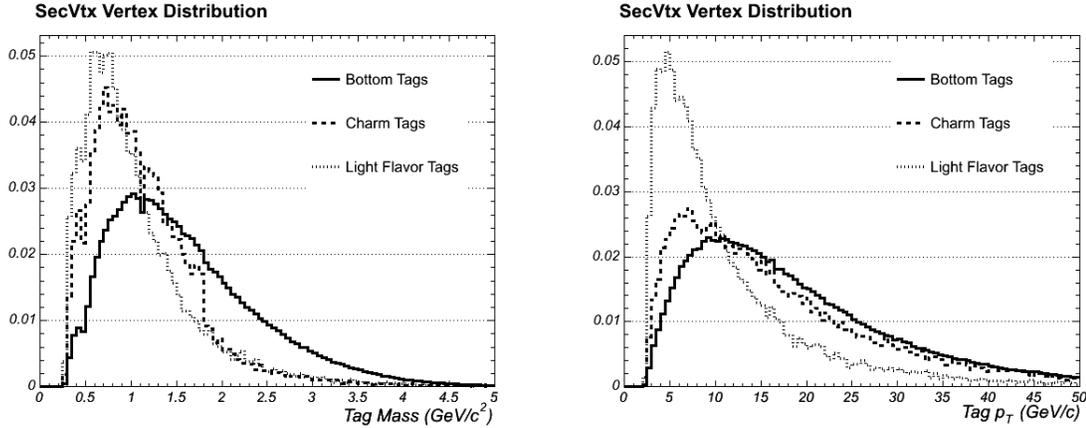


Figure 5.8: Invariant mass (left) and p_T of loose SECVTX-tagged jets. Distributions are taken from $t\bar{t}$ Monte Carlo.

L_{2d} . These vertices are unphysical, but can come from mis-measured tracks or incorrect pairings of real tracks. When these vertices are sufficiently negative, $\frac{L_{2d}}{\sigma_{L_{2d}}} < -7.5$ (-6.0) for the tight (loose) tagger, the jet has a *negative* tag. The effects leading to negative tags are expected to be approximately symmetric with respect to the primary vertex, so the negative tag rate can be used as an estimate of the rate of false positive tags [65]. The details of this method are presented in Section 5.3.

5.2 *b*-Tagging Efficiency

The most important figure of merit for a tagging algorithm is the *b*-tagging efficiency, the fraction of *b*-jets that are correctly identified. The efficiency itself is a strong function of the momentum and lifetime of the heavy flavor hadron, the charged particle multiplicity of its decays, and the efficiency and spatial resolution of the tracking algorithms. Since all these are included in Monte Carlo simulations, we use simulated *b*-tags to derive a first-order estimate of the efficiency.

However, to compensate for imperfections in the simulation, we also calibrate the *b*-tagging efficiency in data, using a sample of dijet events with identified leptons, both electrons and muons. The presence of a lepton inside a jet is often an indication of heavy flavor, so the SECVTX-tagged jets in such events are dominated by real *b*-tags rather than mistags. By checking the efficiency in these samples against the *b*-tagging efficiency in an equivalent simulated sample, we derive a single multiplicative correction to the Monte Carlo efficiency, the *b*-tagging scale factor SF [62]:

$$SF = \frac{\epsilon^{data}}{\epsilon^{MC}}, \quad (5.1)$$

where the ϵ 's are the *b*-tagging efficiencies for heavy flavor jets in data and Monte Carlo.

In a perfect simulation, SF would be unity, but any mis-modeling will impact SF . In Run 1, for instance, SF was measured to be as large as 1.2 due to an incorrect *B* lifetime in Monte Carlo [58]. Early in Run 2, SF was 0.8, when the simulation overestimated the tracking efficiency relative to data [59].

Now, we measure SF using two distinct methods, using samples with electrons

and muons. Results from the two methods are consistent, and are presented in Sections 5.2.1 and 5.2.2.

5.2.1 Electron Method

Overview

We measure the *b*-tagging efficiency in a heavy flavor-enriched dijet sample, where one jet contains an electron (the “electron jet”) and recoils against a second jet (“away jet”). The sample is split into those events where the electron has a conversion partner and those where no partner is found, producing two complementary subsamples with similar event topologies but different heavy flavor content.

It can be shown that:

$$F_{a\text{-tag}}^{hf} = 1 - \frac{\frac{C_+}{N_+} - \frac{C^+ - \alpha C^-}{N^+ - \alpha N^-}}{\frac{C}{N} - \frac{C^+ - \alpha C^-}{N^+ - \alpha N^-}} (1 - F^{hf}), \quad (5.2)$$

where $F_{a\text{-tag}}^{hf}$ and F^{hf} are the fraction of electron jets containing heavy flavor with and without requiring the away jet to be *b*-tagged, and the N 's and C 's correspond to the total number of events and number of conversion events in the sample, with the superscript representing to the electron-jet tag and the subscript representing the away jet (tight) tag [60]. The α terms in the equation are the *mistag asymmetry*, the ratio of positive light flavor tags to negative tags [61]. Since we use negative tags in SECVTX to estimate the fraction of light flavor jets that are mistagged, we need to account for two fundamental issues with this construction. First, even the light flavor tag L_{2d} distribution is not perfectly symmetric; conversions, tracks from material interactions, and K_S and Λ decay products will be highly displaced from the origin,

and will contribute to some real *b*-tags with apparently long lifetime. Secondly, heavy flavor jets will also contribute to the negative tags, due to both mis-reconstructed prompt tracks and combinatorics with real displaced tracks. In a sample with as much heavy flavor as this, we find that only 30% of the observed number of negative tags is needed to account for the expected number of positive light flavor tags, and we take this value as α for this sample. The derivation of a similar quantity in all-jet events (where light flavor *b*-tags dominate the negative tail) is presented in Section 5.3.3.

We assume that SF is the same for bottom and charm jets, and that it is independent of the tag requirement on the away jet. We can then solve for the tagging efficiency in data using Equation 5.2 and the following definitions:

$$\epsilon_1^{data} = (SF)\epsilon_1^{MC} = \frac{N^+ - \alpha N^-}{N} \frac{1}{F^{hf}} \quad (5.3)$$

$$\epsilon_2^{data} = (SF)\epsilon_2^{MC} = \frac{N_+^+ - \alpha N_+^-}{N_+} \frac{1}{F_{a-tag}^{hf}} \quad (5.4)$$

ϵ_1 and ϵ_2 are the efficiencies to tag heavy flavor electron jets without and with the requirement of an away jet tag. Since the Monte Carlo efficiency is known, the above system of equations has three unknowns, and we can directly solve for SF given a set of N 's and C 's [60].

Event Selection

The data were collected with a trigger requiring an 8 GeV electron in the central calorimeter, and we apply a tuned electron selection (different from that in Section 4.2) optimized for low- p_T , non-isolated electrons. We produce a generic PYTHIA

dijet Monte Carlo, which is filtered for a 7-GeV electron at generator level (*i.e.*, before conversions are added). Fewer than 1% of generated events are passed the filter, but the resulting sample is dominated by semileptonic heavy flavor decays, with a charm-to-bottom ratio of 25%. For both data and Monte Carlo, we take the highest E_T electron with:

- $E_T > 9$ GeV
- $p_T > 8$ GeV/ c
- $0.5 < \frac{E}{p} < 2$
- $\frac{Had}{EM} < 0.05$
- $L_{shr} < 0.2$
- CES $\Delta x < 3$ cm
- CES $\Delta z < 5$ cm
- Strip $\chi^2 < 10$
- z_0 within 5 cm of primary vertex
- $\Delta R < 0.4$ (to nearest jet)
- Isolation > 0.1
- Fiducial to SVX

The electron jet is the 15-GeV jet near the identified electron, and the away jet is the highest E_T jet (over 15 GeV) with $\Delta\varphi > 2$ from the electron jet. Conversions are identified with the same algorithm described in Section 4.2, looking for a partner track to the electron with a path tangent to the electron's with similar dip angles; a scatter plot of the position of identified conversions is shown in Figure 5.9. A summary of the data and MC counts for the samples used in the scale factor measurement after event selection is shown in Table 5.2.

	All Data	Conversion Data	Monte Carlo
N	2841530	980713	153147
N_+	92011	18065	15294
Loose SECVTX			
N^+	205614	21948	50123
N^-	20520	6219	158
N_+^+	21142	1475	5595
N_+^-	849	152	130
Tight SECVTX			
N^+	163096	13848	43024
N^-	8257	2354	774
N_+^+	17981	1139	4825
N_+^-	346	60	68

Table 5.2: Summary of the tag totals in the low- p_T electron data and Monte Carlo samples. The superscript refers to the electron-jet tag, and the subscript refers to the away-jet tight tag.

Figures 5.10 and 5.12 show comparisons between data and Monte Carlo for the samples used to calculate SF . Due to the difference in the physics content of the samples, we have further separated data into identified conversions and non-conversions, normalizing the Monte Carlo distribution to the non-conversions. The conversion-finding efficiency for electrons at this p_T is $\tilde{80}\%$, so the contribution from real conversions to the non-conversion distribution is non-negligible. This is likely the cause of some of the disagreement in distributions like the electron jet η , since the photon conversion probability is larger for photons traversing more material.

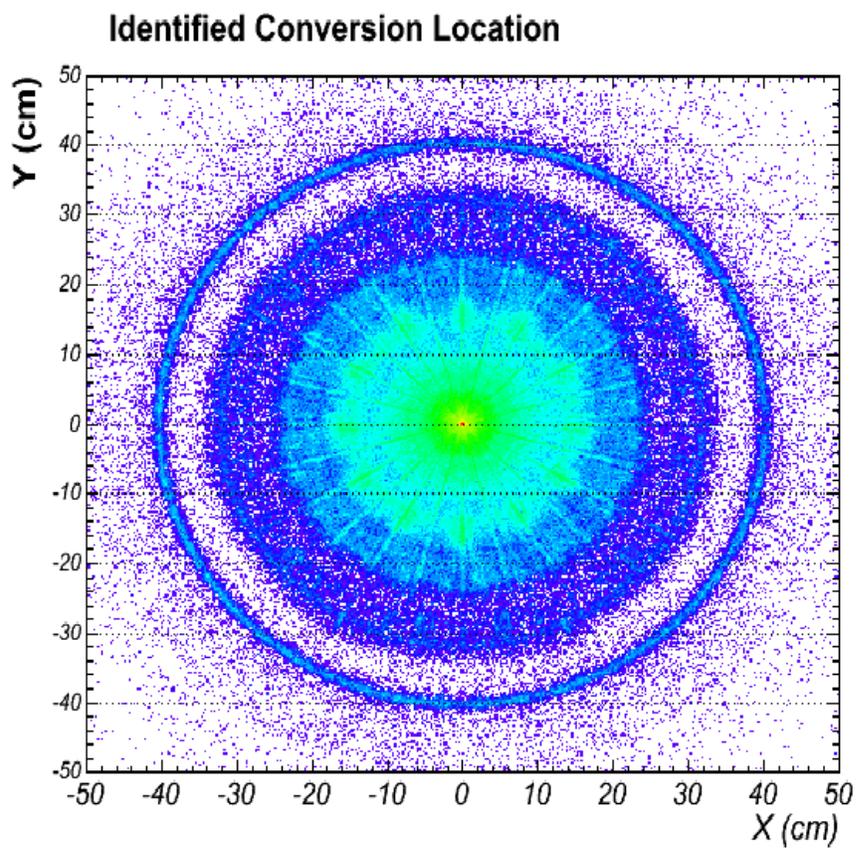


Figure 5.9: The position of identified conversions in the full dataset. Much of the detector structure is visible, including the azimuthal symmetry of the silicon detector at small radii, and the inner wall of the COT at a radius of 40 cm.

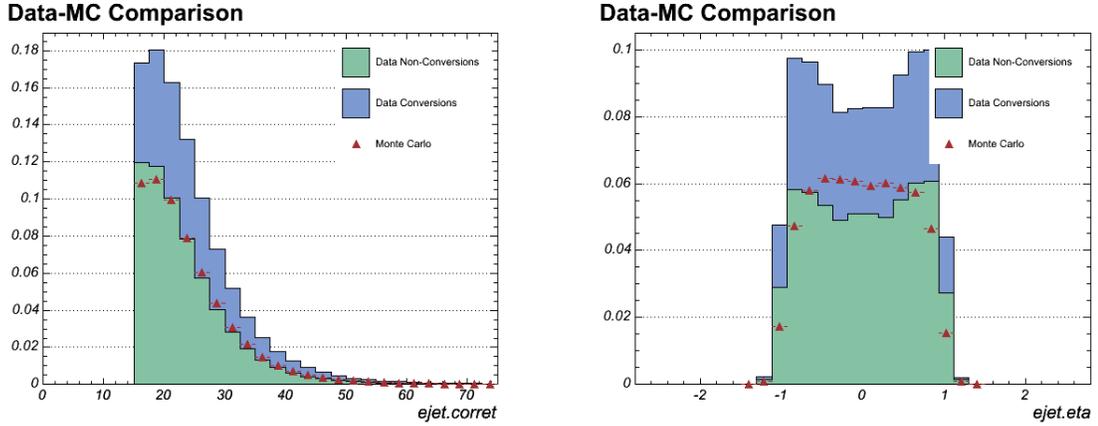


Figure 5.10: Data and Monte Carlo distributions of electron jet E_T and η in the data and Monte Carlo samples used to measure the scale factor SF . The data are taken from the 8-GeV electron sample, and the Monte Carlo is normalized to the non-conversion electron data.

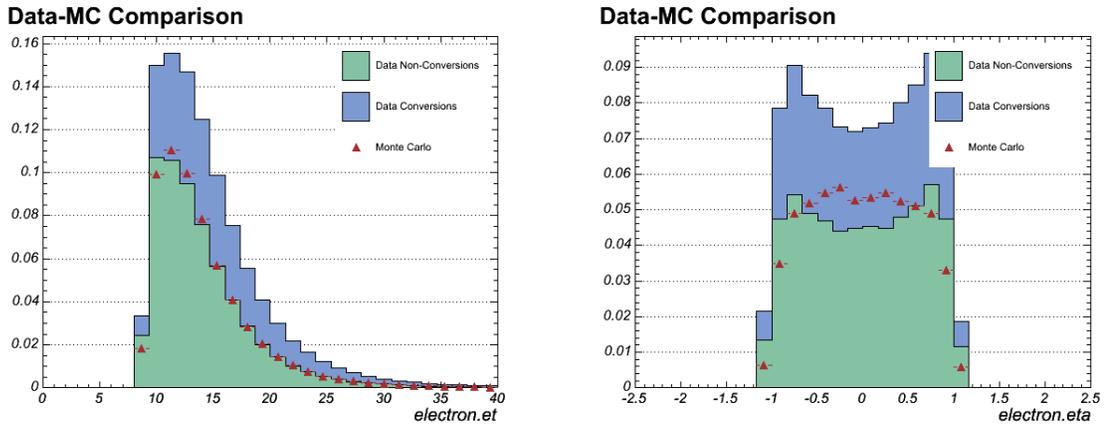


Figure 5.11: Data and Monte Carlo distributions of electron E_T and η in the data and Monte Carlo samples used to measure the scale factor SF . The data are taken from the 8-GeV electron sample, and the Monte Carlo is normalized to the non-conversion electron data.

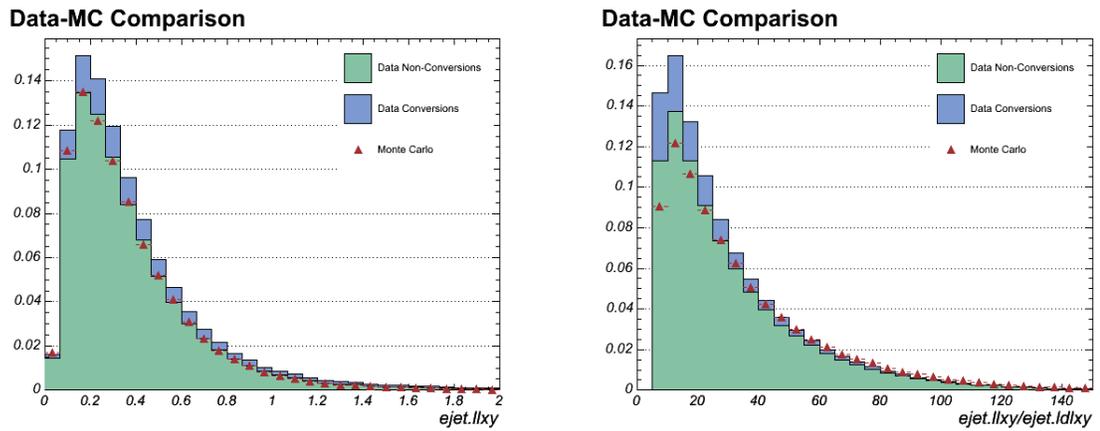


Figure 5.12: Data and Monte Carlo distributions of electron jet L_{2d} and L_{2d} significance for loose-tagged jets in the data and Monte Carlo samples used to measure the scale factor SF . The data are taken from the 8-GeV electron sample, and the Monte Carlo is normalized to the non-conversion electron data.

Systematics

We consider several sources of error, all related to the measurement of the scale factor directly in this electron sample. Other sources of error arise from the application of this result to alternate samples, such as $t\bar{t}$ events, but those are only discussed in Section 5.2.3. We account for the following systematic uncertainties:

- **Mistag Subtraction:** As described above, we use a correction factor α to scale the negative tag yield to the expected positive light flavor tag yield, but this correction is evaluated making some assumptions about the heavy flavor content of the data sample, which is not *a priori* known. We vary this correction within its uncertainties and allow for different values between the two taggers, and take the maximum deviation from the central value as a systematic error, 3.4% (loose) and 1.5% (tight).
- **Charm Scale Factor:** In this measurement, we do not separate charm and bottom tags, though $c\bar{c}$ events should contribute roughly 5% to the double-tagged sample. If the charm content is well-modeled in the simulation, the only repercussion is if the scale factor for bottom and charm is not the same. By adjusting the *b*-tagging efficiency in simulation by $\pm 15\%$, we re-measure the scale factor as is and take the deviation as a systematic uncertainty, 0.4% (loose) and 0.3% (tight).
- **Charm Content:** Similar to the previous point, if the charm efficiency is well-modeled in simulation but not the total charm production rate, we would

also mis-measure the bottom scale factor.³ Here, we fit the distribution of the SECVTX tag invariant mass, using separate templates for light flavor, charm, and bottom, and we use the results to infer the charm-to-bottom ratio before tagging. The fit results are shown in Figure 5.13. We measure a ratio of 0.35 ± 0.15 , consistent with the 0.25 taken from the simulation. If we remove enough *b* events to match this ratio in simulation and remeasure *SF*, we observe systematic shifts of 1.5% (loose) and 1.4% (tight).

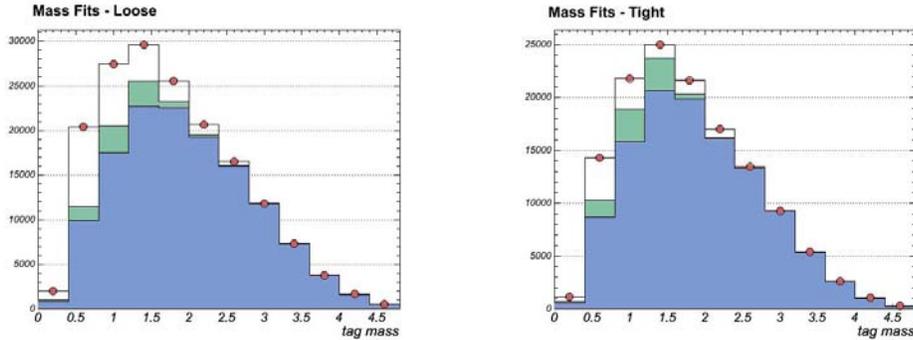


Figure 5.13: Fits to the vertex mass for loose-tagged (left) and tight-tagged (right) electron jets. From the bottom, the templates are for *b*, *c*, and light flavor electron jets.

- **Conversion Bias:** We assume in this method that the *b*-tagging rate in heavy flavor jets is unaffected by the presence of a conversion, though this often indicates the presence of at least one additional displaced track. We find in an unfiltered $b\bar{b}$ simulation that conversion-*b* jets have an efficiency roughly 5% higher than normal, dominated by *misidentified* conversions. Incorporating this

³In fact, since the semi-leptonic branching fractions are very different for D^0 , D^\pm , and D_s , the lepton-filtered Monte Carlo we use here is very sensitive to the relative populations of these hadrons. This has not been explored in depth, but the charm-to-bottom ratio is poorly constrained, and has large variations depending on the MC generator used.

Source	Loose (%)	Tight (%)
c/b Ratio	1.5	1.4
Conversion Bias	1.2	1.1
Semi-Leptonic Bias	1.6	1.4
Mistag Subtraction	3.4	1.5
Charm SF	0.4	0.3
MC Stats	1.3	1.5
Total	4.4	3.1

Table 5.3: Summary of systematic errors on the scale factor for both taggers. All numbers are quoted as relative percentage of the scale factor. Systematics which are relevant only to the application of the scale factor, such as the jet energy dependence, will be evaluated in Section 5.2.3.

5% shift into the scale factor formula, we estimate this bias to be 1.2% (loose) or 1.1% (tight).

- Semi-Leptonic Bias:** Finally, we assume here that the scale factor we measure is independent of the B decay topology, or that b -jets with electrons will have the same scale factor as those without. We assume that the key point here is the track multiplicity, since semi-leptonic decays will in general have fewer charged particles in the final states. If we evaluate the efficiency as a function of the number of tracks, and we re-weight according to the track multiplicity distribution on the away-jet side (which has no such bias), we get an estimate of how large this effect is. We find the maximum shift from this re-weighting procedure is 1.6% (loose) and 1.4% (tight).

A summary of the systematics is included in Table 5.3.

Results

The *b*-tagging scale factor SF can be calculated from the numbers in Table 5.2 and the equations in Section 5.2.1. However, these totals must first be corrected for the prescale of the trigger; since the trigger rate can be very large when the instantaneous luminosity is high, only a fraction of the events that pass the Level 2 requirements are sent to Level 3. This fraction is initially set to $\frac{1}{100}$, and approaches unity as the instantaneous luminosity drops. Since the *b*-tagging efficiency has been shown to fall at high luminosity (see Figure 5.19, for instance), we re-weight events in data only to compensate for the prescale changes. The net effect is roughly a 2% total shift in the observed tag rate [62]. Including systematic errors, the results are:

$$SF_{loose} = 0.954 \pm 0.016(stat) \pm 0.042(syst) \quad (5.5)$$

$$SF_{tight} = 0.977 \pm 0.018(stat) \pm 0.030(syst) \quad (5.6)$$

Figures 5.14 through 5.19 show the dependence of the scale factor and measured efficiency on various jet and event properties. The prescale correction is not taken into account here, and the uncertainties are statistical only. We see no evidence of a non-uniform scale factor, although the uncertainties are large at high E_T . The first and last bins in each plot include the underflows and overflows.

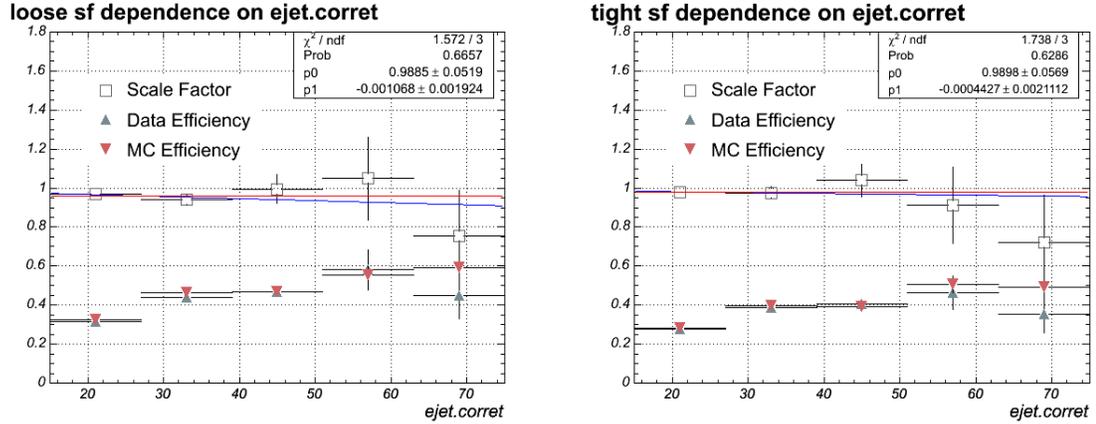


Figure 5.14: Dependence of the loose (left) and tight (right) scale factors and efficiencies on the electron jet corrected E_T . Errors are statistical (data and MC) only, and linear and constant fits are shown.

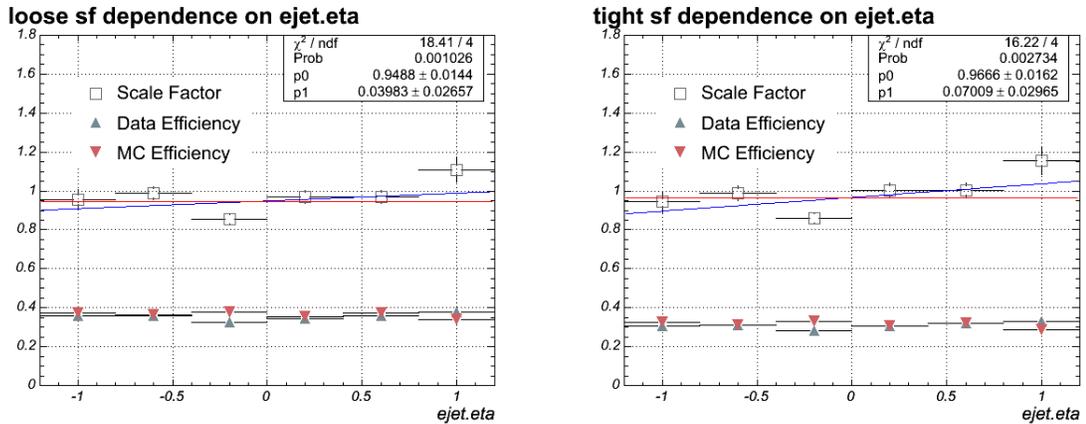


Figure 5.15: Dependence of the loose (left) and tight (right) scale factors and efficiencies on the electron jet η . Errors are statistical (data and MC) only, and linear and constant fits are shown.

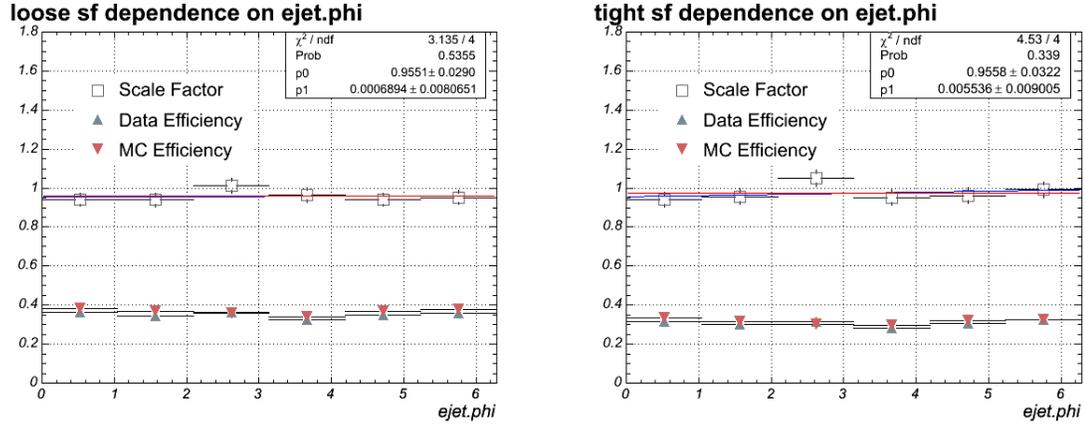


Figure 5.16: Dependence of the loose (left) and tight (right) scale factors and efficiencies on the electron jet ϕ . Errors are statistical (data and MC) only, and linear and constant fits are shown.

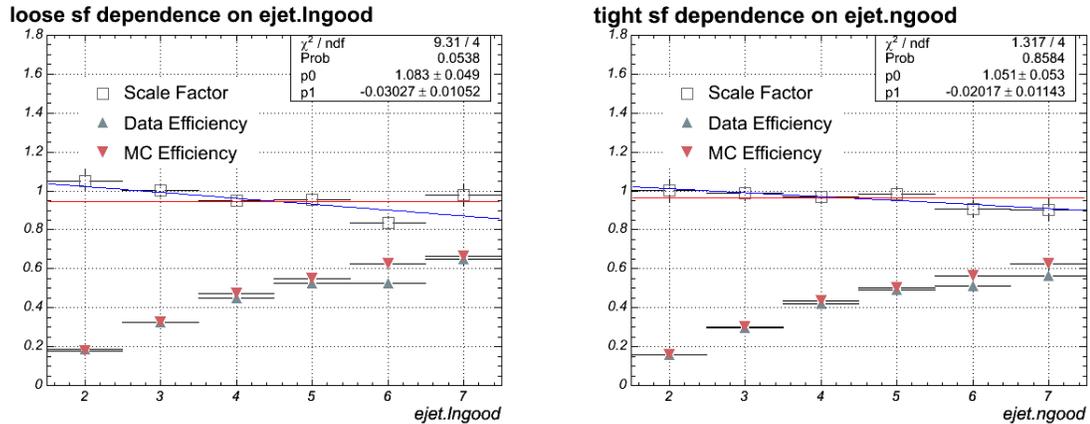


Figure 5.17: Dependence of the loose (left) and tight (right) scale factors and efficiencies on the electron jet good track multiplicity. Errors are statistical (data and MC) only, and linear and constant fits are shown.

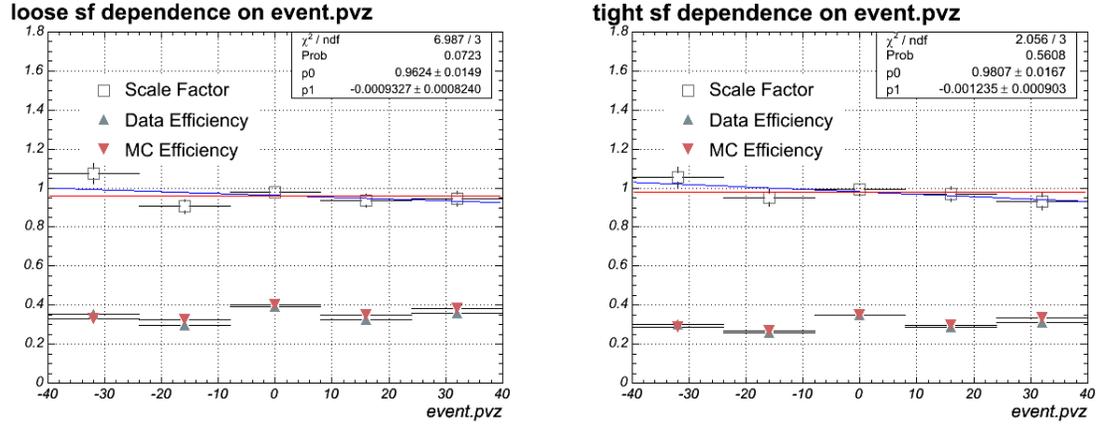


Figure 5.18: Dependence of the loose (left) and tight (right) scale factors and efficiencies on the primary vertex z . Errors are statistical (data and MC) only, and linear and constant fits are shown.

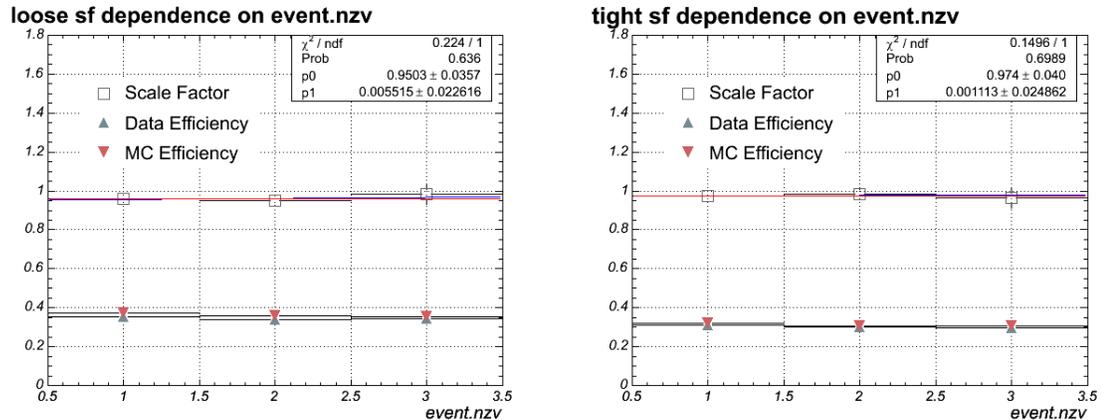


Figure 5.19: Dependence of the loose (left) and tight (right) scale factors and efficiencies on the number of z vertices. Errors are statistical (data and MC) only, and linear and constant fits are shown.

5.2.2 Muon Method

Overview

We use a second method to evaluate SF in dijet events with muons. These events are kinematically and topologically similar to those in the electron sample described above, but are purer in heavy flavor jets since there are no conversions. We fit the distribution of p_{Trel} , the muon momentum perpendicular to the rest of the jet, with Monte Carlo templates for bottom and non-bottom jets, and we determine the heavy flavor content of muon jets before (N_{total}^b) and after (N_{tag}^b) requiring a *b*-tag [63]. Because *B* hadrons are more massive than charm or light flavor, muons from *B* decays will tend to have a larger p_{Trel} . The *b*-tagging efficiency in this sample is:

$$\epsilon^b = \frac{N_{tag}^b}{N_{total}^b} \quad (5.7)$$

As with the electron sample, we define the scale factor SF to be the ratio of efficiencies in data and Monte Carlo ($\frac{\epsilon^{data}}{\epsilon^{MC}}$), and the Monte Carlo efficiency is determined from the generator information.

Event Selection

The data sample is taken from a muon trigger requiring an 8-GeV/*c* muon in the central detectors (CMUP). Much like the electron method, we require the muon to be inside a jet with $E_T > 15$ GeV, and we demand the presence of a second “away” jet. Muons are identified according to the selection listed in Section 4.2, except that the p_T requirement on the muon is lowered to 9 GeV/*c*, and we require the muon isolation I be ≥ 0.1 . To enhance the *b*-jet purity of the sample, the away jet is required to have

a loose SECVTX *b*-tag with an invariant mass greater than $1.5 \text{ GeV}/c^2$.

The Monte Carlo simulation is identical to that used in the electron method, though filtered for a muon at generator level. A comparison of some kinematic quantities in data and Monte Carlo is shown in Figure 5.20.

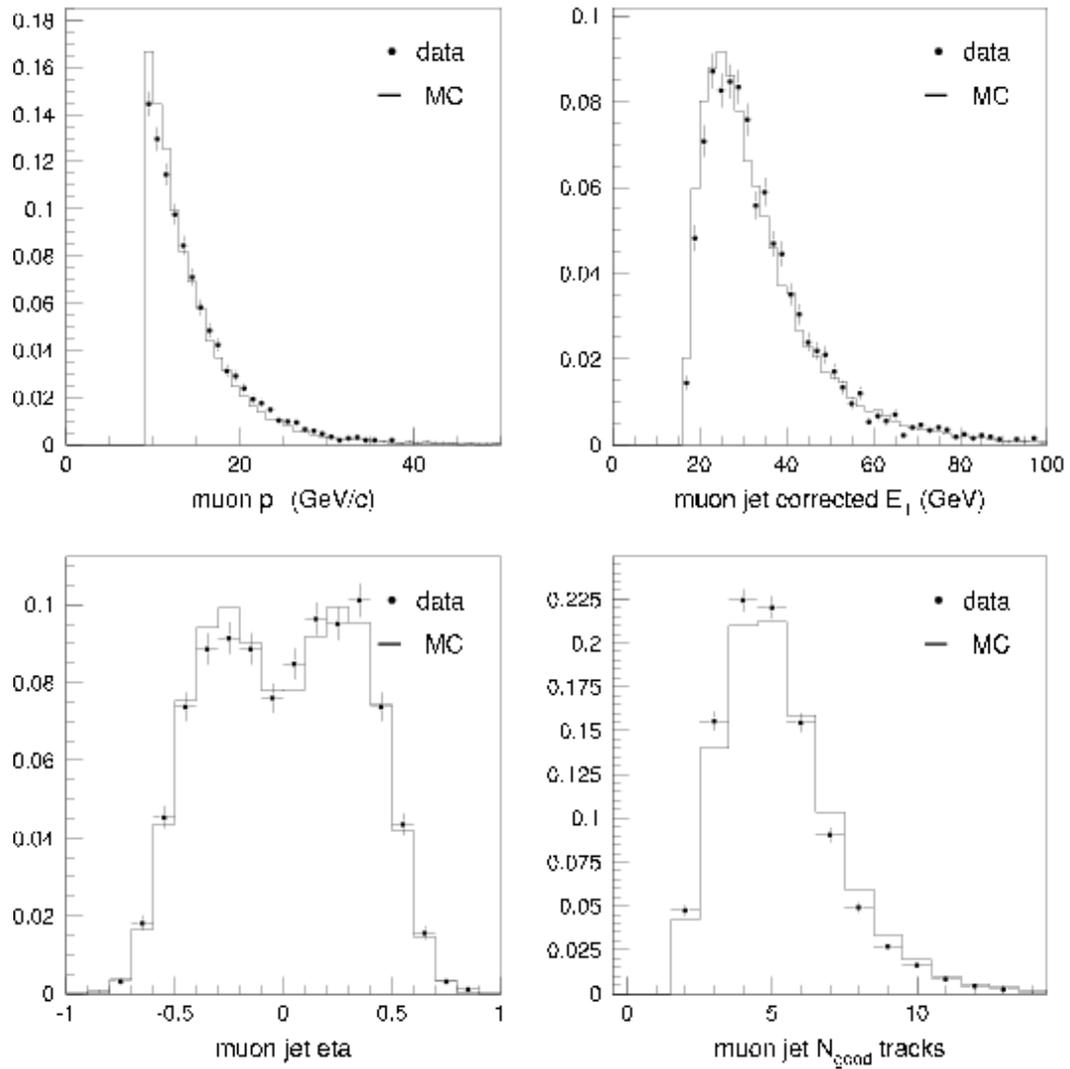


Figure 5.20: Comparison of kinematic quantities between data and PYTHIA Monte Carlo in the 8-GeV muon sample. The muon p_T (top left, mislabeled), muon jet E_T (top right), and muon η (bottom left) agree well, while the tagged jet track multiplicity (bottom right) shows fewer tracks in data than in Monte Carlo.

Fits

The p_{Trel} variable discriminates well between bottom and light flavor jets, but not between charm and light flavor. We therefore include only two components in the fits, *b* and non-*b*. The template for non-*b*'s is determined in four ways: charm jets in simulation, light flavor jets in simulation, jets in data in which the muon fails the CMUP matching requirements, and jets in data containing no displaced (pass 1) tracks. For bottom templates, the shapes both before and after requiring a *b*-tag are taken directly from the simulation. The four non-*b* templates are shown in Figure 5.21. The pretag and tag fits are shown for the loose and tight tagger in Figures 5.22 and 5.23.

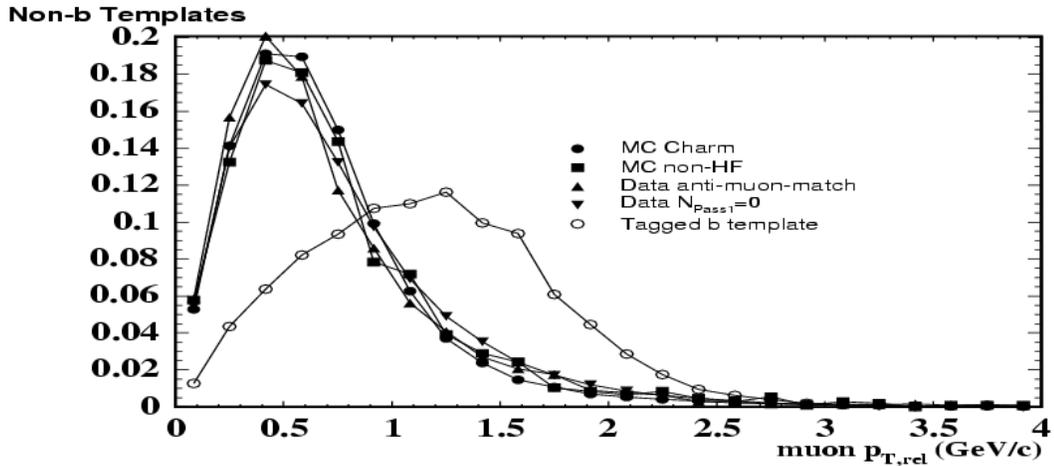


Figure 5.21: Non-*b* and bottom templates for muon p_{Trel} . Each of the four non-*b* shapes is used, and the spread is taken as a systematic error. The tagged *b* shape is shown for comparison.

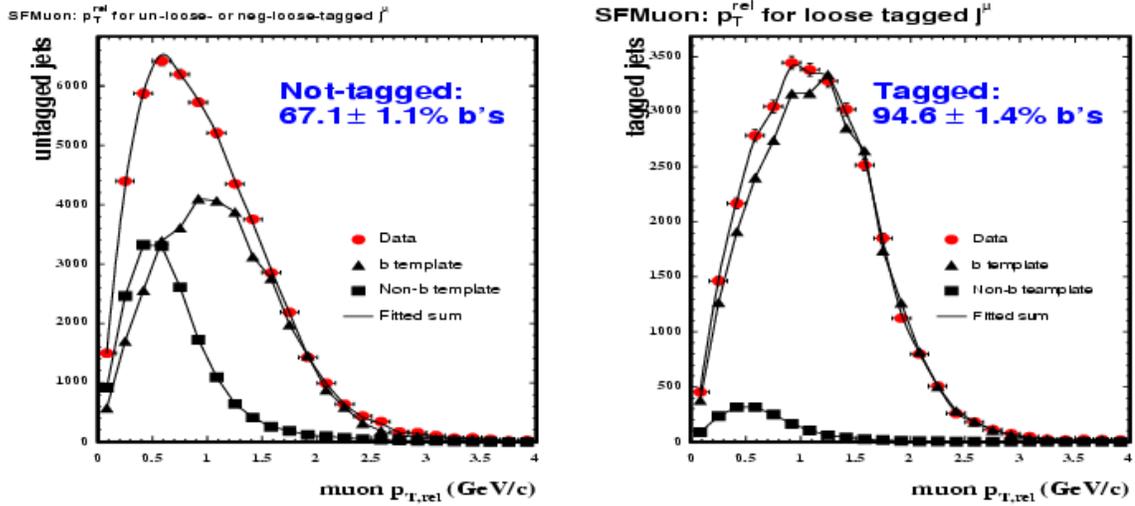


Figure 5.22: Muon $p_{T,rel}$ fits for the loose tagger in the untagged (left) and tagged (right) samples.

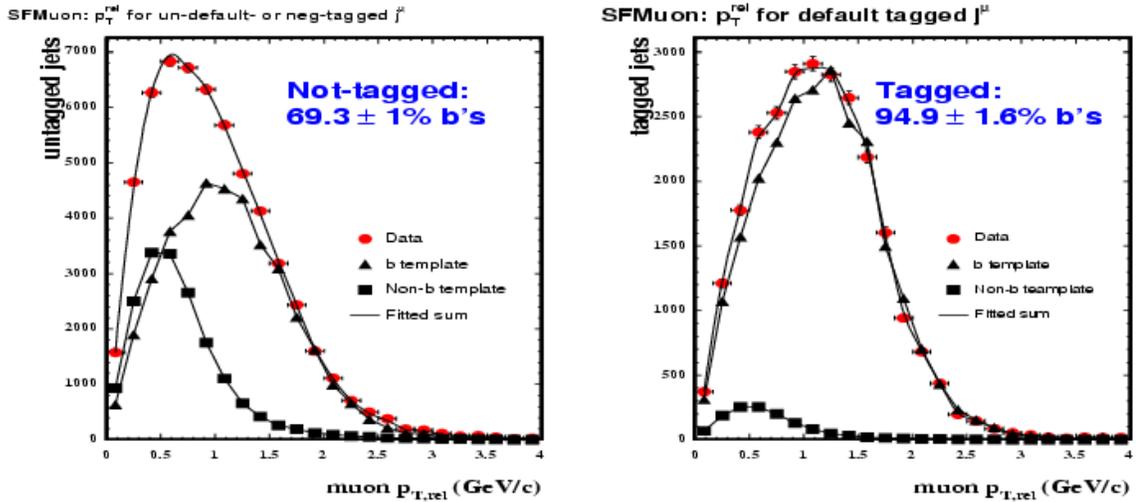


Figure 5.23: Muon $p_{T,rel}$ fits for the tight tagger in the untagged (left) and tagged (right) samples.

Systematic Errors

Like the electron method above, the scale factor measurement is subject to several systematic errors, due to both modeling limitations and method biases. The largest

sources are listed here:

- **Non-*b* Templates:** As mentioned above, the p_{Trel} variable does not distinguish charm from light flavor, and there are multiple choices for non-*b* templates, both from simulation and from data. We perform the fit with four different templates (shown in Figure 5.21), use the average as the central value, and take the largest deviation as a systematic uncertainty.
- **Jet Direction:** The p_{Trel} of the muon is calculated relative to the direction of the jet, after correcting the jet's energy for the muon's expected deposit in the calorimeter, typically 2 GeV. Without this correction, one may also calculate the direction of the jet from the four-momenta of the tracks. This changes the p_{Trel} templates and the fit results, and we take the difference as a systematic uncertainty.
- **Tag Bias in Templates:** We use different bottom templates for the tagged and untagged fits, both of which are taken from the Monte Carlo. If the observed differences between these templates, shown in Figure 5.24, are not the same in the data, the fit results will be biased. We repeat the fit using the same (untagged) template in both fits, and take the difference as a systematic uncertainty.
- **Semileptonic Decay Model:** The shape of the p_{Trel} templates is sensitive to having the correct fragmentation and branching fractions for bottom and charm hadrons. We compare the true Monte Carlo generated fractions with previous experimental results, and we re-weight to correct for the small discrepancies

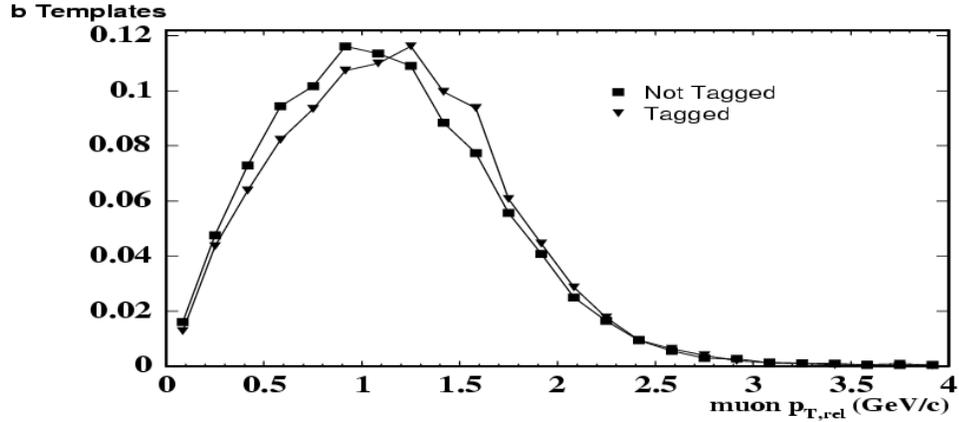


Figure 5.24: Bottom templates used in the muon $p_{T,rel}$ fits, taken directly from the PYTHIA simulation.

seen. The adjusted templates are used in the fit, and the difference is taken as a systematic error.

- Track Multiplicity Bias:** If the scale factor SF is different for semileptonic and hadronic B decays, the measurement in the muon sample will be biased. We assume that the most important difference between the two types of decays is the track multiplicity (hadronic decays will have more tracks), and we re-weight the efficiency versus number of tracks with the distribution for hadronic B decays. The ratio of the re-weighted efficiencies in data and Monte Carlo is an adjusted scale factor, and we use the difference from the central value as an estimate of this uncertainty.

A summary of the systematic errors is shown in Table 5.4

Source	Loose (%)	Tight (%)
Non- <i>b</i>	0.2	0.2
Jet Direction	0.3	0.7
Tag/Untag <i>b</i> -Jet Model	1.1	1.4
Semi-Leptonic Decay	0.8	0.8
Track Multiplicity	0.6	1.7
Total	1.5	2.5

Table 5.4: Summary of systematic errors on the scale factor for both taggers. All numbers are quoted as relative percentage of the scale factor. Systematics which are relevant only to the application of the scale factor in other samples, such as the jet energy dependence, will be evaluated separately in Section 5.2.3.

Results

Including systematic uncertainties, we measure the following scale factors:

$$SF_{loose} = 0.944 \pm 0.021(stat) \pm 0.014(syst) \quad (5.8)$$

$$SF_{tight} = 0.932 \pm 0.022(stat) \pm 0.024(syst) \quad (5.9)$$

Results are consistent with those from the electron method, and Figures 5.25 through 5.28 show the dependence of the scale factor SF and measured efficiency on various jet and event properties. Only statistical errors are included, and a linear fit to the scale factor is shown as well [63].

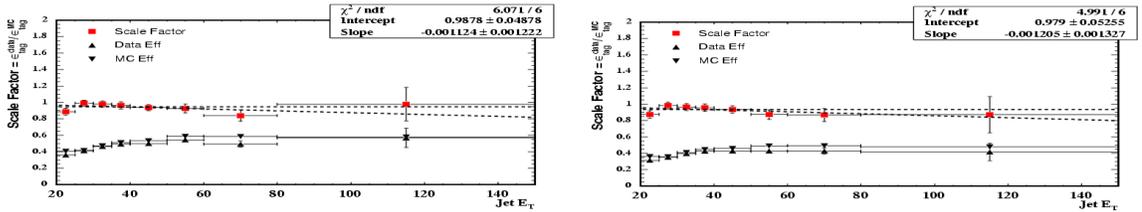


Figure 5.25: Dependence of the loose (left) and tight (right) scale factors and efficiencies on the muon jet E_T . Errors are statistical (data and MC) only.

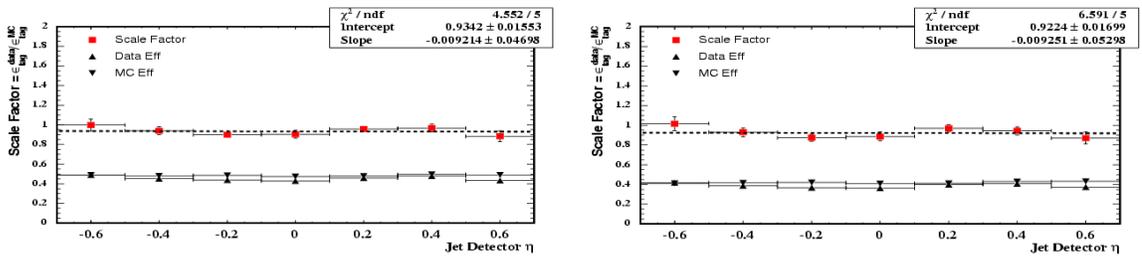


Figure 5.26: Dependence of the loose (left) and tight (right) scale factors and efficiencies on the muon jet η . Errors are statistical (data and MC) only.

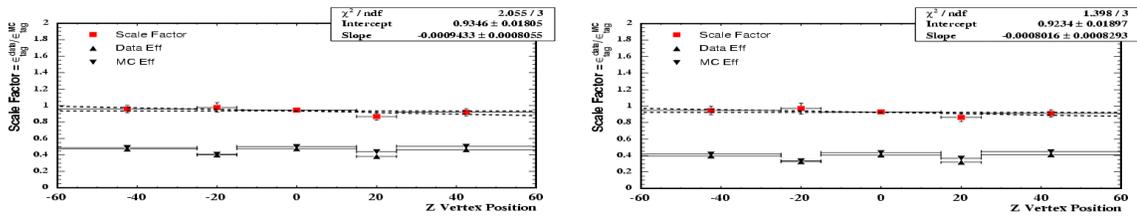


Figure 5.27: Dependence of the loose (left) and tight (right) scale factors and efficiencies on the primary vertex z position. Errors are statistical (data and MC) only.

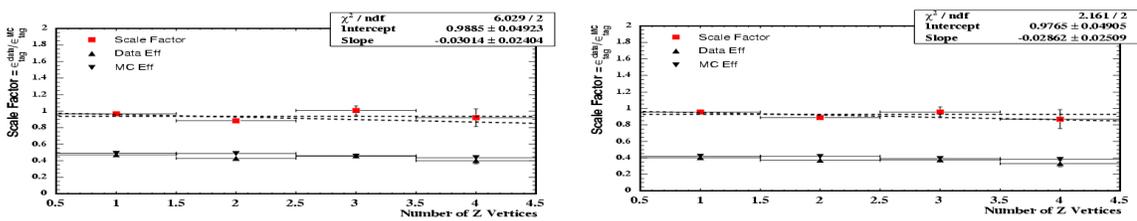


Figure 5.28: Dependence of the loose (left) and tight (right) scale factors and efficiencies on the number of reconstructed $p\bar{p}$ vertices. Errors are statistical (data and MC) only.

5.2.3 Combined Result

The two lepton-based results can be combined to yield a single best SF result for each tagger. We assume the systematics are 100% uncorrelated except in the case of the semi-leptonic biases, and find the best central values to be 0.95 for both the tight and loose SECVTX taggers. More importantly, we combine the statistics of the two samples to evaluate the jet E_T dependence of SF , a potentially large systematic error. Bottom jets in $t\bar{t}$ events have typical transverse energies of 50 GeV, while the measured scale factors are poorly constrained at those energies (see Figure 5.14, for example). By taking the combined, weighted average as a function of jet energy and re-weighting according to the b -jet energy distribution from $t\bar{t}$ Monte Carlo, we derive an additional common E_T dependence systematic to reflect how poorly-constrained the efficiency is at high p_T , a 2.0% effect for the loose tagger and 2.8% for the tight tagger [64]. Altogether, then, we measure scale factors of:

$$SF_{loose} = 0.95 \pm 0.01(stat) \pm 0.05(syst) \quad (5.10)$$

$$SF_{tight} = 0.95 \pm 0.01(stat) \pm 0.04(syst) \quad (5.11)$$

A scale factor SF consistent with unity indicates that the Monte Carlo is modeling the b -tagging efficiency well, but the 5% difference is still an important correction in the $t\bar{t}$ cross section measurement. We use these corrections are used to weight Monte Carlo-tagged b -jets and determine the expected b -tagging efficiency in data. Since neither the electron nor the muon method is sensitive to differences between the charm and bottom scale factor, we simply use the same central value for charm, but with a factor of three larger error.

The E_T and η dependence of the *b*-tagging efficiency, after correcting for the scale factor, are shown in Figure 5.29.

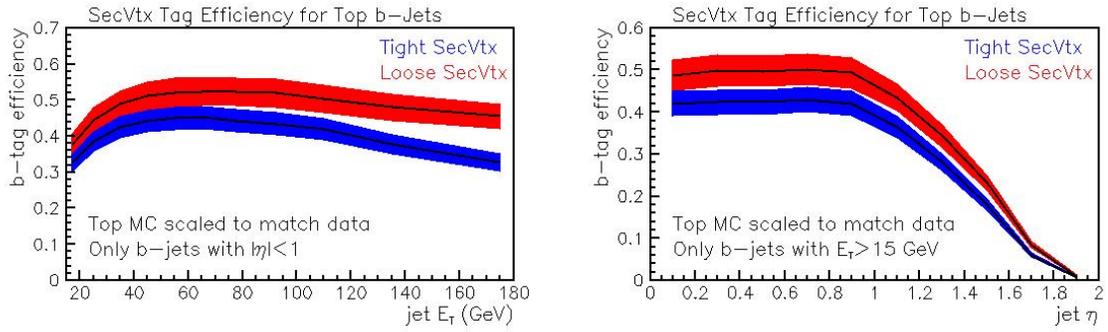


Figure 5.29: Tagging efficiency for *b*-jets in $t\bar{t}$ events, after correcting for the data-to-Monte Carlo scale factor SF , as a function of jet E_T (left) and jet η (right).

5.3 Light Flavor Tagging

Unlike heavy flavor *b*-tagging efficiencies, which are accurate in Monte Carlo to within 5%, the mistag rate must be derived directly in data. In particular, the light flavor tag rate is very sensitive to track mis-measurement, multiple $p\bar{p}$ interactions, and the primary vertex resolution, none of which is perfectly modeled in Monte Carlo. In some samples, the discrepancy in the mistag rates in data and Monte Carlo can be as large as 50%, with data tending to have more mistags.

We determine the average light flavor tag rate in data in two steps: we first measure a parameterized negative tag rate (R_{mistag}^-), the *mistag matrix* [65], in a large data sample of multijet events, and we then derive a multiplicative correction to this tag rate, $\alpha\beta$, which scales from the total negative tag rate to the average positive light flavor rate in that same calibration sample [61].

To be precise:

$$R_{mistag}^- = \frac{N_{light}^- + N_{heavy}^-}{N_{light}^{pre} + N_{heavy}^{pre}} \quad (5.12)$$

$$\alpha = \frac{N_{light}^+}{N_{light}^- + N_{heavy}^-} \quad (5.13)$$

$$\beta = \frac{N_{light}^{pre} + N_{heavy}^{pre}}{N_{light}^{pre}} \quad (5.14)$$

$$\alpha\beta R_{mistag}^- = \frac{N_{light}^+}{N_{light}^{pre}} \quad (5.15)$$

The N 's are numbers of jets in data, where the superscript corresponds to the tag requirement (*pre* is used for *pretag*, before any tag requirement) and the subscript distinguished between light flavor and heavy flavor jets. The term R_{mistag}^- is evaluated as a function of the jet E_T , η , number of good SECVTX tracks (see Section 5.1 for a definition), primary vertex z position, number of reconstructed z vertices, and the sum of all the jet E_T 's in the event (ΣE_T). The mistag asymmetries α and β are evaluated only with respect to jet E_T .

Jet Selection

We evaluate the mistag matrix and asymmetry in a sample comprising four jet triggers, **Jet20**, **Jet50**, **Jet70**, and **Jet100**. Each of these triggers requiring a single calorimeter cluster at Level 2 above a threshold, 20 GeV for **Jet20**, 50 GeV for **Jet50**, and so on. We select all jets with $|\eta| < 2.4$ and $E_T > 10$ GeV, accepting jets both above and below the trigger threshold. The jet E_T 's in this case are *not* corrected as in Section 4.3, we instead use the raw jet E_T . Most events are dijets, though there are often additional jets at low p_T . The total tag rates for each available dataset are shown in Table 5.5. Underlying distributions for each sample are shown in Figures 5.30 to 5.33.

Parameterization

We select eight bins in E_T , nine in N_{trk} , four in ΣE_T , and four in $|\eta|$. We bin the number of z vertices and primary vertex z in five bins each, for a total of 28,800 [65]. The bin boundaries are shown in Table 5.3. Since the tight and loose algorithms select tracks differently, a jet may fall into different bins in the tight and loose matrices.

All Jets	
Loose Negative Rate	0.01382 ± 0.00001
Tight Negative Rate	0.00563 ± 0.00001
Jet20	
Loose Negative Rate	0.00366 ± 0.00001
Tight Negative Rate	0.00137 ± 0.00001
Jet50	
Loose Negative Rate	0.01187 ± 0.00002
Tight Negative Rate	0.00473 ± 0.00001
Jet70	
Loose Negative Rate	0.01729 ± 0.00003
Tight Negative Rate	0.00702 ± 0.00002
Jet100	
Loose Negative Rate	0.02402 ± 0.00003
Tight Negative Rate	0.01003 ± 0.00002

Table 5.5: Summary of tight and loose tag rates for the various components of the mistag matrices.

The content of each bin is the observed negative tag rate for such jets, with binomial statistical errors; these errors are in general small.

E_T (GeV)	N_{trk}	ΣE_T (GeV)	$ \eta $	$N_{Z_{Vertices}}$	PV Z (cm)
0	2	0	0	1	-25
15	3	80	0.4	2	-10
22	4	140	0.8	3	10
30	5	220	1.1	4	25
40	6			6	
60	7				
90	8				
130	10				
	13				

Table 5.6: Low edges for the bins used in the tag matrices. The highest bins are always inclusive, and the z binning is inclusive on both ends.

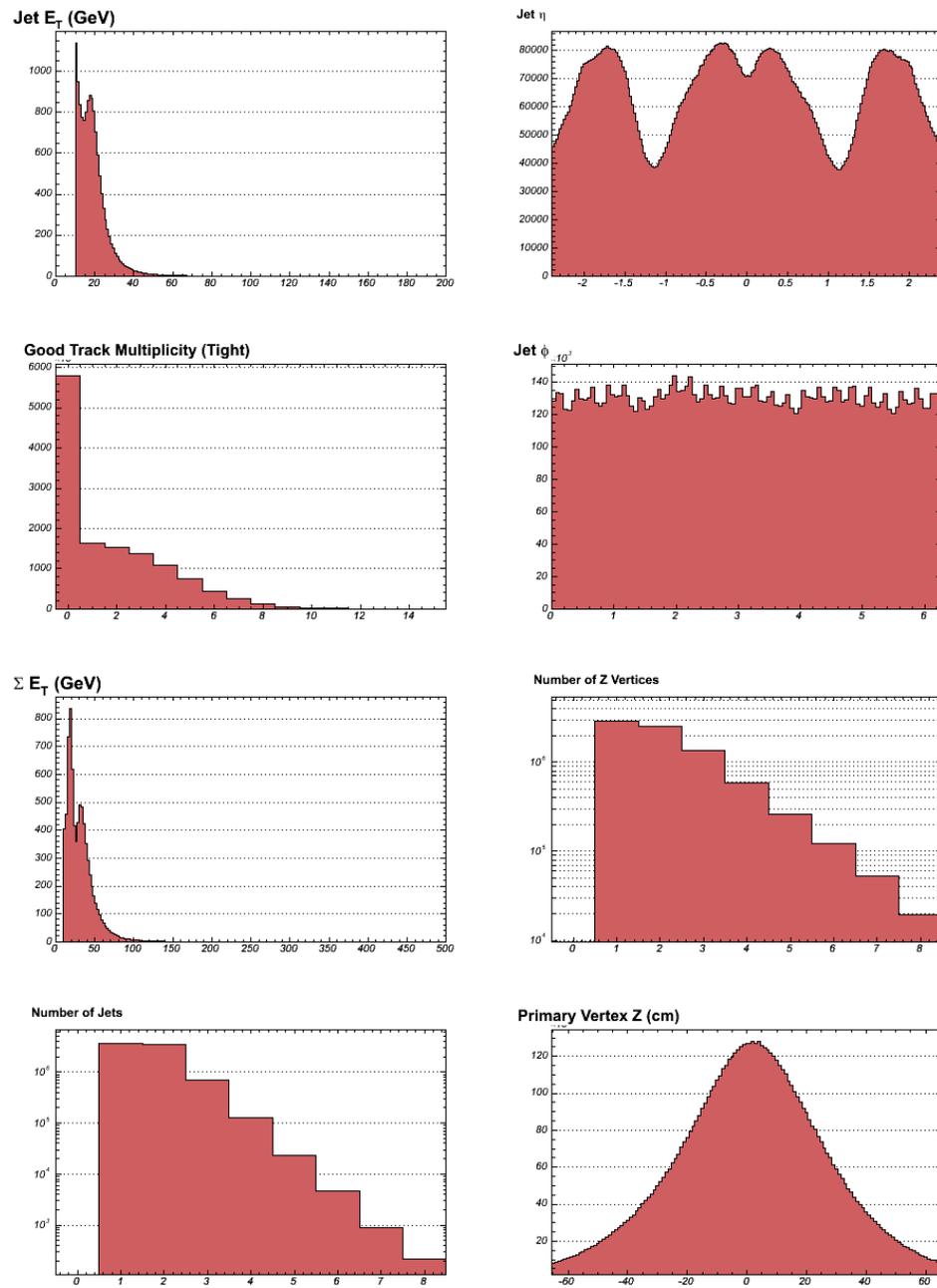


Figure 5.30: Kinematic distributions in the Jet20 dataset, for all events containing at least one 10 GeV jet.

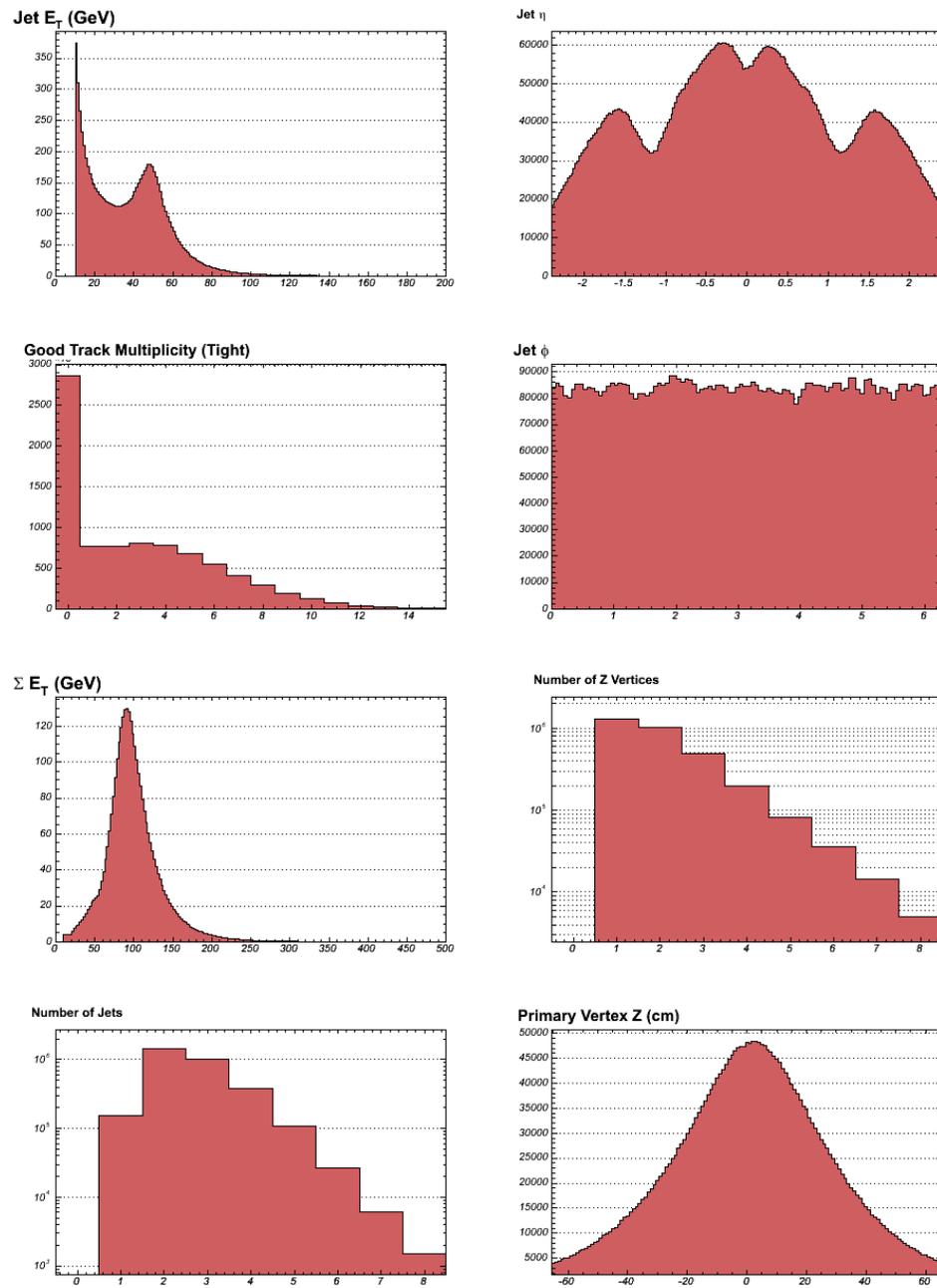


Figure 5.31: Kinematic distributions in the Jet50 dataset, for all events containing at least one 10 GeV jet.

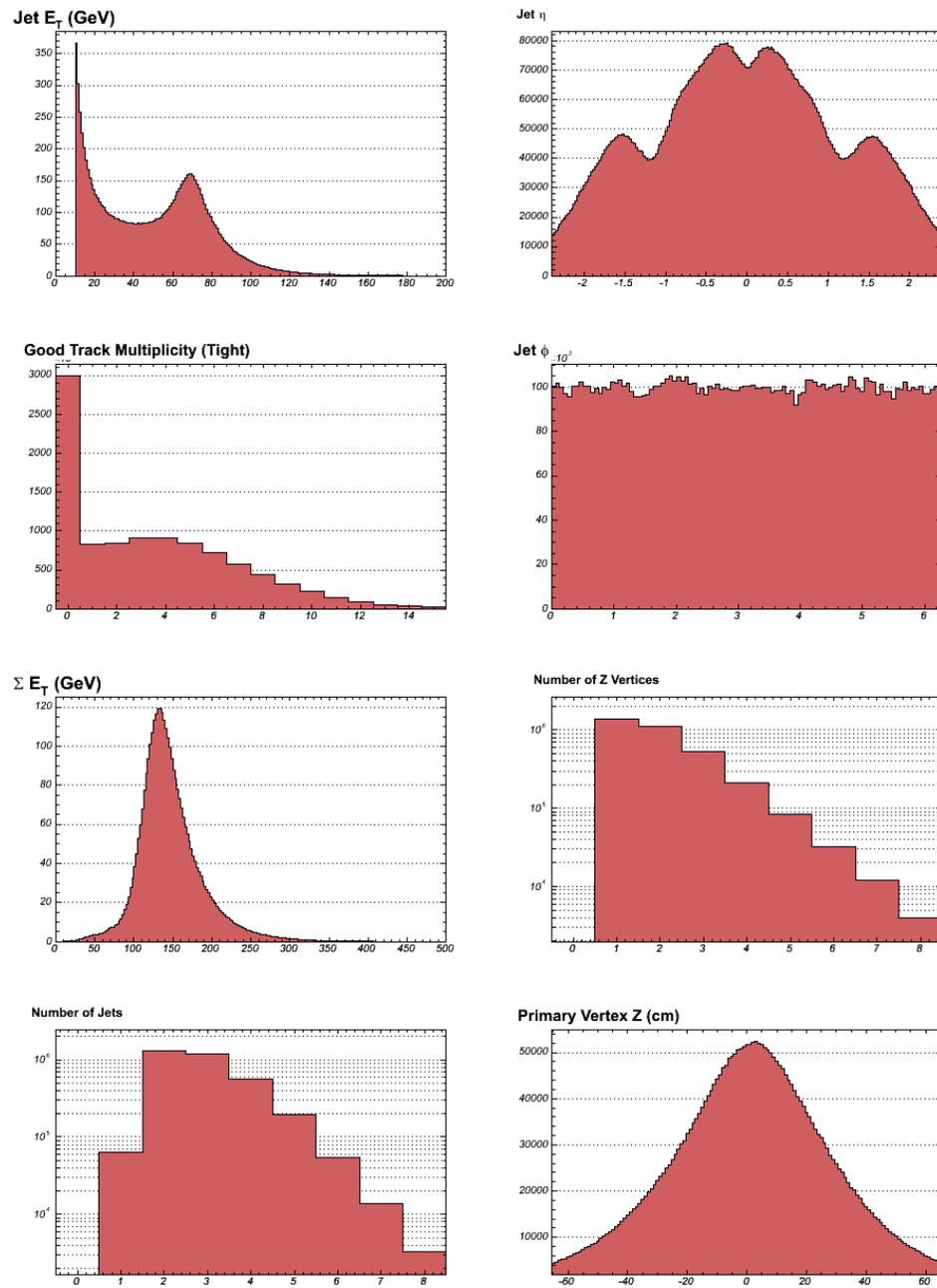


Figure 5.32: Kinematic distributions in the Jet70 dataset, for all events containing at least one 10 GeV jet.

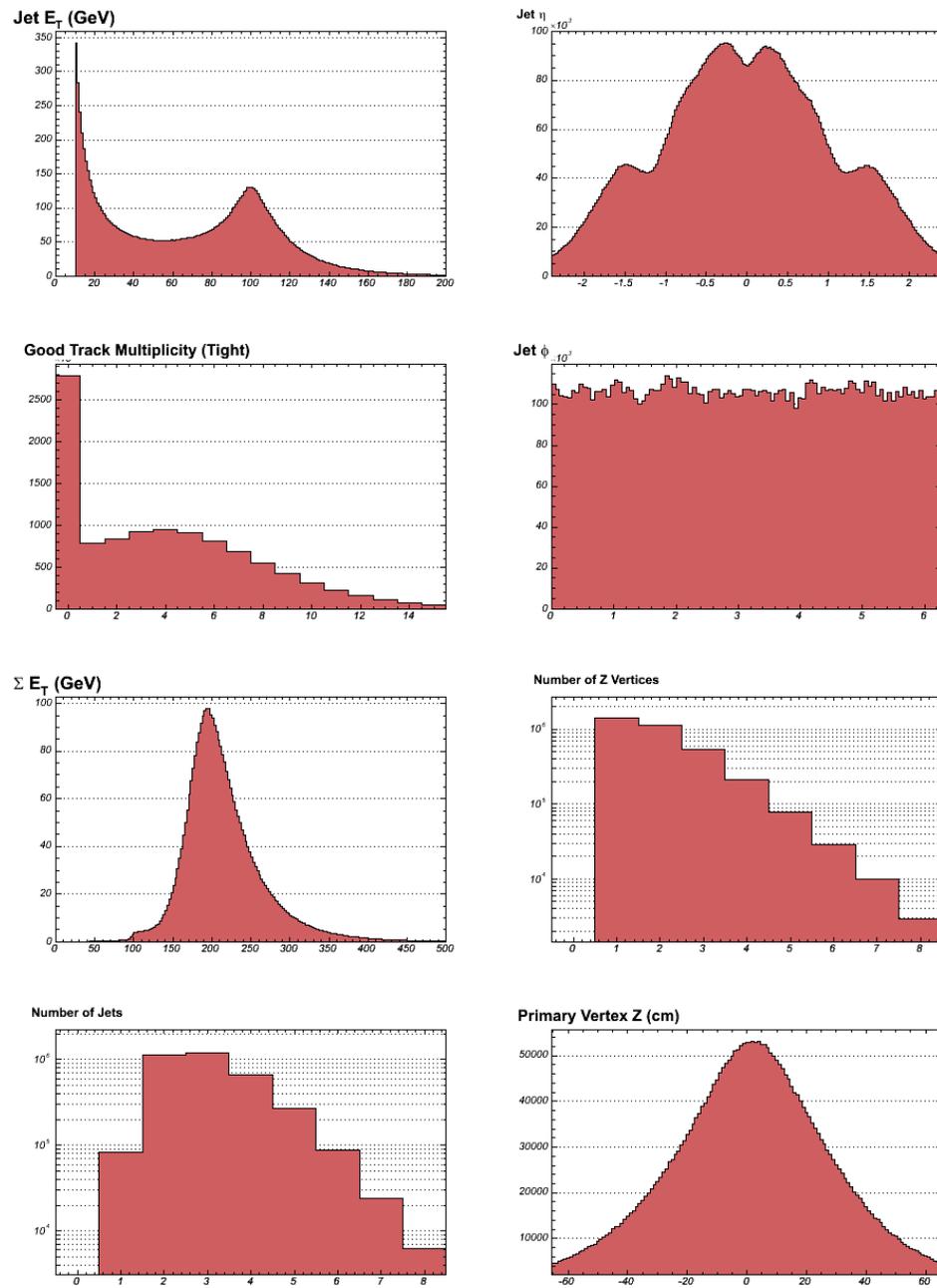


Figure 5.33: Kinematic distributions in the Jet100 dataset, for all events containing at least one 10 GeV jet.

5.3.1 Mistag Matrix Performance

We test the validity of the parameterization by comparing the tag rates predicted by it with those observed in the data. To insure that the jets in the matrix and those in the control samples are statistically independent, we produce a mistag matrix using the same prescription with jets in even-numbered events, and we apply it to jets in odd-numbered events. We compare the observed and predicted tag rates as functions of a number of variables, the six kinematic variables used in the default matrix, as well as the instantaneous luminosity, the jet multiplicity and the event track multiplicity.

The self-consistency cross-check plots using the full data sample are shown in Figures 5.34 through 5.42. As expected, the tag rates increase as the E_T , ΣE_T , and number of tracks increase, both positive and negative. The silicon structure is visible in Figure 5.39, since tracks have fewer silicon hits at the barrel boundaries. Figure 5.38 shows the dependence of the tag rate on the number of reconstructed z vertices; the COT tracking efficiency drops when the occupancy is high, resulting in fewer tracks for finding displaced vertices.

We expect perfect agreement in matrix variables, provided the binning is identical in the plots and the parameterization. (This is intentionally not the case for the primary vertex z plots, since the structure of the silicon detector can only be seen with more bins.) For quantities not used in the parameterization, we expect only the *average* predicted and observed tag rates to be the same; the dependence of the tag rate will be reproduced only when the quantity is highly correlated with the variables used in the parameterization.

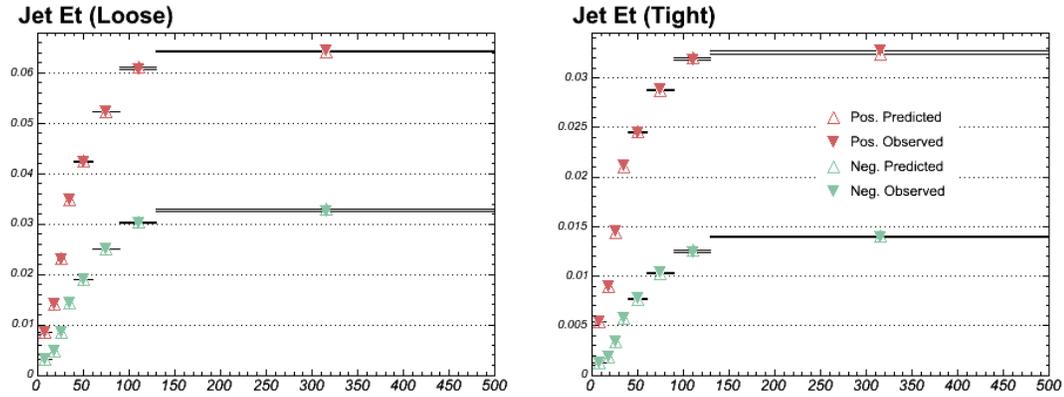


Figure 5.34: Observed and predicted tag rates for the loose (left) and tight (right) SECVTX taggers as a function of jet E_T . The matrix is made from events with even run number and run on odd-numbered events.

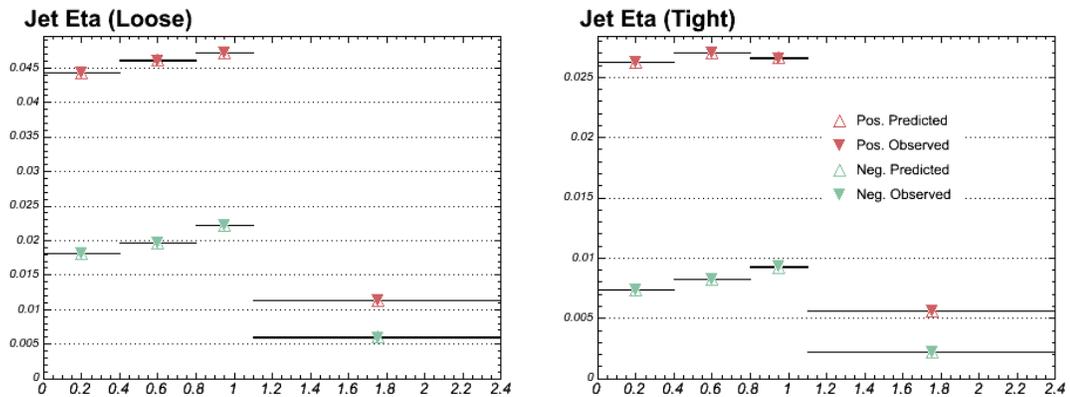


Figure 5.35: Observed and predicted tag rates for the loose (left) and tight (right) SECVTX taggers as a function of jet η . The matrix is made from events with even run number and run on odd-numbered events.

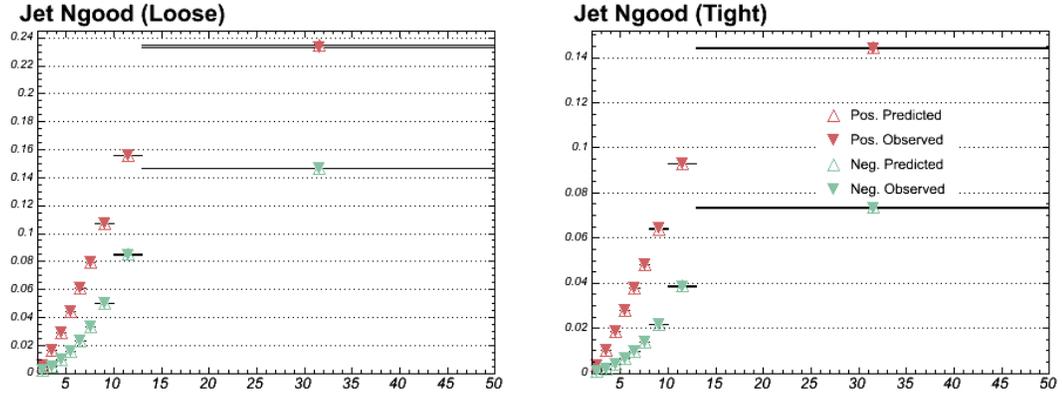


Figure 5.36: Observed and predicted tag rates for the loose (left) and tight (right) SECVTX taggers as a function of good SECVTX track multiplicity. The matrix is made from events with even run number and run on odd-numbered events.

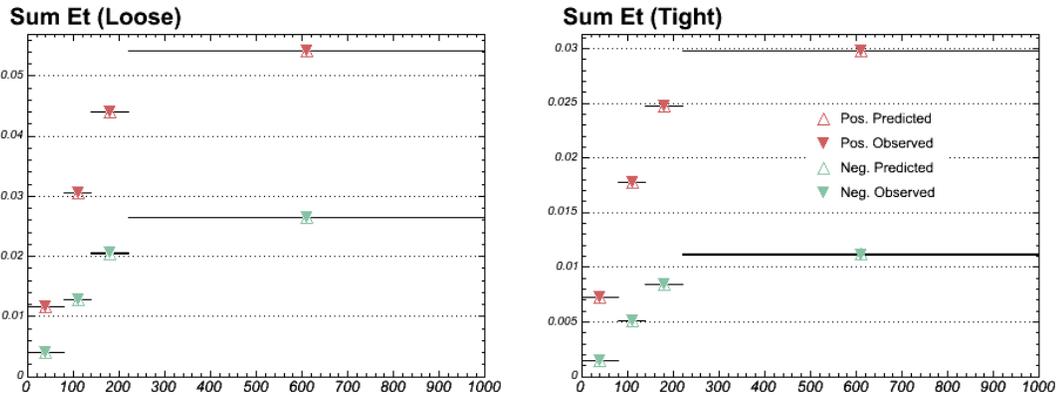


Figure 5.37: Observed and predicted tag rates for the loose (left) and tight (right) SECVTX taggers as a function of event ΣE_T . The matrix is made from events with even run number and run on odd-numbered events.

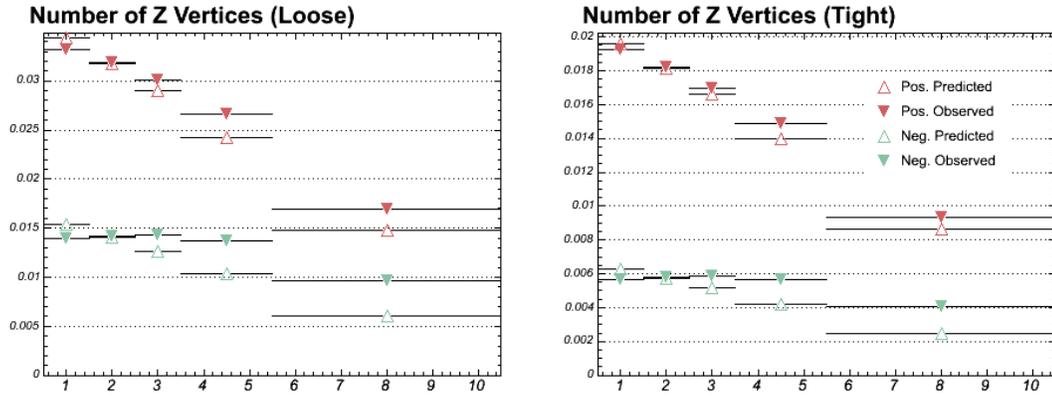


Figure 5.38: Observed and predicted tag rates for the loose (left) and tight (right) SECVTX taggers as a function of number of z vertices. The matrix is made from events with even run number and run on odd-numbered events.

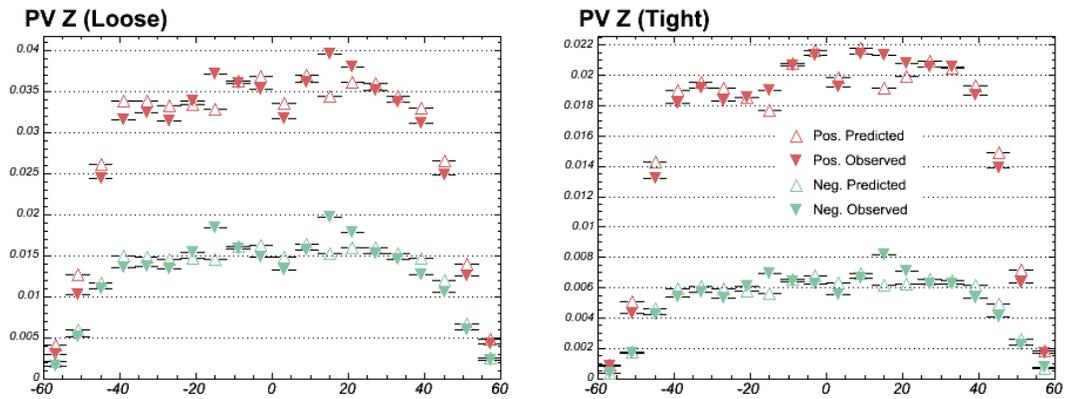


Figure 5.39: Observed and predicted tag rates for the loose (left) and tight (right) SECVTX taggers as a function of primary vertex z . The matrix is made from events with even run number and run on odd-numbered events.

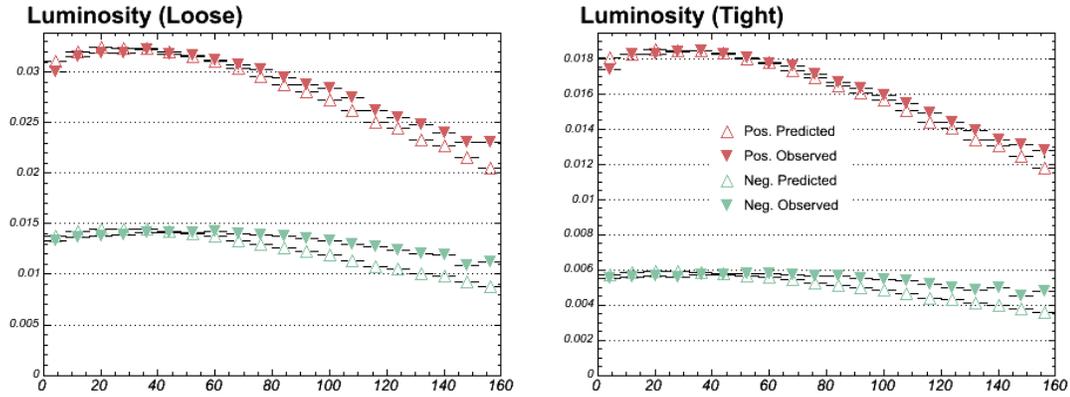


Figure 5.40: Observed and predicted tag rates for the loose (left) and tight (right) SECVTX taggers as a function of instantaneous luminosity. The matrix is made from events with even run number and run on odd-numbered events.

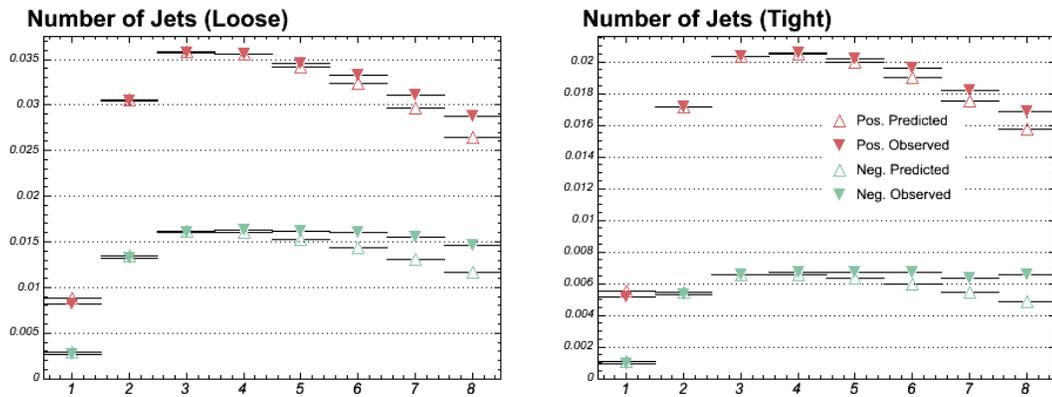


Figure 5.41: Observed and predicted tag rates for the loose (left) and tight (right) SECVTX taggers as a function of the number of jets in the event. The matrix is made from events with even run number and run on odd-numbered events.

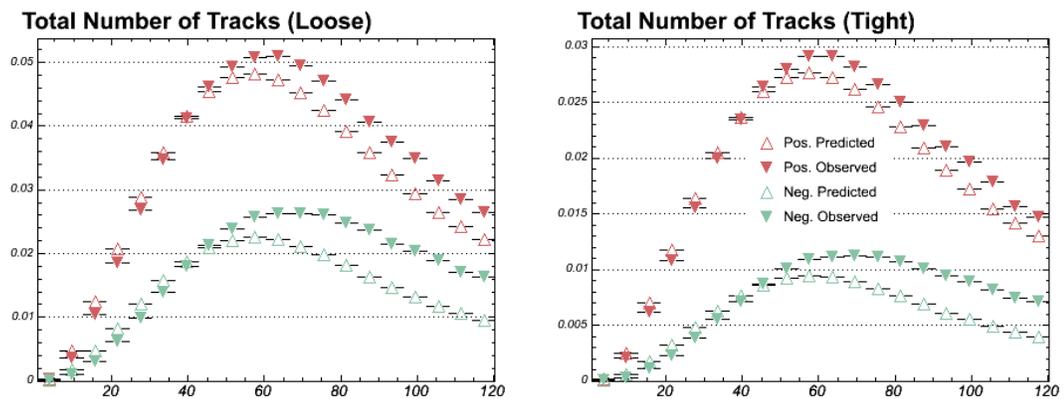


Figure 5.42: Observed and predicted tag rates for the loose (left) and tight (right) SECVTX taggers as a function of total number of tracks in the event. The matrix is made from events with even run number and run on odd-numbered events.

5.3.2 Systematic Uncertainties

The systematic uncertainty on the matrix comprises three sources:

- **Sample Dependence:** We apply the *even-event* matrix to the individual jet samples, as well as to events on a four-jet trigger, and take the largest deviation between the total observed and predicted rates as a systematic error.
- **ΣE_T Bias:** The matrix is sensitive to fluctuations in the clustered jet energy in the event through the ΣE_T variable, which can happen either randomly (up or down) or systematically (down), when a jet falls on a calorimeter crack. We scale the ΣE_T up and down by half the mean jet energy (depending on the trigger sample), and take the deviation between predicted and observed rates in the “scaled-up” check as a systematic error. This uncertainty also covers rate differences in samples where not all the event energy is in jets, for instance those with leptons and missing E_T .
- **Trigger Bias:** We apply the matrix separately to the trigger jets in each event, and take the deviation from the observed rate as a systematic uncertainty. The “trigger jet” is defined to be the jet above threshold nearest the L2 trigger cluster.

A summary of the total tag rates for all checks is shown in Table 5.7. Numbers in boldface are used to determine a systematic uncertainty, all of which are summarized in Table 5.8. The trigger bias rates (observed and predicted) were evaluated for taggable jets (with two or more tracks) only, but the taggability requirement factors out in the ratio.

Sample	Observed Rate (%)	Predicted Rate (%)	Ratio
Loose Negative			
All	0.01397 ± 0.00002	0.01394 ± 0.00000	0.99808 ± 0.00116
Jet20	0.00370 ± 0.00001	0.00382 ± 0.00000	1.03267 ± 0.00417
Jet50	0.01193 ± 0.00003	0.01215 ± 0.00001	1.01876 ± 0.00286
Jet70	0.01748 ± 0.00004	0.01750 ± 0.00001	1.00068 ± 0.00213
Jet100	0.02443 ± 0.00004	0.02400 ± 0.00001	0.98251 ± 0.00169
ΣE_T	0.01311 ± 0.00005	0.01323 ± 0.00001	1.00908 ± 0.00378
ΣE_T Up	0.01397 ± 0.00002	0.01439 ± 0.00000	1.03054 ± 0.00120
ΣE_T Down	0.01397 ± 0.00002	0.01346 ± 0.00000	0.96415 ± 0.00112
Trigger	0.02184 ± 0.00002	0.02240 ± 0.00002	1.02565 ± 0.00158
Tight Negative			
All	0.00568 ± 0.00001	0.00568 ± 0.00000	0.99925 ± 0.00180
Jet20	0.00138 ± 0.00001	0.00144 ± 0.00000	1.03927 ± 0.00683
Jet50	0.00475 ± 0.00002	0.00484 ± 0.00000	1.01810 ± 0.00450
Jet70	0.00709 ± 0.00002	0.00712 ± 0.00000	1.00390 ± 0.00333
Jet100	0.01018 ± 0.00003	0.01001 ± 0.00000	0.98319 ± 0.00260
ΣE_T	0.00532 ± 0.00003	0.00541 ± 0.00000	1.01616 ± 0.00594
ΣE_T Up	0.00568 ± 0.00001	0.00591 ± 0.00000	1.03941 ± 0.00188
ΣE_T Down	0.00568 ± 0.00001	0.00545 ± 0.00000	0.95889 ± 0.00173
Trigger	0.00877 ± 0.00002	0.00903 ± 0.00002	1.02900 ± 0.00265

Table 5.7: Observed and predicted tag rates for the various subsamples and systematic studies. The matrix is made from even-numbered events and run on odd-numbered events. Numbers in boldface are used to evaluate systematic errors. Note that the rates for the trigger bias systematic are for taggable jets ($N_{trk} \geq 2$), but the taggability requirement factors out in the ratio.

Systematic Source	Loose SECVTX	Tight SECVTX
Sample Dependence (%)	3.3	3.9
ΣE_T Dependence (%)	3.1	3.9
Trigger Bias (%)	2.6	2.9
Total (%)	5	6

Table 5.8: Summary of relative systematic uncertainties for the mistag matrices.

5.3.3 Mistag Asymmetry

Method

As mentioned above in Section 5.3, the negative tag rate is only an approximation of the positive light flavor tag rate. Both light flavor tags and negative tags can come from detector resolution effects, but light flavor positive tags may also be due to real displaced tracks from long-lived light particle decays (K_S and Λ) and interactions in the detector material. The overall distribution of light flavor tags is biased positive, and we must account for these effects with a *mistag asymmetry*, used as a correction to the mistag matrix [61].

The correction actually has two components, α and β . α corrects for the heavy flavor contribution to the negative-tagged sample, and β is a smaller correction for the heavy flavor content of the pretag jet sample.

$$\alpha = \frac{N_{light}^+}{N^-} \quad (5.16)$$

$$\beta = \frac{N^{pre}}{N_{light}^{pre}} \quad (5.17)$$

Both α and β are necessary to estimate a true light flavor mistag *probability* from the mistag matrix, as shown algebraically in Section 5.3.

The basic strategy for measuring α and β is to fit the invariant mass distribution of *b*-tags in the jet samples using Monte Carlo templates to extract the bottom, charm, and light flavor fractions in the positive and negative-tagged samples. To preserve the difference between positive and negative tags, we sign the invariant mass according to the sign of the tag. Since the negative L_{2d} tail is poorly modeled in

simulation, and because the statistics are limited in this sample, the fit is performed using templates in the tag *excess*, where the negative distribution is subtracted from the positive. Since the tags from resolution are mostly symmetric, the light flavor template in this case is dominated by material interactions and K_S/Λ decays. We fit using the `TFractionFitter` [66] package in `ROOT` [67], which takes the statistical uncertainties in the templates into account.

After the fit is performed, we separate the positive and negative sides of each template in proportion to the fit fraction, and we determine the ratio of the total negative tail in data to the summed tail in Monte Carlo. This *negative scale factor* is an *ad hoc* correction to improve the fit, rather than a physically meaningful quantity. It represents the extent to which mistags are underestimated in simulation, and it is found to be between 1.2 to 1.5, depending on the sample. For the central value of α , we assume the negative scale factor is the same for all species of jets, and the effect of this assumption is built into the systematic uncertainty. We normalize the negative tail to the data, and we correct *symmetrically* the positive tags for this underestimate. We then determine α directly from the integrals of the total negative tail and the positive light flavor tail.

To determine β , we also must derive the number of light flavor jets in the pre-tag sample. We divide the fit number of tagged heavy flavor jets by the *b*-tagging efficiency, and subtract this number from the total number of jets.

$$N_{heavy}^{pre} = \frac{N_{heavy}^+}{SF \epsilon_{heavy}^{MC}} \quad (5.18)$$

$$N_{light}^{pre} = N^{pre} - N_{heavy}^{pre} \quad (5.19)$$

Monte Carlo Samples

Since the objective of this note is a correction to the mistag probabilities computed with the mistag matrix, we use jet selection criteria identical to that used in the mistag matrix described above. Four Monte Carlo samples are generated with a generic 2-parton process in PYTHIA, where one parton is required to have $p_T > 18$ GeV/ c , 40 GeV/ c , 60 GeV/ c , and 90 GeV/ c to match the four data samples. We further correct the Monte Carlo samples for the trigger efficiency turnon in each of the corresponding data samples, and we weight them according to the statistics of the data sample.

Fit Results

We present in this section the fit results from the individual trigger samples for the loose tagger only, when no additional selection is applied. These fits are not used in the final result, but illustrate the concept and the quality of the Monte Carlo templates.

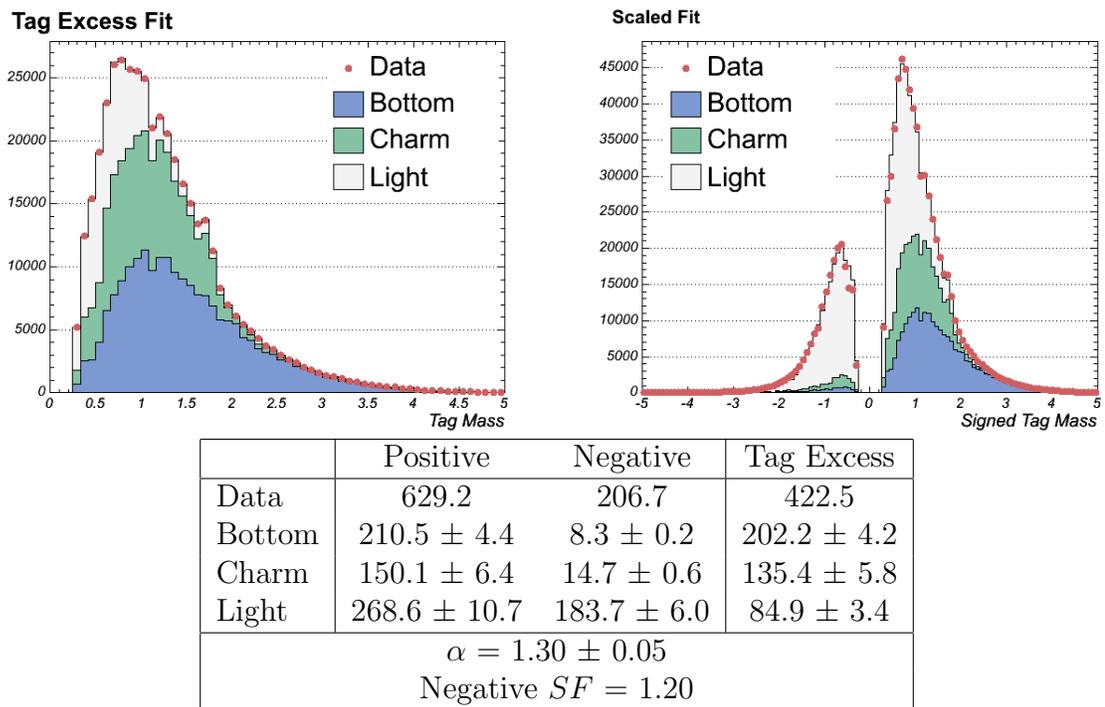


Figure 5.43: The tag excess fit results (top left) and the unfolded, corrected fit result (top right) for the loose tagger in the Jet20 sample. The fit integrals (in kJets) and α result are shown as well.

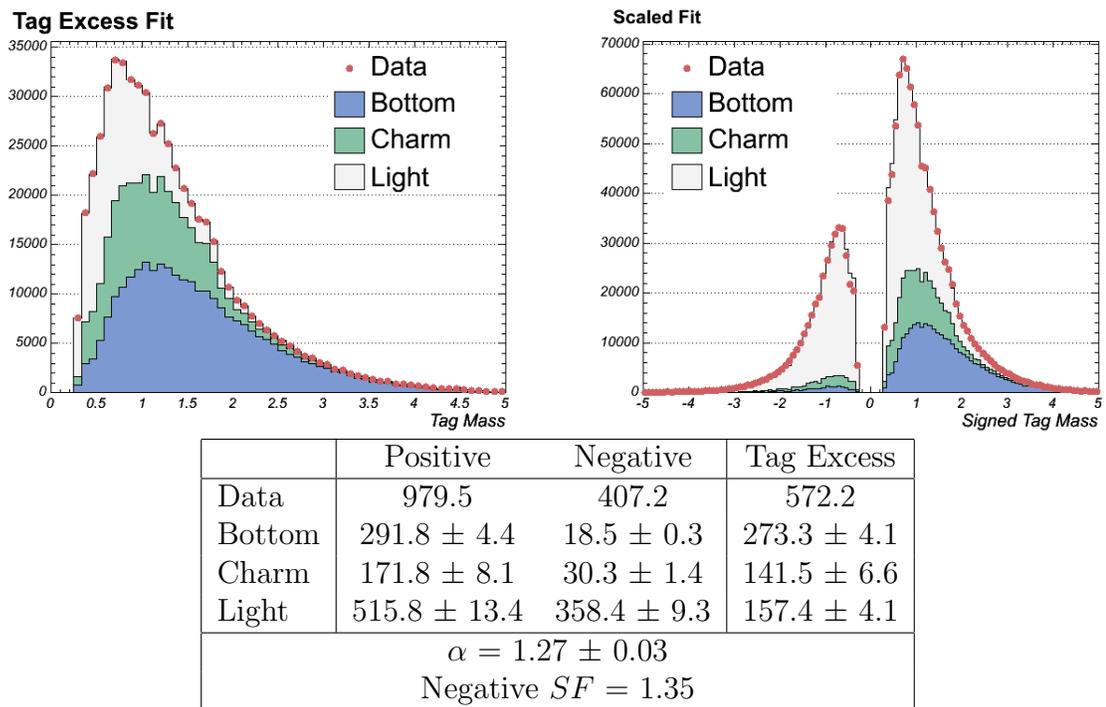


Figure 5.44: The tag excess fit results (top left) and the unfolded, corrected fit result (top right) for the loose tagger in the Jet50 sample. The fit integrals (in kJets) and α result are shown as well.

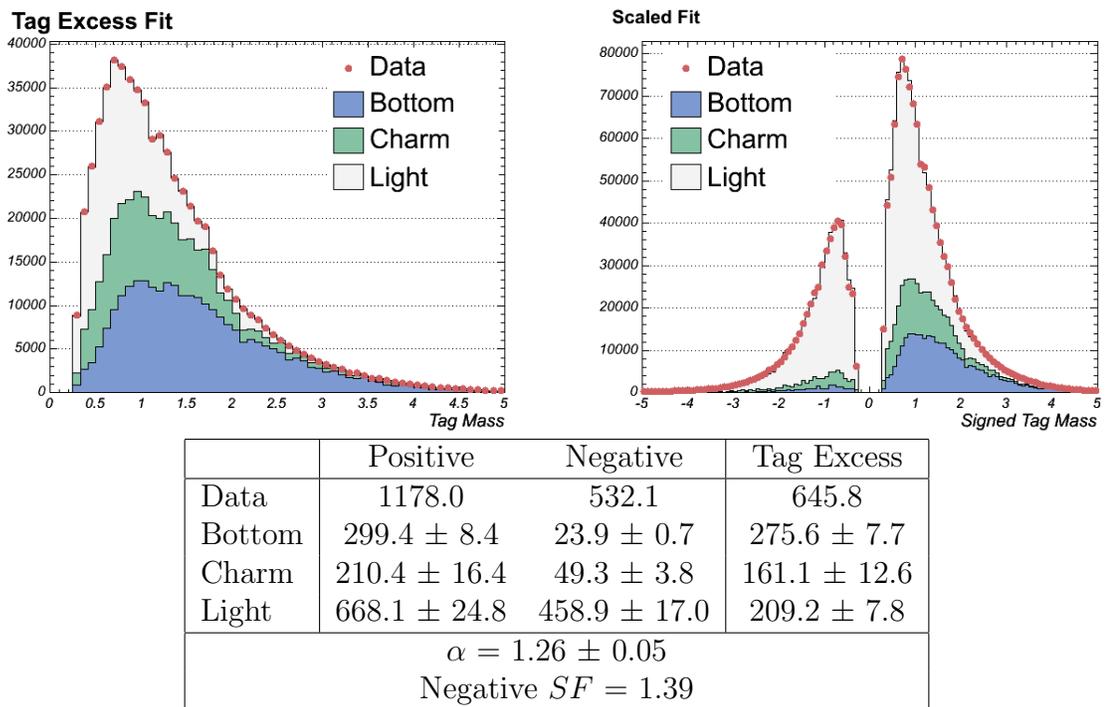


Figure 5.45: The tag excess fit results (top left) and the unfolded, corrected fit result (top right) for the loose tagger in the Jet70 sample. The fit integrals (in kJets) and α result are shown as well.

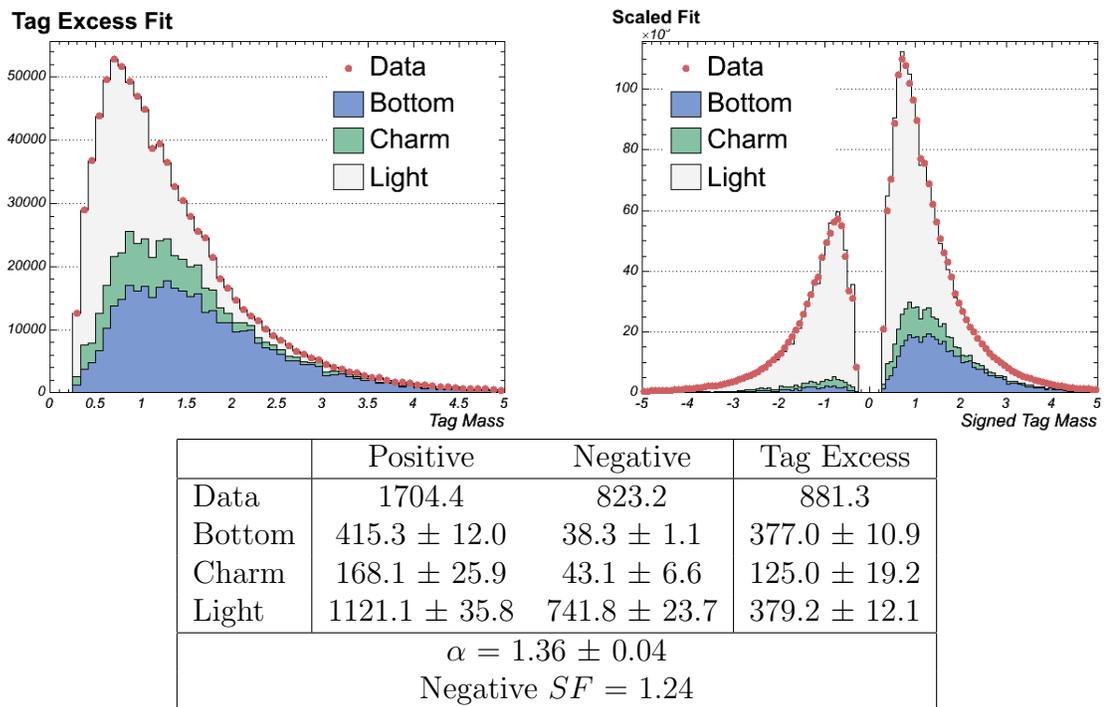


Figure 5.46: The tag excess fit results (top left) and the unfolded, corrected fit result (top right) for the loose tagger in the Jet100 sample. The fit integrals (in kJets) and α result are shown as well.

E_T Dependence of α

The heavy flavor content of the jet-triggered samples is a strong function of E_T , since heavy flavor jets above the trigger threshold are dominantly produced through direct production of $b\bar{b}$ and $c\bar{c}$ pairs, while jets below the threshold are dominated by gluon splitting. To account for this, we derive the α and β corrections as functions of this variable. Since they are corrections to the mistag probabilities, we define our E_T binning to be compatible with the matrix, with boundaries at 10, 22, 40, 60, and 200 GeV, four bins total. In each bin, we make composite templates for bottom, charm, and light by mixing the distributions from each MC sample in proportion equal to the data. The fit procedure is identical to those presented above.

The error included in the plots shows the statistical error on the light flavor fraction only, but the uncertainty is scaled up by the fit χ^2/dof , which varies between 1.2 and 2.0 over the full range. This adjustment is intended to cover some of the systematics involved in mass fitting. The evolution of α with jet E_T for the loose (tight) tagger is shown in Figure 5.47 (5.48).

Systematics

The uncertainties shown in Figures 5.47 and 5.48 include the statistical error inflated by the χ^2/dof of the individual tag excess fits. We consider additional sources of error linked to the choice of the tag variable, extent of E_T sculpting in the simulation, and the application of the negative scale factor.

First, we repeat the fit procedure using pseudo- $c\tau$, $L_{2d} \frac{m^{vtx}}{p_T^{vtx}}$, in place of the tag mass. This has the advantage of being is less sensitive to the Monte Carlo than the

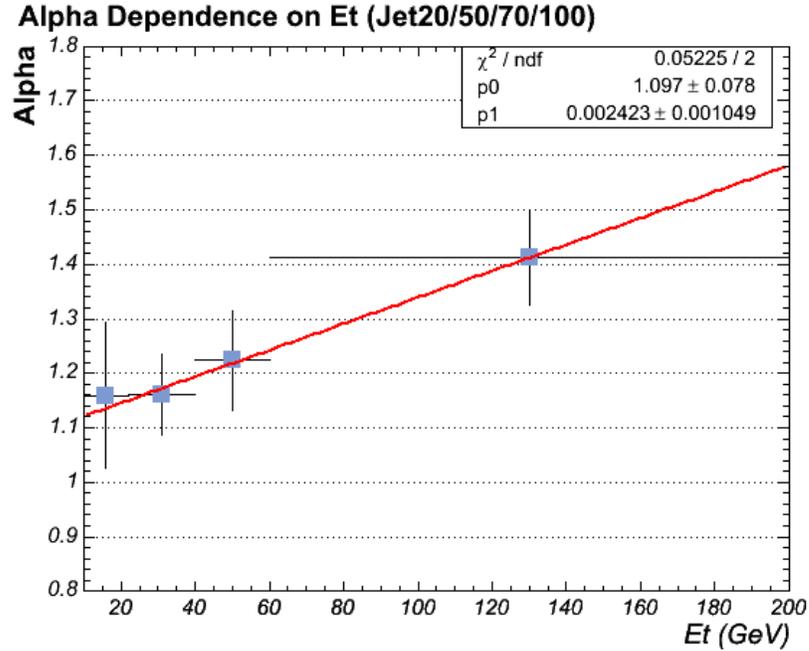


Figure 5.47: Dependence of the mistag asymmetry α on jet E_T for the loose SECVTX tagger. The increase at high E_T corresponds to an equivalent increase in the mistag rate.

tag mass, since the vertex displacement can be the same whether all the decay tracks are included or not. When fitting the full distribution with $c\tau$ rather than the tag mass, we measure α to be 1.41 instead of 1.39 for the tight tagger and 1.27 instead of 1.29 for the loose tagger. We therefore adopt 2% as a systematic uncertainty for both tight and loose.

Second, we test the sensitivity of the fit result to our trigger simulation, in which we subject the leading jet in Monte Carlo to a Gaussian turnon at the trigger threshold. If we instead impose a tight requirement on the leading jet in *both* data and MC (10 GeV above the trigger threshold), we find that fit results are consistent in the combined sample, and we take the mean difference of 2% as a systematic error for

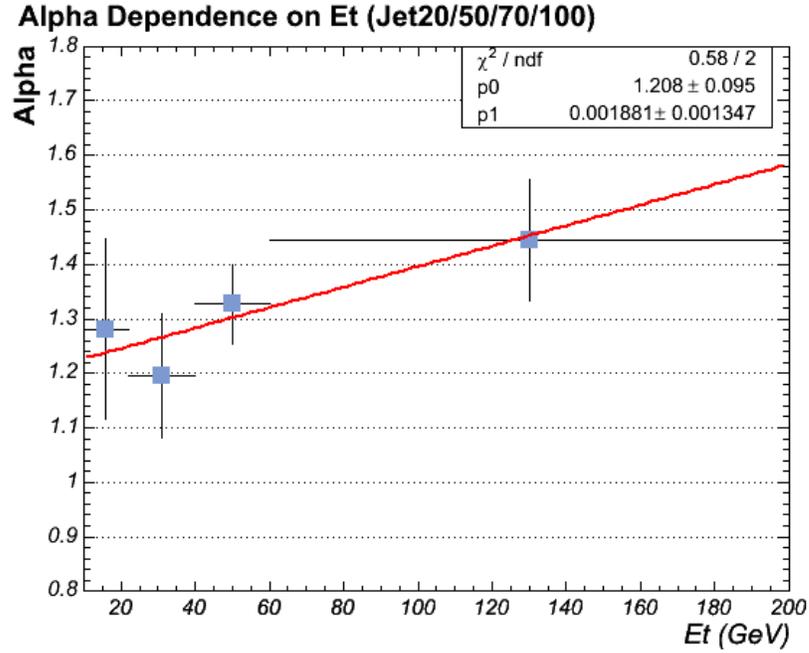


Figure 5.48: Dependence of the mistag asymmetry α on jet E_T for the tight SECVTX tagger. The increase at high E_T corresponds to an equivalent increase in the mistag rate.

both taggers.

Finally, we assume that the underestimate of the negative tags (the negative scale factor) is equally applicable to all species of jets. If we instead interpret this as a feature of light flavor tags only, the measured value of α increases accordingly. The size of this effect for the loose tagger in the four E_T bins is 2%/3%/4%/3%, and for the tight tagger we observe shifts of 1%/3%/2%/3%.

Derivation of β

The β measurement is more straightforward, and the result is only subject to uncertainties from the fit and the systematic error on the *b*-tagging efficiency (or

	10-22 GeV	22-40 GeV	40-60 GeV	60-200 GeV
Pretag Data	68.4 MJets	35.6 MJets	24.2 MJets	36.9 MJets
Loose SECVTX				
Fit <i>b</i>	179.8 kJets	317.4 kJets	248.4 kJets	443.2 kJets
Fit <i>c</i>	180.2 kJets	215.9 kJets	155.5 kJets	184.7 kJets
<i>b</i> Efficiency ($\pm 5\%$)	24.9%	38.5%	43.8%	47.8%
<i>c</i> Efficiency ($\pm 15\%$)	5.9%	9.9%	12.3%	15.4%
Pretag Heavy Flavor	$5.8 \pm 0.5\%$	$8.3 \pm 0.7\%$	$7.5 \pm 0.6\%$	$5.7 \pm 0.4\%$
Tight SECVTX				
Fit <i>b</i>	153.1 kJets	271.4 kJets	196.7 kJets	315.7 kJets
Fit <i>c</i>	125.0 kJets	148.3 kJets	104.2 kJets	181.2 kJets
<i>b</i> Efficiency ($\pm 5\%$)	21.0%	33.0%	37.1%	39.7%
<i>c</i> Efficiency ($\pm 15\%$)	4.3%	7.2%	8.6%	10.3%
Pretag Heavy Flavor	$5.3 \pm 0.5\%$	$8.1 \pm 0.7\%$	$7.2 \pm 0.6\%$	$6.9 \pm 0.6\%$

Table 5.9: Summary of fit heavy flavor counts and the average tagging efficiency corrected by the scale factor. This estimate of the pretag heavy flavor fraction is used to extract β .

SF). In Table 5.9, we show the quantities used to calculate β : the pretag jet counts in the data sample, the fit bottom and charm content of the positive tags in each E_T bin, and the mean *b*-tagging efficiency for each species in that bin. The efficiencies are corrected by the *b*-tagging scale factor, where we triple the uncertainty for charm jets.

Averaging results based on the tight and loose fits and taking the difference as an additional systematic uncertainty, we determine β to be 1.06 ± 0.01 , 1.09 ± 0.02 , 1.08 ± 0.01 , and 1.07 ± 0.01 in the four E_T bins. The systematic uncertainty on the fit fractions is taken to be 100% correlated with that contribution to α . Final results for the asymmetry corrections is shown in Table 5.10. In general, we have found that the true light flavor asymmetry $\frac{N_{light}^+}{N_{light}^-}$ is between 1.4 and 1.6, implying that material tags and long-lived light flavor constitute roughly one third of positive light flavor

	10-22 GeV	22-40 GeV	40-60 GeV	60-200 GeV
Loose α	1.16 ± 0.15	1.16 ± 0.09	1.22 ± 0.10	1.41 ± 0.10
Loose $\alpha\beta$	1.23 ± 0.17	1.26 ± 0.11	1.32 ± 0.12	1.51 ± 0.12
Tight α	1.28 ± 0.18	1.19 ± 0.12	1.33 ± 0.09	1.44 ± 0.11
Tight $\alpha\beta$	1.36 ± 0.20	1.30 ± 0.14	1.44 ± 0.12	1.54 ± 0.13

Table 5.10: Final results for α and $\alpha\beta$ for both taggers.

tags. Heavy flavor jets account for approximately 5-7% of the total pretag sample, and with a negative tag rate nearly twice as high as light flavor, accounts for 10-15% of the negative tags used to construct the mistag matrices [61].

Chapter 6

Composition of the W +Jets

Sample

In this chapter, we present an estimate of the composition of the lepton+jets sample using the “Method 2” prescription. Where Method 1 uses the average b -tagging rate in all-jet data to estimate the tag rate in W +jets, Method 2 employs a combination of data- and MC-driven approaches to evaluate each physics contribution separately [68]. We assume here the theoretical $t\bar{t}$ cross section of 6.7 pb for a top mass of 175 GeV/ c^2 [6]; the direct measurement of the top cross section following from these results is described in the following chapter.

6.1 Method 2 Overview

The fundamental advantage of Method 2 is its use of the data, before any b -tagging requirements (pretag), to constrain the sample composition. We assume

that the pretag sample consists of three distinct components: events without a real leptonic W decay (non- W), low-rate electro-weak processes ($t\bar{t}$, dibosons, single top), and real W 's produced in association with jets (W +jets). This last contribution, which dominates the pretag sample and the b -tagged 1 and 2-jet bins, is not known *a priori*, due to large corrections to the cross section beyond NLO. We therefore take the overall normalization from the pretag data sample, subtract off the non- W and purely electroweak components, and assume the remainder is W +jets. This normalization is performed independently in each jet multiplicity bin.

In order to estimate the b -tag rate for W +jets events, we must additionally evaluate the *heavy flavor fraction* in W +jets, defined as the fraction of events containing bottom or charm jets [69]:

$$F^{hf} \equiv \frac{N^{W+HF}}{N^{W+jets}}, \quad (6.1)$$

where the N 's represent the number of events in data.

We assume that these fractions will be less sensitive to higher-order effects than an absolute cross section. The heavy flavor fractions are derived from Monte Carlo simulations taking into account the $Wb\bar{b}$, $Wc\bar{c}$, and Wc processes with multiple jets in the final state.

We check the overall rate of heavy flavor production directly in data by comparing b -tag yields in high-statistics jet samples, the same as those described in Section 5.3. As shown in Figure 6.1, the main diagram contributing to $Wb\bar{b}$ contains a $g \rightarrow b\bar{b}$ vertex, also present in all-jet heavy flavor events; the same is true for $Wc\bar{c}$. We therefore assume that the ratio of the data and Monte Carlo heavy flavor fractions

measured in all-jet events, called K , is the same for $Wb\bar{b}$ and $Wc\bar{c}$ events as well [70]. Determination of the heavy flavor fractions and the measurement of K are both discussed in depth in Appendix B.

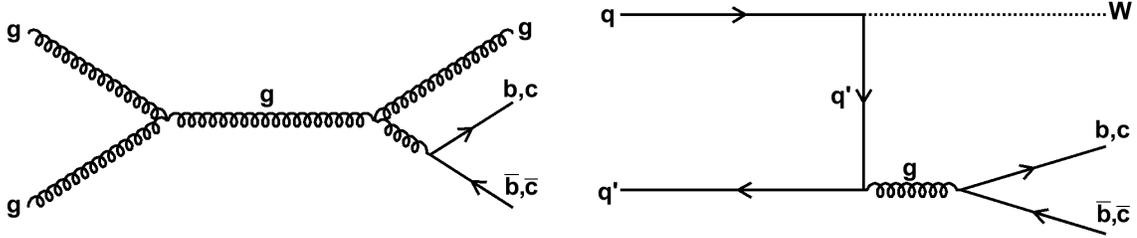


Figure 6.1: Feynman diagrams contributing to heavy flavor production in all-jet (left) and W events (right).

For W +jets, then, the pretag and tag expectations can be written:

$$N_{pre}^W = N_{pre}^{data} (1 - F_{non-W}) - N_{pre}^{t\bar{t}} - N_{pre}^{diboson} - N_{pre}^{singletop} \quad (6.2)$$

$$N_{tag}^W = N_{pre}^W \left(\sum_i \epsilon_i^{hf} F_i^{hf} + \epsilon^{lf} (1 - \sum_i F_i^{hf}) \right) \quad (6.3)$$

Here, the N 's are the numbers of W +jets, $t\bar{t}$, diboson, and single top events, the ϵ 's are b -tagging efficiencies for W +heavy (hf) and light (lf) flavor, and F_i^{hf} are the heavy flavor fractions corrected for K . The sum over i includes different heavy flavor configurations (W +1 b -jet, 2 b -jets, 1 c -jet, and 2 c -jets). The quantity F_{non-W} is the fraction of the pretag sample not attributed to physics process with a real W decay.

It is clear from the above equations that the backgrounds, especially the W +jets normalization (N_{pre}^W), depend on the assumed cross sections for other processes. Method 2 can therefore be implemented iteratively, to solve for the $t\bar{t}$ cross section, for example; this is the approach used in Chapter 7. For now, we assume the $t\bar{t}$ cross

section is known, and we describe the remaining sample composition dictated by this assumption.

6.2 Event Selection

6.2.1 Triggers and Data Event Selection

The data were collected on inclusive high- p_T triggers, requiring an electron in the CEM with E_T larger than 18 GeV (trigger path ELECTRON_CENTRAL_18) or a muon in the CMUP or CMX with p_T greater than 18 GeV/ c (MUON_CMUP18_L2_PT15 and MUON_CMX18_L2_PT15). The trigger is 97% (92%/96%) efficient for leptons in the CEM (CMUP/CMX) [71][46]. In the analysis, we further require leptons to have E_T (electrons) or p_T (muons) above 20 GeV and to be isolated (see Section 4.2), and we accept only tight leptons from their respective trigger samples (i.e. electron events from the muon triggers are rejected, and vice versa). We remove cosmic ray muons, conversion electrons, high- p_T dilepton events, and Z decays. For tight muons, we further require that the track matching χ^2 be <2.75 for the first 300 pb $^{-1}$ and <2.3 for the remaining sample; the cut is set such that the efficiency is 99% for real muons [46]. Standard lepton selection requirements are discussed in Section 4.2.

We require jets to have transverse energy above 20 GeV and absolute $\eta < 2.0$ and the missing E_T to be larger than 20 GeV. Pretag yields for each jet multiplicity are listed in Table 6.1. More detailed tables including the event yield at each selection stage can be found in Appendix A. The $t\bar{t}$ signal region consists of events with *exactly* one tight lepton and three or more jets; we also evaluate the expected Method

	1 jet	2 jet	3 jet	4 jet	≥ 5 jet
Pretag	115313	18922	3136	694	172
Loose ≥ 1 (≥ 2) b -Tags	2729(-)	1277(82)	499(79)	284(79)	90(33)
Tight ≥ 1 (≥ 2) b -Tags	1697(-)	835(46)	358(52)	224(50)	75(22)

Table 6.1: Event yields sorted by the number of jets and b -tags in the event.

2 backgrounds in one-jet and two-jet events, although these events are not directly used in this analysis.

The *single-tag* analysis also requires that *at least one* jet be b -tagged, and the double-tag analysis requires *at least two* b -tagged jets.

6.2.2 b -Tagging

We initially use both the tight and loose SECVTX b -taggers, and we present results in parallel. The tight tuning has been used as the standard b -tagger in past analyses, with high purity and an average b -tagging efficiency of 40% per b -jet. The loose tagger was introduced primarily for double-tagged analyses, and has a b -jet efficiency 20% higher at the expense of a factor of three increase in the light flavor b -tag rate. In double-tagged events, the non-bottom contribution remains quite small. Details on the algorithms and their respective efficiencies and fake rates are presented in the previous chapter. Optimization studies comparing the expected measurement quality for each b -tagger will be shown in Section 7.1. The b -tagged event yields in data for both b -taggers are shown in Table 6.1.

6.2.3 Monte Carlo

The Method 2 signal and backgrounds are modeled in Monte Carlo simulations, and are subjected to the same kinematic event selection as the data. We simulate the beam position and profile and silicon efficiency for all recorded runs, and we include multiple $p\bar{p}$ interactions according to the instantaneous luminosity in those runs. All samples use PYTHIA v6.216 as a generator except for $W+jets$, which uses ALPGEN v2.1.0 patched into PYTHIA v6.326 for the shower (this will be discussed in Appendix B). We use the EvtGen package to decay heavy flavor hadrons.

6.3 Signal Expectation

We quantify the $t\bar{t}$ expectation in two distinct pieces: the acceptance (including the geometric acceptance, branching fraction, and lepton identification and trigger efficiencies) and the b -tagging efficiency. The former determines the top content of the pretag sample, and the latter is the fraction of these events expected to be b -tagged. The next two subsections are devoted to the calculation of these two quantities.

6.3.1 Acceptance

All signal estimates are based on PYTHIA $t\bar{t}$ samples generated with a top mass of $175 \text{ GeV}/c^2$. We apply the event selection discussed in Section 6.2.1 directly to the simulation to determine the acceptance, which is calculated separately in each jet multiplicity bin (merging ≥ 5 jets) and sorted by the lepton subdetector; we accept only events where the lepton is a muon in the CMUP or CMX or an electron in

the CEM. Electrons in the PEM and muons in the BMU are rejected. The muon matching χ^2 requirement is not applied to the Monte Carlo, but rather we scale the acceptance to account for the cut efficiency (see Table 6.3).

The Monte Carlo sample is $t\bar{t}$, with no restriction placed on the top decay channel at simulation. A nominal $t\bar{t}$ lepton+jets event will have exactly one lepton and four jets, but other decays may enter the pretag sample (such as a dilepton event where one lepton is lost, or a τ +jets event with a leptonic τ decay). The yield of events at each stage of event selection in MC is shown in Table 6.2.

The total acceptance is the fraction of simulated events that reach the pretag stage in our event selection. Depending on the lepton detector, this rate is corrected for known differences between data and Monte Carlo; the efficiencies to reconstruct and identify leptons and the primary vertex-finding efficiency, for example, are adjusted by applying additional *scale factors*, while data-only effects such as the trigger efficiency are applied directly as corrections to the acceptance. A summary of these scale factors is in Table 6.3, sorted by the appropriate run range.¹ Most lepton efficiencies are measured by reconstructing Z boson decays to lepton pairs, and therefore may be biased due to the relatively low detector occupancy in such events, though the size of the effect is small compared to its uncertainty. We do not adjust the acceptance corrections for this effect, but do include an additional systematic uncertainty. Before correction, the raw acceptance for ≥ 3 -jet events in Monte Carlo is 4.7% for electrons, 4.2% for muons; after correction, the acceptances are 4.3% and 3.5%, respectively.

The total number of Monte Carlo events before event selection is scaled to $\sigma_{t\bar{t}} \int \mathcal{L} dt$

¹We validate the data in sections of roughly 300 pb^{-1} , including measuring these efficiencies and scale factor. While the values between run ranges are in general consistent with one another, we use the measured value corresponding to the specific run range.

	W+0 jets	W+1 jet	W+2 jets	W+3 jets	W+4 jets	W+ \geq 5 jets	Total
CEM							
Initial	2840	53562	296504	720647	1118487	1801727	3993767
Silicon Good Run	2840	53555	296449	720557	1118330	1801471	3993202
PV $ z < 60\text{cm}$	2735	51620	285515	690362	1069112	1722986	3822330
≥ 1 Lepton	859	15495	70322	110763	108104	53337	358880
Missing $E_{T\cancel{l}} > 20$ GeV	787	14364	65090	100587	96926	45728	323482
Lepton Iso.	787	14132	62291	91310	82146	30195	280861
Dilepton Veto	368	7952	42941	85405	80488	29796	246950
Z Mass Veto	330	7406	40215	81176	78218	29000	236345
Conversion Veto	330	7380	39773	79217	75047	25008	226755
Lepton z_0	330	7380	39766	79211	75036	25005	226728
CMUP							
Initial	2840	53562	296504	720647	1118487	1801727	3993767
Silicon Good Run	2840	53555	296449	720557	1118330	1801471	3993202
PV $ z < 60\text{cm}$	2735	51620	285515	690362	1069112	1722986	3822330
≥ 1 Lepton	612	10268	46000	69921	65496	25649	217946
Missing $E_{T\cancel{l}} > 20$ GeV	554	9553	42449	63294	58357	21998	196205
Lepton Iso.	523	9010	38984	55637	48975	16174	169303
Dilepton Veto	238	4903	26023	51732	47932	15955	146783
Z Mass Veto	225	4629	24864	50186	47182	15677	142763
Cosmic Ray Veto	225	4629	24864	50186	47182	15677	142763
Lepton z_0	225	4628	24864	50177	47169	15671	142734
CMX							
Initial	2840	53562	296504	720647	1118487	1801727	3993767
Silicon Good Run	2840	53555	296449	720557	1118330	1801471	3993202
PV $ z < 60\text{cm}$	2735	51620	285515	690362	1069112	1722986	3822330
≥ 1 Lepton	287	4546	19887	29116	27201	10502	91539
Missing $E_{T\cancel{l}} > 20$ GeV	266	4231	18383	26383	24339	8968	82570
Lepton Iso.	254	3985	16835	23349	20825	6753	72001
Dilepton Veto	111	2112	11147	21671	20358	6648	62047
Z Mass Veto	102	1984	10618	21014	20012	6553	60283
Cosmic Ray Veto	102	1984	10618	21014	20012	6553	60283
Lepton z_0	102	1983	10616	21008	20006	6547	60262

Table 6.2: Event yields at each stage of event selection in $t\bar{t}$ Monte Carlo. Yields are sorted by number of jets, and results are presented for the three lepton types.

Correction	2/02-8/04	9/04-8/05	9/05-2/06	3/06-8/06
CEM				
Trigger Efficiency	0.962(7)	0.976(6)	0.979(4)	0.959(7)
Electron ID SF	0.991(5)	0.985(5)	0.974(4)	
Lepton Isolation	1.000(8)			
CMUP				
Trigger Efficiency	0.902(4)	0.919(4)	0.918(5)	0.913(6)
Muon ID SF	0.985(4)	0.989(4)	0.975(5)	0.975(6)
Muon Reconstruction	0.951(4)	0.939(4)	0.941(4)	0.955(5)
χ^2 Cut	0.990(12)			
Lepton Isolation	1.000(10)			
CMX Arches				
Trigger Efficiency	0.967(4)	0.955(4)	0.954(5)	0.947(6)
Muon ID SF	1.014(4)	1.000(5)	1.004(6)	1.000(8)
Muon Reconstruction	0.996(2)	0.993(2)	0.989(3)	0.991(3)
χ^2 Cut	0.989(17)			
Lepton Isolation	1.000(14)			
CMX Miniskirt/Keystone				
Trigger Efficiency	-	0.772(14)	0.744(19)	0.755(23)
Muon ID SF	-	0.979(11)	0.990(13)	1.001(15)
Muon Reconstruction	-	0.933(9)	0.939(11)	0.902(16)
χ^2 Cut	0.989(17)			
Lepton Isolation	1.000(14)			
Common				
Z Vertex < 60cm	0.958(2)			
Isolated Track Efficiency	1.014(2)			

Table 6.3: A summary of scale factors and efficiencies used to correct the pretag acceptances. Systematic errors are shown in parentheses.

Uncorrected Acceptance (%)					
	1 jet	2 jet	3 jet	4 jet	≥ 5 jet
2/02-8/04					
CEM	0.197 ± 0.004	1.039 ± 0.009	2.047 ± 0.013	1.950 ± 0.012	0.652 ± 0.007
CMUP	0.118 ± 0.003	0.649 ± 0.007	1.302 ± 0.010	1.224 ± 0.010	0.398 ± 0.006
CMX	0.045 ± 0.002	0.232 ± 0.004	0.439 ± 0.006	0.417 ± 0.006	0.137 ± 0.003
9/04-8/05					
CEM	0.194 ± 0.004	1.040 ± 0.008	2.079 ± 0.012	1.963 ± 0.011	0.657 ± 0.007
CMUP	0.125 ± 0.003	0.646 ± 0.006	1.312 ± 0.009	1.244 ± 0.009	0.415 ± 0.005
CMX	0.042 ± 0.002	0.231 ± 0.004	0.463 ± 0.005	0.440 ± 0.005	0.143 ± 0.003
CMX (Mini/Key)	0.012 ± 0.001	0.070 ± 0.002	0.143 ± 0.003	0.130 ± 0.003	0.043 ± 0.002
9/05-2/06					
CEM	0.187 ± 0.004	1.043 ± 0.010	2.094 ± 0.015	1.980 ± 0.014	0.652 ± 0.008
CMUP	0.118 ± 0.003	0.659 ± 0.008	1.326 ± 0.012	1.232 ± 0.011	0.417 ± 0.006
CMX	0.044 ± 0.002	0.233 ± 0.005	0.465 ± 0.007	0.457 ± 0.007	0.148 ± 0.004
CMX (Mini/Key)	0.013 ± 0.001	0.068 ± 0.003	0.141 ± 0.004	0.133 ± 0.004	0.045 ± 0.002
3/06-8/07					
CEM	0.194 ± 0.005	1.024 ± 0.013	2.075 ± 0.018	1.950 ± 0.017	0.669 ± 0.010
CMUP	0.117 ± 0.004	0.649 ± 0.010	1.302 ± 0.014	1.227 ± 0.014	0.410 ± 0.008
CMX	0.047 ± 0.003	0.232 ± 0.006	0.468 ± 0.009	0.455 ± 0.008	0.148 ± 0.005
CMX (Mini/Key)	0.011 ± 0.001	0.069 ± 0.003	0.141 ± 0.005	0.136 ± 0.005	0.042 ± 0.003

Table 6.4: Uncorrected acceptances for the various run ranges and sub-detectors based on $175 \text{ GeV}/c^2$ $t\bar{t}$ Monte Carlo.

for each run range separately, assuming a cross section of 6.7 pb ; multiplying by the corrected acceptance and summing over run ranges gives an estimate of the number of events in the pretag sample. The total luminosity times acceptance in the signal region is $52.3 \pm 3.2 \text{ events/pb}$. A summary of raw and corrected acceptances is given in Tables 6.4 and 6.5, sorted by jet multiplicity.

6.3.2 b -Tagging Efficiency

For reasons outlined in Section 5.2, we do not assume the b -tagging efficiency in Monte Carlo accurately represents the efficiency in data. The b -tagging scale factor (SF) described there is the measured ratio of per-jet b -tagging efficiencies in data and Monte Carlo, and is determined to be 0.95 ± 0.05 . To measure the tagging efficiency in events with multiple heavy flavor jets, though, a single multiplicative correction is insufficient.

Corrected Acceptance (%)					
	1 jet	2 jet	3 jet	4 jet	≥ 5 jet
2/02-8/04					
CEM	0.183 ± 0.006	0.962 ± 0.020	1.896 ± 0.035	1.806 ± 0.034	0.604 ± 0.014
CMUP	0.096 ± 0.004	0.528 ± 0.016	1.058 ± 0.028	0.994 ± 0.027	0.324 ± 0.011
CMX	0.042 ± 0.003	0.217 ± 0.009	0.412 ± 0.015	0.391 ± 0.014	0.129 ± 0.006
9/04-8/05					
CEM	0.181 ± 0.006	0.972 ± 0.020	1.942 ± 0.035	1.833 ± 0.033	0.614 ± 0.014
CMUP	0.103 ± 0.004	0.530 ± 0.015	1.077 ± 0.028	1.021 ± 0.027	0.341 ± 0.011
CMX	0.038 ± 0.002	0.211 ± 0.008	0.422 ± 0.015	0.400 ± 0.014	0.130 ± 0.006
CMX (Mini/Key)	0.008 ± 0.001	0.047 ± 0.003	0.097 ± 0.005	0.088 ± 0.005	0.029 ± 0.002
9/05-2/06					
CEM	0.173 ± 0.006	0.966 ± 0.021	1.940 ± 0.037	1.834 ± 0.036	0.604 ± 0.015
CMUP	0.096 ± 0.005	0.534 ± 0.017	1.074 ± 0.030	0.998 ± 0.028	0.338 ± 0.012
CMX	0.040 ± 0.003	0.212 ± 0.009	0.424 ± 0.016	0.416 ± 0.016	0.135 ± 0.007
CMX (Mini/Key)	0.009 ± 0.001	0.045 ± 0.003	0.093 ± 0.006	0.088 ± 0.005	0.030 ± 0.003
3/06-8/07					
CEM	0.176 ± 0.007	0.929 ± 0.023	1.882 ± 0.040	1.770 ± 0.038	0.607 ± 0.017
CMUP	0.096 ± 0.006	0.530 ± 0.019	1.064 ± 0.032	1.004 ± 0.031	0.335 ± 0.013
CMX	0.043 ± 0.004	0.210 ± 0.010	0.422 ± 0.017	0.411 ± 0.017	0.133 ± 0.007
CMX (Mini/Key)	0.007 ± 0.001	0.045 ± 0.004	0.092 ± 0.006	0.089 ± 0.006	0.028 ± 0.003

Table 6.5: Corrected acceptances for the various run ranges and sub-detectors based on $175 \text{ GeV}/c^2$ $t\bar{t}$ Monte Carlo.

In this analysis, we take advantage of the fact that the scale factor is smaller than one, and we simply ignore 5% ($1-SF$) of the positive heavy flavor b -tags in the simulation. Light flavor b -tags in $t\bar{t}$ events also contribute a non-negligible component of the total event b -tagging efficiency. We estimate this contribution by applying the mistag parameterization to light flavor jets in the simulation, and assuming that these jets will be tagged at a rate consistent with the mistag matrix prediction.² Including these b -tags with the re-scaled heavy flavor b -tags, we derive an estimate for the total event efficiency expected in data.

These efficiencies for the tight and loose b -taggers are shown in Table 6.6. The total acceptance and b -tag expectations for this analysis are shown in Table 6.7. Uncertainties in the tables include MC statistics, luminosity, and scale factors.

²This method is only reasonable for incremental contributions to the event b -tagging efficiency. The distributions of quantities used in the parameterization (especially N_{tracks}) are sufficiently different in data and simulation that the expected b -tag rate for $W+4$ light jets, for instance, is not reliable.

	≥ 1 b -Tag (%)	≥ 2 b -Tags (%)
Tight	59 ± 3	15 ± 2
Loose	68 ± 4	22 ± 3

Table 6.6: Total expected b -tagging efficiency for $t\bar{t}$ events with 3 or more jets for both b -taggers.

	1 jet	2 jet	3 jet	4 jet	≥ 5 jet
Pretag	24.29 ± 1.50	130.50 ± 8.07	261.01 ± 16.14	246.90 ± 15.27	82.13 ± 5.08
Tight-Tagged	8.17 ± 0.83	65.62 ± 6.68	146.30 ± 14.90	151.86 ± 15.46	51.33 ± 5.23
Loose-Tagged	9.78 ± 1.00	76.54 ± 7.79	169.78 ± 17.29	174.14 ± 17.73	58.99 ± 6.01

Table 6.7: Expected number of $t\bar{t}$ events, assuming a top cross section of 6.7 pb. The luminosity uncertainty and the systematic errors on all scale factors are included here. No uncertainty is taken on the top cross section.

6.4 Backgrounds

In addition to real $t\bar{t}$ production, other physics processes have a signature consistent with our event selection. We consider four main categories of background for this analysis: diboson and single top production (*MC backgrounds*), generic QCD jets with a fake W signature (*non- W*), W 's produced in association with real heavy flavor jets (*$W+Heavy Flavor$*), and real W 's with misidentified light flavor b -tags (*mistags*). These backgrounds and the methods for their evaluation are described in detail in the following subsections.

6.4.1 Electroweak & Single Top (MC Backgrounds)

Several distinct physics processes involving real W 's can reproduce the signature of top pair production. In this analysis, we account for contributions from diboson production (WW , WZ , and ZZ), single top production, and $Z \rightarrow \tau\tau$. These

processes have final states including leptons and missing E_T , and they have a high probability of producing b -tagged jets. WW and WZ events can result in a leptonic W and a heavy flavor decay of the other boson; single top yields at least one b -jet and a real W ; ZZ requires one Z to decay leptonically with one missed (or misidentified) lepton, with the other going to $b\bar{b}$ or $c\bar{c}$; and $Z \rightarrow \tau\tau$ events may have one τ decay leptonically with a b -tag on the opposite-side three-prong hadronic decay, reconstructed as a jet.

The diboson and single top background calculations are performed analogously to the signal estimation. Again, b and c -jet tags are adjusted to account for the SF , and light flavor tags are included at a rate consistent with data. The exceptional background is $Z \rightarrow \tau\tau$, since most b -tags are on three-prong τ decays rather than heavy flavor. Tags in τ jets have not been studied in depth, but Monte Carlo suggests that hadronic τ decays will only be tagged with 5% efficiency for tight SECVTX, smaller even than the charm tagging efficiency. Rather than specifically isolating τ tags and correcting them, we let the simulation model the b -tag rate, and we add a 20% uncertainty to this background, comparable to the uncertainty assigned to charm jets.

The assumed cross sections are listed in Table 6.8. Diboson and single top cross sections are recent NLO calculations [72][9], while the $Z \rightarrow \tau\tau$ cross section is taken from a direct CDF measurement [73]. The diboson simulations are leading-order PYTHIA, the single top MC uses a MADEVENT/PYTHIA combination [74], and $Z \rightarrow \tau\tau$ is simulated with a combination of three ALPGEN/PYTHIA samples, for $Z+0$, 1, and 2 light partons. The expected single- and double-tag contributions from each

Process	Cross Section (pb)
WW	12.4 ± 0.25
WZ	3.96 ± 0.06
ZZ	1.58 ± 0.05
t (s-ch)	0.88 ± 0.05
t (t-ch)	1.98 ± 0.08
$Z \rightarrow \tau\tau$	265 ± 30

Table 6.8: Cross sections used to scale MC-derived backgrounds.

background are shown in Tables 6.9 and 6.10.

Process	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
Loose, ≥ 1 b -tag					
WW	12.11 ± 1.27	29.74 ± 3.12	9.89 ± 1.04	2.46 ± 0.26	0.67 ± 0.07
WZ	5.29 ± 0.55	10.88 ± 1.13	3.08 ± 0.32	0.73 ± 0.08	0.18 ± 0.02
ZZ	0.15 ± 0.02	0.41 ± 0.05	0.23 ± 0.03	0.06 ± 0.01	0.02 ± 0.00
Single Top (s-ch)	5.64 ± 1.27	20.08 ± 4.54	6.28 ± 1.42	1.32 ± 0.30	0.28 ± 0.06
Single Top (t-ch)	19.56 ± 3.51	25.03 ± 4.49	5.44 ± 0.97	0.80 ± 0.14	0.09 ± 0.02
$Z \rightarrow \tau\tau$	33.45 ± 8.10	21.90 ± 5.30	10.09 ± 2.44	2.21 ± 0.53	0.35 ± 0.08
Tight, ≥ 1 b -tag					
WW	7.30 ± 0.77	18.86 ± 1.98	5.99 ± 0.63	1.51 ± 0.16	0.41 ± 0.04
WZ	4.06 ± 0.42	8.68 ± 0.90	2.34 ± 0.24	0.55 ± 0.06	0.13 ± 0.01
ZZ	0.12 ± 0.01	0.33 ± 0.04	0.18 ± 0.02	0.05 ± 0.01	0.01 ± 0.00
Single Top (s-ch)	4.75 ± 1.07	17.49 ± 3.95	5.46 ± 1.23	1.15 ± 0.26	0.25 ± 0.06
Single Top (t-ch)	16.69 ± 2.99	20.90 ± 3.75	4.59 ± 0.82	0.70 ± 0.12	0.08 ± 0.01
$Z \rightarrow \tau\tau$	22.30 ± 5.40	13.40 ± 3.24	6.08 ± 1.47	1.33 ± 0.32	0.19 ± 0.05

Table 6.9: Expected number of ≥ 1 -tag events from diboson, single top, and $Z \rightarrow \tau\tau$ backgrounds.

6.4.2 Non-W

The pretag requirements of a single identified lepton and large missing E_T are chosen to enhance the W purity of the triggered sample, but both can be faked. Lepton signatures may arise from conversions or misidentified jets and photons (electrons) or misidentified pions/kaons (muons), and additional real leptons from semi-leptonic

Process	2 jets	3 jets	4 jets	≥ 5 jets
Loose, ≥ 2 b -tags				
WW	0.33 ± 0.06	0.37 ± 0.07	0.18 ± 0.03	0.08 ± 0.01
WZ	1.92 ± 0.35	0.56 ± 0.10	0.13 ± 0.02	0.04 ± 0.01
ZZ	0.06 ± 0.01	0.05 ± 0.01	0.01 ± 0.00	0.00 ± 0.00
Single Top (s-ch)	5.75 ± 1.76	1.96 ± 0.60	0.45 ± 0.14	0.10 ± 0.03
Single Top (t-ch)	0.57 ± 0.15	0.67 ± 0.17	0.16 ± 0.04	0.03 ± 0.01
$Z \rightarrow \tau\tau$	0.29 ± 0.09	0.40 ± 0.13	0.09 ± 0.03	0.05 ± 0.02
Tight, ≥ 2 b -tags				
WW	0.10 ± 0.02	0.13 ± 0.02	0.09 ± 0.02	0.03 ± 0.01
WZ	1.40 ± 0.26	0.42 ± 0.08	0.07 ± 0.01	0.02 ± 0.00
ZZ	0.04 ± 0.01	0.03 ± 0.01	0.01 ± 0.00	0.00 ± 0.00
Single Top (s-ch)	3.97 ± 1.21	1.34 ± 0.41	0.30 ± 0.09	0.07 ± 0.02
Single Top (t-ch)	0.27 ± 0.07	0.43 ± 0.11	0.10 ± 0.03	0.02 ± 0.00
$Z \rightarrow \tau\tau$	0.12 ± 0.04	0.16 ± 0.05	0.04 ± 0.01	0.03 ± 0.01

Table 6.10: Expected number of ≥ 2 -tag events from diboson, single top, and $Z \rightarrow \tau\tau$ backgrounds.

heavy flavor decays. Missing E_T can arise without a neutrino anytime jets are mis-measured.

Since the missing E_T distribution from W decays is fairly well-modeled, we identify non- W events by looking for deviations in the missing E_T distribution from that predicted in W and $t\bar{t}$ Monte Carlo, particularly at low missing E_T . We fit the pretag and b -tagged missing E_T distributions in the range 0 to 120 GeV separately for each jet multiplicity, in each instance fixing the $t\bar{t}$ component to the value expected for the assumed cross section. Due to limited statistics in the b -tagged sample, we perform a single fit for all events with ≥ 3 jets. We minimize the χ^2 with `Minuit`, with three floating components constrained to sum to the data sample.

The templates used in the fit (other than $t\bar{t}$) are derived from ALPGEN W Monte Carlo (the same used to measure the heavy flavor fractions described in Appendix B),

lepton-filtered dijets (dominated by $b\bar{b}$), and *antielectrons*, events in the high- p_T electron samples in which the trigger electron fails two or more tight lepton selection requirements (as outlined in Section 4.2) [75]. Antielectrons are expected to be dominated by fakes rather than real electrons, so the missing E_T is typically due to resolution rather than an accompanying neutrino. The dijet MC sample is the same as is used to measure the b -tagging scale factor, and is described in Section 5.2. We measure a *fraction* of non- W events, F_{non-W} , defined as the integral of the antielectron and $b\bar{b}$ shapes in the fit divided by the total number of pretag events in the data.

A summary of the results (assuming a $t\bar{t}$ cross section of 6.7 pb) including the fit quality is shown in Table 6.11. The pretag and single, loose-tagged fits are shown in Figures 6.2 through 6.4. We derive the uncertainty from the relative error returned by the fitter, and we scale this error by the fit χ^2 per degree of freedom (*dof*) to account for our assumptions about the validity of the fit templates. In the one-jet bin, the fit is very poor ($\chi^2/dof = 9.6$), and the relative error is nearly 100%. While it is possible that there is an additional source of background not represented in the template fits, most likely this effect arises from missing jet corrections in the antielectrons; we expect one-jet antielectron events to be dominated by dijets, but we do not correct the antielectron energy, leading to an inherent asymmetry in the jet and electron E_T . The effect is small in events with multiple jets, and the χ^2/dof are quite reasonable there.

In addition to the fit error, we apply a 15% systematic error to the antielectron component, an upper limit on the contamination from real W electrons [75].

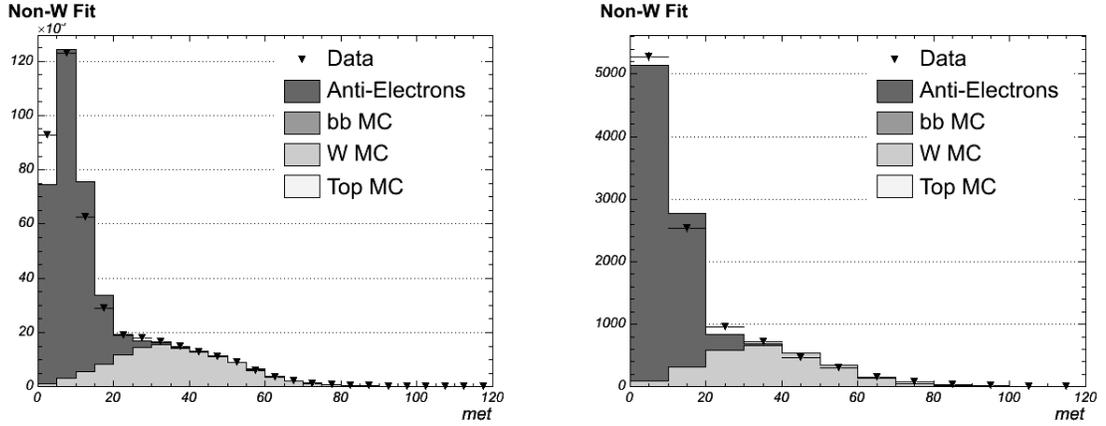


Figure 6.2: Missing E_T fits for the non- W component of the background for the pretag (left) and loose b -tagged (right) sample in the 1-jet bin.

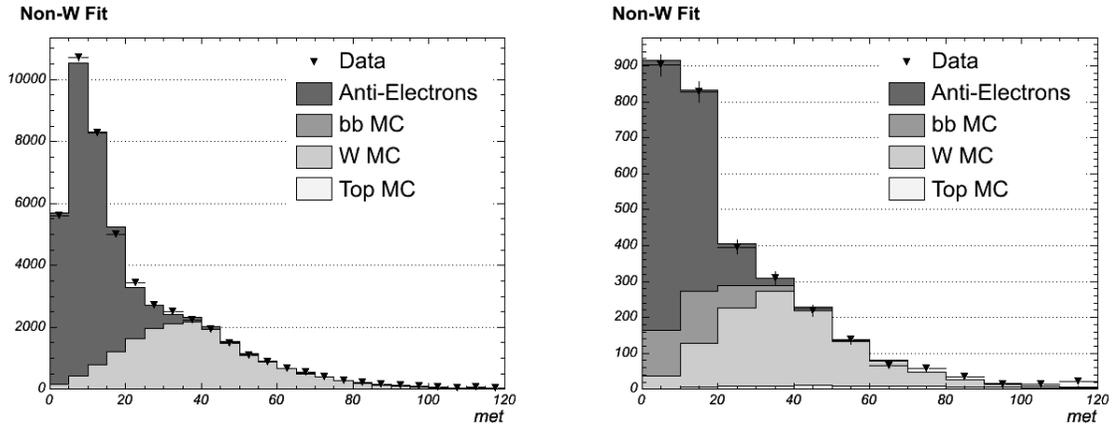


Figure 6.3: Missing E_T fits for the non- W component of the background for the pretag (left) and loose b -tagged (right) sample in the 2-jet bin.

Check of Non- W Background

We check the results of the missing E_T fits using the method employed in Run 1, in which the sidebands in both missing E_T and lepton isolation are used to infer the non- W content of the signal region [76]. The primary assumptions of this method are that the lepton's (or fake's) direction and the missing E_T are independent of the

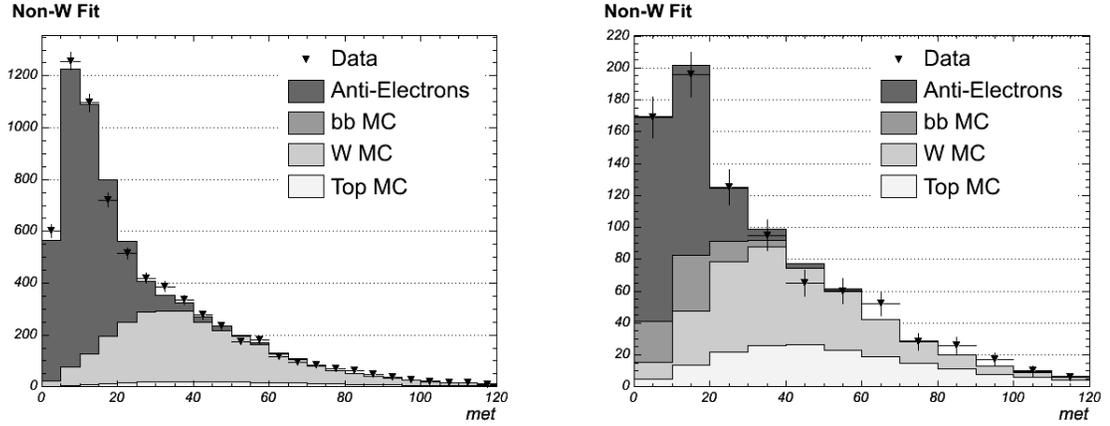


Figure 6.4: Missing E_T fits for the non- W component of the background for the pretag (left) and loose b -tagged (right) sample in the ≥ 3 -jet bin.

directions of nearby jets. Under these circumstances, the missing E_T and the lepton isolation (*i.e.*, the amount of *additional* energy – as a fraction of the lepton energy – surrounding it in a cone of 0.4 in ΔR) are uncorrelated. We divide the missing E_T -isolation plane into four separate regions and establish a similarity relationship between their contents. The region definitions are as follows:

- **A:** Missing $E_T < 15$ GeV, Isolation > 0.2
- **B:** Missing $E_T < 15$ GeV, Isolation < 0.1
- **C:** Missing $E_T > 20$ GeV, Isolation > 0.2
- **D:** Missing $E_T > 20$ GeV, Isolation < 0.1 (Signal Region)

We estimate the background contribution to the pretag sample (Region D) by assuming the ratio of non- W events in regions B and A is equal to that between regions D and C . Since the signal region is dominated by real W 's, we extract

	1-jet	2-jet	3-jet	4-jet	≥ 5 -jet
F_{non-W}^{pre} (%)	9.5 ± 6.4 (335)	16.0 ± 2.5 (5.6)		17.3 ± 2.8 (1.5)	
Loose					
F_{non-W}^{tag} (%) (χ^2/dof)	0.25 ± 0.05 (9.6)	1.2 ± 0.4 (2.7)		1.7 ± 0.3 (1.4)	
F_{non-W}^{2tag} (%) (χ^2/dof)	-	0.05 ± 0.10 (0.3)		0.13 ± 0.09 (2.6)	
N_{non-W}^{tag}	288.15 ± 43.93	227.48 ± 66.73	54.23 ± 7.02	12.00 ± 1.55	2.97 ± 0.38
N_{non-W}^{2tag}	-	9.14 ± 19.33	4.11 ± 2.95	0.91 ± 0.65	0.23 ± 0.16
Tight					
F_{non-W}^{tag} (%) (χ^2/dof)	0.18 ± 0.16 (8.2)	0.76 ± 0.21 (1.7)		1.08 ± 0.29 (2.78)	
F_{non-W}^{2tag} (%) (χ^2/dof)	-	0.03 ± 0.04 (0.3)		0.08 ± 0.11 (3.8)	
N_{non-W}^{tag}	211.99 ± 190.65	144.05 ± 39.07	34.96 ± 7.17	7.74 ± 1.59	1.92 ± 0.39
N_{non-W}^{2tag}	-	5.09 ± 7.14	2.38 ± 3.44	0.53 ± 0.76	0.13 ± 0.19

Table 6.11: Expected contribution from non- W backgrounds assuming a cross section of 6.7pb. The $t\bar{t}$ contribution is removed from the missing E_T distribution. Statistics are merged for all events with three or more jets in the fit, and errors include the systematics. The fit χ^2 is shown in parentheses.

F_{non-W} , the fraction of the pretag sample from such events. The pretag and tagged distributions in the missing E_T -isolation plane are shown in Figures 6.5 and 6.6. Electrons and muons are treated separately, and we assign a 25% systematic to this method by varying the borders for the four regions.

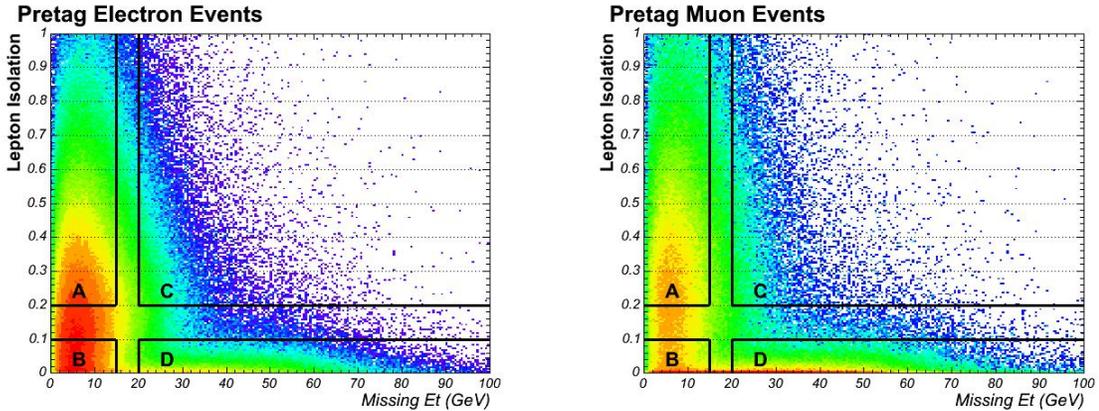


Figure 6.5: Missing E_T vs. Isolation distribution in electron (left) and muon (right) pretag events.

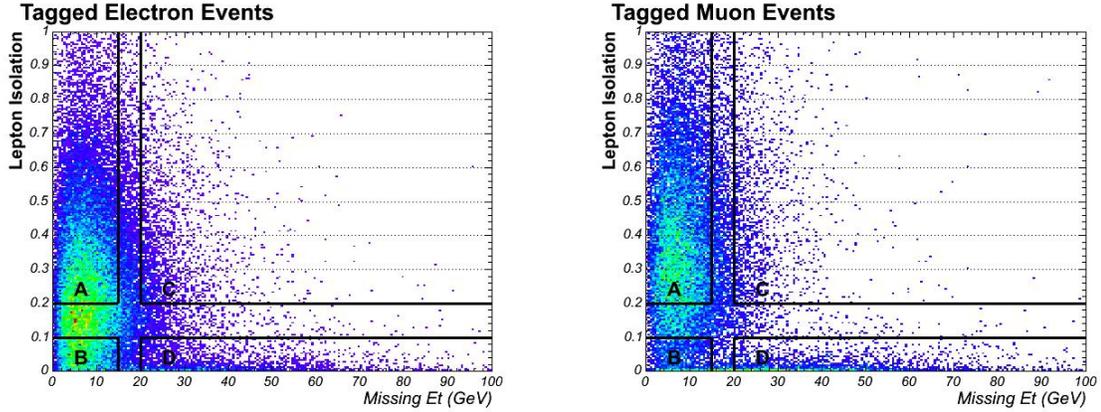


Figure 6.6: Missing E_T vs. Isolation distribution in electron (left) and muon (right) loose b -tagged events.

In the check, the non- W contribution to the b -tagged samples is calculated in two ways: the *tag* and *pretag* methods. The former assumes the same similarity relationship holds for tagged events, or, equivalently, that the ratio of b -tag rates in regions B and A is the same as the ratio of non- W b -tag rates in regions D and C . A systematic error of 33% is assigned to this method, derived by varying the boundaries of the four regions. The *pretag* method simply assumes that the non- W b -tag rate in regions D and B are the same, and the region B tag rate is applied to F_{non-W} . An additional systematic uncertainty of 20% is placed on this assumption, derived from the difference in tag rates between regions A and C , which is combined in quadrature with the initial uncertainty on F_{non-W} itself. In both cases, jets within a ΔR cone of 0.4 of the lepton are not counted. This effect is only relevant in Regions C and A .

The weighted average of these two methods is used for the single-tag estimate, but only the *pretag* method is used in the double-tag measurement, since statistics are poor. The double-tag estimate has a 60% systematic error (also derived from

comparing tag rates in Regions A and C).

A comparison of the non- W fractions for the pretag and tagged samples using the two methods is shown in Table 6.12. We find acceptable agreement, and use the missing E_T fit results for our central value.

	1-jet	2-jet	3-jet	≥ 4 -jet
Electron				
Pretag Fit (%)	14.3 ± 10.3	23.4 ± 3.7	24.8 ± 4.0	
Pretag MetIso (%)	20.9 ± 5.2	29.4 ± 7.4	29.7 ± 7.5	28.0 ± 7.4
Tag Fit (%)	0.24 ± 0.05	1.1 ± 0.3	1.6 ± 0.3	
Tag MetIso(%)	0.38 ± 0.09	1.2 ± 0.3	1.9 ± 0.5	3.9 ± 1.1
Muon				
Pretag Fit (%)	4.3 ± 17.5	10.1 ± 2.9	6.3 ± 5.2	
Pretag MetIso (%)	6.8 ± 1.7	9.2 ± 2.3	10.6 ± 2.7	7.7 ± 2.3
Tag Fit (%)	0.13 ± 0.07	0.37 ± 0.03	0.9 ± 3.0	
Tag MetIso(%)	0.13 ± 0.03	0.44 ± 0.11	0.86 ± 0.23	1.9 ± 0.6

Table 6.12: Comparison of the new fitting technique with the older missing E_T vs. Isolation method for calculating non- W . These results are for the tight b -tagger and assume a cross section of 6.7 pb. Systematic errors are included.

6.4.3 W + Heavy Flavor

The linchpin of the Method 2 cross section measurement is the assumption that after subtracting off the non- W , purely electroweak, and signal contributions to the pretag and b -tagged samples, what remains is W +jets. This assumption alleviates the necessity of calculating absolute cross sections for all W processes. Instead, we measure the *heavy flavor fractions*, the fractions of W +jets contains b and c jets [69]. The method is partially Monte Carlo-driven, but the final measurements are checked against generic jet data. We consider contributions from $Wc\bar{c}$, $Wb\bar{b}$, and Wc , and we explicitly account for all possible final state multiplicities with independent ALPGEN samples. The various samples and their cross sections (used to weight each sample before merging) are listed in Table 6.13, where a p denotes a light flavor parton in the final state; the samples listed cover events with final state parton multiplicities from zero to four. Using a sample of all-jet events, we find that the heavy flavor fractions predicted by ALPGEN are accurate to 30%, and so we scale the W +Heavy Flavor fractions by a K factor of 1.0 ± 0.3 [70]. The derivation of the heavy flavor fractions and K , and the construction of a true, weighted W +Heavy Flavor sample from these simulations is described in depth in Appendix B.

We show the final results for the fractions in Table 6.14. We assume that the fraction is the same for events with 4 and ≥ 5 jets. The b -tagging efficiency for each class of event is measured in Monte Carlo, then corrected for data-Monte Carlo differences (SF), as we do for the signal estimate. A summary of the efficiencies is shown in Table 6.15. The efficiency times the heavy flavor fraction, summed over the four classes of heavy flavor, yields the aggregate b -tag rate for W +Heavy Flavor; this

total rate times the corrected pretag sample provides the estimate of the b -tagged background expectation:

$$N_{tag}^{W+HF} = N_{pre}^W \left(\sum_i \epsilon_i^{hf} F_i^{hf} \right), \quad (6.4)$$

where the F^{hf} are again the heavy flavor fractions, and the sum i is over the number and type of heavy flavor jets.

Sample	Cross Section (pb)
$W+0p$	1810
$W+1p$	225
$W+2p$	35.3
$W+3p$	5.59
$W+4p$	1.03
$Wb\bar{b}+0p$	2.98
$Wb\bar{b}+1p$	0.888
$Wb\bar{b}+2p$	0.287
$Wc\bar{c}+0p$	5.00
$Wc\bar{c}+1p$	1.79
$Wc\bar{c}+2p$	0.628
$Wc+0p$	17.1
$Wc+1p$	3.39
$Wc+2p$	0.507
$Wc+3p$	0.083

Table 6.13: ALPGEN Monte Carlo samples and the generated cross sections used to weight the W +Heavy Flavor samples.

6.4.4 Mistags

After accounting for all non- W events, heavy- and light flavor b -tags in top and other electroweak processes, and heavy flavor b -tags in W +jets, the only remaining contribution is due to b -tags in W +Light Flavor, or *mistags*. This background is

	1-jet	2-jet	3-jet	≥ 4 -jet
Corrected Fractions (%)				
$Wb\bar{b}$, 1 b	0.7 ± 0.3	1.4 ± 0.6	2.4 ± 1.0	3.3 ± 1.4
$Wb\bar{b}$, 2 b		0.9 ± 0.4	1.8 ± 0.7	3.3 ± 1.3
$Wc\bar{c}$ or Wc , 1 c	5.5 ± 1.7	8.7 ± 3.2	11.7 ± 3.8	13.3 ± 4.9
$Wc\bar{c}$, 2 c		1.4 ± 0.6	3.3 ± 1.3	6.0 ± 2.4

Table 6.14: The heavy flavor fractions for W +Heavy Flavor background, derived in Appendix B. These are the fractions of generic W +jets (in %) containing heavy flavor jets, sorted by the amount of heavy flavor and the physical process.

expected to dominate the lower jet bins with the loose b -tagger, as outlined in Section 5.3. To determine the mistag component, we apply the parameterized *negative mistag matrix* [65] on the pretag data sample, then correct the total predicted number of tags for events which are not W +Light Flavor. The matrix applies the generic jet b -tag rate to jets in the pretag sample and determines a probability for each jet to be b -tagged; the sum of these probabilities is the expected number of mistagged jets. Once each jet has been assigned a mistag probability, we can trivially construct the probability for each event to have ≥ 1 or ≥ 2 mistagged jets.

This estimate is corrected by the jet E_T -dependent *mistag asymmetry*, as outlined in Section 6.2.2, to account for the imbalance in positive and negative b -tags for light jets, as well as the effect of heavy flavor b -tags in the samples where the matrix is created [61]. The derivation of these corrections can be found in Section 5.3.3.

We multiply the total prediction by the fraction of the pretag sample not attributed to $t\bar{t}$, dibosons and single top, non- W , and W +Heavy Flavor. In general, however, heavy flavor jets from top will have higher mistag probabilities than light jets (b -jets have higher E_T and a larger track multiplicity), meaning that an average $t\bar{t}$ event will contribute more to the total prediction than a typical W +4 jets event.

HF jets	1-jet	2-jet	3-jet	≥ 4 -jet
Single Loose Efficiency (%)				
1 b	36 ± 2	40 ± 3	43 ± 3	47 ± 3
2 b		62 ± 2	64 ± 3	65 ± 3
1 c	9 ± 2	12 ± 2	14 ± 2	18 ± 2
2 c		19 ± 2	21 ± 3	25 ± 3
Single Tight Efficiency (%)				
1 b	31 ± 2	34 ± 2	36 ± 2	37 ± 3
2 b		55 ± 2	56 ± 2	57 ± 3
1 c	7 ± 2	8 ± 2	9 ± 2	11 ± 2
2 c		14 ± 2	15 ± 2	17 ± 3
Double Loose Efficiency (%)				
1 b		$1 \pm <1$	$2 \pm <1$	4 ± 1
2 b		16 ± 2	19 ± 2	19 ± 3
1 c		$0 \pm <1$	$1 \pm <1$	2 ± 1
2 c		$1 \pm <1$	$2 \pm <1$	2 ± 1
Double Tight Efficiency (%)				
1 b		$0 \pm <1$	$1 \pm <1$	$2 \pm <1$
2 b		12 ± 1	13 ± 1	13 ± 2
1 c		$0 \pm <1$	$0 \pm <1$	$0 \pm <1$
2 c		$1 \pm <1$	1 ± 1	$1 \pm <1$

Table 6.15: b -Tagging efficiency for the various classes of heavy flavor event in $W+jets$. Efficiencies are corrected for the scale factor SF .

The per-event mistag probabilities for 3-jet events in the data and Monte Carlo are shown in Figures 6.7 and 6.8. The average value in the data is slightly lower than Monte Carlo, even without heavy flavor jets, but we assume the ratio between top and W +jets is reasonably well-determined in simulation. We find that top events have per-event probabilities that are 25% higher than W +jets for single-tags, both loose and tight, and 35% (loose) and 40% (tight) higher for double-tags. Therefore, in the pretag correction, we weight top events by factor of 1.25 in the ≥ 1 -tag correction and 1.35 or 1.40 in the ≥ 2 -tag analysis, depending on the b -tagger. We ignore the smaller difference between $Wb\bar{b}$ and W +light flavor.

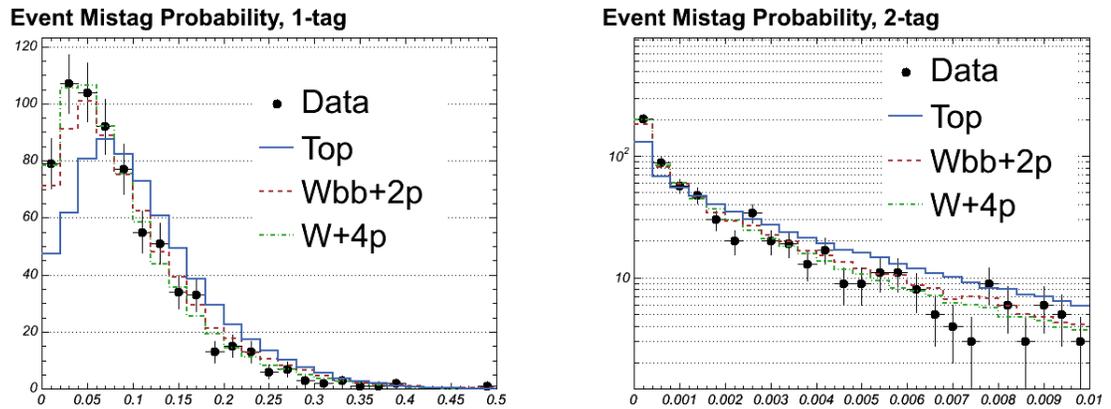


Figure 6.7: Per-event mistag probabilities derived by applying the loose corrected mistag matrix to 3-jet events in data and simulation. We correct for the higher probabilities in $t\bar{t}$ by taking the ratio of these distributions in MC, 1.25 for single-tags and 1.35 for double-tags.

The raw single-tag rates are shown in Table 6.16. A summary of the mistag background corrected for top and heavy flavor is in Table 6.17.

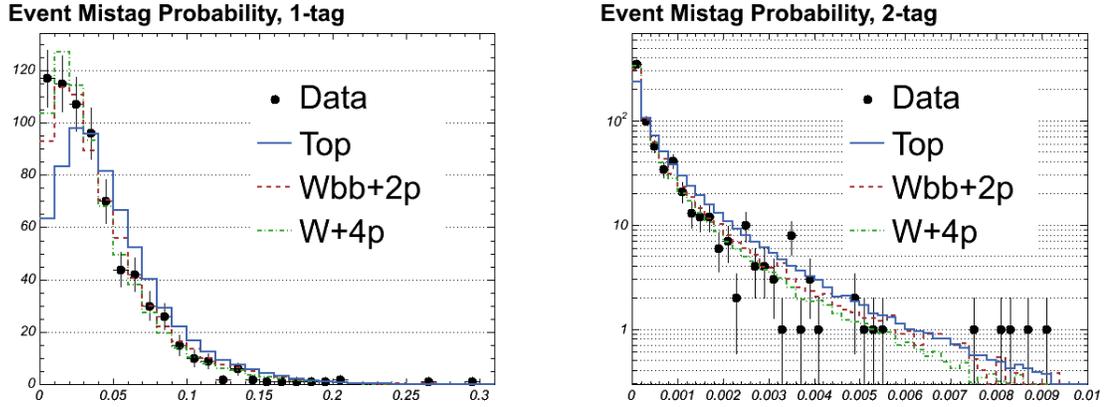


Figure 6.8: Per-event mistag probabilities derived by applying the tight corrected mistag matrix to 3-jet events in data and simulation. We correct for the higher probabilities in $t\bar{t}$ by taking the ratio of these distributions in MC, 1.25 for single-tags and 1.40 for double-tags.

Total Mistags in Data		
1	517 ± 108	1320 ± 290
2	226 ± 43	555 ± 120
3	79.2 ± 15.2	182 ± 35
4	27.3 ± 5.8	62.0 ± 14.6
≥ 5	9.3 ± 2.1	20.4 ± 4.6

Table 6.16: The mistag matrix prediction from the pretag data set. The mistag asymmetries are taken into account, and we combine the matrix values from each jet to produce an event tagging probability.

6.5 Summary Tables

Merging all the background methods described in the text, we get a full estimate of the tagged $W+jets$ sample content. Summary tables for the two taggers and single- and double-tags are presented in Tables 6.18 to 6.21, and stacked plots illustrating this composition are shown in Figures 6.9 and 6.10. The excess of b -tagged events with ≥ 3 jets points to a $t\bar{t}$ cross section higher than 6.7 pb, the subject of Chapter 7.

	1-jet	2-jet	3-jet	4-jet	≥ 5 -jet
Tight, ≥ 1 -tag	431.54 ± 94.94	153.48 ± 33.76	41.50 ± 9.13	7.62 ± 1.68	1.88 ± 0.41
Tight, ≥ 2 -tag	-	0.68 ± 0.20	0.46 ± 0.13	0.13 ± 0.04	0.04 ± 0.01
Loose, ≥ 1 -tag	1102.29 ± 242.50	377.75 ± 83.11	95.36 ± 20.98	17.30 ± 3.81	4.19 ± 0.92
Loose, ≥ 2 -tag	-	3.85 ± 1.12	2.35 ± 0.68	0.68 ± 0.20	0.22 ± 0.06

Table 6.17: Expected contribution from tagged W +Light Flavor events after correcting for the light flavor asymmetry ($\alpha\beta$). The raw expectation is derived from the full dataset, and the corrected total accounts for $t\bar{t}$ with a cross section of 6.7 pb.

Process	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	115313	18922	3136	694	172
Pretag Top (6.7 pb)	24.29 ± 1.50	130.50 ± 8.07	261.01 ± 16.14	246.90 ± 15.27	82.13 ± 5.08
WW	7.30 ± 0.77	18.86 ± 1.98	5.99 ± 0.63	1.51 ± 0.16	0.41 ± 0.04
WZ	4.06 ± 0.42	8.68 ± 0.90	2.34 ± 0.24	0.55 ± 0.06	0.13 ± 0.01
ZZ	0.12 ± 0.01	0.33 ± 0.04	0.18 ± 0.02	0.05 ± 0.01	0.01 ± 0.00
Single Top (s-ch)	4.75 ± 1.07	17.49 ± 3.95	5.46 ± 1.23	1.15 ± 0.26	0.25 ± 0.06
Single Top (t-ch)	16.69 ± 2.99	20.90 ± 3.75	4.59 ± 0.82	0.70 ± 0.12	0.08 ± 0.01
$Z \rightarrow \tau\tau$	22.30 ± 5.40	13.40 ± 3.24	6.08 ± 1.47	1.33 ± 0.32	0.19 ± 0.05
Wbb	225.88 ± 101.02	139.93 ± 62.58	37.97 ± 16.98	8.70 ± 3.89	1.63 ± 0.73
$Wc\bar{c}, Wc$	396.84 ± 177.47	133.43 ± 59.67	32.37 ± 14.48	7.08 ± 3.17	1.33 ± 0.59
W +Light Flavor	431.54 ± 94.94	153.48 ± 33.76	41.50 ± 9.13	7.62 ± 1.68	1.88 ± 0.41
Non- W	211.99 ± 190.65	144.05 ± 39.07	34.96 ± 7.17	7.74 ± 1.59	1.92 ± 0.39
Background	1321.47 ± 352.23	650.54 ± 134.28	171.43 ± 35.78	36.41 ± 9.23	7.83 ± 1.99
Top (6.7 pb)	8.17 ± 0.83	65.62 ± 6.68	146.30 ± 14.90	151.86 ± 15.46	51.33 ± 5.23
Tags	1697	835	358	224	75

Table 6.18: Summary of signal and backgrounds for the tight, single-tag sample. We assume a top cross section of 6.7 pb.

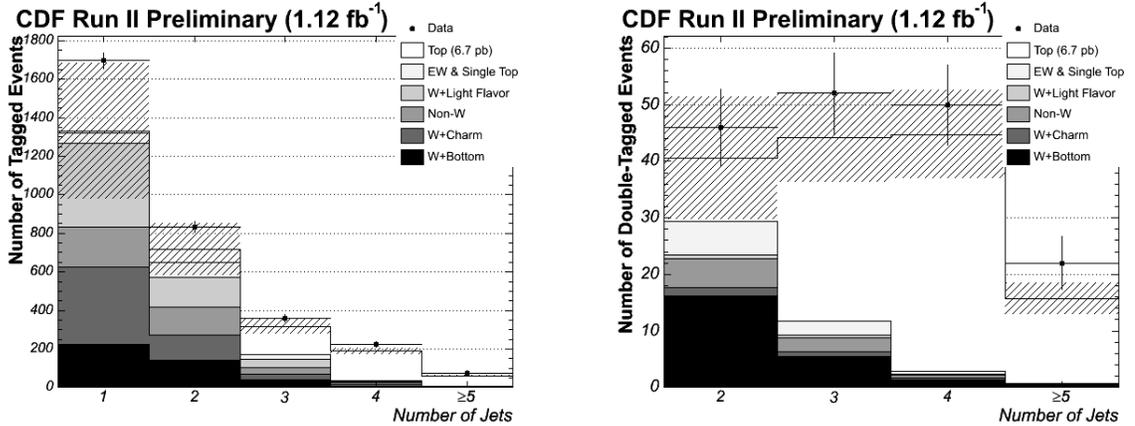


Figure 6.9: Expected and observed numbers of tagged events sorted by jet multiplicity, after requiring ≥ 1 (left) and ≥ 2 (right) tight tags. Normalizations here are the same as in Tables 6.18 and 6.19. The hashed region corresponds to the systematic uncertainty on the background.

Process	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	18922	3136	694	172
Pretag Top (6.7 pb)	130.50 ± 8.07	261.01 ± 16.14	246.90 ± 15.27	82.13 ± 5.08
WW	0.10 ± 0.02	0.13 ± 0.02	0.09 ± 0.02	0.03 ± 0.01
WZ	1.40 ± 0.26	0.42 ± 0.08	0.07 ± 0.01	0.02 ± 0.00
ZZ	0.04 ± 0.01	0.03 ± 0.01	0.01 ± 0.00	0.00 ± 0.00
Single Top (s-ch)	3.97 ± 1.21	1.34 ± 0.41	0.30 ± 0.09	0.07 ± 0.02
Single Top (t-ch)	0.27 ± 0.07	0.43 ± 0.11	0.10 ± 0.03	0.02 ± 0.00
$Z \rightarrow \tau\tau$	0.12 ± 0.04	0.16 ± 0.05	0.04 ± 0.01	0.03 ± 0.01
$Wb\bar{b}$	16.19 ± 7.24	5.43 ± 2.43	1.35 ± 0.60	0.25 ± 0.11
$Wc\bar{c}, Wc$	1.51 ± 0.67	0.93 ± 0.42	0.29 ± 0.13	0.05 ± 0.02
W +Light Flavor	0.68 ± 0.20	0.46 ± 0.13	0.13 ± 0.04	0.04 ± 0.01
Non-W	5.09 ± 7.14	2.38 ± 3.44	0.53 ± 0.76	0.13 ± 0.19
Background	29.37 ± 10.74	11.71 ± 5.13	2.90 ± 1.86	0.65 ± 0.55
Top (6.7 pb)	11.17 ± 2.03	32.45 ± 5.90	41.82 ± 7.60	15.08 ± 2.74
Tags	46	52	50	22

Table 6.19: Summary of signal and backgrounds for the tight, double-tag sample. We assume a top cross section of 6.7 pb.

Process	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	115313	18922	3136	694	172
Pretag Top (6.7 pb)	24.29 ± 1.50	130.50 ± 8.07	261.01 ± 16.14	246.90 ± 15.27	82.13 ± 5.08
WW	12.11 ± 1.27	29.74 ± 3.12	9.89 ± 1.04	2.46 ± 0.26	0.67 ± 0.07
WZ	5.29 ± 0.55	10.88 ± 1.13	3.08 ± 0.32	0.73 ± 0.08	0.18 ± 0.02
ZZ	0.15 ± 0.02	0.41 ± 0.05	0.23 ± 0.03	0.06 ± 0.01	0.02 ± 0.00
Single Top (s-ch)	5.64 ± 1.27	20.08 ± 4.54	6.28 ± 1.42	1.32 ± 0.30	0.28 ± 0.06
Single Top (t-ch)	19.56 ± 3.51	25.03 ± 4.49	5.44 ± 0.97	0.80 ± 0.14	0.09 ± 0.02
$Z \rightarrow \tau\tau$	33.45 ± 8.10	21.90 ± 5.30	10.09 ± 2.44	2.21 ± 0.53	0.35 ± 0.08
$Wb\bar{b}$	266.15 ± 119.02	162.82 ± 72.82	44.70 ± 19.99	10.33 ± 4.62	1.94 ± 0.87
$Wc\bar{c}, Wc$	545.64 ± 244.02	189.76 ± 84.86	48.97 ± 21.90	10.86 ± 4.86	2.03 ± 0.91
W +Light Flavor	1102.29 ± 242.50	377.75 ± 83.11	95.36 ± 20.98	17.30 ± 3.81	4.19 ± 0.92
Non-W	288.15 ± 56.35	227.48 ± 72.31	54.23 ± 9.66	12.00 ± 2.14	2.97 ± 0.53
Background	2278.43 ± 442.45	1065.85 ± 194.18	278.28 ± 50.50	58.07 ± 12.59	12.73 ± 2.74
Top (6.7 pb)	9.78 ± 1.00	76.54 ± 7.79	169.78 ± 17.29	174.14 ± 17.73	58.99 ± 6.01
Tags	2729	1277	499	284	90

Table 6.20: Summary of signal and backgrounds for the loose, single-tag sample. We assume a top cross section of 6.7 pb.

Process	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	18922	3136	694	172
Pretag Top (6.7 pb)	130.50 ± 8.07	261.01 ± 16.14	246.90 ± 15.27	82.13 ± 5.08
WW	0.33 ± 0.06	0.37 ± 0.07	0.18 ± 0.03	0.08 ± 0.01
WZ	1.92 ± 0.35	0.56 ± 0.10	0.13 ± 0.02	0.04 ± 0.01
ZZ	0.06 ± 0.01	0.05 ± 0.01	0.01 ± 0.00	0.00 ± 0.00
Single Top (s-ch)	5.75 ± 1.76	1.96 ± 0.60	0.45 ± 0.14	0.10 ± 0.03
Single Top (t-ch)	0.57 ± 0.15	0.67 ± 0.17	0.16 ± 0.04	0.03 ± 0.01
$Z \rightarrow \tau\tau$	0.29 ± 0.09	0.40 ± 0.13	0.09 ± 0.03	0.05 ± 0.02
$Wb\bar{b}$	23.51 ± 10.51	8.11 ± 3.63	2.16 ± 0.97	0.40 ± 0.18
$Wc\bar{c}, Wc$	3.96 ± 1.77	2.41 ± 1.08	0.82 ± 0.37	0.15 ± 0.07
W +Light Flavor	3.85 ± 1.12	2.35 ± 0.68	0.68 ± 0.20	0.22 ± 0.06
Non-W	9.14 ± 19.33	4.11 ± 2.95	0.91 ± 0.65	0.23 ± 0.16
Background	49.37 ± 22.56	20.99 ± 6.54	5.60 ± 2.75	1.30 ± 0.83
Top (6.7 pb)	16.58 ± 3.01	48.31 ± 8.78	61.62 ± 11.21	22.36 ± 4.07
Tags	82	79	79	33

Table 6.21: Summary of signal and backgrounds for the loose, double-tag sample. We assume a top cross section of 6.7 pb.

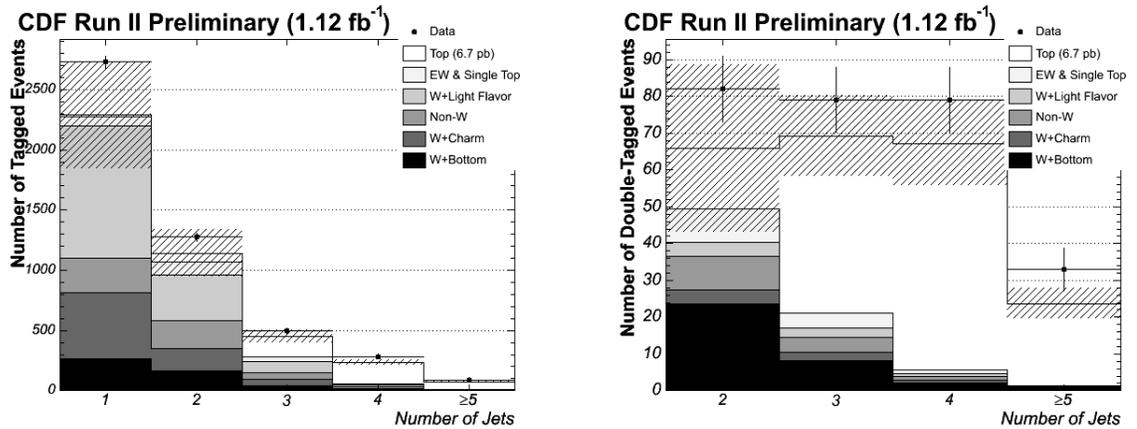


Figure 6.10: Expected and observed numbers of tagged events sorted by jet multiplicity, after requiring ≥ 1 (left) and ≥ 2 (right) loose tags. Normalizations here are the same as in Tables 6.20 and 6.21. The hashed region corresponds to the systematic uncertainty on the background.

Chapter 7

Results

In this chapter, we present the measurement of the top quark pair production cross section ($\sigma_{t\bar{t}}$) in the lepton+jets decay channel using SECVTX b -tags. The result follows directly from Chapter 6, in which $\sigma_{t\bar{t}}$ is fixed at 6.7 pb and the non-top contribution to the lepton+jets sample is evaluated with Method 2 [68]. The cross section is ultimately calculated using the following formula:

$$\sigma_{t\bar{t}} = \frac{N_{obs} - N_{bkg}}{\epsilon_{tag}\epsilon_{pretag} \int \mathcal{L} dt} \quad (7.1)$$

- N_{obs} : Number of events in data passing event selection
- N_{bkg} : Number of non- $t\bar{t}$ events expected to pass event selection
- ϵ_{tag} : Efficiency to b -tag a $t\bar{t}$ event in data
- ϵ_{pretag} : $t\bar{t}$ acceptance (geometric acceptance and event selection efficiency)
- $\int \mathcal{L} dt$: Integrated luminosity

As described in Chapter 6, N_{bkg} is dependent on $\sigma_{t\bar{t}}$, in that the normalization for W +jets backgrounds is derived from the pretag data sample after removing top and other known contributions. We therefore determine the cross section through an iterative procedure, in which we require that the observed and predicted numbers of b -tagged events are the same. If more b -tagged events are observed in data than are expected, the assumed value of $\sigma_{t\bar{t}}$ is scaled up to include the excess, and the backgrounds are re-evaluated.¹

The signal and background contributions are evaluated using a procedure identical to that described in Chapter 6. In this chapter, we start by using those backgrounds (for $\sigma_{t\bar{t}} = 6.7$ pb) to optimize the event selection for a cross section measurement, and the results of this are presented in Section 7.1. We present the signal and background expectations for the optimized selection in Sections 7.2 and 7.3, and results of the measurement are presented in Section 7.4. We discuss additional sources of systematic error in Section 7.5.

7.1 Optimization

The appropriate event selection for top quark measurements varies with the analysis, depending on the tolerance for backgrounds and other sources of uncertainty. This analysis is a counting experiment, sensitive more to the overall signal purity than to the kinematics of signal and backgrounds. A previous, similar measurement was limited by systematic uncertainties with one-quarter of this dataset [77]. Therefore,

¹Since the total background gets smaller with a larger assumed cross section, the predicted number of b -tagged events in this case will remain lower than the observed at each iterative step. We simply stop when the result is stable to $<0.01\%$.

Cut	0σ (%)	$+1\sigma$ (%)	-1σ (%)	Deviation
Baseline	8.91 ± 0.01	9.17 ± 0.01	8.64 ± 0.01	3.0%
$N_{jet} > 3$	4.97 ± 0.01	5.38 ± 0.01	4.55 ± 0.01	8.4%
$H_T > 200$ GeV	8.80 ± 0.01	9.07 ± 0.01	8.49 ± 0.01	3.3%
$H_T > 250$ GeV	7.97 ± 0.01	8.39 ± 0.01	7.50 ± 0.01	5.6%
$H_T > 300$ GeV	5.85 ± 0.01	6.46 ± 0.01	5.24 ± 0.01	10.4%
Missing $E_T > 30$ GeV	7.72 ± 0.01	8.01 ± 0.01	7.42 ± 0.01	3.8%
Jet $E_T^1 > 30$ GeV	8.90 ± 0.01	9.16 ± 0.01	8.62 ± 0.01	3.0%
Jet $E_T^1 > 50$ GeV	8.32 ± 0.01	8.65 ± 0.01	7.98 ± 0.01	4.0%

Table 7.1: Effect on the $t\bar{t}$ acceptance, in percent, of the systematic error in the jet energy scale. This error increases with tightened event selection, and therefore figures into the event selection tuning.

we can improve the precision of the result by tightening the event selection, which suppresses the background and reduces the systematic error at the cost of statistical power.

We scan cuts on the missing E_T , the leading jet E_T , the number of jets, and H_T , the scalar sum of the jet and lepton p_T 's and the missing E_T . The optimization is performed with respect to two figures of merit: the statistical power, $\frac{S}{\sqrt{S+B}}$, and the total significance or ‘‘Measurement Quality,’’ $\frac{S}{\sqrt{S+B+\Delta B^2}}$. The second of these includes the effect of systematic uncertainties as well, and is therefore a better indicator of the expected precision. We include the effect of the jet energy correction scale (JES) uncertainty in the denominator, since the JES systematic error becomes large with tighter energy cuts [48]. Table 7.1 shows the effect on the raw top acceptance of 1σ fluctuations in the energy scale for a few indicative optimization cuts.

We find the most improvement in the total significance by imposing a tight requirement on H_T , since $t\bar{t}$ events will typically be more energetic than background. Additionally, by moving the missing E_T cut up from 20 GeV, we can reject more non-

W background, which has a relative uncertainty of 30%. In Figures 7.1 to 7.4, we show the optimization curves for H_T and, sequentially, missing E_T . For consistency between tight and loose, we choose cuts that are approximately optimal for both, **requiring $H_T > 250$ ($H_T > 200$) GeV for single (double) tags and missing $E_T > 30$ GeV for all.** With these requirements, we reduce the total background by 70% and lose only 20% of the $t\bar{t}$ signal. From the maximum value of the “Measurement Quality” in the plots, we expect *a priori* that the inclusive cross section with the tight tagger will be the most precise result by about 10%, with the loose, double the better double-tagged result by about 20%. The pretag and tagged data yields for the optimized samples are shown in Table 7.2.

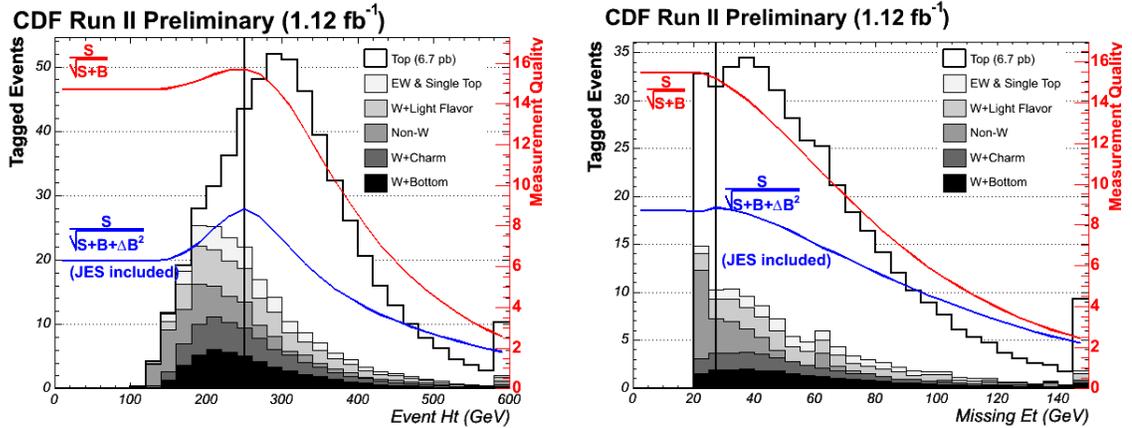


Figure 7.1: Optimization curves for the tight, inclusive tag analysis assuming a cross section of 6.7 pb. The top curve shows the statistical power of the measurement as a function of the cut value, and the bottom shows the power when including systematics from the background and due to the jet energy scale. A vertical line is drawn at the peak of the lower curve (if one exists), showing the optimal cut value. The left plot uses the baseline cuts, and the right is after a cut on $H_T > 250$ GeV.

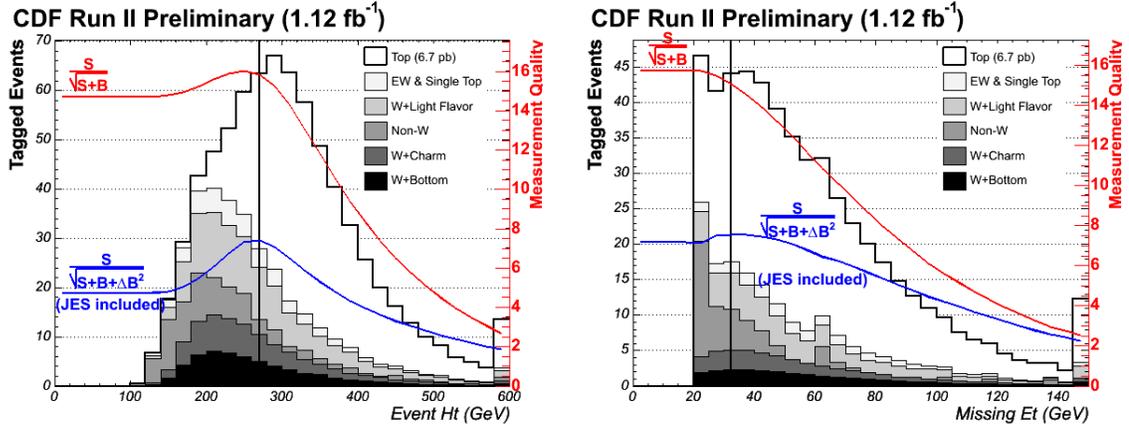


Figure 7.2: Optimization curves for the loose, inclusive tag analysis assuming a cross section of 6.7 pb. The top curve shows the statistical power of the measurement as a function of the cut value, and the bottom shows the power when including systematics from the background and due to the jet energy scale. A vertical line is drawn at the peak of the lower curve (if one exists), showing the optimal cut value. The left plot uses the baseline cuts, and the right is after a cut on $H_T > 250$ GeV.

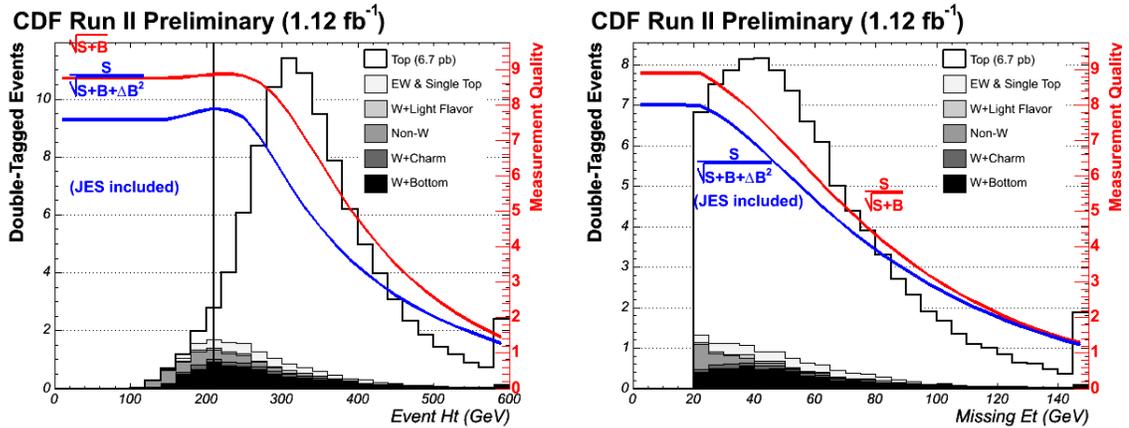


Figure 7.3: Optimization curves for the tight, double tag analysis assuming a cross section of 6.7 pb. The top curve shows the statistical power of the measurement as a function of the cut value, and the bottom shows the power when including systematics from the background and due to the jet energy scale. A vertical line is drawn at the peak of the lower curve (if one exists), showing the optimal cut value. The left plot uses the baseline cuts, and the right is after a cut on $H_T > 200$ GeV.

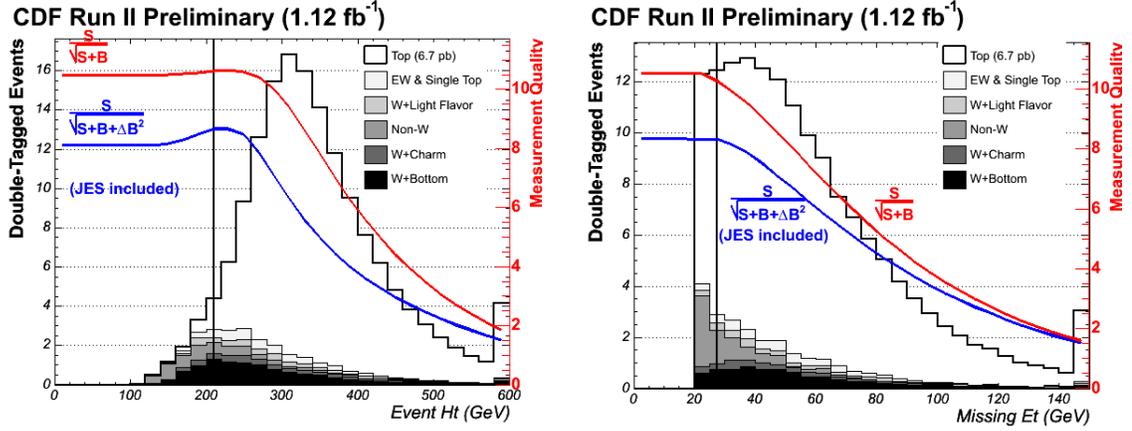


Figure 7.4: Optimization curves for the loose, double tag analysis assuming a cross section of 6.7 pb. The top curve shows the statistical power of the measurement as a function of the cut value, and the bottom shows the power when including systematics from the background and due to the jet energy scale. A vertical line is drawn at the peak of the lower curve (if one exists), showing the optimal cut value. The left plot uses the baseline cuts, and the right is after a cut on $H_T > 200$ GeV.

	1 jet	2 jet	3 jet	4 jet	≥ 5 jet
Tight, Inclusive					
Pretag	78903	12873	870	413	130
Tagged	1067	585	185	169	62
Tight, Double					
Pretag	78903	12873	1515	507	132
Tagged	-	34	41	48	20
Loose, Inclusive					
Pretag	78903	12873	870	413	130
Tagged	1788	888	240	209	72
Loose, Double					
Pretag	78903	12873	1515	507	132
Tagged	-	63	64	72	29

Table 7.2: Pretag and tag event yields after optimization cuts are applied, sorted by the number of tight jets in the event.

7.2 Top Signal

The top signal acceptance and efficiency are calculated using identical samples and techniques to those presented in Chapter 6. The only differences are the optimization cuts on missing E_T and H_T and the normalization of the top expectation. Tables 7.3 and 7.4 show the acceptances for the optimized single-tag and double-tag samples, respectively. These have been corrected for the appropriate lepton identification scale factors, isolation systematics, trigger efficiencies, etc. The total acceptance for ≥ 3 -jet events in the single-tag analysis is $6.2 \pm 0.1\%$, and for double-tags it is $6.7 \pm 0.1\%$, compared to $7.9 \pm 0.2\%$ for the baseline cuts.

The tagging efficiency is also calculated according to the prescription from Chapter 6. A summary of the tight and loose single-tag and double-tag efficiencies after optimization cuts is shown in Table 7.5.

Corrected Acceptance (%)					
	1 jet	2 jet	3 jet	4 jet	≥ 5 jet
2/02-8/04					
CEM	0.167 ± 0.006	0.865 ± 0.019	1.379 ± 0.027	1.476 ± 0.028	0.514 ± 0.012
CMUP	0.087 ± 0.004	0.473 ± 0.015	0.776 ± 0.022	0.814 ± 0.023	0.274 ± 0.010
CMX	0.039 ± 0.003	0.195 ± 0.008	0.296 ± 0.011	0.320 ± 0.012	0.108 ± 0.005
9/04-8/05					
CEM	0.164 ± 0.005	0.871 ± 0.018	1.414 ± 0.027	1.504 ± 0.028	0.522 ± 0.012
CMUP	0.093 ± 0.004	0.476 ± 0.014	0.796 ± 0.022	0.839 ± 0.022	0.290 ± 0.010
CMX	0.035 ± 0.002	0.187 ± 0.008	0.305 ± 0.011	0.325 ± 0.012	0.112 ± 0.005
CMX (Mini/Key)	0.008 ± 0.001	0.042 ± 0.003	0.067 ± 0.004	0.072 ± 0.004	0.025 ± 0.002
9/05-2/06					
CEM	0.159 ± 0.006	0.864 ± 0.020	1.424 ± 0.029	1.513 ± 0.030	0.516 ± 0.013
CMUP	0.086 ± 0.004	0.482 ± 0.016	0.788 ± 0.023	0.819 ± 0.024	0.288 ± 0.011
CMX	0.035 ± 0.003	0.192 ± 0.009	0.307 ± 0.012	0.346 ± 0.013	0.115 ± 0.006
CMX (Mini/Key)	0.008 ± 0.001	0.040 ± 0.003	0.070 ± 0.005	0.073 ± 0.005	0.025 ± 0.002
3/06-8/07					
CEM	0.160 ± 0.007	0.828 ± 0.021	1.384 ± 0.031	1.461 ± 0.033	0.522 ± 0.015
CMUP	0.086 ± 0.005	0.477 ± 0.017	0.789 ± 0.025	0.828 ± 0.026	0.284 ± 0.012
CMX	0.039 ± 0.003	0.189 ± 0.010	0.307 ± 0.014	0.336 ± 0.015	0.116 ± 0.007
CMX (Mini/Key)	0.006 ± 0.001	0.040 ± 0.004	0.067 ± 0.005	0.072 ± 0.005	0.023 ± 0.003

Table 7.3: Corrected acceptances (ϵ_{pretag}) for the various run ranges and subdetectors based on $175 \text{ GeV}/c^2$ $t\bar{t}$ Monte Carlo, after applying a 30 GeV missing E_T cut and a 250 GeV H_T cut for ≥ 3 -jet events. This acceptance is selected for events with one or more tags.

Corrected Acceptance (%)					
	1 jet	2 jet	3 jet	4 jet	≥ 5 jet
2/02-8/04					
CEM	0.167 ± 0.006	0.865 ± 0.019	1.611 ± 0.031	1.552 ± 0.030	0.519 ± 0.013
CMUP	0.087 ± 0.004	0.473 ± 0.015	0.905 ± 0.025	0.852 ± 0.023	0.276 ± 0.010
CMX	0.039 ± 0.003	0.195 ± 0.008	0.352 ± 0.013	0.338 ± 0.013	0.109 ± 0.005
9/04-8/05					
CEM	0.164 ± 0.005	0.871 ± 0.018	1.652 ± 0.030	1.577 ± 0.029	0.526 ± 0.012
CMUP	0.093 ± 0.004	0.476 ± 0.014	0.924 ± 0.024	0.877 ± 0.023	0.292 ± 0.010
CMX	0.035 ± 0.002	0.187 ± 0.008	0.361 ± 0.013	0.344 ± 0.012	0.113 ± 0.005
CMX (Mini/Key)	0.008 ± 0.001	0.042 ± 0.003	0.081 ± 0.005	0.076 ± 0.004	0.025 ± 0.002
9/05-2/06					
CEM	0.159 ± 0.006	0.864 ± 0.020	1.660 ± 0.033	1.583 ± 0.032	0.520 ± 0.013
CMUP	0.086 ± 0.004	0.482 ± 0.016	0.918 ± 0.026	0.857 ± 0.025	0.289 ± 0.011
CMX	0.035 ± 0.003	0.192 ± 0.009	0.359 ± 0.014	0.363 ± 0.014	0.116 ± 0.006
CMX (Mini/Key)	0.008 ± 0.001	0.040 ± 0.003	0.081 ± 0.005	0.076 ± 0.005	0.026 ± 0.002
3/06-8/07					
CEM	0.160 ± 0.007	0.828 ± 0.021	1.613 ± 0.035	1.525 ± 0.034	0.526 ± 0.015
CMUP	0.086 ± 0.005	0.477 ± 0.017	0.916 ± 0.028	0.864 ± 0.027	0.287 ± 0.012
CMX	0.039 ± 0.003	0.189 ± 0.010	0.358 ± 0.015	0.353 ± 0.015	0.117 ± 0.007
CMX (Mini/Key)	0.006 ± 0.001	0.040 ± 0.004	0.081 ± 0.006	0.076 ± 0.006	0.023 ± 0.003

Table 7.4: Corrected acceptances (ϵ_{pretag}) for the various run ranges and subdetectors based on $175 \text{ GeV}/c^2$ $t\bar{t}$ Monte Carlo, after applying a 30 GeV missing E_T cut and a 200 GeV H_T cut for ≥ 3 -jet events. This acceptance is appropriate for events with two or more tags.

	Expected Tagging Efficiency (%)	
	≥ 1 Tag	≥ 2 Tags
Tight	60 ± 3	15 ± 2
Loose	69 ± 4	22 ± 3

Table 7.5: Event tagging efficiencies in the optimized samples. The ≥ 1 -tag and ≥ 2 -tag are subjected to different H_T cuts.

Expected Number of Tagged Events ($\sigma_{t\bar{t}} = 6.7$ pb)					
	1 jet	2 jet	3 jet	4 jet	≥ 5 jet
Missing $E_T > 30$ GeV, $H_T > 250$ GeV					
Tight, ≥ 1 -Tag	7.47 ± 0.76	59.01 ± 6.01	109.22 ± 11.12	125.44 ± 12.77	43.78 ± 4.46
Loose, ≥ 1 -Tag	8.96 ± 0.91	68.86 ± 7.01	126.81 ± 12.91	143.79 ± 14.64	50.33 ± 5.13
Missing $E_T > 30$ GeV, $H_T > 200$ GeV					
Tight, ≥ 2 -Tag	-	10.20 ± 1.86	28.04 ± 5.10	35.90 ± 6.53	12.98 ± 2.36
Loose, ≥ 2 -Tag	-	15.13 ± 2.75	41.61 ± 7.57	52.93 ± 9.62	19.20 ± 3.49

Table 7.6: Number of expected tagged events assuming a $t\bar{t}$ cross section of 6.7 pb. The error includes the luminosity, acceptance, and tagging efficiency. Each set of numbers is listed for the optimized selection for that measurement.

7.3 Backgrounds

Like the $t\bar{t}$ expectation, the backgrounds are recalculated identically to those presented in Chapter 6, and more detail on the samples and techniques can be found there. The purely electroweak backgrounds (dibosons and single top), the least significant non-top component of the sample, are treated precisely the same as $t\bar{t}$, but are reduced by 50% by the H_T and missing E_T cuts. The non- W background is again evaluated by fitting the missing E_T distribution; this background is reduced by a factor of 5 by the tighter missing E_T cut.

The W +jets backgrounds require more effort, since the heavy flavor fractions also change with the kinematic cuts. Table 7.7 lists the heavy flavor fractions for the single-tag and double-tag optimized selections, including the 1.0 ± 0.3 K correction. The W +light flavor mistag background is accounted for in the same way as for the baseline analysis, but the reduced acceptance and the heavy flavor fractions change the overall expectation.

Again assuming a $t\bar{t}$ cross section of 6.7 pb, we repeat the procedures from Chapter 6 to derive the expected sample composition for the optimized cross section measurement. These results are summarized in Tables 7.8 to 7.11 and Figures 7.5 and 7.6.

	1-jet	2-jet	3-jet	≥ 4 -jet
Missing $E_T > 20$ GeV, $H_T > 0$ GeV				
$Wb\bar{b}$, 1 b	0.7 ± 0.3	1.4 ± 0.6	2.4 ± 1.0	3.3 ± 1.4
$Wb\bar{b}$, 2 b		0.9 ± 0.4	1.8 ± 0.7	3.3 ± 1.3
$Wc\bar{c}$ or Wc , 1 c	5.5 ± 1.7	8.7 ± 3.2	11.7 ± 3.8	13.3 ± 4.9
$Wc\bar{c}$, 2 c		1.4 ± 0.6	3.3 ± 1.3	6.0 ± 2.4
Missing $E_T > 30$ GeV, $H_T^{\geq 3jet} > 250$ GeV				
$Wb\bar{b}$, 1 b	0.7 ± 0.3	1.4 ± 0.6	2.9 ± 1.2	3.6 ± 1.5
$Wb\bar{b}$, 2 b		0.9 ± 0.4	2.2 ± 0.9	3.5 ± 1.4
$Wc\bar{c}$ or Wc , 1 c	5.4 ± 1.7	8.5 ± 3.2	11.7 ± 3.8	13.2 ± 4.9
$Wc\bar{c}$, 2 c		1.4 ± 0.6	3.5 ± 1.3	6.2 ± 2.4
Missing $E_T > 30$ GeV, $H_T^{\geq 3jet} > 200$ GeV				
$Wb\bar{b}$, 1 b	0.7 ± 0.3	1.4 ± 0.6	2.9 ± 1.2	3.6 ± 1.5
$Wb\bar{b}$, 2 b		0.9 ± 0.4	2.1 ± 0.9	3.4 ± 1.4
$Wc\bar{c}$ or Wc , 1 c	5.4 ± 1.7	8.5 ± 3.2	11.7 ± 3.8	13.1 ± 4.9
$Wc\bar{c}$, 2 c		1.4 ± 0.6	3.4 ± 1.3	6.0 ± 2.4

Table 7.7: The heavy flavor fractions for the W +Heavy Flavor background as a function of the optimized selection. The 40% error is dominated by the K correction.

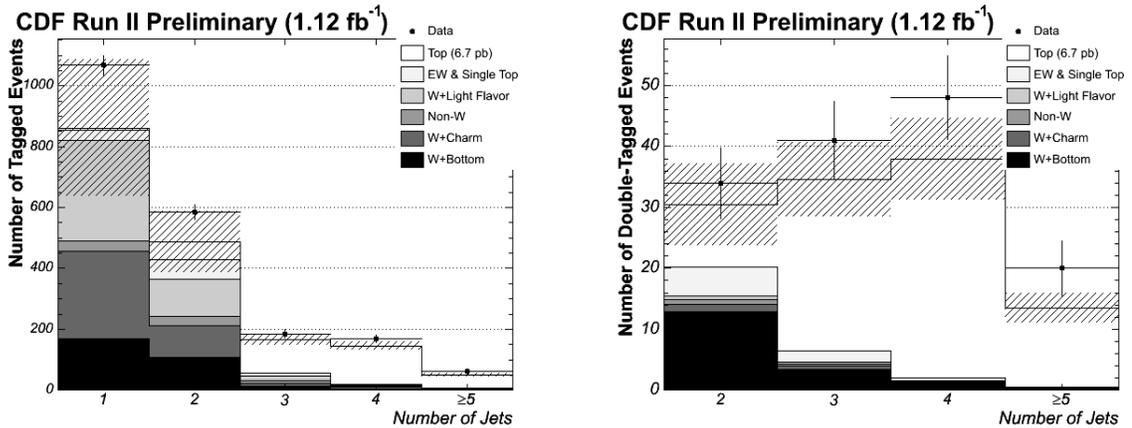


Figure 7.5: Expected and observed numbers of tagged events sorted by jet multiplicity, after requiring ≥ 1 (left) and ≥ 2 (right) tight tags with the optimized event selection. Normalizations here are the same as in Tables 7.8 and 7.9. The hashed region corresponds to the systematic uncertainty on the background.

Process	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	78903	12873	870	413	130
Pretag Top (6.7 pb)	22.11 ± 1.37	117.03 ± 7.24	190.84 ± 11.80	202.68 ± 12.53	69.94 ± 4.32
WW	6.21 ± 0.65	15.29 ± 1.61	2.18 ± 0.23	0.95 ± 0.10	0.36 ± 0.04
WZ	3.20 ± 0.33	6.96 ± 0.72	0.74 ± 0.08	0.28 ± 0.03	0.10 ± 0.01
ZZ	0.08 ± 0.01	0.21 ± 0.02	0.05 ± 0.01	0.03 ± 0.00	0.01 ± 0.00
Single Top (s-ch)	3.98 ± 0.90	14.77 ± 3.34	2.72 ± 0.61	0.83 ± 0.19	0.20 ± 0.05
Single Top (t-ch)	13.80 ± 2.47	17.56 ± 3.15	1.99 ± 0.36	0.47 ± 0.08	0.06 ± 0.01
$Z \rightarrow \tau\tau$	6.19 ± 1.50	7.83 ± 1.90	1.77 ± 0.43	0.71 ± 0.17	0.16 ± 0.04
$Wb\bar{b}$	168.06 ± 75.16	108.26 ± 48.41	13.21 ± 5.91	5.14 ± 2.30	1.47 ± 0.66
$Wc\bar{c}, Wc$	289.92 ± 129.66	103.15 ± 46.13	10.42 ± 4.66	4.13 ± 1.85	1.19 ± 0.53
W+Light Flavor	330.98 ± 72.82	123.34 ± 27.14	15.26 ± 3.36	4.44 ± 0.98	1.48 ± 0.33
Non-W	31.61 ± 58.59	30.25 ± 8.85	7.34 ± 2.73	3.49 ± 1.29	1.10 ± 0.41
Background	854.03 ± 225.28	427.62 ± 100.00	55.66 ± 13.11	20.48 ± 5.93	6.12 ± 1.73
Top (6.7 pb)	7.47 ± 0.76	59.01 ± 6.01	109.22 ± 11.12	125.44 ± 12.77	43.78 ± 4.46
Tags	1067	585	185	169	62

Table 7.8: Summary of signal and backgrounds for the tight, single-tag sample. We assume a top cross section of 6.7 pb.

Process	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	12873	1515	507	132
Pretag Top (6.7 pb)	117.03 ± 7.24	222.83 ± 13.78	212.46 ± 13.14	70.53 ± 4.36
WW	0.08 ± 0.01	0.10 ± 0.02	0.08 ± 0.02	0.03 ± 0.01
WZ	1.11 ± 0.20	0.27 ± 0.05	0.06 ± 0.01	0.01 ± 0.00
ZZ	0.02 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Single Top (s-ch)	3.36 ± 1.03	1.03 ± 0.32	0.25 ± 0.08	0.06 ± 0.02
Single Top (t-ch)	0.23 ± 0.06	0.34 ± 0.09	0.08 ± 0.02	0.01 ± 0.00
$Z \rightarrow \tau\tau$	0.07 ± 0.02	0.10 ± 0.03	0.04 ± 0.01	0.03 ± 0.01
$Wb\bar{b}$	12.80 ± 5.73	3.42 ± 1.53	1.09 ± 0.49	0.22 ± 0.10
$Wc\bar{c}, Wc$	1.26 ± 0.56	0.55 ± 0.25	0.23 ± 0.10	0.05 ± 0.02
W+Light Flavor	0.57 ± 0.17	0.31 ± 0.09	0.11 ± 0.03	0.03 ± 0.01
Non-W	0.77 ± 0.20	0.39 ± 1.77	0.13 ± 0.59	0.03 ± 0.15
Background	20.27 ± 6.43	6.52 ± 3.30	2.07 ± 1.55	0.49 ± 0.48
Top (6.7 pb)	10.20 ± 1.86	28.04 ± 5.10	35.90 ± 6.53	12.98 ± 2.36
Tags	34	41	48	20

Table 7.9: Summary of signal and backgrounds for the tight, double-tag sample. We assume a top cross section of 6.7 pb.

Process	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	78903	12873	870	413	130
Pretag Top (6.7 pb)	22.11 ± 1.37	117.03 ± 7.24	190.84 ± 11.80	202.68 ± 12.53	69.94 ± 4.32
WW	10.25 ± 1.08	24.36 ± 2.56	3.79 ± 0.40	1.54 ± 0.16	0.57 ± 0.06
WZ	4.24 ± 0.44	8.71 ± 0.90	1.02 ± 0.11	0.40 ± 0.04	0.14 ± 0.01
ZZ	0.09 ± 0.01	0.25 ± 0.03	0.07 ± 0.01	0.03 ± 0.00	0.01 ± 0.00
Single Top (s-ch)	4.72 ± 1.07	16.98 ± 3.84	3.17 ± 0.72	0.95 ± 0.22	0.23 ± 0.05
Single Top (t-ch)	16.16 ± 2.90	21.02 ± 3.77	2.37 ± 0.42	0.54 ± 0.10	0.07 ± 0.01
$Z \rightarrow \tau\tau$	10.17 ± 2.46	13.50 ± 3.27	3.23 ± 0.78	1.21 ± 0.29	0.27 ± 0.07
$Wb\bar{b}$	198.63 ± 88.83	126.23 ± 56.45	15.55 ± 6.96	6.12 ± 2.74	1.76 ± 0.79
$Wc\bar{c}, Wc$	399.68 ± 178.74	145.74 ± 65.18	16.00 ± 7.15	6.43 ± 2.88	1.85 ± 0.83
W +Light Flavor	843.58 ± 185.59	302.75 ± 66.60	34.39 ± 7.57	10.06 ± 2.21	3.30 ± 0.73
Non- W	43.74 ± 8.35	48.19 ± 16.51	12.73 ± 6.53	6.04 ± 3.10	1.90 ± 0.98
Background	1531.26 ± 325.92	707.73 ± 141.12	92.32 ± 19.20	33.34 ± 8.42	10.10 ± 2.51
Top (6.7 pb)	8.96 ± 0.91	68.86 ± 7.01	126.81 ± 12.91	143.79 ± 14.64	50.33 ± 5.13
Tags	1788	888	240	209	72

Table 7.10: Summary of signal and backgrounds for the loose, single-tag sample. We assume a top cross section of 6.7 pb.

Process	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	12873	1515	507	132
Pretag Top (6.7 pb)	117.03 ± 7.24	222.83 ± 13.78	212.46 ± 13.14	70.53 ± 4.36
WW	0.28 ± 0.05	0.25 ± 0.05	0.14 ± 0.03	0.08 ± 0.01
WZ	1.54 ± 0.28	0.34 ± 0.06	0.11 ± 0.02	0.04 ± 0.01
ZZ	0.03 ± 0.01	0.02 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Single Top (s-ch)	4.86 ± 1.49	1.53 ± 0.47	0.38 ± 0.11	0.08 ± 0.03
Single Top (t-ch)	0.48 ± 0.12	0.53 ± 0.14	0.14 ± 0.04	0.02 ± 0.01
$Z \rightarrow \tau\tau$	0.16 ± 0.05	0.23 ± 0.08	0.08 ± 0.03	0.04 ± 0.01
$Wb\bar{b}$	18.56 ± 8.30	5.27 ± 2.36	1.75 ± 0.78	0.36 ± 0.16
$Wc\bar{c}, Wc$	3.09 ± 1.38	1.52 ± 0.68	0.69 ± 0.31	0.14 ± 0.06
W +Light Flavor	3.20 ± 0.93	1.57 ± 0.46	0.56 ± 0.16	0.18 ± 0.05
Non- W	1.45 ± 2.18	1.19 ± 1.90	0.40 ± 0.64	0.10 ± 0.17
Background	33.64 ± 9.60	12.46 ± 4.68	4.25 ± 2.33	1.04 ± 0.73
Top (6.7 pb)	15.13 ± 2.75	41.61 ± 7.57	52.93 ± 9.62	19.20 ± 3.49
Tags	63	64	72	29

Table 7.11: Summary of signal and backgrounds for the loose, double-tag sample. We assume a top cross section of 6.7 pb.

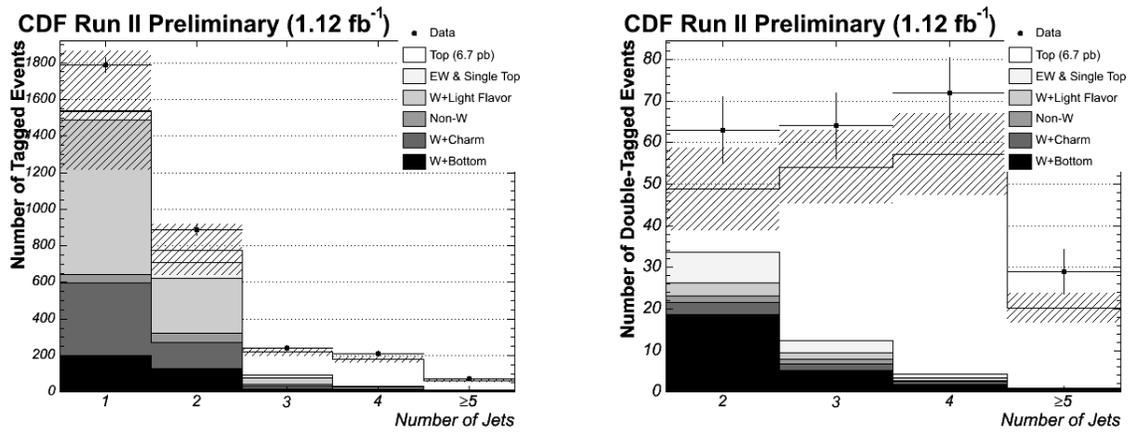


Figure 7.6: Expected and observed numbers of tagged events sorted by jet multiplicity, after requiring ≥ 1 (left) and ≥ 2 (right) loose tags with the optimized event selection. Normalizations here are the same as in Tables 7.10 and 7.11. The hashed region corresponds to the systematic uncertainty on the background.

7.4 Results

To determine the top pair production cross section, we require that the total number of b -tagged events expected from Method 2 is equal to the number observed in the data sample. Since we fit the non- W component rather than deriving it from a formula, the Method 2 expectation can not be written analytically as a function of $\sigma_{t\bar{t}}$. Instead, we iteratively select cross sections, attributing the excess or deficit to $t\bar{t}$ until the cross section is stable. In the sample of events with ≥ 1 tight b -tag, we measure a cross section of 8.2 ± 0.5 (stat) pb, and in the double loose-tagged sample, we measure 8.8 ± 0.8 (stat) pb. The loose, single-tag result of 8.3 ± 0.5 pb and the tight, double-tag result of 8.8 ± 0.9 pb are consistent with each other and with the lead results. All four are summarized in Tables 7.12 to 7.15 and Figures 7.7 and 7.10.

We cross check the main result in several subsamples, looking for dependence on run number, lepton type, or optimization cuts. For the optimized, tight result, we measure cross sections of 9.0 ± 1.1 pb, 8.4 ± 1.0 pb, 7.8 ± 1.1 pb, and 7.8 ± 1.3 pb in the four run ranges, all statistically consistent with 8.2. Using only electrons, we measure 9.0 ± 0.7 pb, while for muons we measure 7.2 ± 0.7 pb, a 1.5σ discrepancy. With only the base selection (no H_T cut and missing $E_T > 20$ GeV) and the tight tagger, we measure 8.5 ± 0.6 pb. All results are consistent with the main results.

Process	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	78903	12873	870	413	130
Pretag Top (8.2 pb)	27.06 ± 1.67	143.23 ± 8.86	233.56 ± 14.44	248.06 ± 15.34	85.60 ± 5.29
WW	6.21 ± 0.65	15.29 ± 1.61	2.18 ± 0.23	0.95 ± 0.10	0.36 ± 0.04
WZ	3.20 ± 0.33	6.96 ± 0.72	0.74 ± 0.08	0.28 ± 0.03	0.10 ± 0.01
ZZ	0.08 ± 0.01	0.21 ± 0.02	0.05 ± 0.01	0.03 ± 0.00	0.01 ± 0.00
Single Top (s-ch)	3.98 ± 0.90	14.77 ± 3.34	2.72 ± 0.61	0.83 ± 0.19	0.20 ± 0.05
Single Top (t-ch)	13.80 ± 2.47	17.56 ± 3.15	1.99 ± 0.36	0.47 ± 0.08	0.06 ± 0.01
$Z \rightarrow \tau\tau$	6.19 ± 1.50	7.83 ± 1.90	1.77 ± 0.43	0.71 ± 0.17	0.16 ± 0.04
$Wb\bar{b}$	168.05 ± 75.16	108.01 ± 48.30	12.09 ± 5.41	3.55 ± 1.59	0.93 ± 0.41
$Wc\bar{c}, Wc$	289.90 ± 129.65	102.91 ± 46.02	9.53 ± 4.26	2.85 ± 1.27	0.74 ± 0.33
W+Light Flavor	330.95 ± 72.81	123.00 ± 27.06	13.81 ± 3.04	2.99 ± 0.66	0.91 ± 0.20
Non-W	31.68 ± 58.22	30.98 ± 8.03	8.47 ± 6.03	4.02 ± 2.86	1.27 ± 0.90
Background	854.04 ± 225.17	427.52 ± 99.85	53.34 ± 13.75	16.69 ± 5.95	4.72 ± 1.74
Top (8.2 pb)	9.15 ± 0.93	72.22 ± 7.35	133.68 ± 13.61	153.53 ± 15.63	53.58 ± 5.46
Tags	1067	585	185	169	62

Table 7.12: Summary of signal and backgrounds for the tight, single-tag sample. We assume the measured top cross section of 8.2 pb.

Process	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	12873	1515	507	132
Pretag Top (8.8 pb)	153.71 ± 9.50	292.67 ± 18.10	279.06 ± 17.25	92.64 ± 5.73
WW	0.08 ± 0.01	0.10 ± 0.02	0.08 ± 0.02	0.03 ± 0.01
WZ	1.11 ± 0.20	0.27 ± 0.05	0.06 ± 0.01	0.01 ± 0.00
ZZ	0.02 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Single Top (s-ch)	3.36 ± 1.03	1.03 ± 0.32	0.25 ± 0.08	0.06 ± 0.02
Single Top (t-ch)	0.23 ± 0.06	0.34 ± 0.09	0.08 ± 0.02	0.01 ± 0.00
$Z \rightarrow \tau\tau$	0.07 ± 0.02	0.10 ± 0.03	0.04 ± 0.01	0.03 ± 0.01
$Wb\bar{b}$	12.76 ± 5.71	3.18 ± 1.42	0.75 ± 0.34	0.11 ± 0.05
$Wc\bar{c}, Wc$	1.26 ± 0.56	0.51 ± 0.23	0.16 ± 0.07	0.02 ± 0.01
W+Light Flavor	0.57 ± 0.16	0.28 ± 0.08	0.07 ± 0.02	0.02 ± 0.00
Non-W	0.85 ± 1.64	0.62 ± 1.47	0.21 ± 0.49	0.05 ± 0.13
Background	20.30 ± 6.73	6.45 ± 3.35	1.71 ± 1.65	0.36 ± 0.52
Top (8.8 pb)	13.40 ± 2.44	36.83 ± 6.70	47.15 ± 8.57	17.05 ± 3.10
Tags	34	41	48	20

Table 7.13: Summary of signal and backgrounds for the tight, double-tag sample. We assume the measured top cross section of 8.8 pb.

Process	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	78903	12873	870	413	130
Pretag Top (8.3 pb)	27.39 ± 1.69	144.98 ± 8.96	236.41 ± 14.62	251.08 ± 15.52	86.64 ± 5.36
WW	10.25 ± 1.08	24.36 ± 2.56	3.79 ± 0.40	1.54 ± 0.16	0.57 ± 0.06
WZ	4.24 ± 0.44	8.71 ± 0.90	1.02 ± 0.11	0.40 ± 0.04	0.14 ± 0.01
ZZ	0.09 ± 0.01	0.25 ± 0.03	0.07 ± 0.01	0.03 ± 0.00	0.01 ± 0.00
Single Top (s-ch)	4.72 ± 1.07	16.98 ± 3.84	3.17 ± 0.72	0.95 ± 0.22	0.23 ± 0.05
Single Top (t-ch)	16.16 ± 2.90	21.02 ± 3.77	2.37 ± 0.42	0.54 ± 0.10	0.07 ± 0.01
$Z \rightarrow \tau\tau$	10.17 ± 2.46	13.50 ± 3.27	3.23 ± 0.78	1.21 ± 0.29	0.27 ± 0.07
$Wb\bar{b}$	198.61 ± 88.82	125.91 ± 56.31	14.14 ± 6.33	4.10 ± 1.83	1.06 ± 0.47
$Wc\bar{c}, Wc$	399.65 ± 178.73	145.38 ± 65.02	14.55 ± 6.51	4.31 ± 1.93	1.11 ± 0.50
W +Light Flavor	843.51 ± 185.57	301.84 ± 66.41	30.91 ± 6.80	6.57 ± 1.44	1.94 ± 0.43
Non- W	43.74 ± 8.32	49.12 ± 15.19	15.41 ± 6.66	7.32 ± 3.16	2.30 ± 1.00
Background	1531.15 ± 325.90	707.08 ± 140.78	88.65 ± 18.44	26.97 ± 7.53	7.71 ± 2.20
Top (8.3 pb)	11.10 ± 1.13	85.31 ± 8.69	157.09 ± 16.00	178.12 ± 18.14	62.35 ± 6.35
Tags	1788	888	240	209	72

Table 7.14: Summary of signal and backgrounds for the loose, single-tag sample. We assume the measured top cross section of 8.3 pb.

Process	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	12873	1515	507	132
Pretag Top (8.8 pb)	153.71 ± 9.50	292.67 ± 18.10	279.06 ± 17.25	92.64 ± 5.73
WW	0.28 ± 0.05	0.25 ± 0.05	0.14 ± 0.03	0.08 ± 0.01
WZ	1.54 ± 0.28	0.34 ± 0.06	0.11 ± 0.02	0.04 ± 0.01
ZZ	0.03 ± 0.01	0.02 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Single Top (s-ch)	4.86 ± 1.49	1.53 ± 0.47	0.38 ± 0.11	0.08 ± 0.03
Single Top (t-ch)	0.48 ± 0.12	0.53 ± 0.14	0.14 ± 0.04	0.02 ± 0.01
$Z \rightarrow \tau\tau$	0.16 ± 0.05	0.23 ± 0.08	0.08 ± 0.03	0.04 ± 0.01
$Wb\bar{b}$	18.50 ± 8.27	4.90 ± 2.19	1.21 ± 0.54	0.18 ± 0.08
$Wc\bar{c}, Wc$	3.08 ± 1.38	1.42 ± 0.63	0.48 ± 0.21	0.07 ± 0.03
W +Light Flavor	3.18 ± 0.92	1.44 ± 0.42	0.37 ± 0.11	0.08 ± 0.02
Non- W	1.73 ± 2.19	1.52 ± 1.75	0.51 ± 0.59	0.13 ± 0.15
Background	33.84 ± 9.75	12.18 ± 4.93	3.42 ± 2.52	0.73 ± 0.81
Top (8.8 pb)	19.87 ± 3.61	54.65 ± 9.94	69.52 ± 12.64	25.22 ± 4.59
Tags	63	64	72	29

Table 7.15: Summary of signal and backgrounds for the loose, double-tag sample. We assume the measured top cross section of 8.8 pb.

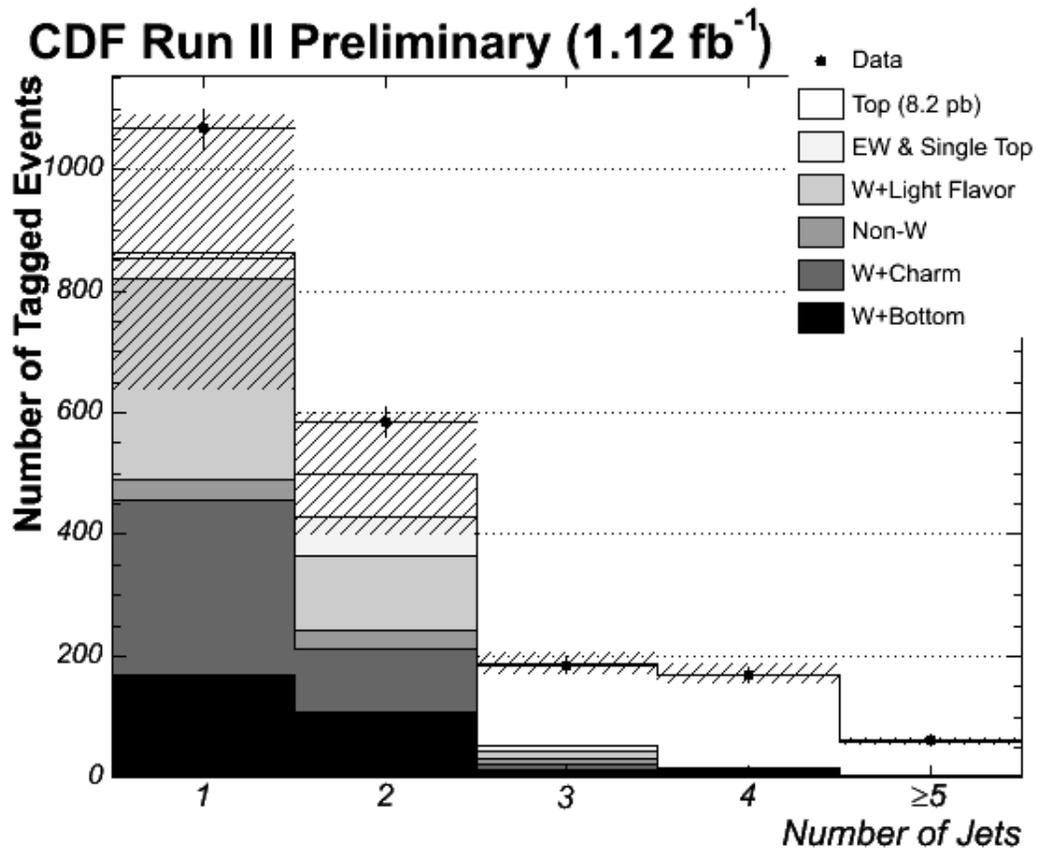


Figure 7.7: Expected and observed numbers of tagged events sorted by jet multiplicity, requiring ≥ 1 tight tag. Normalizations here are the same as in Table 7.12. The hashed region corresponds to the systematic uncertainty on the background.

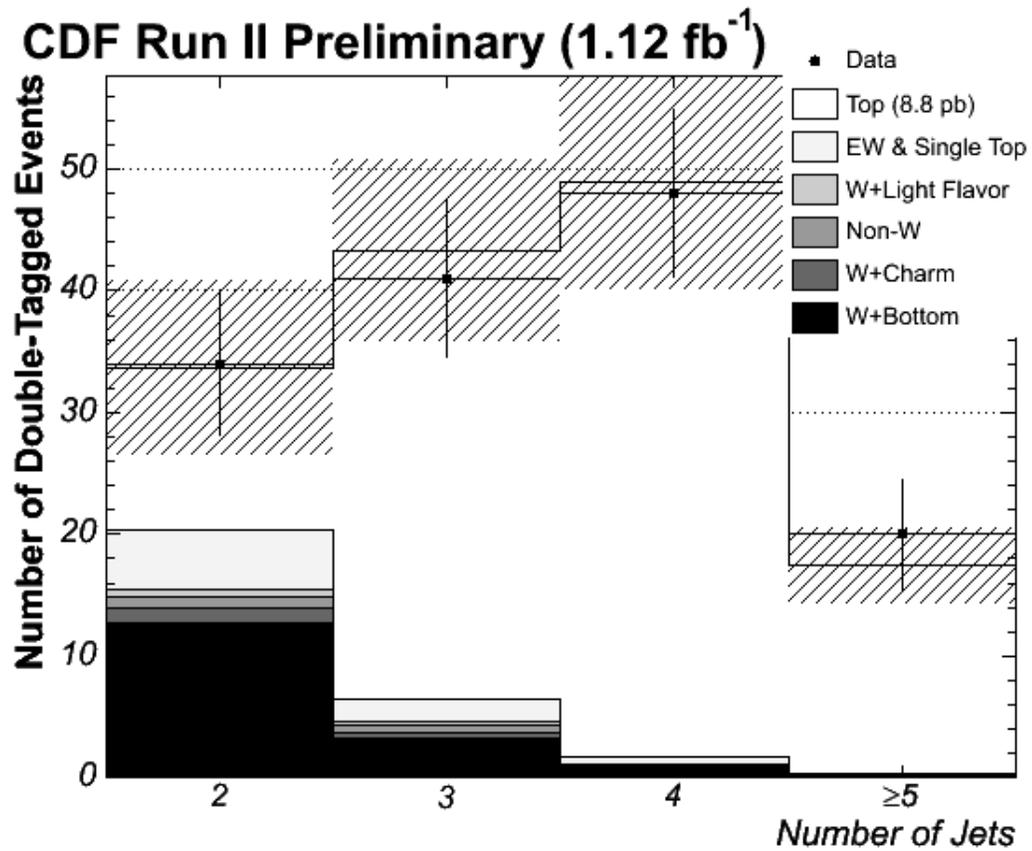


Figure 7.8: Expected and observed numbers of tagged events sorted by jet multiplicity, requiring ≥ 2 tight tags. Normalizations here are the same as in Table 7.13. The hashed region corresponds to the systematic uncertainty on the background.

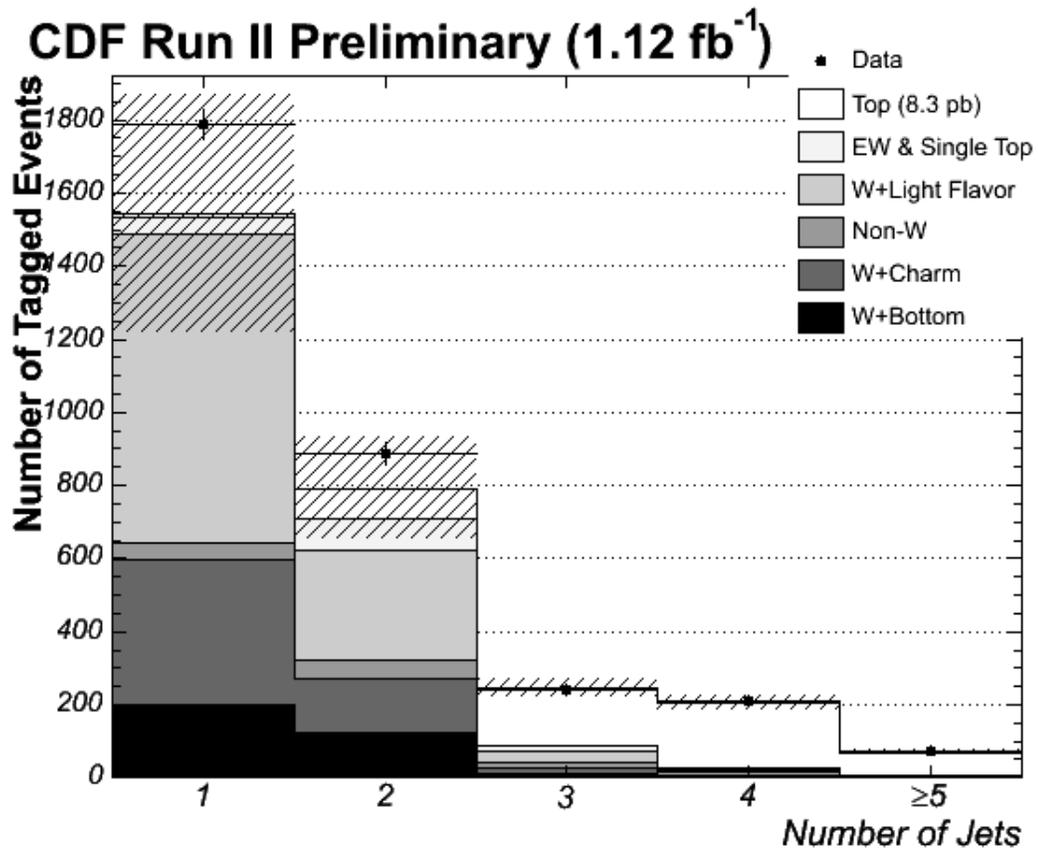


Figure 7.9: Expected and observed numbers of tagged events sorted by jet multiplicity, requiring ≥ 1 loose tag. Normalizations here are the same as in Table 7.14. The hashed region corresponds to the systematic uncertainty on the background.

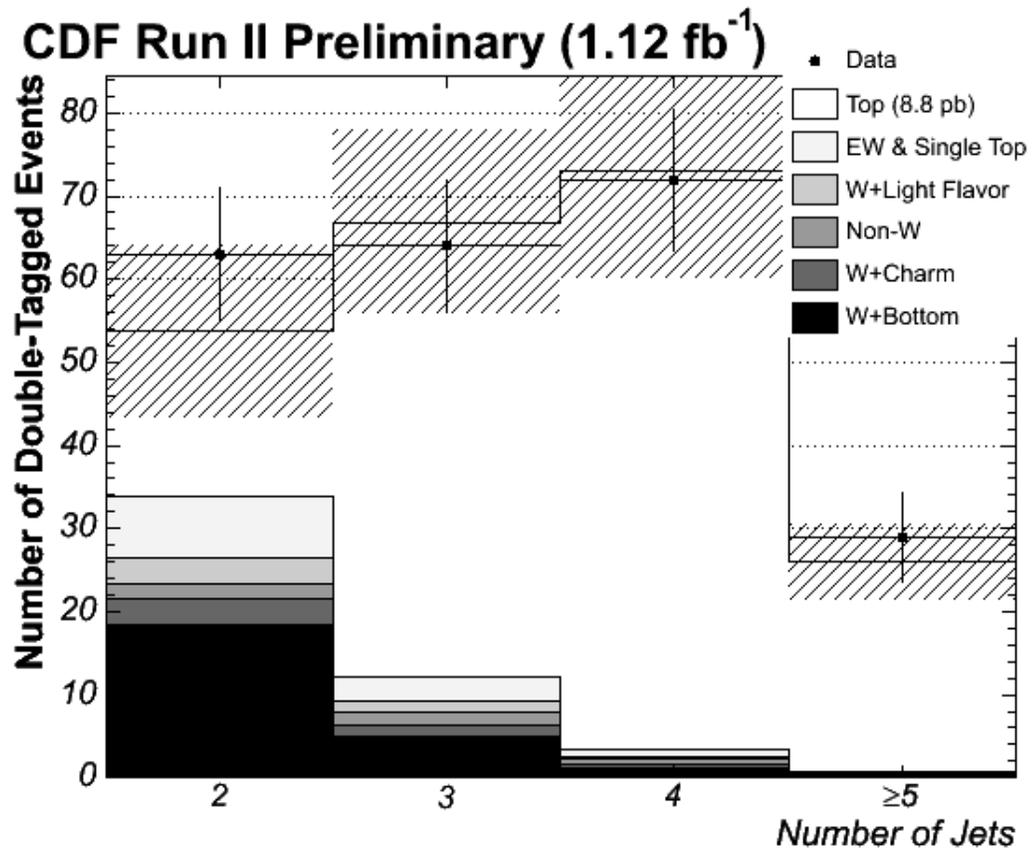


Figure 7.10: Expected and observed numbers of tagged events sorted by jet multiplicity, requiring ≥ 2 loose tag. Normalizations here are the same as in Table 7.15. The hashed region corresponds to the systematic uncertainty on the background.

7.5 Uncertainties

The strength of Method 2 is its use of the data to constrain as much of the sample composition as possible, such that the result is not subject to large theoretical uncertainties. The downside is that the background and signal uncertainties are correlated in complicated ways; the W +jets normalization is anti-correlated with both the signal and the non- W component, and the tagging efficiencies are 100% correlated (through the scale factor [64]) for all signal and backgrounds depending on Monte Carlo. For these reasons, error propagation is nearly impossible to carry out analytically.

For systematic uncertainties, then, the simplest approach is to vary each contributing term by $\pm 1\sigma$ and repeat the full procedure to quantify the effect on the cross section. We separate these errors into systematics on the acceptance, b -tagging efficiency, and backgrounds. We evaluate errors here only for the tight inclusive and loose double-tagged cross sections. All acceptance systematics are summarized in Table 7.16. The errors relevant to the cross section are described below and summarized in Table 7.19.

- **Luminosity:** CDF assigns a 5.9% systematic uncertainty to the integrated luminosity (1.12 fb^{-1}). This includes the uncertainty in the $p\bar{p}$ inelastic cross section and the precision of the CLC [32]. It affects the pretag expectation for signal, and also is included in the expectations for the MC backgrounds. Since the effect on the signal and backgrounds are correlated, we vary the luminosity $\pm 1\sigma$ and re-measure the cross section to determine the actual error. The net effect is a 6.2% shift for both the single and double-tag cross section.

Sample	Single-Tags		Double-Tags	
	Acceptance (%)	Difference (%)	Acceptance (%)	Difference (%)
Main Sample	7.00 ± 0.01	-	7.64 ± 0.01	-
Herwig	7.34 ± 0.03	4.8	7.92 ± 0.03	3.7
ALPGEN+PYTHIA $t\bar{t} + 0p$	6.90 ± 0.03	1.5	7.58 ± 0.03	0.8
ALPGEN+PYTHIA $t\bar{t} + 1p$	8.58 ± 0.03	22.6	8.92 ± 0.03	16.8
ALPGEN+PYTHIA Weighted	7.29 ± 0.03	4.3	7.90 ± 0.03	3.4
JES $+1\sigma$	7.42 ± 0.01	5.9	7.95 ± 0.01	4.0
JES -1σ	6.56 ± 0.01	6.3	7.33 ± 0.01	4.1
More ISR	7.07 ± 0.03	1.0	7.66 ± 0.03	0.2
Less ISR	7.00 ± 0.03	0.0	7.63 ± 0.03	0.2
More FSR	6.95 ± 0.03	0.8	7.56 ± 0.03	1.0
Less FSR	7.04 ± 0.03	0.5	7.65 ± 0.03	0.1
MRST75	7.05 ± 0.03	0.6	7.71 ± 0.03	0.8
MRST72	7.02 ± 0.03	0.2	7.67 ± 0.03	0.3

Table 7.16: Raw acceptances (in %) for the two optimized selections in various systematic samples. The error is statistical only.

- Jet Energy Scale:** The raw calorimeter energy of the jets is corrected to better represent the relevant jet physics quantities, and this rescaling carries a systematic uncertainty (see Figure 4.15) [48]. We recalculate the $t\bar{t}$ acceptance with the jet corrections scaled up and down by the JES uncertainty, and the difference is taken as an uncertainty on the cross section. The acceptances are shown in Table 7.16. The effect is 6.1% for the single-tags, 4.1% for double-tags.
- Lepton ID:** We correct the MC acceptance for known differences in lepton identification, such as the trigger efficiency and the reconstruction efficiency. These corrections are different for each tight lepton type (CEM, CMUP, and CMX), and are listed in Table 6.3 [44][46]. Like the luminosity error above, this affects the top signal and the MC backgrounds. The error is sufficiently small that we ignore the correlations, and apply the uncertainty only to the signal. The common systematics (primary vertex efficiency, COT track efficiency), contribute a total of 0.3% to the acceptance error. The lepton-specific terms (trigger efficiencies, muon reconstruction and χ^2 efficiency) are 1.3% for

CEM, 2.7% for CMUP, and 2.5% for CMX, including the lepton isolation. The pretag-weighted average of these uncertainties is 1.8%.

- **ISR/FSR:** We measure the effect on the acceptance of using simulations with more or less initial and final state radiation. We determine the $t\bar{t}$ acceptance using four Monte Carlo samples, each with the initial or final state radiation doubled or halved. Acceptances for each sample are listed in Table 7.16. We take the mean deviation from the default setting as a systematic for each. The result is a 0.5% error for ISR and 0.6% for FSR for the single-tag result, and 0.2% and 0.6% for the double-tag result.
- **Pythia vs. Herwig:** We check the difference in acceptance when using an alternate Monte Carlo generator for our $t\bar{t}$ signal [78]. This covers differences between the two in the generated kinematics, shower model, and fragmentation. This number is corrected by 2.5% for the different W branching ratios in the two MC generators. We observe an overall shift of 2.2% (1.1%) in the acceptance for the single-tag (double-tag) selection, and we take this as a systematic uncertainty.
- **PDF:** Uncertainties in the proton parton distribution function (PDF) can also propagate into the acceptance. By default, the Monte Carlo is generated using the CTEQ5L parameterization of the parton distribution functions [4]. For the incoming parton momenta and Q^2 in each simulated event, we can determine the likelihood for such an event to be generated given a different PDF set. Comparing the expected acceptance for 46 different PDF parameterizations, we use

the root-mean-square deviation from CTEQ5L as a systematic uncertainty [79]. The results of this study are shown in Table 7.17 for the different CTEQ and MRST parameterizations. We additionally measure directly the acceptance in generated samples with alternate PDFs, as shown in Table 7.16. The difference is somewhat bigger than that reported by the re-weighting scheme (the α_s error is 0.5% from the above), though the total error is still quite small. If we substitute 0.5% for the 0.1 in the Table, we reach a total uncertainty of 0.9%, ignoring asymmetric terms.

Source	Relative Error (%)
CTEQ uncertainty	+0.27, -0.62
MRST uncertainty	± 0.33
α_s uncertainty	± 0.11
Total uncertainty	+0.44, -0.71

Table 7.17: Itemization of the PDF uncertainty on the cross section from the sample re-weighting method.

- Background Errors:** The systematic uncertainties associated with each background also contribute to the uncertainty on the cross section, but not in a predictable way. For instance, if the non- W background is measured to have a tag rate identical to W +jets, the systematic becomes negligible. We evaluate the effect on the cross section of varying each of the three key ingredients in evaluating the background: the heavy flavor fractions (globally by 40%, since the error is dominated by the K correction 1.0 ± 0.3), the non- W fraction (by its statistical and systematic error), and the per-jet mistag rate (an E_T -dependent systematic between 15% and 22%, as outlined in Section 5.3.3). The results of these tests are shown in Table 7.18; the systematic error is fairly small, owing

in large part to the tightness of the optimization cuts.

Source	Relative Error on $\sigma_{t\bar{t}}$	
	Single-Tag	Double-Tag
F_{non-W} (tag and pretag) $\pm 1\sigma$	1.7%	1.3%
Heavy flavor fraction $\pm 1\sigma$	3.3%	2.0%
Per-jet mistag rate $\pm 1\sigma$	1.0%	0.3%

Table 7.18: Variation in the measured cross section after introducing a variation in the background. Since all the backgrounds are correlated in complicated ways, explicit analytical calculation of the errors is quite difficult.

- Heavy Quark Tagging:** Our measurement of the b -tagging scale factor (SF) has a 5% error, 0.95 ± 0.05 for both taggers [64]. We vary the scale factor by its uncertainty and continue applying it to both charm and bottom, and we take the mean absolute deviation as the systematic. For inclusive tight tags, we derive an error of 5.8%, and for double loose tags, the error is 12.1%.
- Charm Tagging:** We additionally assume that the charm scale factor is the same as for bottom, though there has not been precisely measured in Run 2. We triple the uncertainty on SF (0.95 ± 0.15) for charm, and re-evaluate the cross section at $\pm 1\sigma$ (while keeping the b -tagging SF fixed at 0.95). We determine this uncertainty to be 1.1% (2.1%) for the inclusive (double-tag) measurement.
- Light Flavor Tagging:** The method of using the mistag matrix from data to predict the light flavor tag rate carries with it a large systematic as well, though we only use this method in events with at least one heavy flavor jet. We vary the light flavor tag probability in MC by 20% (of the same order as α and β [61]), and we recompute the cross section. The variation is 0.3% and 0.7% for

the single and double-tagged samples, respectively.

Systematic	Relative Error (%)	
	Inclusive (Tight)	Double (Loose)
Lepton ID (CEM)	1.3	
Lepton ID (CMUP)	2.7	
Lepton ID (CMX)	2.5	
ISR	0.5	0.2
FSR	0.6	0.6
PDFs	0.9	
PYTHIA vs. Herwig	2.2	1.1
Luminosity	6.2	
JES	6.1	4.1
b -Tagging	5.8	12.1
Charm Tagging	1.1	2.1
Light Flavor Tagging	0.3	0.7
Non- W	1.7	1.3
HF Fractions	3.3	2.0
Mistag Matrix	1.0	0.3
Total	11.5	14.8

Table 7.19: Summary of relative systematic uncertainties (in %) on the cross section measurements.

7.5.1 Statistical Errors

We have thusfar evaluated the statistical error assuming a fixed background, or rather that the $t\bar{t}$ signal is a true excess over a pre-determined background. With Method 2, however, the background estimate has some dependence on the assumed $t\bar{t}$ cross section, making the signal and background yields anti-correlated. We account for this fact by determining the background as a linear function of $\sigma_{t\bar{t}}$ and including it in a Poisson likelihood:

$$-2 \ln \mathcal{L} = -2(N_{data} \ln(D\sigma + B) + N_{data}! + (D\sigma + B)), \quad (7.2)$$

where N_{data} is the number of observed events, σ is the cross section, D is the denominator of the cross section formula (number of events expected per pb), and B is the background. For the single-tags, we use $D=41.559$ and $B = 74.70 - 4.03$ ($\sigma=8.21$), and for double-tags we use $D=16.975$ and $B = 16.36 - 0.65$ ($\sigma=8.76$).

This method will exaggerate the statistical error, since it assumes that the number of tagged events is distributed as a Poisson; in fact, the *pretag* yields are Poisson-distributed, and the tagged event yields are binomially distributed given the pretag total and the expected tag rates. Since the backgrounds are normalized to the pretag sample, the cross section measurement is fairly stable against background fluctuations. From pseudo-experiment tests near the measured cross section, we find that the error predicted by the above method will *overestimate* the uncertainty by roughly 8% for single-tags; the pull distribution for these pseudo-experiments is shown in Figure 7.11. The effect on the double-tags is only 3%, since the signal-to-background ratio is much higher. We ultimately derive statistical errors of ± 0.5 pb for the single-tag result and $^{+0.8}_{-0.7}$ pb for the double-tag result.

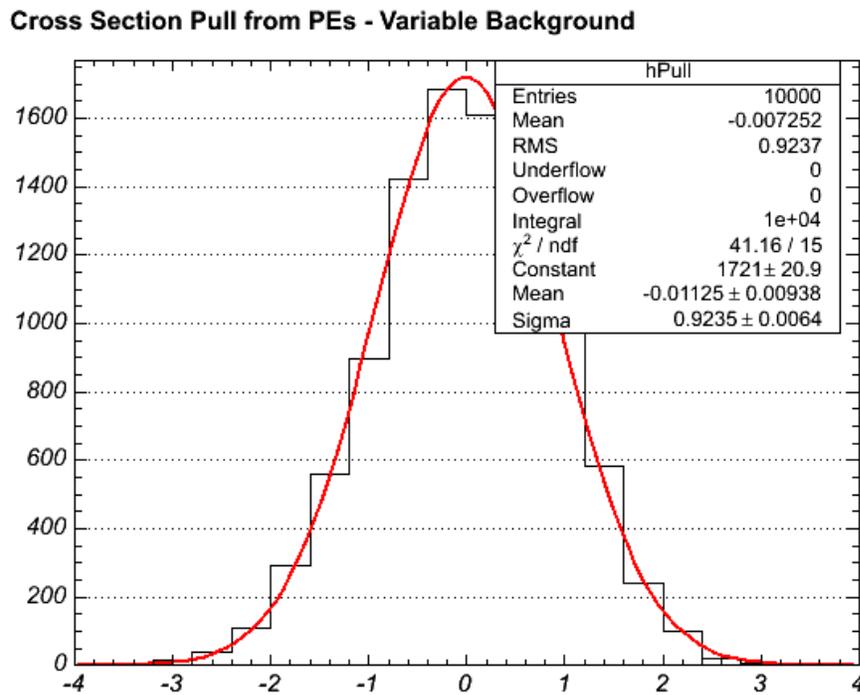


Figure 7.11: Pull distribution for 10,000 pseudo-experiments assuming a cross section of 8.2 pb. We assume the pretag yield is distributed as a Poisson, while the tagged event yields are binomial, and find that the statistical error evaluated using a Poisson likelihood overestimates the statistical error by roughly 8%.

Chapter 8

Conclusion and Outlook

We have measured the top pair production cross section in $p\bar{p}$ collisions with $\sqrt{s} = 1.96$ TeV using 1.12 fb^{-1} of data collected with the CDF II detector. The most precise result, calculated with events containing ≥ 1 tight b -tag, is 8.2 ± 0.5 (stat) ± 0.8 (syst) ± 0.5 (lum) pb. This is currently the world's most precise single measurement of $\sigma_{t\bar{t}}$. Due to the large statistics of the lepton+jets dataset, the kinematic selection is extremely tight, and the resulting $t\bar{t}$ purity is greater than 80%. In a 90% pure sample of events with two or more loose b -tags, we measure $8.8^{+0.8}_{-0.7}$ (stat) ± 1.2 (syst) ± 0.5 (lum) pb.

The results are higher than expected, but consistent with the theoretical prediction of $6.7^{+0.7}_{-0.9}$ pb for a top mass of $175 \text{ GeV}/c^2$. Since the acceptance is correlated with the top mass, the measured single-tag result varies by ± 0.08 pb for every $\mp 1 \text{ GeV}/c^2$ in the assumed top quark mass; the theoretical prediction varies by ± 0.16 for the same shift in the true top mass. Therefore, for a mass of $171 \text{ GeV}/c^2$, the current world average, the new measured value of 8.4 ± 0.5 (stat) ± 0.8 (syst) ± 0.5 (lum)

pb is in better agreement with the predicted value of 7.8 ± 0.9 pb.

As the Tevatron continues to operate well and the CDF and DØ collaborations acquire larger statistics samples, the analysis strategy will necessarily evolve to trade statistical power for better understanding of systematics. The two largest *reducible* sources of uncertainty, the jet energy corrections and the b -tagging scale factor, will both become better constrained as they are determined directly in $t\bar{t}$ events. By constraining the pair of untagged jets in doubly-tagged lepton+jets events to the W mass, the overall JES corrections can be checked. Similarly, requiring that the cross section measurement in single-tags and double-tags agree puts a direct constraint on the b -tagging scale factor; we expect that this method will be as precise as the current low- p_T measurements described in Section 5.2 with a dataset of 2 fb^{-1} .

Naturally, further understanding of the backgrounds and W +jets sample composition would benefit the cross section measurement, though the need is more pressing for analyses with worse signal-to-background ratios, such as single top and WH . In particular, a direct measurement of Wc is critical, since this background is quite large, and is not calibrated in generic jets, as $Wb\bar{b}$ and $Wc\bar{c}$ have been.

With the turn-on of the Large Hadron Collider at CERN scheduled for the beginning of 2008, results from the Tevatron will be critical in informing the methods and strategies used there to study top and other high- p_T physics. There is still potential for discovery at the Tevatron, but the advanced techniques being developed now to overcome systematic limitations are more critical in preparing for the next phase of particle physics at the LHC.

Appendix A

Data Selection Details

This chapter provides a detailed list of event yields at each stage of event selection, from the trigger to the pretag sample. Yields are sorted by the number of reconstructed jets, and divided into four run ranges. We require that the event be taken in a run with the silicon detector on, and that the lepton pass the tight selection described in Section 4.2. We require the missing E_T be greater than 20 GeV and that leptons be isolated, and we reject events with additional leptons or tracks which form the Z invariant mass with the lepton. Events with conversion electrons and cosmic muons are also removed, as are events in which the lepton z_0 is more than 5 cm from the primary vertex. Event yields for events with CEM electrons, CMUP muons, and CMX muons are shown in Tables A.1 to A.3.

	W+0 jets	W+1 jet	W+2 jets	W+3 jets	W+4 jets	W+ \geq 5 jets	Total
Feb. 2002 - Aug. 2004							
Initial	489321	380224	339727	48578	7632	1224	1266706
Silicon Good Run	322786	249180	222959	31917	5066	812	832720
≥ 1 Lepton	293209	218840	197924	26863	4057	637	741530
Missing $E_T > 20$ GeV	185438	37444	12857	3259	679	173	239850
Lepton Iso.	185438	20444	3636	645	141	34	210338
Dilepton Veto	185327	20362	3599	627	139	34	210088
Z Mass Veto	183660	19178	3305	553	126	32	206854
Conversion Veto	183660	19158	3293	553	125	32	206821
Lepton z_0	183500	19153	3292	553	125	31	206654
Sep. 2004 - Aug. 2005							
Initial	438276	361169	329437	46552	6961	1193	1183588
Silicon Good Run	337925	278921	254288	35966	5402	943	913445
≥ 1 Lepton	321471	249618	228848	30628	4360	721	835646
Missing $E_T > 20$ GeV	203163	45339	16423	3992	780	179	269876
Lepton Iso.	203163	22954	4084	697	154	35	231087
Dilepton Veto	203023	22833	4034	680	153	34	230757
Z Mass Veto	201024	21489	3702	616	139	33	227003
Conversion Veto	201024	21466	3686	614	138	32	226960
Lepton z_0	200939	21461	3685	614	137	32	226868
Sep. 2005 - Feb. 2006							
Initial	269514	224167	205330	29817	4475	787	734090
Silicon Good Run	235078	195661	179039	26050	3947	670	640445
≥ 1 Lepton	223463	175294	161096	22267	3197	536	585853
Missing $E_T > 20$ GeV	141928	32912	12003	2844	534	170	190391
Lepton Iso.	141928	16226	2745	537	113	31	161580
Dilepton Veto	141836	16124	2718	529	111	30	161348
Z Mass Veto	140324	15112	2501	469	102	29	158537
Conversion Veto	140324	15097	2491	465	102	29	158508
Lepton z_0	140239	15091	2490	465	102	29	158416
March 2006 - Aug. 2006							
Initial	194481	163873	149814	21586	3347	609	533710
Silicon Good Run	150013	126501	115887	16695	2584	476	412156
≥ 1 Lepton	139767	111833	103474	14214	2071	357	371716
Missing $E_T > 20$ GeV	88803	21551	8236	1934	404	84	121012
Lepton Iso.	88803	10378	1857	315	64	15	101432
Dilepton Veto	88734	10329	1836	310	63	15	101287
Z Mass Veto	87710	9699	1682	286	57	14	99448
Conversion Veto	87710	9685	1676	286	57	14	99428
Lepton z_0	87658	9680	1676	285	57	14	99370

Table A.1: Event yields at each stage of event selection for CEM electrons. The data are sorted by number of jets, and results are presented for four run ranges.

	W+0 jets	W+1 jet	W+2 jets	W+3 jets	W+4 jets	W+ \geq 5 jets	Total
Feb. 2002 - Aug. 2004							
Initial	387035	122514	47405	9789	1591	305	568639
Silicon Good Run	259037	81928	31579	6519	1041	202	380306
≥ 1 Lepton	144293	48483	17917	3682	579	120	215074
Missing $E_T > 20$ GeV	102532	11784	3020	788	186	52	118362
Lepton Iso.	98759	9714	1516	245	62	14	110310
Dilepton Veto	98593	9655	1490	235	62	14	110049
Z Mass Veto	95580	9320	1435	228	60	14	106637
Cosmic Ray Veto	94186	9306	1431	228	60	14	105225
Lepton z_0	93728	9285	1429	228	60	14	104744
Sep. 2004 - Aug. 2005							
Initial	403468	128616	48870	10004	1629	277	592864
Silicon Good Run	310412	98989	37533	7668	1265	216	456083
≥ 1 Lepton	159373	54507	19925	4062	682	111	238660
Missing $E_T > 20$ GeV	113415	13156	3448	842	203	37	131101
Lepton Iso.	108352	10609	1785	281	65	14	121106
Dilepton Veto	108192	10542	1761	269	65	14	120843
Z Mass Veto	104829	10133	1682	262	62	13	116981
Cosmic Ray Veto	103863	10119	1678	262	62	13	115997
Lepton z_0	103709	10108	1678	262	62	13	115832
Sep. 2005 - Feb. 2006							
Initial	251138	79914	30742	6144	1001	168	369107
Silicon Good Run	218735	69527	26767	5356	849	145	321379
≥ 1 Lepton	111394	37668	14155	2805	421	71	166514
Missing $E_T > 20$ GeV	79211	9340	2409	564	119	25	91668
Lepton Iso.	75286	7580	1195	176	40	6	84283
Dilepton Veto	75174	7539	1184	173	40	6	84116
Z Mass Veto	72747	7245	1133	168	38	5	81336
Cosmic Ray Veto	72134	7235	1132	167	38	5	80711
Lepton z_0	72010	7228	1132	167	38	5	80580
March 2006 - Aug. 2006							
Initial	183034	58237	22399	4551	753	153	269127
Silicon Good Run	142801	45681	17491	3538	578	125	210214
≥ 1 Lepton	72450	24888	9251	1778	299	60	108726
Missing $E_T > 20$ GeV	51564	6247	1595	396	88	22	59912
Lepton Iso.	48691	4973	785	129	24	11	54613
Dilepton Veto	48627	4948	775	127	24	11	54512
Z Mass Veto	47020	4771	746	122	24	11	52694
Cosmic Ray Veto	46637	4760	745	122	23	11	52298
Lepton z_0	46572	4756	744	122	23	11	52228

Table A.2: Event yields at each stage of event selection for CMUP muons. The data are sorted by number of jets, and results are presented for four run ranges.

	W+0 jets	W+1 jet	W+2 jets	W+3 jets	W+4 jets	W+ \geq 5 jets	Total
Feb. 2002 - Aug. 2004							
Initial	387035	122514	47405	9789	1591	305	568639
Silicon Good Run	259037	81928	31579	6519	1041	202	380306
≥ 1 Lepton	77576	26703	9976	2026	299	56	116636
Missing $E_T > 20$ GeV	54589	6026	1591	368	79	18	62671
Lepton Iso.	52610	4880	767	116	18	6	58397
Dilepton Veto	52566	4855	751	112	18	6	58308
Z Mass Veto	51113	4684	727	110	17	6	56657
Cosmic Ray Veto	50903	4679	726	110	17	6	56441
Lepton z_0	50559	4661	721	109	17	6	56073
Sep. 2004 - Aug. 2005							
Initial	403468	128616	48870	10004	1629	277	592864
Silicon Good Run	310412	98989	37533	7668	1265	216	456083
≥ 1 Lepton	101572	36796	13428	2636	425	65	154922
Missing $E_T > 20$ GeV	71015	8280	2161	526	101	23	82106
Lepton Iso.	68041	6605	1030	163	37	7	75883
Dilepton Veto	67981	6572	1014	159	37	7	75770
Z Mass Veto	66002	6328	974	149	36	6	73495
Cosmic Ray Veto	65873	6322	973	149	36	6	73359
Lepton z_0	65711	6312	970	149	35	6	73183
Sep. 2005 - Feb. 2006							
Initial	251138	79914	30742	6144	1001	168	369107
Silicon Good Run	218735	69527	26767	5356	849	145	321379
≥ 1 Lepton	72460	26035	9373	1870	296	46	110080
Missing $E_T > 20$ GeV	50694	5963	1591	379	85	21	58733
Lepton Iso.	48379	4775	706	124	25	8	54017
Dilepton Veto	48349	4751	698	121	25	8	53952
Z Mass Veto	46908	4568	663	111	24	8	52282
Cosmic Ray Veto	46787	4563	663	110	24	8	52155
Lepton z_0	46670	4558	659	109	24	8	52028
March 2006 - Aug. 2006							
Initial	183034	58237	22399	4551	753	153	269127
Silicon Good Run	142801	45681	17491	3538	578	125	210214
≥ 1 Lepton	46190	16678	6082	1234	193	40	70417
Missing $E_T > 20$ GeV	32463	4014	992	263	51	16	37799
Lepton Iso.	30824	3171	487	77	15	3	34577
Dilepton Veto	30793	3160	474	76	14	3	34520
Z Mass Veto	29840	3032	448	73	14	3	33410
Cosmic Ray Veto	29771	3024	447	73	14	3	33332
Lepton z_0	29700	3020	446	73	14	3	33256

Table A.3: Event yields at each stage of event selection for CMX muons. The data are sorted by number of jets, and results are presented for four run ranges.

Appendix B

Calibration of Heavy Flavor in $W + \text{Jets}$

We present in this appendix the determination of the heavy flavor content of the $W + \text{jets}$ data sample. As described in Chapter 6, heavy flavor b -tags from processes like $Wb\bar{b}$ constitute a significant background to $t\bar{t}$ in the lepton+jets decay channel. While the cross sections for these background processes can be calculated, large NLO and NNLO effects lead to large theoretical uncertainties. Rather than relying on these calculations, we instead normalize the $W + \text{jets}$ background to the data sample and measure the *heavy flavor fraction*, the fraction of $W + \text{jets}$ events containing heavy flavor jets [80]:

$$F^{hf} \equiv \frac{N^{W+HF}}{N^{W+jets}}, \quad (\text{B.1})$$

where the N 's represent the number of events in data. Since the heavy flavor fraction is effectively a ratio of cross sections, we expect many of the large systematic

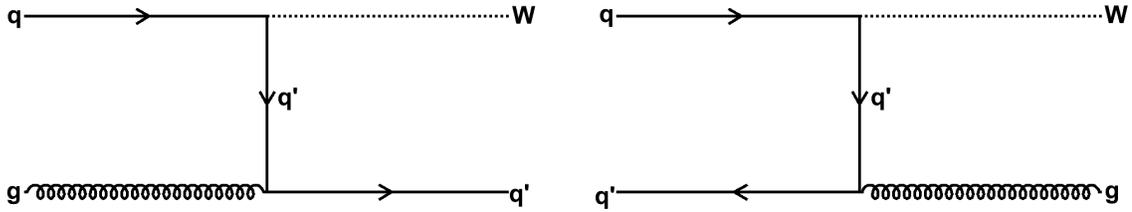
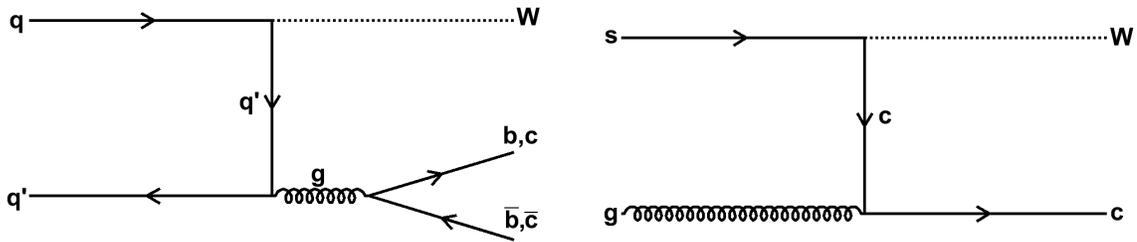
uncertainties on the cross sections to cancel out.

We calculate the heavy flavor fractions in two parts: we first determine the fractions in $W+jets$ Monte Carlo, and we then check the method by comparing the measured heavy flavor content of all-jet data and Monte Carlo samples. In the second step, we measure a multiplicative correction to the $W+heavy$ flavor fractions, K , accounting for mis-modeling of heavy flavor production modes common to all-jet and $W+jet$ processes, especially gluon splitting [70]. Details on this correction can be found in Section B.3.

Details on the usage of the fractions in Method 2 analyses can be found in Chapter 6.

B.1 $W+jets$ Monte Carlo

We use the ALPGEN generator, which calculates exact leading-order matrix elements for high-multiplicity processes ($W+4$ jets, for instance). ALPGEN only generates particles in the matrix element, the W (and its decay products), quarks, and gluons, so we must pass the output to a *parton showering* Monte Carlo package, which handles fragmentation, gluon splitting, and particle decays. In this case, events are showered with PYTHIA, and are simulated with the standard CDF package (see Section 4.5). We produce separate samples for each class of events, sorted by the number and type of heavy flavor quarks (denoted “Q”) and light flavor partons (“p”) in the matrix element, up to a final-state multiplicity of four partons. We account for contributions from W , $Wb\bar{b}$, $Wc\bar{c}$, and Wc with additional jets. Some example diagrams are shown in Figures B.1 and B.2.


 Figure B.1: Example Feynman diagrams contributing to W +jets with no heavy flavor.

 Figure B.2: Example Feynman diagrams contributing to $Wb\bar{b}$ and $Wc\bar{c}$ (left) and Wc (right).

Because each final state is explicitly generated, events must be matched to their respective samples to avoid double counting. For instance, an additional jet produced by PYTHIA in the parton shower may occupy the same phase space as an event generated in a higher multiplicity sample. We therefore impose kinematic restrictions on light partons generated in each sample, and we reject events with more or fewer “jets” than expected.¹ The generator and matching parameters used in ALPGEN are listed in Table B.2. In the highest-multiplicity samples, events with extra jets are allowed.

We simulate 15 samples to represent the W sample out to four jets: $W+[0-4]p$,

¹These “jets” are clustered before detector simulation, so are not equivalent to the jets used in the analysis. A four-parton final state, for instance, can be reconstructed with fewer than four jets.

Process	Cross Section
$W+0p$	1.81 nb
$W+1p$	225 pb
$W+2p$	35.3 pb
$W+3p$	5.59 pb
$W+4p$	1.03 pb
$Wb\bar{b}+0p$	2.98 pb
$Wb\bar{b}+1p$	0.888 pb
$Wb\bar{b}+2p$	0.287 pb
$Wc\bar{c}+0p$	5.00 pb
$Wc\bar{c}+1p$	1.79 pb
$Wc\bar{c}+2p$	0.628 pb
$Wc+0p$	17.1 pb
$Wc+1p$	3.39 pb
$Wc+2p$	0.507 pb
$Wc+3p$	83 fb

Table B.1: ALPGEN Monte Carlo samples used for the W heavy flavor fractions with cross sections.

$Wb\bar{b}+[0-2]p$, $Wc\bar{c}+[0-2]p$, and $Wc+[0-3]p$. A list of the W samples, including cross sections, is presented in Table B.1.

The heavy flavor fractions are measured in the W +jets channel, and we therefore subject the Monte Carlo to the same event selection as is used for the $t\bar{t}$ cross section measurement. We require a tight, isolated, 20-GeV lepton (electron or muon) in the CEM, CMUP, or CMX. Jets are required to have $E_T > 20$ GeV and have pseudo-rapidity less than 2, and the missing energy is required to exceed 20 GeV after being adjusted for jet energy corrections and muon momenta.

Parameter	Setting	Meaning
Gen. p_{Tj}	>15	Minimum parton p_T
Gen. η_j	<3	Maximum light parton η
Gen. ΔR_{jj}	>0.4	Minimum parton-parton (and Q-parton) ΔR
Gen. heavy flavor p_T	>8	Minimum b/c parton p_T or $b\bar{b}/c\bar{c}$ p_T
Gen. heavy flavor η	<3	Maximum b/c η
Gen. heavy flavor ΔR	>0	Minimum ΔR between heavy flavor quarks
Gen. p_{Tl}	>1	Minimum lepton p_T
Gen. η_l	<5	Maximum lepton η
Gen. ΔR_{lj}	>0	Minimum ΔR between lepton and jets
Match jet p_T	>15	Matching cluster E_T
Match jet η	<3	Matching η
Match jet ΔR	>0.4	Matching parton-parton (and Q-parton) ΔR

Table B.2: Settings used in ALPGEN 2.1 by default for the W Monte Carlo samples.

B.2 Raw Heavy Flavor Fractions

To merge the 15 samples, events are weighted according to their cross sections, defined as the full sample cross section divided by the total number of generated events. We take a weighted sum of these events, sorted by the number of reconstructed jets, to produce the denominator of the fraction. We count the (weighted) number of events with bottom-quark or charm-quark jets² to determine the numerators. Because the event weight is derived using a total cross section and number of events before event selection, this method incorporates the relative acceptance for each sample. These acceptances are shown in Table B.3.

Precisely, we calculate the fractions according to the following formulae, where the sum is over all 15 W samples:

²A bottom- or charm-quark jet is defined as a reconstructed jet with a bottom or charm hadron within a cone of $\Delta R < 0.4$.

$$w^i = \frac{\sigma_{gen}^i}{N_{gen}^i} \quad (\text{B.2})$$

$$F_{2b} = \frac{\sum w^i N_{\geq 2b}^i}{\sum w^i N^i} \quad (\text{B.3})$$

$$F_{1b} = \frac{\sum w^i N_{1b}^i}{\sum w^i N^i} \quad (\text{B.4})$$

$$F_{2c} = \frac{\sum w^i N_{\geq 2c,0b}^i}{\sum w^i N^i} \quad (\text{B.5})$$

$$F_{1c} = \frac{\sum w^i N_{1c,0b}^i}{\sum w^i N^i} \quad (\text{B.6})$$

Here, the N_{gen}^i are determined after filtering for good runs with a well-defined Z vertex and rejecting events which fail the parton matching described in Section B.1. In rare cases, where multiple species of heavy flavor jets are present, events will count in both the bottom and charm groups. The effect of this classification is less than 0.1%.

B.2.1 Heavy Flavor Overlap Removal

In ALPGEN, the final-state light flavor parton multiplicity is required to match the generated multiplicity, as described in Section B.1; however, there is no equivalent matching procedure for final-state heavy flavor quarks, such that there are cases where the same heavy flavor final states can be produced in different samples after the parton shower. A $Wb\bar{b}$ event, for instance, can be produced in both $Wb\bar{b}+0p$ and $W+1p$ samples, when the outgoing parton is a gluon that splits to a $b\bar{b}$ pair, as in Figure B.1 (right) and Figure B.2 (left). When we combine multiple ALPGEN samples, we therefore reject some heavy flavor production modes in each of the samples to avoid double counting.

The simplest approach is to require the appropriate heavy flavor content for the dedicated heavy samples ($Wb\bar{b}$, $Wc\bar{c}$, and Wc), allowing *additional* heavy flavor from the shower only when it is lighter than the primary generated partons (i.e., charm pairs in $Wb\bar{b}$ are allowed, not vice versa) or when it fails the kinematic requirements used at generation (p_T of either heavy quark *or* the pair must be greater than 8 GeV/ c). The problem with this approach is a mismatch in the heavy flavor production rates from the shower (generated by PYTHIA) and the matrix element, which in turn gives rise to kinematic discontinuities in the merged, weighted $W+jets$ sample. An illustration of this is shown in Figure B.3, where the p_T of $c\bar{c}$ pairs is shown from the weighted ALPGEN sample in the stacked solid histogram, and the dashed line shows the same distribution for $c\bar{c}$ pairs from the shower in the $W+Np$ samples. While some disagreement between the two models is expected, the excess production of charm pairs at low p_T leaves a discontinuity in the distribution when the samples are combined in this way.

An alternative approach, the one adopted in this analysis, is to select heavy flavor production from the shower or matrix element based on reconstructed quantities, like jets. As shown in Figure B.4, we find that the shower model and ALPGEN predict very different ΔR (separation) distributions between heavy quarks, where the shower is tuned on the more collinear gluon splitting pairs but fails for large opening angles. We explicitly reject events from ALPGEN where the matrix-element heavy flavor quarks fall in the same reconstructed jet, and we also remove heavy flavor from the shower in which only one of the two heavy quarks is inside the jet cone. In both cases, this distinction is made with simple geometric cuts, defining quarks with $\Delta R < 0.4$ from

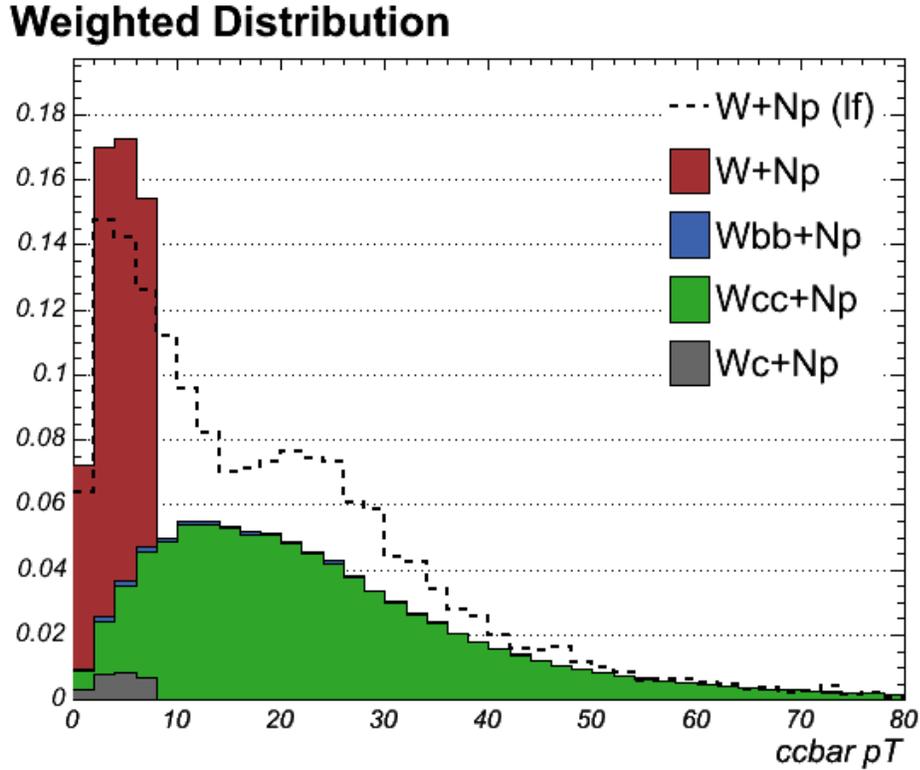


Figure B.3: Distribution of p_T of charm pairs from the dedicated ALPGEN samples (stacked histogram) and from only the shower in PYTHIA. Though the ALPGEN samples are generated with a p_T cut on the heavy flavor quark momentum, the shower does not produce charm at a rate consistent with the matrix element.

the jet axis to be a part of the jet. The fractions of accepted events kept for all subsamples are listed in Table B.3.

This strategy ignores one of the objectives of the p_T -based overlap removal described above: we make no effort to fill in the phase space at low p_T . However, the heavy flavor we ignore falls below the jet threshold, and does not contribute to the numerator of the heavy flavor fractions. The size of the effect is 2-5%, depending on the sample, and is corrected by the K correction described in Section B.3. Figures B.5 to B.8 show comparisons of p_T and ΔR distributions from both methods.

Weighted Distribution

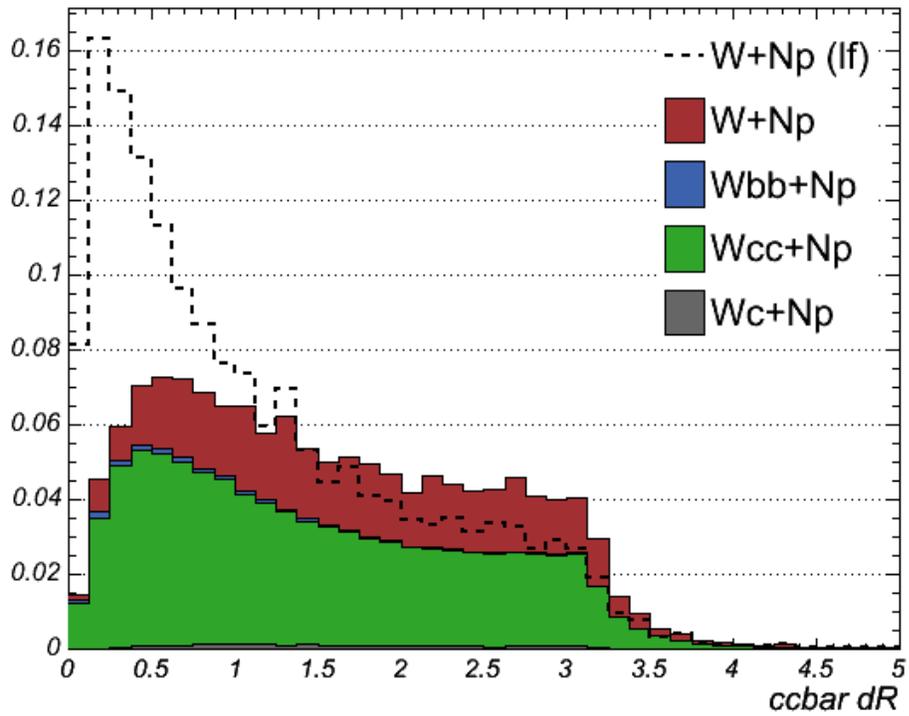


Figure B.4: Distribution of ΔR between charm pairs from the dedicated ALPGEN samples (stacked histogram) and from only the shower in PYTHIA. The shower model is expected to be inferior for back-to-back heavy flavor production, but the large discrepancy at small angles is unexpected.

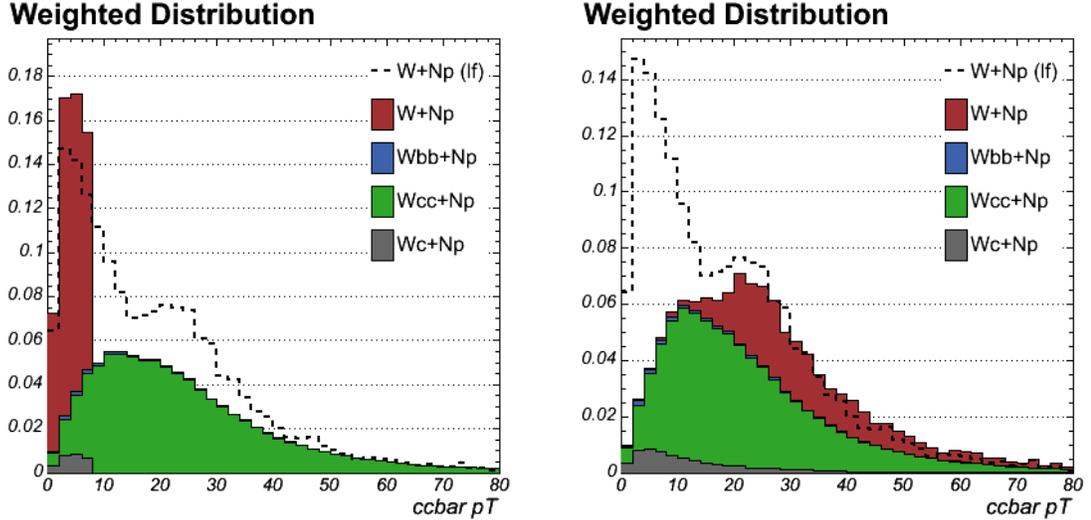


Figure B.5: Distribution of p_T for charm pairs from the dedicated ALPGEN samples (stacked histogram) and from only the shower in PYTHIA. The plot on the left uses the kinematic overlap removal scheme, and the plot on the right uses the new ΔR scheme to combine samples. The discrepancy at low p_T only contributes to the denominator of the fractions, not the numerator.

Process	Total Events	Pretag Events (≥ 1 Jet)	Kept Events	Acceptance ($\sigma_A < 0.001$)	Kept Fraction
$W+0p$	942k	2562	2330	0.003	0.909 ± 0.003
$W+1p$	957k	135542	117977	0.142	0.870 ± 0.001
$W+2p$	947k	195163	160622	0.206	0.823 ± 0.001
$W+3p$	948k	211304	168047	0.223	0.795 ± 0.001
$W+4p$	935k	209037	160867	0.224	0.770 ± 0.001
$Wb\bar{b}+0p$	1458k	129006	111967	0.088	0.868 ± 0.001
$Wb\bar{b}+1p$	1461k	283983	257195	0.194	0.906 ± 0.001
$Wb\bar{b}+2p$	1415k	316030	280263	0.223	0.887 ± 0.001
$Wc\bar{c}+0p$	1897k	153223	132766	0.081	0.866 ± 0.001
$Wc\bar{c}+1p$	1877k	350948	316932	0.187	0.903 ± 0.001
$Wc\bar{c}+2p$	1819k	399549	351450	0.220	0.880 ± 0.001
$Wc+0p$	1820k	150225	150203	0.081	1.000 ± 0.001
$Wc+1p$	1752k	335470	334143	0.191	0.996 ± 0.001
$Wc+2p$	1818k	425248	421732	0.234	0.992 ± 0.001
$Wc+3p$	1814k	443731	435903	0.245	0.982 ± 0.001

Table B.3: Acceptance for events from each physics process and the fraction of events in each Monte Carlo sample left after removal of overlap events. The uniqueness requirement is applied here after event selection, *requiring at least one jet*. The numbers here are specific to CEM electrons.

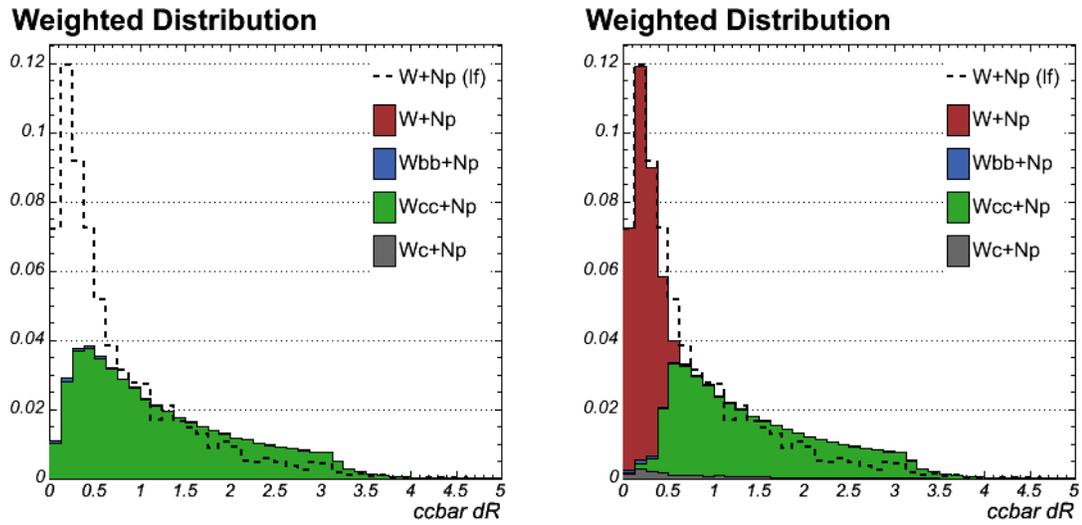


Figure B.6: Distribution of ΔR between charm pairs from the dedicated ALPGEN samples (stacked histogram) and from only the shower in PYTHIA, for pairs with a combined p_T greater than 20 GeV. The plot on the left uses the kinematic overlap removal scheme, and the plot on the right uses the new ΔR scheme to combine samples.

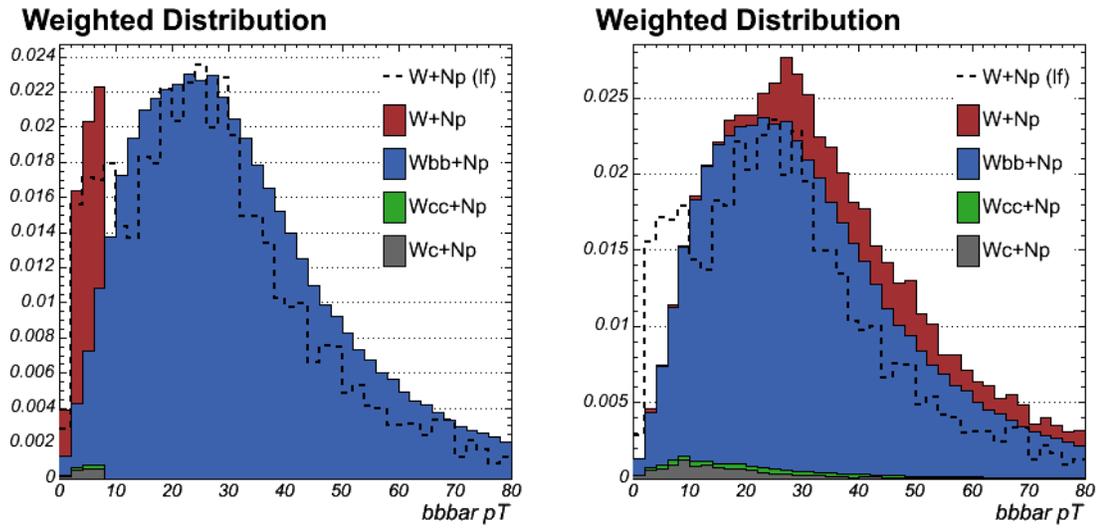


Figure B.7: Distribution of p_T between bottom pairs from the dedicated ALPGEN samples (stacked histogram) and from only the shower in PYTHIA. The plot on the left uses the kinematic overlap removal scheme, and the plot on the right uses the new ΔR scheme to combine samples. The discrepancy at low p_T only contributes to the denominator of the fractions, not the numerator.

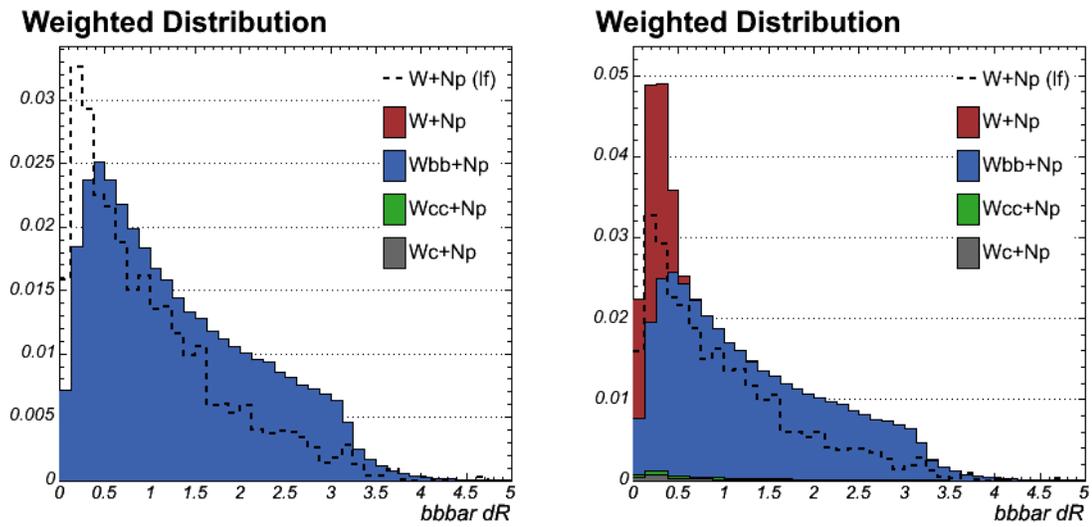


Figure B.8: Distribution of ΔR between bottom pairs from the dedicated ALPGEN samples (stacked histogram) and from only the shower in PYTHIA, for pairs with a combined p_T greater than 20 GeV. The plot on the left uses the kinematic overlap removal scheme, and the plot on the right uses the new ΔR scheme to combine samples.

B.2.2 Initial Results

Following the event selection and overlap removal procedure described above, we derive the heavy flavor fractions directly from Monte Carlo. These are shown in Table B.4 with statistical errors only. We observe fractions which are somewhat higher than those previously derived using an older version of ALPGEN, although the K correction then boosted the fractions by an additional 50% [76]. These results are consistent with those measured in Run I.

	1-jet	2-jet	3-jet	≥ 4 -jet
Monte Carlo Heavy Flavor Fractions (%)				
$Wb\bar{b}$, 1 b	0.68 ± 0.01	1.36 ± 0.04	2.43 ± 0.08	3.26 ± 0.10
$Wb\bar{b}$, 2 b		0.86 ± 0.02	1.84 ± 0.07	3.28 ± 0.11
$Wc\bar{c}$ or Wc , 1 c	5.5 ± 0.1	8.7 ± 0.2	11.7 ± 0.3	13.3 ± 0.4
$Wc\bar{c}$, 2 c		1.38 ± 0.03	3.27 ± 0.13	5.99 ± 0.14

Table B.4: The heavy flavor fractions from ALPGEN W +jets Monte Carlo. Errors are statistical only.

B.3 K -Factor Determination

Though we expect heavy flavor fractions to be more stable against higher-order corrections, this assumption can not be tested directly in W +jets data. Instead, we check heavy flavor production from ALPGEN in an independent sample of all-jet QCD events, where statistics are much higher and the kinematics are well-modeled. In particular, we assume that the rates of gluon splitting to bottom and charm pairs will be (mis)estimated consistently in W and all-jet Monte Carlo, since gluon splitting contributes to both through the diagrams shown in Figure B.9.

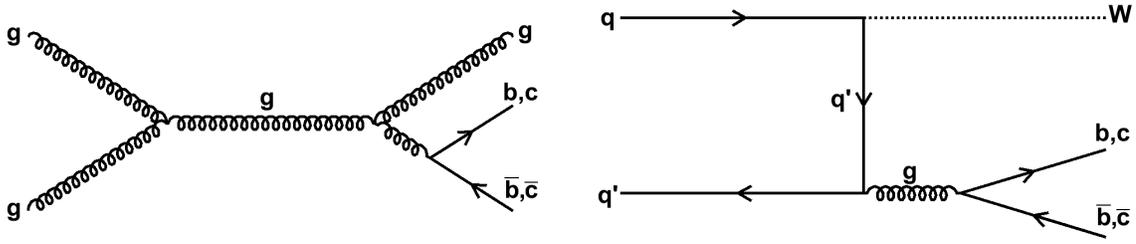


Figure B.9: Feynman diagrams contributing to heavy flavor production in all-jet (left) and W events (right).

We define K to be the ratio of heavy flavor fractions in QCD data and Monte Carlo ($\frac{F_{data}^{hf}}{F_{MC}^{hf}}$), and we use the result as a multiplicative correction to the $Wb\bar{b}$ and $Wc\bar{c}$ fractions shown in Table B.4. The sample generation and preparation techniques are necessarily identical to those used for the W samples, though here we use ALPGEN QCD samples with light flavor ($[2-5]p$), $b\bar{b}+[0-3]p$, and $c\bar{c}+[0-3]p$ final states. The heavy quarks and light partons are also generated with the same kinematic selection at generation, requiring light flavor partons to have a transverse momentum of 15 GeV/ c , while $b\bar{b}$ and $c\bar{c}$ pairs are only required to exceed 8 GeV/ c .

The data samples used are **Jet20** and **Jet50**, the same as those used in the mistag parameterization described in Section 5.3. These are simple calorimeter triggers that require local clusters³ with energy exceeding 15 GeV and 40 GeV at Level 2 and clustered jets above 20 GeV and 50 GeV at Level 3. We require offline that all jets have corrected transverse energy over 20 GeV and an absolute pseudo-rapidity η less than 2. Further, to avoid the effects of the trigger turn-on, we require at least one jet to have E_T over 70 GeV.

³The Level 2 trigger does not use sophisticated jet clustering; rather, the trigger only requires adjacent towers above threshold and sums their deposited energies.

Overall, the kinematics of the weighted and merged ALPGEN samples compare quite well with data. Some examples are shown in Figures B.10 to B.11. For comparison, an additional sample of PYTHIA dijets is also shown; the PYTHIA sample is a generic $2 \rightarrow 2$ simulation, requiring at least one of the outgoing particles to have transverse momentum larger than $40 \text{ GeV}/c$. Of particular interest are the plots in Figure B.11, which are sensitive to the relative rates of gluon splitting and direct production of heavy flavor jets.

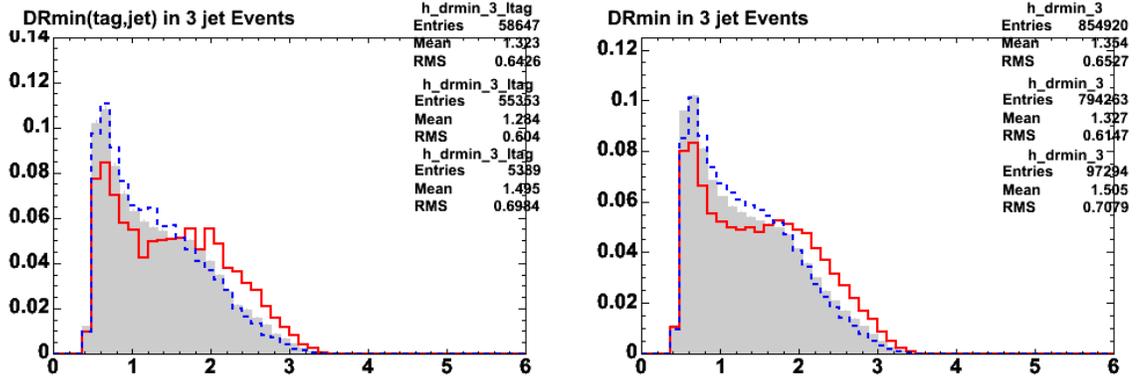


Figure B.10: Smallest ΔR between jets in all (left) and b -tagged (right) 3-jet events. The data is represented by the shaded histogram, while ALPGEN (dashed) and PYTHIA (solid) are represented by single lines.

We initially measure the relative fractions of bottom, charm, and light flavor among the set of *non-leading* b -tagged jets. The leading jet is removed from the fits due to the large mistag probability at high energy and the minimal presence of gluon splitting, the heavy flavor production mode that dominates $W+jets$. We discriminate between species by fitting the distributions of the tag mass (invariant mass of all tracks in the b -tagged vertex) and $c\tau$ (two-dimensional displacement of the vertex times the tag mass divided by the vertex transverse momentum). We fit

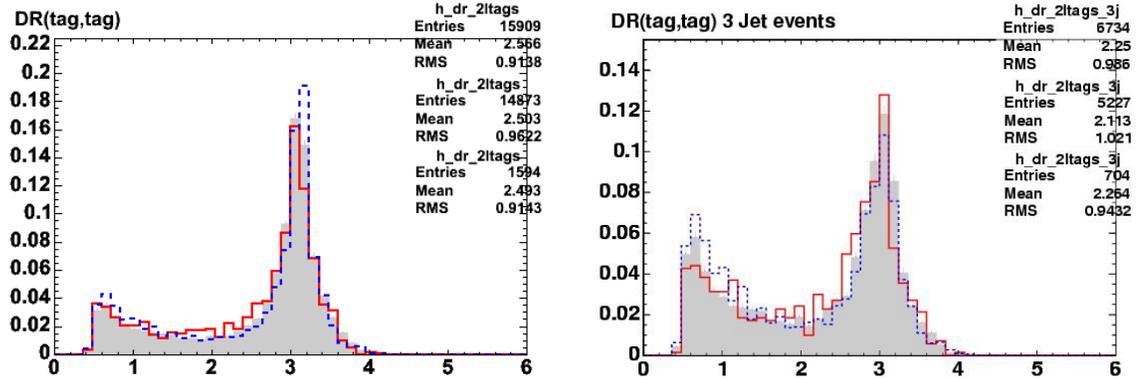


Figure B.11: Smallest ΔR between b -tagged jets in all (left) and 3-jet (right) double loose-tagged events. The data is represented by the shaded histogram, while ALPGEN (dashed) and PYTHIA (solid) are represented by single lines. The peaks at 0.5 and 3 correspond to gluon splitting ($g \rightarrow b\bar{b}$) and direct $b\bar{b}$ production.

both distributions and use both tunings of the SECVTX b -tagger, taking the spread as a systematic uncertainty. Since the efficiency to b -tag bottom and charm jets is in principle known from the Monte Carlo rates and the b -tagging scale factor SF (see Section 5.2), the *pretag* heavy flavor content of the data samples can be inferred from the fit results:

$$F_{pre}^{hf} = \frac{N_{tag}^{hf}}{SF \epsilon_{MC}^{hf} N_{pre}}, \quad (\text{B.7})$$

where N_{tag}^{hf} is the fit number of heavy flavor jets, ϵ_{MC}^{hf} is the b -tagging efficiency for heavy flavor jets in Monte Carlo, and N_{pre} is the total number of jets in the *pretag* sample. The K factor is the ratio of *pretag* heavy flavor fractions in data and simulation, which may depend on jet multiplicity and energy, and also may be different for charm and bottom.

Example template fits to data for the loose SECVTX vertex mass and the tight

SECVTX pseudo- $c\tau$ are shown in Figure B.13. Overall fit fractions in data and true fractions from Monte Carlo are presented in Table B.5, and the final K results (and those sorted by jet multiplicity) are shown in Table B.6. Figure B.12 shows the observed dependence on the jet energy.

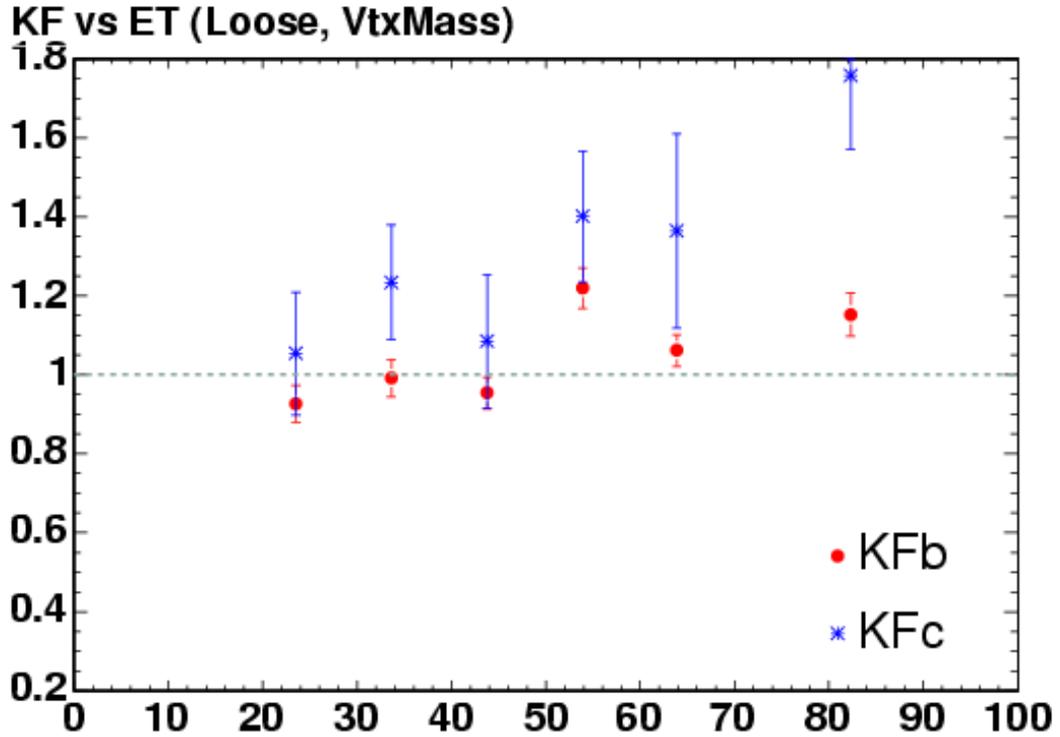
Fraction	Data	ALPGEN
b	0.527 ± 0.008	0.558 ± 0.004
c	0.266 ± 0.011	0.243 ± 0.002
light	0.207 ± 0.006	0.199 ± 0.002

Table B.5: Fitted fraction of heavy flavor with the vertex mass distribution using loose SECVTX, with statistical errors only.

Tagger, Var.	All	1-jet	2-jet	3-jet	≥ 4 -jet
K_b					
Loose, Mass	1.01 ± 0.06	1.12 ± 0.09	1.05 ± 0.09	0.97 ± 0.14	0.94 ± 0.31
Loose, $c\tau$	1.19 ± 0.07	1.48 ± 0.09	1.01 ± 0.08	0.90 ± 0.15	1.06 ± 0.32
Tight, Mass	0.98 ± 0.07	1.24 ± 0.10	0.94 ± 0.09	1.08 ± 0.15	0.80 ± 0.28
Tight, $c\tau$	1.17 ± 0.08	1.63 ± 0.10	1.12 ± 0.10	0.99 ± 0.18	1.10 ± 0.37
K_c					
Loose, Mass	1.17 ± 0.12	1.53 ± 0.19	0.91 ± 0.18	0.98 ± 0.33	0.99 ± 0.73
Loose, $c\tau$	1.08 ± 0.10	1.65 ± 0.11	1.04 ± 0.16	0.89 ± 0.26	0.66 ± 0.49
Tight, Mass	1.24 ± 0.18	1.26 ± 0.28	0.84 ± 0.21	1.04 ± 0.39	1.37 ± 0.87
Tight, $c\tau$	1.19 ± 0.10	1.23 ± 0.12	0.97 ± 0.15	0.67 ± 0.28	0.48 ± 0.52

Table B.6: Dependence of K on the number of jets for both taggers, two discriminating variables, and bottom and charm separately. The errors are statistical only.

We measure a K factor consistent with unity for both bottom and charm which appears to be largely independent of jet multiplicity and transverse energy. The value in the 1-jet bin for charm comes out somewhat higher, between 1.3 for tight SECVTX and 1.5 for loose SECVTX. K should be independent of the tagger, but the different tunings are differently sensitive to the various jet species. Since the leading

Figure B.12: K_b and K_c dependence on jet E_T .

jet is removed from the fits, and tagged dijets are dominated by direct heavy flavor production, this deviation is likely unique to the all-jet sample. We consider the departure in assigning systematics, but we do not incorporate the dependence into the K correction. We therefore take a K correction of 1.0 universally. The systematic error, including fit uncertainties, b -tagging scale factors for bottom and charm jets, and the spread in fit results with different tunings of `SECVTX`, is estimated to be 30%.

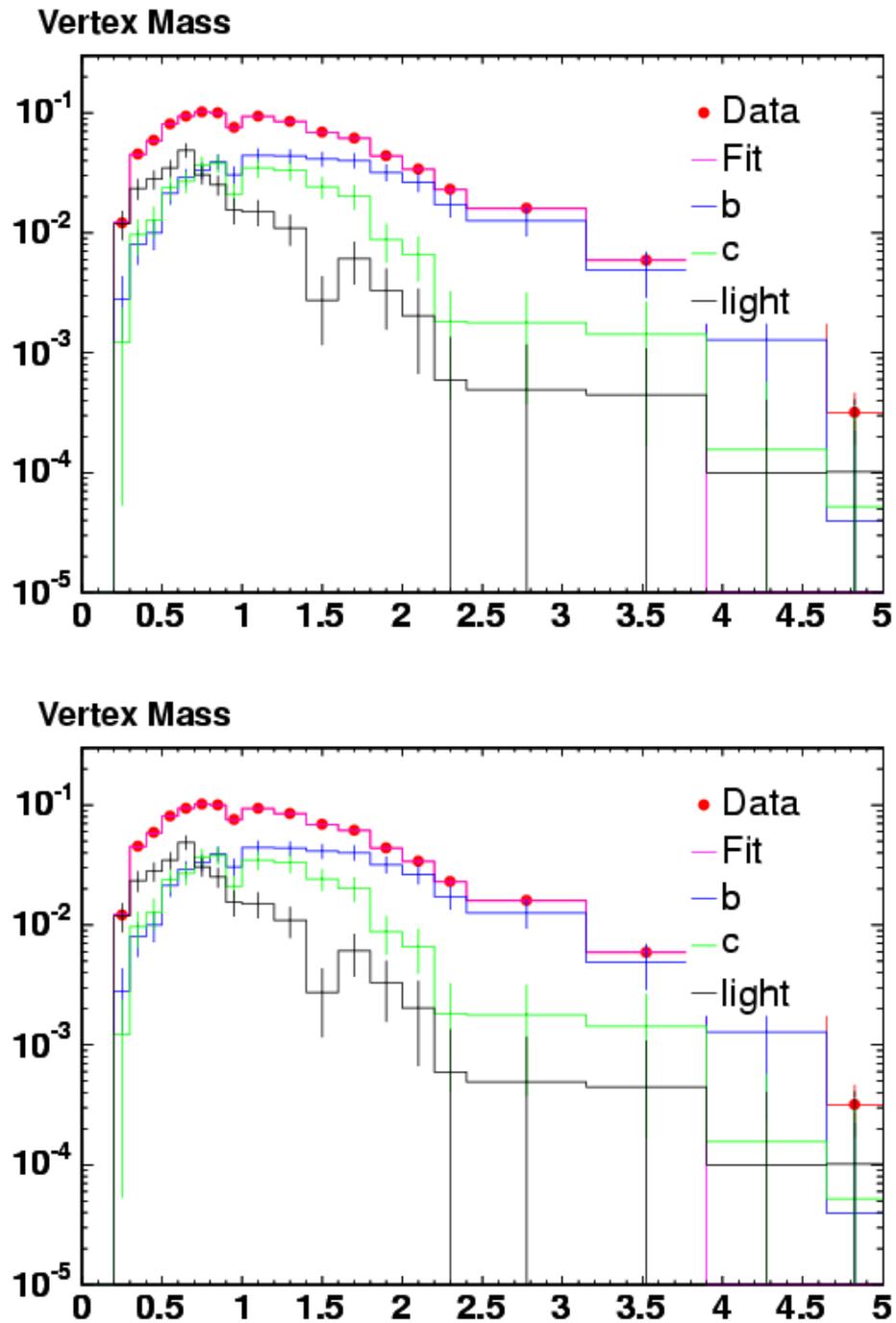


Figure B.13: Template fits to extract the fractions of b -tagged jets that are bottom and charm using the vertex mass and loose SECVTX (left) and using pseudo- $c\tau$ and tight SECVTX (right).

B.4 Systematic Uncertainties

In addition to the large systematic uncertainty on the K correction, we consider several sources of error due to settings in ALPGEN. Most relevant systematic errors can be evaluated directly by varying the parameters of ALPGEN and determining the heavy flavor fractions for these new sets. However, due to the difficulty of generating a full suite of samples for each set of parameters, we instead cut out the highest multiplicity samples and assign systematic errors based on the changes on the merged 2-jet and 3-jet bins. We generate small (15,000 events) samples for $W+0-3p$, $Wc+0-2p$, $Wb\bar{b}+0-1p$, and $Wc\bar{c}+0-1p$, and additionally save computing time by skipping detector simulation (*i.e.* only producing lists of generated particle momenta). We apply jet clustering directly to the Monte Carlo particle lists, and we approximate lepton+jets event selection by requiring central leptons and neutrinos. A comparison of this simplified generation and production scheme with real production is shown in Figure B.14; we find that our clustering tool works quite well on average, even getting the average jet E_T correct within a few percent.

We re-weight the default samples with the generated cross sections and acceptances, and we take the average shift in the heavy flavor production rates as a systematic. We vary the online parton matching parameters (minimum p_T of light partons and matched jet p_T), heavy quark masses, Q^2 scale, parton distribution function, and initial-state and final-state radiation to estimate these systematics. The sets of varied parameters and the two-jet heavy flavor fractions are summarized in Table B.7.⁴ For

⁴We simplify the overlap removal scheme when doing these studies, both because we lack real jets and also to conserve statistics. Therefore, the “default” value is slightly different from the results presented earlier. We evaluate systematic errors relative to the default, and apply the same relative error to the main results.

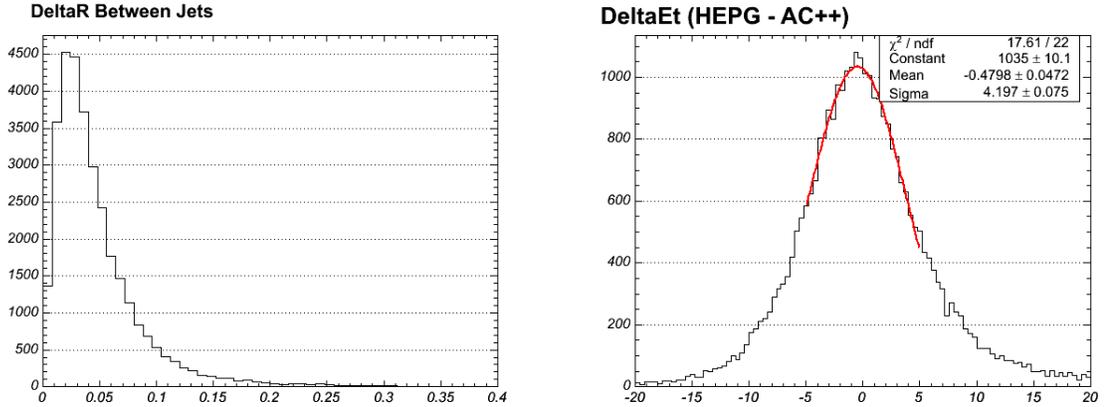


Figure B.14: Performance of the Monte Carlo-clustering tool used to evaluate systematic uncertainties in $W+2p$ Monte Carlo. The left plot shows ΔR between nearest simulated and Monte Carlo-only jets, and the right plot shows the difference in E_T between nearest jets. In both cases, the MC-only clustering tool produces results quite similar to fully simulated events.

each setting, we take the RMS deviation over the full set as the uncertainty. The systematic errors derived from this study are listed in Table B.8.

B.5 Results

After applying the K correction to the $Wb\bar{b}$ and $Wc\bar{c}$ fractions, and including all sources of systematic uncertainty, we derive the final W +heavy flavor fractions. The results are shown in Table B.9. The use of these fractions in evaluating the Method 2 W +jets sample composition is the subject of Chapter 6.

Setting	Fraction				Relative Shift			
	$1b$	$2b$	$1c$	$2c$	$1b$	$2b$	$1c$	$2c$
Default	0.0124	0.0068	0.0640	0.0104	0	0	0	0
Matching 10-10	0.0135	0.0070	0.0681	0.0103	0.088	0.015	0.064	-0.009
Matching 20-20	0.0121	0.0065	0.0637	0.0097	-0.024	-0.056	-0.004	-0.067
Double Q^2	0.0098	0.0058	0.0540	0.0081	-0.209	-0.158	-0.156	-0.221
Half Q^2	0.0156	0.0087	0.0739	0.0126	0.258	0.262	0.154	0.211
m_c low (-0.3 GeV)	0.0117	0.0066	0.0642	0.0097	-0.056	-0.042	0.003	-0.067
m_c high (+0.3 GeV)	0.0121	0.0070	0.0585	0.0099	-0.024	0.015	-0.085	-0.048
m_b low (-0.3 GeV)	0.0125	0.0068	0.0597	0.0099	0.008	-0.013	-0.067	-0.048
m_b high (+0.3 GeV)	0.0114	0.0063	0.0600	0.0097	-0.080	-0.085	-0.062	-0.067
Less ISR	0.0123	0.0073	0.0666	0.0105	-0.008	0.059	0.040	0.009
More ISR	0.0125	0.0068	0.0658	0.0099	0.008	-0.013	0.028	-0.048
Less FSR	0.0126	0.0073	0.0673	0.0107	0.016	0.059	0.051	0.028
More FSR	0.0120	0.0072	0.0634	0.0095	-0.032	0.044	-0.009	-0.086
PDF CTEQ6M	0.0119	0.00711	0.062	0.0103	-0.040	0.031	-0.032	-0.010

Table B.7: Observed shifts in two-jet heavy flavor fractions from systematic variation of ALPGEN parameters.

Systematic Source	Relative Error (%)
Matching	6
Q^2	20
Quark Masses (c and b)	6
ISR	4
FSR	5
PDF	3
K	30
Total	38

Table B.8: Relative systematic uncertainties derived from the deviations in Table B.7. The error is dominated by the K correction and the Q^2 scale. For Wc , we leave out the K factor error.

	Heavy Flavor Fraction (%)			
	1-jet	2-jet	3-jet	≥ 4 -jet
$Wb\bar{b}$, 1 b	0.7 ± 0.3	1.4 ± 0.6	2.4 ± 1.0	3.3 ± 1.4
$Wb\bar{b}$, 2 b		0.9 ± 0.4	1.8 ± 0.7	3.3 ± 1.3
$Wc\bar{c}$ or Wc , 1 c	5.5 ± 1.7	8.7 ± 3.2	11.7 ± 3.8	13.3 ± 4.9
$Wc\bar{c}$, 2 c		1.4 ± 0.6	3.3 ± 1.3	6.0 ± 2.4

Table B.9: The final heavy flavor fractions with systematic uncertainties and the K correction (1.0 ± 0.3) applied to $Wb\bar{b}$ and $Wc\bar{c}$.

Appendix C

Kinematics of the W +Jets Sample

We present in this section a series of stacked histograms displaying the properties of the ≥ 1 and ≥ 2 tag samples, assuming the measured value of the cross section in each sample. The templates for $t\bar{t}$ signal, diboson/single top/ $Z \rightarrow \tau\tau$, W +Heavy Flavor, and W +Light Flavor are taken directly from the corresponding Monte Carlo. The non- W template is taken entirely from the pretag antielectron sample described in Section 6.4.2. The contributions are normalized to the expectation by lepton type and number of jets. We use the baseline selection for both single- and double-tags.

Altogether, the kinematics of $t\bar{t}$ events are well-modeled in simulation. The event H_T , lepton p_T , and jet E_T and η distributions in data and Monte Carlo qualitatively agree quite well. On the other hand, the missing E_T and transverse mass distributions (Figures C.2 and C.3) show some discrepancies. The single-tag missing E_T distribution is broader in data than in Monte Carlo, having an excess between 60 and 80 GeV and a deficit between 30 and 50 GeV. There is also an excess in the data at low transverse mass, when the missing E_T and lepton are collinear.

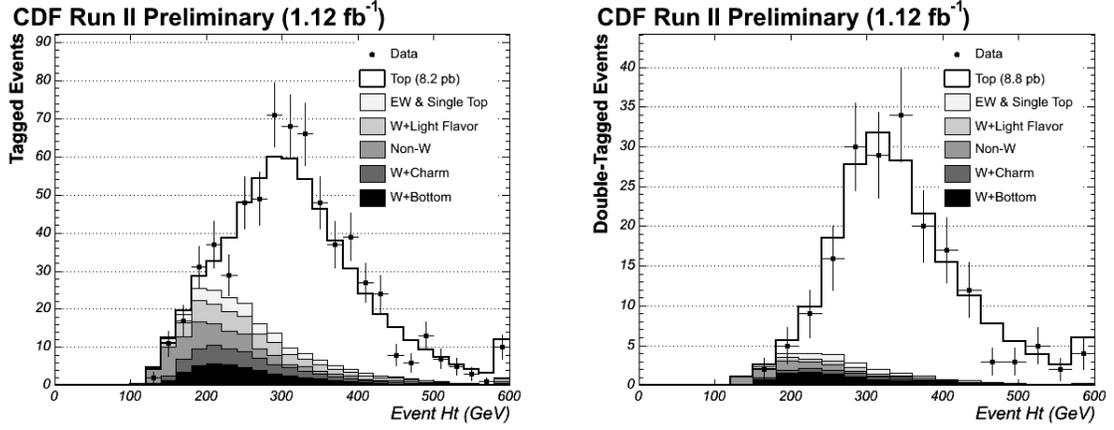


Figure C.1: Expected and observed distribution of H_T for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

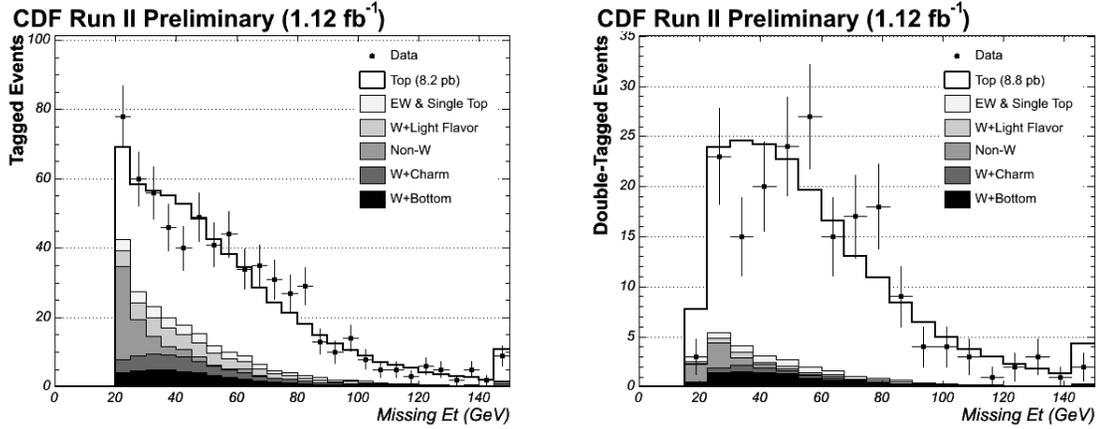


Figure C.2: Expected and observed distribution of missing E_T for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

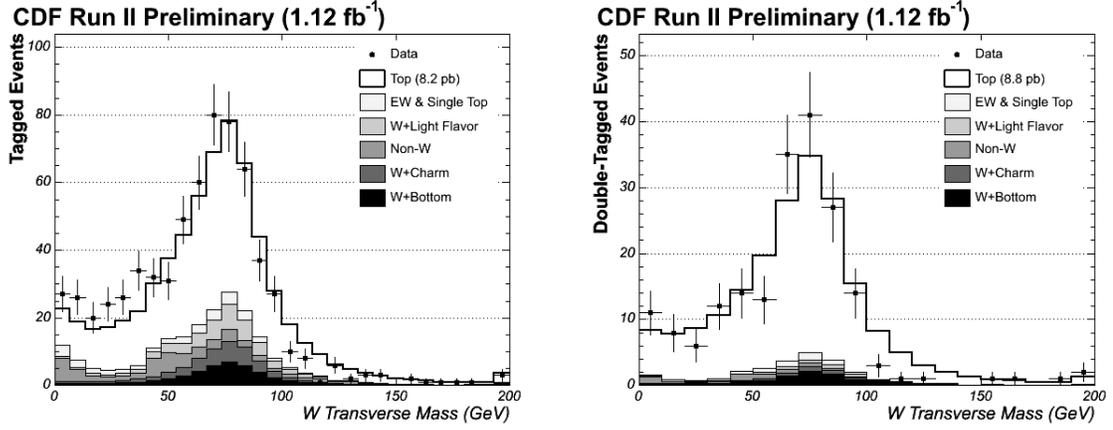


Figure C.3: Expected and observed distribution of m_T^W for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

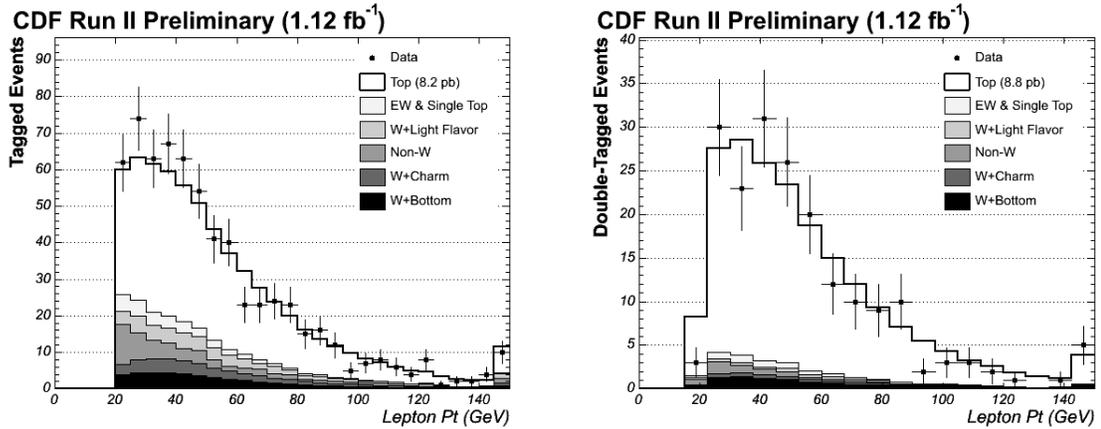


Figure C.4: Expected and observed distribution of lepton p_T for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

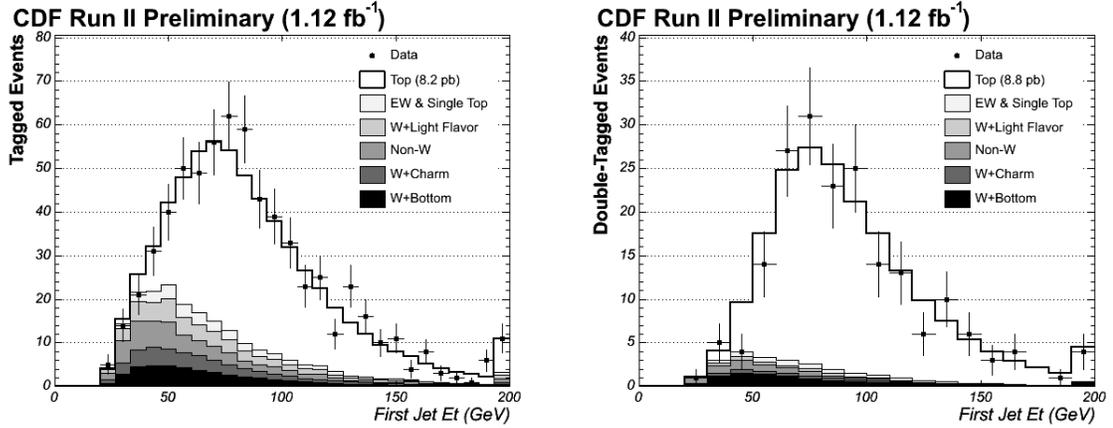


Figure C.5: Expected and observed distribution of leading jet E_T for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

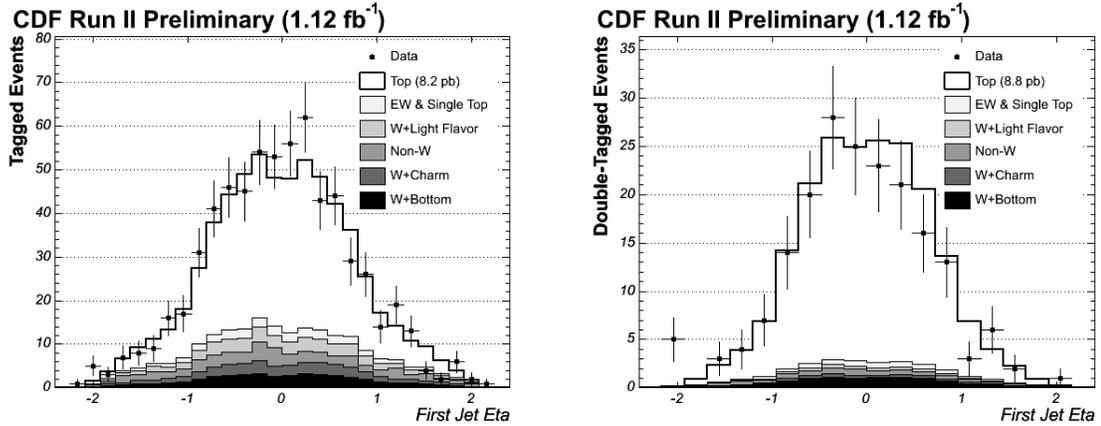


Figure C.6: Expected and observed distribution of leading jet η for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

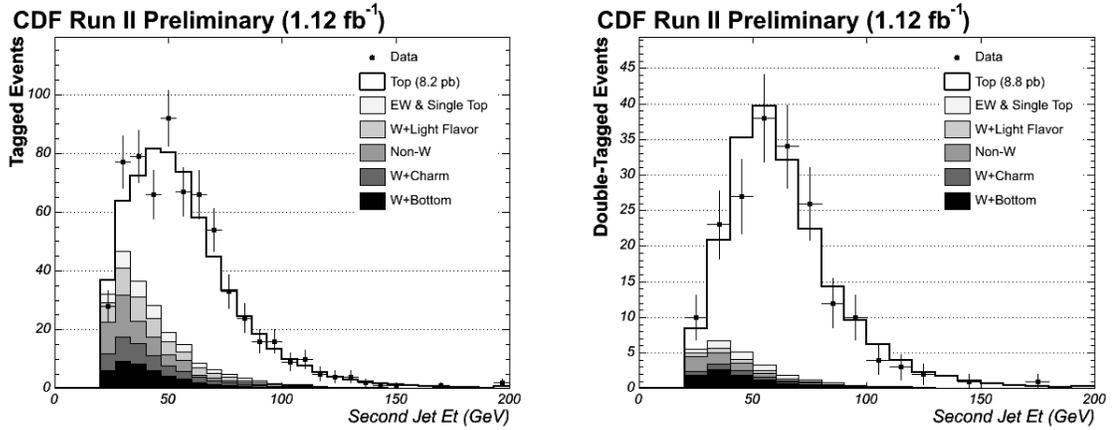


Figure C.7: Expected and observed distribution of second jet E_T for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

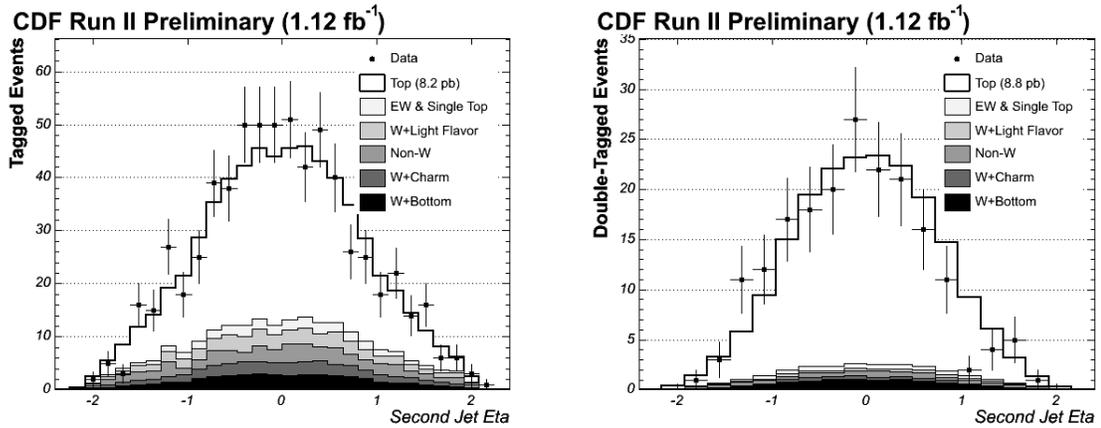


Figure C.8: Expected and observed distribution of second jet η for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

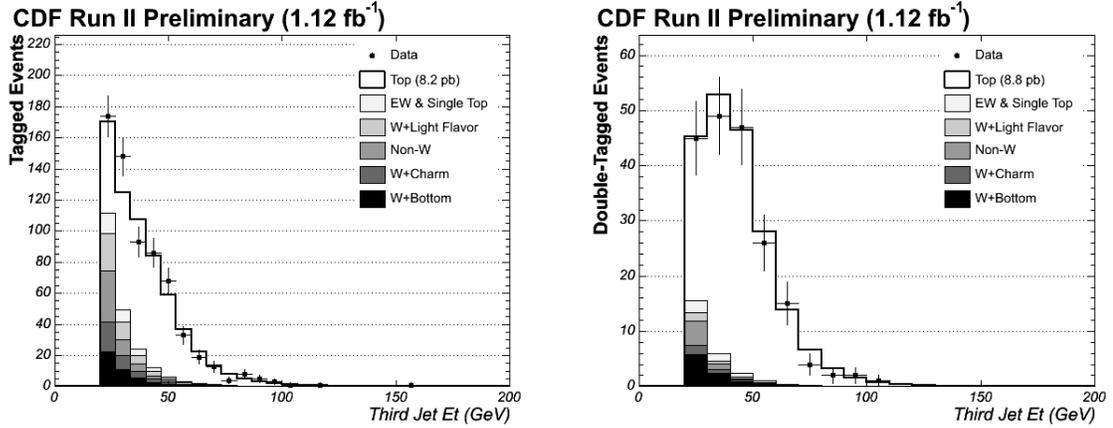


Figure C.9: Expected and observed distribution of third jet E_T for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

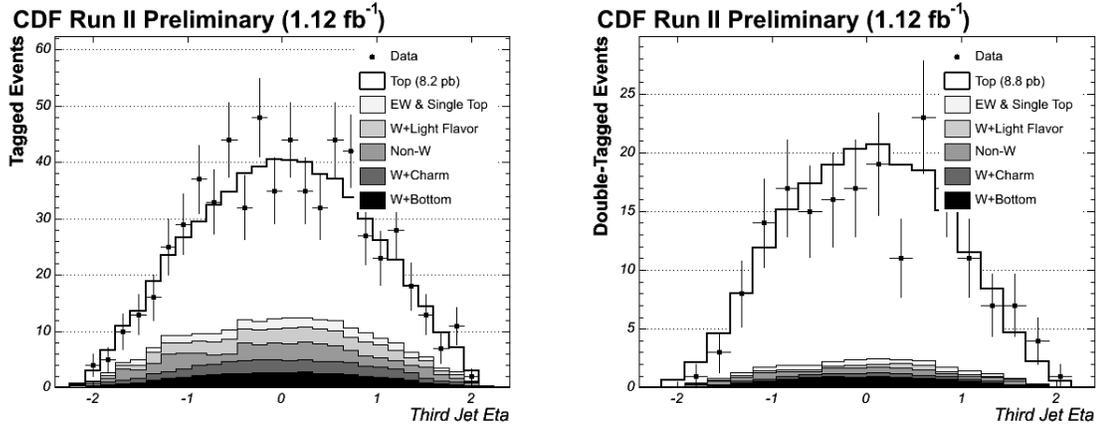


Figure C.10: Expected and observed distribution of third jet η for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

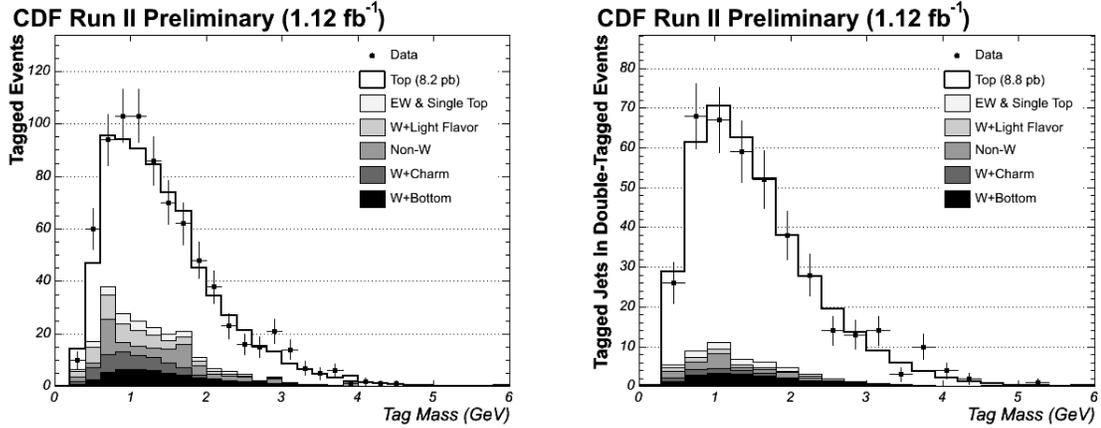


Figure C.11: Expected and observed distribution of tag mass for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

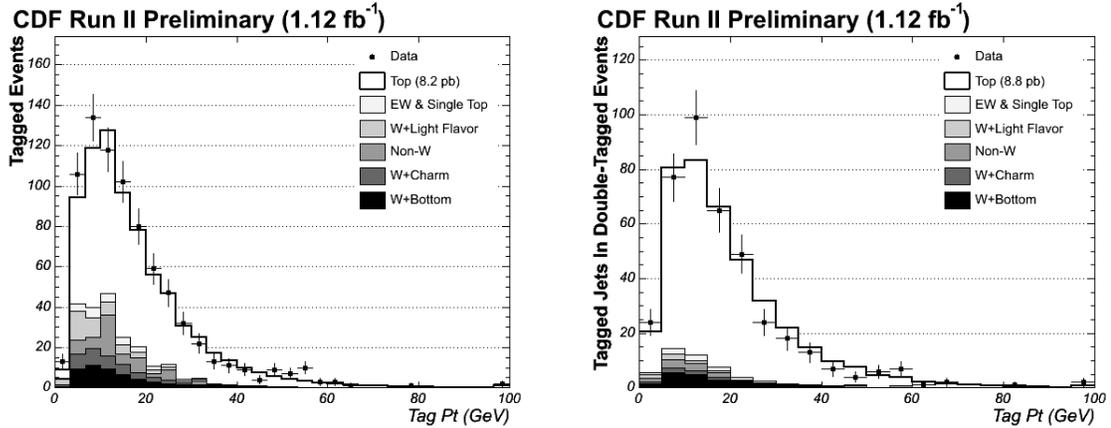


Figure C.12: Expected and observed distribution of tag p_T for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

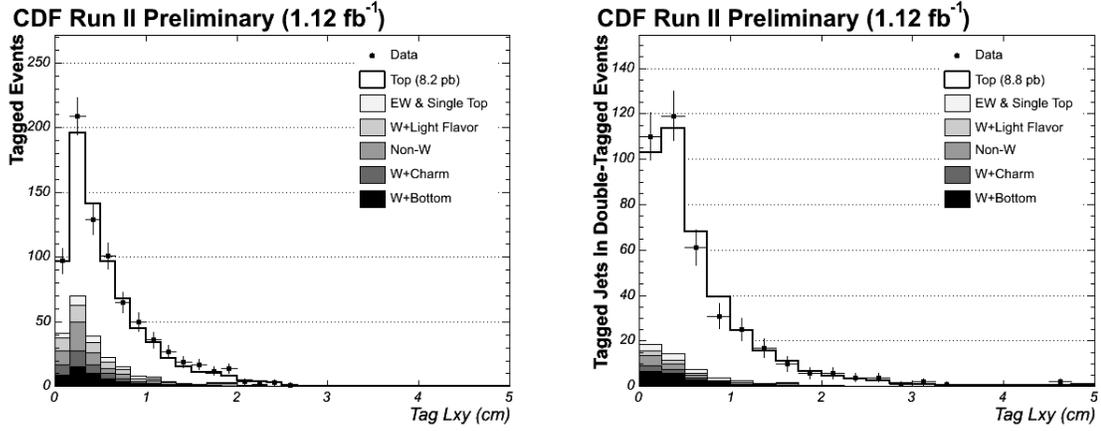


Figure C.13: Expected and observed distribution of tag L_{2d} for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

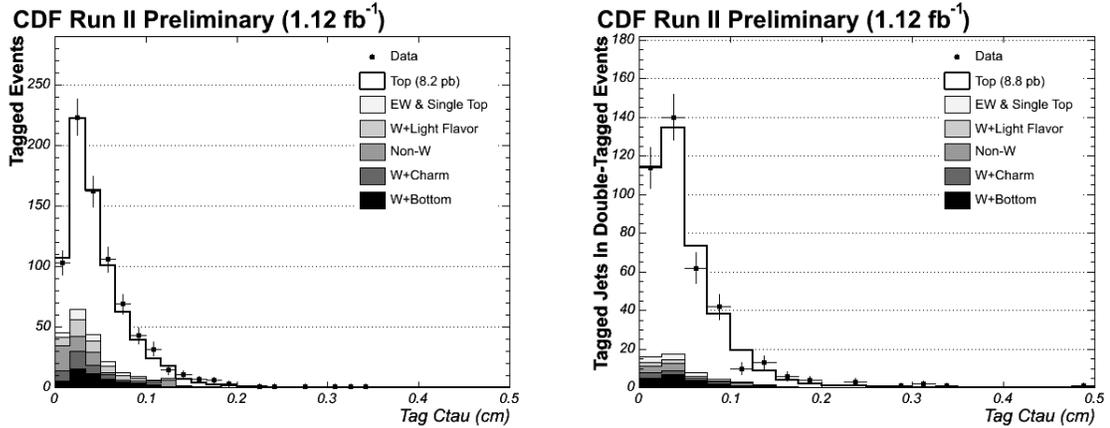


Figure C.14: Expected and observed distribution of tag $c\tau$ for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

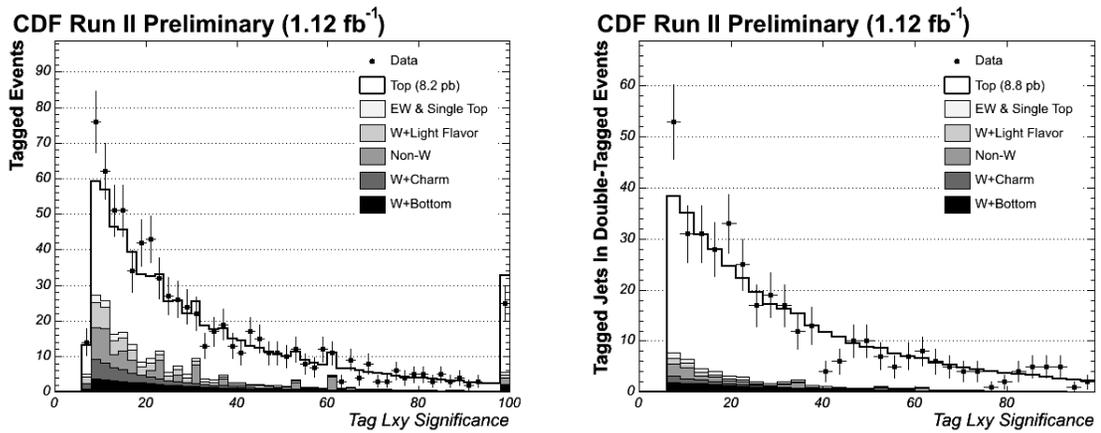


Figure C.15: Expected and observed distribution of tag L_{2d} significance for the inclusive-tight-tagged sample (left) and double-loose-tagged sample (right). The signal and background are normalized to the expectations for the measured cross sections of 8.2 pb and 8.8 pb on the left and right, and the tightened H_T and missing E_T requirements are released.

Bibliography

- [1] F. Abe *et al.* (CDF Collaboration), “Evidence for top quark production in anti-p p collisions at $s^{*}(1/2) = 1.8\text{-TeV}$.” *Phys.Rev.D*50:2966-3026,1994.
- [2] W.-M. Yao *et al.*, “Review of Particle Physics.” *J.Phys.G*33:1-1232,2006.
- [3] Kidonakis, Nikolaos and Vogt, Ramona, “Next-to-next-to-leading order soft-gluon corrections in top quark hadroproduction.” *Phys.Rev.D*68:114014,2003.
- [4] J. Pumplin *et al.*, “New generation of parton distributions with uncertainties from global QCD analysis.” *JHEP* 0207:012,2002.
- [5] C. Campagnari and M. Franklin, “The Discovery of the top quark.” *Rev.Mod.Phys.*69:137-212,1997.
- [6] M. Cacciari *et al.*, “The t anti-t cross-section at 1.8-TeV and 1.96-TeV: A Study of the systematics due to parton densities and scale dependence.” *JHEP* 0404:068,2004.
- [7] D. Chakraborty *et al.*, “Review of top quark physics.” *Ann.Rev.Nucl.Part.Sci.*53:301-351,2003.

-
- [8] R. Bonciani *et al.*, “NLL resummation of the heavy quark hadroproduction cross-section.” Nucl.Phys.B529:424-450,1998.
- [9] J. Campbell *et al.*, “Single top production and decay at next-to-leading order.” Phys.Rev.D70:094012,2004.
- [10] I. Bigi *et al.*, “Production and Decay Properties of Ultraheavy Quarks.” Phys.Lett.B181:157,1986.
- [11] A. Leike, “The Phenomenology of extra neutral gauge bosons.” Phys.Rept.317:143-250,1999.
- [12] H.-C. Cheng *et al.*, “Top partners in little Higgs theories with T-parity.” Phys.Rev.D74:055001,2006.
- [13] R. Eusebi *et al.*, “Updated search for charged Higgs in the tt decay products.” CDF Internal Note 7485.
- [14] J.S. Lee *et al.*, hep-ph/0307377,2003.
- [15] Tevatron Electroweak Working Group, <http://tevewwg.fnal.gov/top/> (2007).
- [16] CDF public top quark results are listed at: <http://www-cdf.fnal.gov/physics/new/top/top.html>
- [17] P. Fernandez *et al.*, “Combined Upper Limit on Standard Model Higgs Boson Production at CDF for 1 fb⁻¹ Data.” CDF Internal Note 8491.
- [18] <http://www.fnal.gov>

-
- [19] Descriptions of each accelerator component are linked from: <http://www-bdnew.fnal.gov/operations/rookie.books/rbooks.html>
- [20] Up-to-date luminosity plots can be found at: <http://www-cdfonline.fnal.gov/ops/opshelp/stores/>
- [21] All CDF II technical publications are linked from: http://www-cdf.fnal.gov/physics/techpub_run2.html
- [22] C.S. Hill, Nucl.Inst.Meth.A530:1,2004.
- [23] A. Sill, *et al.*, Nucl.Inst.Meth.A447:1,2000.
- [24] A. Affolder, *et al.*, Nucl.Inst.Meth.A453:84,2000.
- [25] T. Affolder *et al.*, Nucl.Inst.Meth.A526:249,2004.
- [26] L. Balka, *et al.*, Nucl.Inst.Meth.A267:272-279,1988.
- [27] S. Bertolucci, *et al.*, Nucl.Inst.Meth.A267:301-314,1988.
- [28] S.R. Hahn, *et al.*, Nucl.Inst.Meth.A267:351-366,1988.
- [29] M. Albrow *et al.*, Nucl.Inst.Meth.A480:524-545,2002.
- [30] G. Ascoli *et al.*, Nucl.Inst.Meth.A268:33,1988.
- [31] D. Acosta *et al.*, Nucl.Inst.Meth.A494:57,2002.
- [32] S. Klimenko *et al.*, FERMILAB-FN-0741,2003.
- [33] E.J. Thomson *et al.*, IEEE Trans. on Nucl. Sci. 49:1063,2002.

-
- [34] R. Downing *et al.*, Nucl.Inst.Meth.A570:36-50,2007.
- [35] W. Ashmanskas *et al.*, Nucl.Inst.Meth.A518:532,2004.
- [36] P. Gatti, “Performance of the new tracking system at CDF II.” CDF Internal Note 5561.
- [37] P. Azzi *et al.*, “Histogram Tracking in the COT.” CDF Internal Note 5562.
- [38] M. Bishai *et al.*, “COT and SVX II Tracking Performance.” CDF Internal Note 5931.
- [39] S. Nahn, “Status of CDF Silicon Tracking.” CDF Internal Note 6159.
- [40] S. Lai *et al.*, “An Evaluation of the Beam Width at CDF.” CDF Internal Note 6311.
- [41] <http://www-cdf.fnal.gov/upgrades/computing/projects/reconstruction/tracking/vertex/vertex>
- [42] K. Burkett and J. Guimarães da Costa, “Development of a CTVMFT-based primary vertex finder: PrimeVtxFinder.” CDF Internal Note 6430.
- [43] J. Marriner, “SECONDARY VERTEX FIT WITH MASS AND POINTING CONSTRAINTS (CTVMFT).” CDF Internal Note 1996.
- [44] D. Hare *et al.*, “Electron ID Efficiency and Scale Factors for Winter 2007 Analyses.” CDF Internal Note 8614.
- [45] E. Halkiadakis *et al.*, “A Conversion Removal Algorithm for the 2003 Winter Conferences.” CDF Internal Note 6250.

-
- [46] U. Grundler *et al.*, “High-Pt muons recommended cuts and efficiencies for Winter 2007.” CDF Internal Note 8618.
- [47] B. Flaughner and J. Mueller, “A Guide To JetClu: The CDF Jet Cluster Algorithm.” CDF Internal Note 1814.
- [48] A. Bhatti *et al.*, “Determination of the jet energy scale at the Collider Detector at Fermilab.” Nucl.Inst.Meth.A566:375-412,2006.
- [49] T. Sjostrand *et al.*, “PYTHIA 6.4 Physics and Manual.” JHEP 0605:026,2006.
- [50] M.L. Mangano *et al.*, “ALPGEN, a generator for hard multiparton processes in hadronic collisions.” JHEP 0307:001,2003.
- [51] D.J. Lange, “The EvtGen particle decay simulation package.” Nucl.Inst.Meth.A462:152-155,2001.
- [52] R. Brun *et al.*, “Geant: Simulation Program For Particle Physics Experiments. User Guide And Reference Manual.” CERN-DD-78-2-REV.
- [53] G. Grindhammer and S. Peters, “The Parameterized simulation of electromagnetic showers in homogeneous and sampling calorimeters.” hep-ex/0001020.
- [54] A Abulencia *et al.*, “t-tbar cross section with SLT in 760pb-1.” CDF Internal Note 8482.
- [55] C. Neu *et al.*, “SecVtx Optimization Studies for 5.3.3 Analyses.” CDF Internal Note 7578.

-
- [56] D. Glenzinski *et al.*, “A DETAILED STUDY OF THE SECVTX ALGORITHM.” CDF Internal Note 2925.
- [57] J. Guimarães da Costa *et al.*, “Measurement of the $t\bar{t}$ Production Cross Section using a Loose SecVtx Algorithm.” CDF Internal Note 6983.
- [58] M. Barone *et al.*, “A MEASUREMENT OF THE TOP PRODUCTION CROSS SECTION USING SECVTX TAGS.” CDF Internal Note 3516.
- [59] H. Bachacou *et al.*, “Efficiency of SecVtx B-Tag Used for 2004 Winter Conferences.” CDF Internal Note 6904.
- [60] A. Holloway *et al.*, “Alternate Estimate of the Winter 2004 SecVtx Tagging Scale Factor.” CDF Internal Note 6912.
- [61] S. Grinstein *et al.*, “SecVtx Mistag Asymmetry for Winter 2007.” CDF Internal Note 8626.
- [62] S. Grinstein *et al.*, “Electron-Method SecVtx Scale Factor for Winter 2007.” CDF Internal Note 8625.
- [63] F. Garberson *et al.*, “SECVTX b-Tag Efficiency Measurement Using Muon Transverse Momentum for 1.2/fb Analyses.” CDF Internal Note 8640.
- [64] F. Garberson *et al.*, “Combination of the SecVtx 1.2/fb b-Tagging Scale Factors.” CDF Internal Note 8666.
- [65] S. Budd *et al.*, “Tight, Loose and Ultratight SECVTX Tag Rate Matrix with 1.2fb-1.” CDF Internal Note 8519.

-
- [66] R. Barlow and C. Beeston, *Comp.Phys.Comm.*77:219-228,1993.
- [67] <http://root.cern.ch/>
- [68] S. Miller and D. Amidei, “Method 2 Secvtx Background Estimate and Measurement of the $t\bar{t}$ Cross Section.” CDF Internal Note 6907.
- [69] S. Grinstein *et al.*, “Heavy-Flavor Content of the W+Jets Sample.” CDF Internal Note 8765.
- [70] S. Grinstein *et al.*, “Calibration of Heavy-Flavor Production in QCD Data” CDF Internal Note 8768.
- [71] B.-Y. Han and V. Boisvert, “Trigger Efficiencies for the High Et Central Electron in the Gen6 data(run:217990 - 222426).” CDF Internal Note 8629.
- [72] B. Harris *et al.*, *Phys.Rev.D*66:054024,2002.
- [73] A. Abulencia *et al.* (CDF Collaboration), *Phys.Rev.D* 75:092004,2007.
- [74] T. Stelzer and W.F. Long, *Comput.Phys.Commun.* 81:337,1994.
- [75] B. Cooper *et al.*, “Estimation of the Background to $W \rightarrow e\nu + n$ Jet Events.” CDF Internal Note 7760.
- [76] D. Acosta *et al.* (CDF Collaboration), *Phys.Rev.D* 71:052003,2005.
- [77] A. Abulencia *et al.* (CDF Collaboration), “Measurement of the $t\bar{t}$ -bar Production Cross Section in p anti-p Collisions at $\sqrt{s} = 1.96$ TeV.” *Phys.Rev.Lett.* 97:082004,2006.

-
- [78] G. Marchesini and B.R. Webber, “Herwig: A New Monte Carlo Event Generator For Simulating Hadron Emission Reactions With Interfering Gluons.” Cavendish-HEP-87/9, Dec 1987.
- [79] W.J. Stirling *et al.*, “MRST parton distributions.” AIP Conf.Proc.747:16-21,2005.
- [80] G. Unal, “WBBAR AND WCCBAR BACKGROUNDS IN TOP SVX ANALYSIS.” CDF Internal Note 3389.