Measurement of the ratio of inclusive cross sections $\frac{\sigma(p\bar{p}\rightarrow Z+b-jet)}{\sigma(p\bar{p}\rightarrow Z+jet)}$ in the dilepton final states

Kenneth James Smith

A dissertation submitted to the

Faculty of the Graduate School of

the University at Buffalo, State University of New York

in partial fulfillment of the requirements for the

degree of Doctor of Philosophy

DEPARTMENT OF PHYSICS

Advisor: Avto Kharchilava

October 2010

To Amelia, for having faith in me.

Abstract

The inclusive production of b-jets with a Z boson is an important background to searches for the Higgs boson in associated $ZH \rightarrow llb\bar{b}$ production at the Fermilab Tevatron collider. This thesis describes the most precise measurement to date of the ratio of inclusive cross sections $\sigma(p\bar{p} \rightarrow Z + b\text{-jet})/\sigma(p\bar{p} \rightarrow Z\text{+jet})$ when a Z boson decays into two electrons or muons. The measurement uses a data sample from $p\bar{p}$ collisions at the center of mass energy $\sqrt{s} = 1.96$ TeV corresponding to an integrated luminosity of 4.2 fb^{-1} collected by the D0 detector. The measured ratio $\sigma(Z + b\text{-jet})/\sigma(Z\text{+jet})$ is $0.0187 \pm 0.0021(stat) \pm 0.0015(syst)$ for jets with transverse momentum $p_T > 20$ GeV and pseudorapidity $|\eta| \leq 2.5$. The measurement is compared with the next-to-leading order theoretical predictions from MCFM and is found to be consistent within uncertainties.

The work described in this dissertation
has been published in the following journals:

arXiv:1010.6203

FERMILAB-PUB-10-435-E

Submitted to Physical Review Letters

Acknowledgments

To my family, thank you for always being there. Dad, thanks for always giving me someone to look up to and for teaching me to never stop asking questions. Mom, thanks for teaching me to follow my heart and that the only obstacles between the man that I am and the man that I want to be are internal. April, thank you for always looking out for me, even when I thought that I didn't need it. Mandy, thank you for inspiring me. Jon, thanks for being Jon. To Joey, Brady, Brogan, Nicholas and Chris thanks for pushing me to be a better man. Grandma, thanks for Zelda and being awesome. To the rest of my family, thank you so much for being there, I am a truly lucky man and I love you all.

Professionally, I'd like to start by thanking my advisor, Avto Kharchilava. Without your help and wisdom, this wouldn't have been possible. I'd also like to thank my other two committee members, Ia Iashvili and Doreen Wackeroth. I would also like to thank the entire faculty and staff at SUNY Buffalo. Next, I'd like to thank Michael Strang. There is no way I would've made it through the first year and a half at the lab without you being at my back, thank you. Thank you to Thomas Hemmick, my first professor of physics, and the person who taught me to be passionate about it. To Ashish Kumar, thank you for being such a great research partner. Thank you to the entire DØ collaboration, specifically Joe Haley, Satish Desai and John BackusMayes for your friendship and guidance. I would also like to thank Dan Duggan, Dmitry Bandurin and Don Lincoln for fruitful conversations regarding the analysis.

To Andy, KB, Nick, and Caesar thank you for making me a better man and to all of my

other friends in Buffalo, New York and Chicago, thank you for your friendship and support.

Finally, to Amelia. You moved up to Chicago from Texas so that I didn't have to choose between the love of my life and the dream that I have had since I was a child. For that, I am eternally grateful. I love you.

Contents

Abstract **iv**

Acknowledgments **vi**

List of Tables **xiii**

List of Figures **xvi**

List of Abbreviations **xxiv**

1 Introduction **1**

1.1 Standard Model of Particle Physics 1

1.2 Z Boson in Association with b-quark Production 5

2 Apparatus **9**

2.1 The Fermilab Accelerator 9

2.2 DØ Detector 13

2.2.1 DØ Coordinate System 14

2.2.2 Central Tracking System 17

2.2.3 CFT 18

2.2.4 Solenoid 19

2.3 Preshower Detectors 20

2.3.1 Calorimeter 21

2.3.2 Muon System . . . 25
2.3.3 Luminosity Monitor . . . 29
2.4 Trigger . . . 31
2.4.1 Single Muon and Electron Triggers . . . 32

3 Event Reconstruction **34**
3.1 Track Reconstruction . . . 34
3.2 Primary Vertex Reconstruction . . . 36
3.3 Electron Reconstruction . . . 37
3.3.1 Electron Quality . . . 39
3.3.2 Electron Energy Correction . . . 42
3.4 Jet Reconstruction . . . 43
3.4.1 Jet Quality . . . 44
3.4.2 Jet Energy Scale . . . 45
3.4.3 Jet Energy Resolution . . . 46
3.5 Muon Reconstruction . . . 47
3.5.1 Muon Quality . . . 48
3.6 Missing Energy . . . 49
3.6.1 MET Reconstruction . . . 50

4 Data **51**
4.1 Trigger Requirements . . . 51

5 MC Samples **52**

6 Event Selection **54**
6.1 Primary Vertex . . . 54
6.2 Muons . . . 54
6.2.1 Reconstructed $Z \rightarrow \mu\mu$ boson . . . 55

6.3 Electrons . . . 56
6.3.1 Reconstructed $Z \rightarrow ee$ boson . . . 56
6.4 Jets . . . 56
6.5 Tagged Sample . . . 57
6.6 MET Cut . . . 57

7 Multijet Background **60**
7.1 Normalizing to Data . . . 61
7.1.1 Multijet Background Contributions for Different Topologies in the ee Channel . . . 62

8 B-tagging **65**
8.1 The Neural Network Tagger . . . 65
8.2 Secondary Vertex Mass . . . 67
8.3 Jet Lifetime Probability Tagger . . . 68
8.4 Negative Tagged Jets . . . 69
8.5 Determination of NN b-Tagging Efficiency and Data to MC Scale Factor . . . 70

9 Corrections to Monte Carlo **72**
9.1 Z+jets Cross Sections . . . 72
9.2 Cross Sections for background Processes . . . 73
9.3 Luminosity Reweighting . . . 73
9.4 Primary Vertex Reweighting . . . 74
9.5 Trigger Corrections . . . 74
9.6 Lepton Energy and Identification Corrections . . . 75
9.7 Z p_T Reweighting . . . 75
9.8 Jet Shifting Smearing and Removal (JSSR) . . . 76
9.9 Vertex Confirmation for Data and MC . . . 77
9.10 Scale Factors for b-tagging . . . 78

10 Z+jet Selection **82**

10.1 Dimuon Channel 82

10.2 Dielectron Channel 87

11 Measurement **91**

11.1 Identifying Heavy Flavor Jets 91

11.2 Efficiencies 92

11.3 Jet Taggability Efficiency 93

11.4 MJL Tagger 94

11.4.1 MJL η and p_T Dependence 96

11.4.2 MJL Efficiency 97

11.5 Increased Template Statistics 97

11.5.1 Negative Tagged Jets 97

11.5.2 Increased Statistics for c-jet Template 99

11.5.3 Uncertainty Due to Template Statistics 101

11.6 Background Subtraction 101

11.7 Jet Reconstruction Efficiency, b-jet vs. light-jet 104

11.8 Determination of Jet Flavor Fractions 105

11.8.1 Additional Cross Check 107

11.9 Final Result 109

11.9.1 Comparison to Previous Version of Analysis 114

11.9.2 Cross Check with Alternative NN Operating Point 116

12 Systematic Uncertainties **118**

13 MCFM **124**

13.1 Comparison with Theoretical Predictions 124

14 Conclusions **129**

A Maximum Likelihood Fitter **130**

B Medium Operating Point **134**

B.1 MJL Dependence on NN Scale Factors 138

C Study of p_T Dependence of MJL Templates for b-jets **140**

D Jet p_T distributions Across Channels **143**

D.1 Distributions from Monte Carlo 143

D.2 Distributions from Data 145

E Negative Tagged Jets **147**

F Single Muon Trigger List **151**

G Single Electron Trigger List **153**

H MC Samples **156**

Bibliography **159**

Bibliography **159**

List of Tables

1.1 Table of the ratio measurements and theoretical expectations for different jet kinematics. 7

2.1 Properties of CC layers. 23

2.2 Properties of EC layers. 24

7.1 Results of the multijet and background normalization in the dimuon sample. 61

9.1 The reconstructed jet matching efficiency with particle jets in $Z + b$ jets and Z+light jets as a function of the number of primary vertices in the event. . . 78

10.1 The data, backgrounds, and expected Z+jets events are listed in the preselected event sample. 86

11.1 Efficiencies used in ratio calculation. 92

11.2 Jet flavor fractions in the tagged Z+jets sample. 108

11.3 Results for alternative fitting option in $\mu\mu$ and ee channels. 109

11.4 Jet flavor fractions $\mu\mu + ee$ combined channel. 110

11.5 $\frac{\sigma(z+b)}{\sigma(z+jet)}$ ratio in both $\mu\mu$ and ee channels along with uncertainty due to template fitting. 114

11.6 $\frac{\sigma(z+b)}{\sigma(z+jet)}$ ratio in both $\mu\mu$ and ee channels combined along with uncertainty due to template fitting.. 114

11.7 $\sigma(z+b)/\sigma(z+jet)$ Ratio for $|\eta| < 2.5$ region and NN > 0.5 and NN > 0.65 operating points. 116

12.1 Systematic uncertainties and their contribution to the ratio 121

13.1 The cross section (fb) evaluated by MCFM for Z+heavy flavor and light jet production at the Tevatron. A jet lies in the range p_T >20 GeV and $|\eta|$ <2.5 and cone size of ΔR =0.5 has been used. The cross sections have been obtained for the scale $Q^2 = m_Z^2$ and using MSTW2008 PDF. The labels in the columns have the following meaning :$ZQ \equiv$ exactly one jet, which contains a heavy quark;$Z(Q\bar{Q}) \equiv$ exactly one jet which contains two heavy quarks;$ZQj \equiv$ exactly two jets, one of which contains a heavy quark; $ZQ\bar{Q} \equiv$exactly two jets, both of which contain a heavy quark. For last set of processes, the labels mean: $Zj \equiv$ exactly one jet, which does not contain a heavy quark; $Zjj \equiv$ exactly two jets, none of which contain a heavy quark. The uncertainties on the ratio are from the variation of renormalization scale, factorization scale and the PDFs. 125

13.2 MCFM calculation for $|\eta| < 2.5$ 126

B.1 Fractions from maximum likelihood fitter for ee and $\mu\mu$ and combined channels for $|\eta| < 2.5$ region and NN > 0.65 operating point. 134

B.2 $\sigma(z+b)/\sigma(z+jet)$ ratio for $|\eta| < 2.5$ for region and NN > 0.65 operating point. 135

E.1 Fractions from maximum likelihood fitter for NT b-, NT c- and NT light-jets in NT jets from data. The NT b-, NT c- and NT light-jet templates were taken from MC. 147

F.1 Single muon triggers used in analysis. 152

G.1 Single electron triggers used in analysis. 154

G.2 Single electron triggers used in analysis, continued. 155

H.1 $Z \to \mu\mu$ MC Samples 157

H.2 $Z \to ee$ MC Samples 158

List of Figures

1.1 The fundamental particles in matter. 2

1.2 Simplified schematic of gluon splitting. 3

1.3 The Higgs boson branching ratios for different channels w.r.t. mass 6

1.4 qq Feynman diagrams. 6

1.5 gQ Feynman diagrams 6

2.1 Schematic of Fermilab accelerator chain. 10

2.2 The Cockroft-Walton pre-accelerator. 11

2.3 The Lithium Lens. 12

2.4 Diagram of DØ detector. 15

2.5 Diagram of DØ coordinates. 16

2.6 Schematic of the central tracker. 17

2.7 Schematic of the SMT. 18

2.8 Schematic of the CFT. 19

2.9 Cross section diagram of a scintillating fiber used in the preshower detectors. 20

2.10 Illustration of the DØ calorimeter. 21

2.11 Diagram showing calorimeter towers in η. 22

2.12 Unit cell in the calorimeter. 24

2.13 Muon System drift tubes. 26

2.14 Cross section of PDTs used in WAMUS. 26

2.15 Cross section of MDTs used in FAMUS (dimensions in millimeters). 27
2.16 Illustration showing the orientation of MDTs in each layer of the FAMUS. 28
2.17 Muon System scintillators. 28
2.18 Photograph of the scintillating tiles in the C layer of the FAMUS. 29
2.19 Schematic of luminosity monitor in y-z plane. 30
2.20 Schematic of luminosity monitor scintillators. The red spots represent the PMTs. 30

3.1 Illustration showing the clusters of hits left in each layer of the CFT from a charged particle. 35
3.2 Example of track reconstruction with Hough transformation. The top left shows family of tracks that contain a given hit in x-y space. The top right shows these tracks in ρ-ϕ_0 space. The bottom left shows the intersection for each hit on a track, which indicates the parameter value for that track. The bottom right, the lines are used to fill a histogram and the intersection is identified as a maximum. 36
3.3 Schematic of f_{iso} for EM cluster. 39
3.4 Illustration of the MPF method. 46
3.5 JES correction as a function of η. 47

6.1 MET plot for data and expected physics contributions. 58
6.2 MET plot for b and t-jet events. 58

7.1 0 jet data and αQCD + βMC distributions 62
7.2 1 jet data and αQCD + βMC distributions 62
7.3 2 jet data and αQCD + βMC distributions 63
7.4 The dielectron invariant mass distributions in the complete Z+jets sample (a) and for the CC-CC (b) topology. 64

7.5 The dielectron invariant mass distributions in the complete Z+jets sample for CC-EC(a) and for the EC-EC (b) topologies. 64

8.1 Distributions of b-jet Neural Net outputs for the b-, c- and light-jets. . . . 66
8.2 M_{svt} distributions for b-, c- and light-jets. 67
8.3 Determination of sign of IP. 68
8.4 JLIP distributions for b-, c- and light-jets. 69
8.5 Negative decay length. 70

9.1 (a) The instantaneous luminosity distributions of data and MC samples for Z inclusive sample is shown in the left plot. The MC has been reweighted so that the minimum bias overlay in the generated samples mirrors the instantaneous luminosity of the data sample. (b) The primary vertex Z distribution for data and MC samples. 74
9.2 Reconstruction of Z mass peak in (a) muon and (b) electron channels. . . 75
9.3 The transverse momentum of the dimuon Z candidates for (a) 0 jet inclusive bin and (b) 0 jet exclusive bin. 79
9.4 The transverse momentum of dimuon Z candidates for the (a) 1 jet bin and (b) 2 jet. 79
9.5 The log scale transverse momentum of the dimuon Z candidates for (a) 0 jet inclusive bin and (b) 0 jet exclusive bin. 80
9.6 The log scale transverse momentum of dimuon Z candidates for the (a) 1 jet bin and (b) 2 jet bin. 80
9.7 Distribution of ΔS for CC jets in data and 18 GeV $< p_T^{\gamma} <$ 23 GeV, fit to a Gaussian times an error function. 80
9.8 Top: b and light z vertex acceptance with respect to p_T. Bottom: ratio of b and light acceptances 81

10.1 The leading and second electron p_T spectrum in $Z+ \geq 1$ jets sample before any b-tagging is applied. 83
10.2 The leading and second electron η spectrum in $Z+ \geq 1$ jets sample before any b-tagging is applied. 83
10.3 The rapidity and mass of the Z candidate in the $Z+ \geq 1$ jets sample before any b-tagging is applied. 84
10.4 The p_T Z candidate and missing transverse energy in the $Z+ \geq 1$ jets sample before any b-tagging is applied. 84
10.5 The $\Delta R(\mu\mu)$ for the Z candidate and jet multiplicity in the $Z+ \geq 1$ jets sample before any b-tagging is applied. 85
10.6 The leading jet (a) p_T and (b) η for data and background in the $Z+ \geq 1$ jets sample before any b-tagging is applied. 85
10.7 The second leading jet (a) p_T and (b) η for data and background in the $Z+ \geq 1$ jets sample before any b-tagging is applied. 86
10.8 The leading and second electron p_T spectrum in $Z+ \geq 1$ jets sample before any b-tagging is applied. 87
10.9 The leading and second electron η spectrum in $Z+ \geq 1$ jets sample before any b-tagging is applied. 87
10.10 The rapidity and mass of the Z candidate in the $Z+ \geq 1$ jets sample before any b-tagging is applied. 88
10.11 The p_T Z candidate and missing transverse energy in the $Z+ \geq 1$ jets sample before any b-tagging is applied. 88
10.12 The $\Delta R(ee)$ for the Z candidate and jet multiplicity in the $Z+ \geq 1$ jets sample before any b-tagging is applied. 89
10.13 The leading jet (a) p_T and (b) η for data and background in the $Z+ \geq 1$ jets sample before any b-tagging is applied. 89

10.14 The leading jet (a) p_T and (b) η for data and background in the $Z+ \geq 2$ jets sample before any b-tagging is applied. 90

11.1 The spectrum of b-tagging efficiency as a function of jet transverse momentum measured in $Z + b$ MC for the NN operating point: oldLOOSE (NN>0.5) 92

11.2 b-jet taggability efficiency as a function of jet p_T 93

11.3 Templates for equation 11.1. $i =$ odd on the left and even on the right both increasing on the way down. 95

11.4 MJL template for b-jets, light-jets and c-jets derived from $Z+$ jets MC samples and negative tagged jets from data for NN > 0.5 (a) linear and (b) log scale 96

11.5 The MJL templates in different η regions. 98

11.6 The MJL templates in different p_T regions. 98

11.7 b-jet MJL efficiency with respect to p_T 99

11.8 MJL templates for negative tagged jets in data and light-jets in MC 100

11.9 $Z + c$-jet template compared to Pythia c-jet template. 100

11.10 a) The leading jet p_T for both the $Z + c$-jet and Pythia c-jet samples. b) The ratio of the two samples 102

11.11 a) MJL templates for Pythia c-jet, reweighted Pythia c-jet and $Z + c$-jet samples b) p_T spectra after reweight 102

11.12 $Z+c$ template along with even and odd templates. The even (odd) template adds the $Z + c$-jet bin value to the bin uncertainty for the even (odd) bins and subtracts the bin uncertainty from the $Z+c$-jet bin value for odd (even) bins. 103

11.13 c-jet template along with even and odd templates. The even (odd) template adds the c-jet bin value to the bin uncertainty for the even (odd) bins and subtracts the bin uncertainty from the c-jet bin value for odd (even) bins. 103

11.14 MJL templates for different backgrounds 104
11.15 Ratio of b-jet reconstruction efficiency to light-jet reconstruction efficiency vs. p_T 105
11.16 Dimuon and dielectron data after background subtraction, normalized to unity. 107
11.17 Both likelihood and $Z + b$ fraction results for different c/light values. . . . 110
11.18 Distribution of observed events for MJL discriminant with NN > 0.5 and $|\eta| < 2.5$ cuts. The distributions for the b-, c- and light-jet templates are shown normalized to their fitted fraction. Error on the templates represent combined uncertainties from statistics of MC and the fitted jet flavor fractions. 111
11.19 Distribution of observed events for MJL discriminant with NN > 0.5 and $|\eta| < 1.1$ cuts. 111
11.20 The p_T distribution of b-tagged jets with NN > 0.5 and $|\eta| < 2.5$ cuts in each channel. 112
11.21 The p_T distribution of b-tagged jets with NN > 0.5 and $|\eta| < 1.1$ cuts in each channel. 112
11.22 Distribution of observed events in combined sample for MJL discriminant with NN > 0.5 in (a) $|\eta| < 2.5$ and (b) $|\eta| < 1.1$ regions. 113
11.23 The combined p_T distributions of b-tagged jets with NN > 0.5 in (a) $|\eta| <$ 2.5 cuts and (b) $|\eta| < 1.1$ regions. 113
11.24 rJLIP distribution for light, b-, c- and NT jets. 115
11.25 The combined results for (a) rJLIP distribution of tagged jets and (b) jet p_T of tagged jets measured with the rJLIP templates for jets with $|\eta_{det}| < 1.1$. 116
11.26 Physics η acceptance when a $\eta_{det} < 1.1$ cut is applied 117

12.1 MJL template of c- and b-jets as well as collapsed c- and b-quark jets. . . . 121

12.2 The distributions of $\frac{p_{T\_reco} - p_{T\_particle}}{p_{T\_particle}}$ for both light jets and b-jets before b-jet JES correction in the full p_T range (left) and in the low p_T range (right) of 20-30 GeV. 122

12.3 The effect of additional JES correction on the MJL distributions for b-jets. 123

12.4 The MJL distributions for two different b-fragmentation schemes. 123

13.1 Ratio of b/light spectra for ALPGEN MC partons and ALPGEN MC partons matched to reconstructed jets. 127

13.2 Folded and unfolded MCFM light and b parton p_T 128

A.1 Simple example of distribution and its two components, all normalized to unity. 131

B.1 MJL template for NT data, MC b-, c- and light-jets for NN $> .65$ operating point 135

B.2 MJL distribution from maximum likelihood fitter for $|\eta| < 2.5$ and NN > 0.65 region 136

B.3 Jet p_T distribution from maximum likelihood fitter for $|\eta| < 2.5$ and NN > 0.65 region 136

B.4 MJL distributions for ee and $\mu\mu$ channels in the $|\eta| < 2.5$ region for NN > 0.65 cut. 137

B.5 MJL distributions for ee and $\mu\mu$ channels in the $|\eta| < 2.5$ region for NN > 0.65 cut. 137

B.6 MJL template for b-jets with NN > 0.5 scale factor and exclusive scale factor ("reweighed") applied. 139

C.1 a) MJL and b) p_T distributions for b-jets coming from $Z + b$, diboson and $t\bar{t}$ samples. The jets pass the $NN > 0.5$ requirement. 140

C.2 MJL template for b-jets with (a) $p_T > 20$ GeV and (b) $p_T < 35$ GeV . . . 142

C.3 MJL template for b-jets with (a) $p_T > 20$ GeV and (b) $p_T < 50$ GeV . . . 142

D.1 The b-jet p_T distributions in the dielectron and dimuon channels before (left) and after tagging (right). For the tagging, $NN > 0.775$ requirement has been used. 143

D.2 The c-jet p_T distributions in the dielectron and dimuon channels before (left) and after tagging (right). For the tagging, $NN > 0.775$ requirement has been used. 144

D.3 The light-jet p_T distributions in the dielectron and dimuon channels before (left) and after tagging (right). For the tagging, $NN > 0.775$ requirement has been used. 144

D.4 The Zp_T distributions in the dielectron and dimuon channels in data before tagging. 145

D.5 The leading jet p_T distributions in the dielectron and dimuon channels in data before (left) and after tagging (right). For the tagging, $NN > 0.5$ requirement has been used. 146

D.6 The leading jet η distributions in the dielectron and dimuon channels in data before (left) and after tagging (right). For the tagging, $NN > 0.5$ requirement has been used. 146

E.1 The MJL distributions for b-, c- and light-jets obtained from Monte Carlo negative tagged jets along with negative tagged jets in data. 148

E.2 The MJL distribution in NT data along with contributions from NT b-, c- and light-jets. 149

E.3 The MJL distributions for NT MC light-jets and tagged MC light-jets. . . 150

E.4 The MJL distributions for NT MC multijet sample and NT jets in data. . 150

List of Abbreviations

E_T	Transverse Energy
M_{svt}	Invariant mass of secondary vertex
CC	Central Calorimeter
CFT	Central Fiber Tracker
CH	Course Hadronic
CPS	Central Preshower detector
dca	distance of closest approach
EC	End Calorimeter
EM	Electromagnetic
FAMUS	Forward Angle Muon System
FH	Fine Hadronic
FPS	Forward Preshower detector
JER	Jet Energy Resolution
JES	Jet Energy Scale
JLIP	Jet Lifetime Probability Tagger
MDT	Mini Drift Tube
MET	Missing Transverse Energy
MJL	M_{svt} + JLIP combined discriminant
MPF	Missing E_T Projection Fraction
PDF	Parton Distribution Function

PDT Proportional Drift Tube

PV Primary Vertex

PV Primary Vertex

SMT Silicon Microstrip Tracker

WAMUS Wide Angle Muon System

Chapter 1

Introduction

"We never cease to stand like curious children before the great mystery into which we were born." This quote from Albert Einstein sums up the essence of particle physics. Particle physics is merely the attempt to answer questions that every child has wondered at some point in their life: "Where did we come from?", "What is everything made of?" and "How did we get here?". These are all questions that particle physicists set out to answer. As the search for answers has progressed, an elegant, beautiful and mystifying universe has been uncovered.

1.1 Standard Model of Particle Physics

To help describe the universe, a theoretical framework called the Standard Model of Particle Physics (SM) has been created. The SM combines quantum mechanics and special relativity to describe the interactions of fundamental particles. The SM includes three of the four known natural forces[1]: the nuclear strong force, which is responsible for holding a nucleus together; the nuclear weak force, which powers the sun; and the electromagnetic force, which rules electricity. These forces are mediated via particles called gauge bosons, with different ones belonging to each force. The weak force has two bosons, called the W and Z bosons.

[1] Gravity is the only natural force not included in the SM.

Z has no electromagnetic charge, and W can carry an elementary charge of either +1 or −1. The boson that mediates the electromagnetic force is the photon (γ), which has an electromagnetic charge of 0. The boson that carries the strong force is called the gluon (g), which has no electromagnetic charge, but does carry color charge. Any particles with a color charge can feel the nuclear strong force, any particles with an electromagnetic charge can feel the electromagnetic force, and any particles with a non-zero weak charge (also known as "isospin") can feel the nuclear weak force.

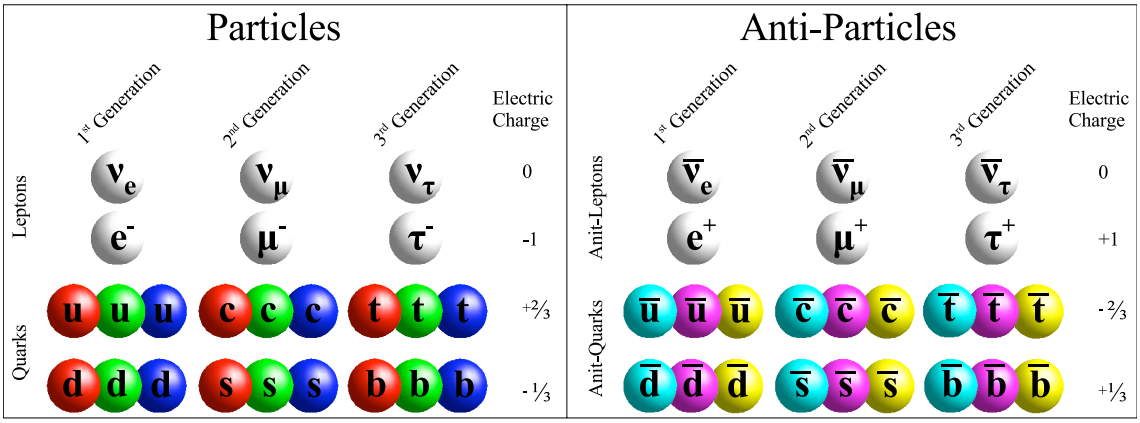


Figure 1.1: The fundamental particles in matter.

Figure 1.1 shows the fundamental particles called fermions that the forces act upon. As the figure shows, there are three generations for each particle type. The only difference between particles in different generations is mass. For example, an electron has all of the same qualities as a muon except for the fact that it is lighter. Most matter in nature is made up of particles from the first generation. This is because the more massive partners decay into the lighter ones. Figure 1.1 also shows three of each type of quark to represent the three colors a quark can have – red, green or blue. Only a particle with color can feel the strong force, and quarks and gluons are the only known particles to carry color. The strong force acts differently than the electromagnetic and nuclear weak force. Both the electromagnetic and nuclear weak forces weaken as the distance between two particles increases. In contrast, as the distance between two quarks increases, the force between them gets stronger and stronger, like a rubber band when it gets stretched. Also like a rubber band, there

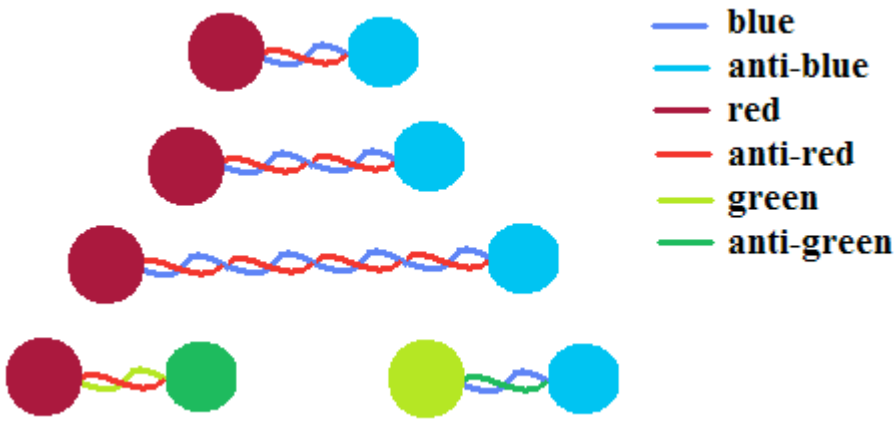


Figure 1.2: Simplified schematic of gluon splitting.

reaches a point where the gluon is stretched too far and it snaps. In the "snapping" process, two new quarks are created, making two pair of quarks instead of just one. Figure 1.2 shows a simplified[2] schematic of gluon splitting. There are two important features to notice about gluon splitting. The first is that the bound quark states must be colorless (also called white). In other words, the sum of the colors must equal zero; where $red + anti - red = 0$ but $red + anti - blue \neq 0$. The second important feature is that gluons carry one color and one anti-color. A result of this splitting feature is that there are no free (or single) quarks. Trying to free a quark is analogous to trying to cut a string until it has one end.

There are two types of fermions – quarks and leptons. Six different types of quarks exist (up (u), down (d), charm (c), strange (s), top (t) and bottom (b)), and each has their own antipartner (denoted by a "bar" over the name, *e.g.* the antipartner of the up quark is $\bar{u}$). Particles that are made up of quarks are called hadrons. Mesons are a type of hadron consisting of a quark-antiquark pair. A pion (π) is an example of a meson – π^0 is either a $u\bar{u}$ or $d\bar{d}$ pair, π^+ is $u\bar{d}$ and π^- is $\bar{u}d$. Baryons are hadrons that consist of three quarks. Both a neutron (udd) and a proton (uud) are examples of baryons. Leptons are the other

[2]It is simplified because gluons carry color, and as a consequence, instead of one gluon connecting two quarks, a "string" of gluons connects two quarks.

type of fundamental fermion. There are six leptons (and their respective antipartners), three that carry electromagnetic charge and three that do not. The three that carry charge are the electron, the muon and the tau particle. The three without electromagnetic charge are called the electron neutrino, the muon neutrino and tau neutrino. Neutrinos only interact via the weak force, and for that reason, they are very difficult to detect.

The final piece to the standard model is the yet-to-be discovered Higgs boson. Interaction with the Higgs boson is believed to be what gives matter mass. The Higgs boson is the only piece of the SM left to be verified experimentally. For this reason, it is the driving force behind many analyses in experimental high energy physics.

Making a measurement in high energy physics is often about exploiting the different behaviors in particles. A hadron containing a b-quark takes longer to decay than a hadron containing light-quarks, making the b-quark an important particle for detection. Because they are incredibly short lived, a Z boson cannot be directly detected, however its signature can be, and some of those are easier to detect than others. A Z boson can decay leptonically (Z decays into two muons or two electrons) or it can decay hadronically (Z decays into two quarks). The leptonic decay channels are advantageous because they are cleaner[3] than the hadronic alternative. This analysis measures the production rate of a Z boson in conjunction with a b-quark, where the Z boson decays leptonically. An important motivation for this analysis is its role in the search for the Higgs boson. The signature for this analysis mimics a signature for one of the Higgs production channels, thus making it an important background. Searching for the Higgs boson is like looking for a needle in a haystack, and this analysis is an attempt to better understand the hay. The motivation for this measurement is discussed in greater detail in the next section.

[3]Cleaner meaning that leptons do not have the splitting feature that accompanies hadron decays and the strong force.

1.2 Z Boson in Association with b-quark Production

The measurement of the production cross section of Z bosons in association with one or more b-jets provides an important test of quantum chromodynamics (QCD) calculations [1]. The understanding of this process and its description by current theoretical calculations is important because it provides a major background to various processes, *e.g.* to the search for the SM Higgs boson in the $ZH \to Zb\bar{b}$ decay mode and to the searches for supersymmetric partners of b-quarks. Figure 1.3 shows the branching ratio for the Higgs boson with respect to its mass [2]. For a lower mass Higgs ($m_H < 135$ GeV[4]), the channel with the highest branching ratio is $H \to b\bar{b}$. For this reason, $Z + b$ is a very important background to Higgs in association with a Z boson production. The production of a Z boson in association with one or more b-jets is also sensitive to the b-quark density in the proton. This density, which there is no direct measurement of, is usually derived perturbatively from the gluon distribution function [3]. A precise knowledge of the b-quark density is necessary to accurately predict processes that strongly depend on it. These processes include the electroweak production of single top quark [4] and the production of non-SM Higgs bosons [5, 6] in association with b-quarks.

The typical parton-level subprocesses expected to contribute to the $Z + b$-jet final states are $gg \to bg \to Zb$, shown in figure 1.5 and $q\bar{q} \to Zb\bar{b}$, shown in figure 1.4. In the first process a b-quark from the proton (evaluated from the gluon distribution function (PDF), which describes the momentum of partons) undergoes a hard scattering and a $\bar{b}$-quark typically remains soft, forward and therefore undetected. In the second process, only one b-jet may be reconstructed. This is because the $b\bar{b}$ quark pair can be produced close to each other and be reconstructed within the same jet, or one of the b-quarks does not get accepted by the detector or the analysis selection criteria. Theoretical calculations for $Z + b$ production are currently available at next-to-leading order (NLO) in QCD [1, 7]. According to QCD

[4]For this analysis, the $c = 1$ standard is used. Thus, the unit GeV is used for energy, mass and momentum.

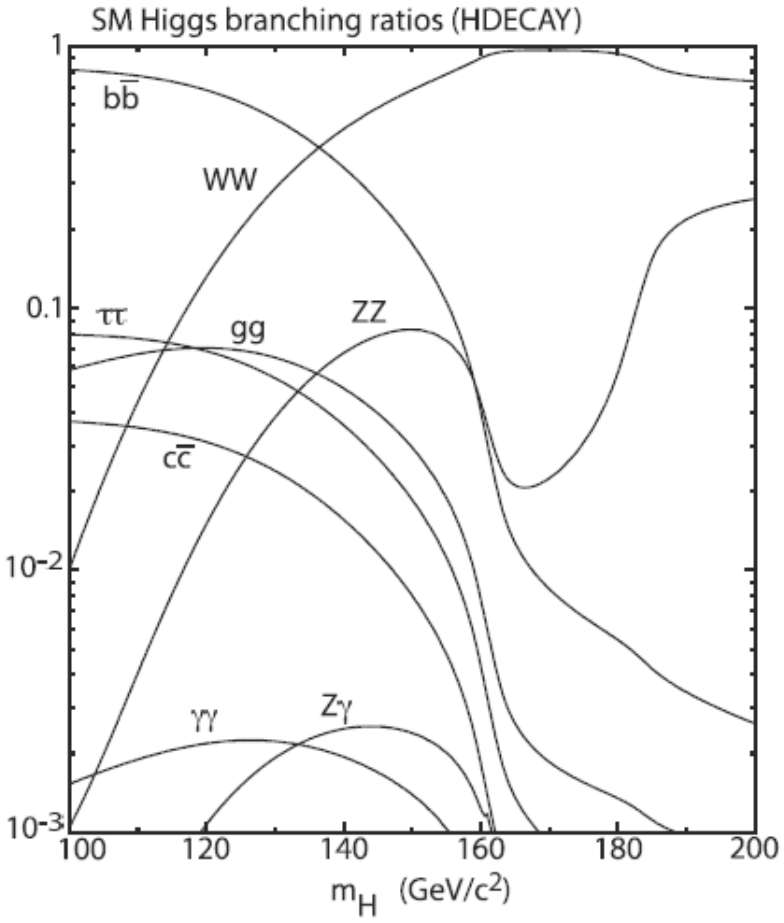


Figure 1.3: The Higgs boson branching ratios for different channels w.r.t. mass

Figure 1.4: qq Feynman diagrams.

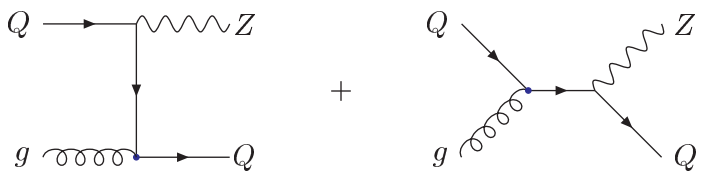


Figure 1.5: gQ Feynman diagrams

calculations, both subprocesses are expected to contribute almost equally to the $Z + b$-jet production at Tevatron, and the production of Z bosons with b-jets is dominated by the single b-jet production.

The cross section of the Z+jets production has been measured at the Tevatron and found to agree well with QCD calculations [11]. The DØ experiment provided the first measurement of the ratio of cross sections for $Z + b$ jet to Z+jets production to be 0.021 ± 0.005 [8] consistent within uncertainties with the theoretical prediction of 0.018 ± 0.004 [1]. This result was based on the 180 pb^{-1} of data and the analysis assumed the ratio of the $Z + b$ jet to $Z + c$ jets cross section from NLO calculations. The CDF experiment has measured the same ratio, as well as the $Z + b$-jet inclusive cross section using 2 fb^{-1} of data [9]. The CDF analysis does not use the theoretical prediction for N_c/N_b, and in fact extracts the fractions of different jet flavors from data. The ratio of $Z + b$-jet to inclusive Z+ jets production measured by CDF is $0.0208 \pm 0.0033 \pm 0.0034$. Table 1.1 shows the various measurements and theoretical expectations for different jet kinematics.

Jet Kinematics	$\sigma(Z + b)/\sigma(Z + j)$ (%)	Reference
$p_T > 20$ GeV,$\|\eta\| \leq 2.5$	2.1±0.4±0.3	DØ:0.18 fb^{-1} (PRL 94, 161801, 2005)
$p_T > 20$ GeV,$\|\eta\| \leq 1.5$	2.36±0.74±0.53	CDF:0.33 fb^{-1} (PRD 74, 032008, 2006)
$p_T > 20$ GeV,$\|\eta\| \leq 1.5$	2.08±0.33±0.34	CDF:2.0 fb^{-1} (PRD 79, 052008, 2009)
$p_T > 20$ GeV,$\|\eta\| \leq 1.1$	1.76±0.24±0.23	DØ:4.2 fb^{-1} (Prelim Result, DØ Note 6053)
Theoretical Calculations		
Jet Kinematics	$\sigma(Z + b)/\sigma(Z + j)$ (%)	Reference
$p_T > 20$ GeV,$\|\eta\| \leq 1.5$	1.88±0.23	MCFM (DØ Note 6052)
$p_T > 20$ GeV,$\|\eta\| \leq 1.1$	1.84±0.22	MCFM (DØ Note 6052)
$p_T > 20$ GeV,$\|\eta\| \leq 2.5$	1.85±0.22	MCFM (Current Analysis)

Table 1.1: Table of the ratio measurements and theoretical expectations for different jet kinematics.

This analysis set out to measure the $\sigma(Z+b)/\sigma(Z+j)$ ratio with a larger dataset and to a higher precision than was done before. The kinematic selections were intentionally aligned to that of the latest $ZH \rightarrow llbb$ analysis at D0 [10]. This alignment allows for a direct one-to-one comparison between the two analyses.

Chapter 2

Apparatus

Fermi National Accelerator lab is a high-energy physics laboratory located in Batavia, IL. It is home to what was, until only recently, the world's most energetic particle accelerator, the Tevatron [12]. The Tevatron collides bunches of protons and antiprotons at a center of mass energy, $\sqrt{s}$, of 1.96 TeV. The Tevatron is an accelerating ring of 3.9 miles in circumference. Located at two points on this ring are particle detectors, DØ and CDF. All of the data used for this analysis was taken by DØ.

2.1 The Fermilab Accelerator

In order to get protons and antiprotons to the desired energy, Fermilab employs a series of accelerators. Figure 2.1 shows the different components of the accelerator.

A proton begins its path through the accelerator as part of a hydrogen molecule (H_2). These hydrogen molecules are then converted to two H^- ions by a magnetron [17]. After the magnetron, the H^- ions are sent to a Cockcroft-Walton pre-accelerator. Figure 2.2 shows a picture of the pre-accelerator at Fermilab as well as a schematic of the voltage multiplier. This device uses its legs to charge the dome to a voltage of 750 kV. The hydrogen ions are then sent into the dome and are accelerated to an energy of 750 keV. These ions are then sent through a linear accelerator (linac [18]). The linac is about 150 m long and uses

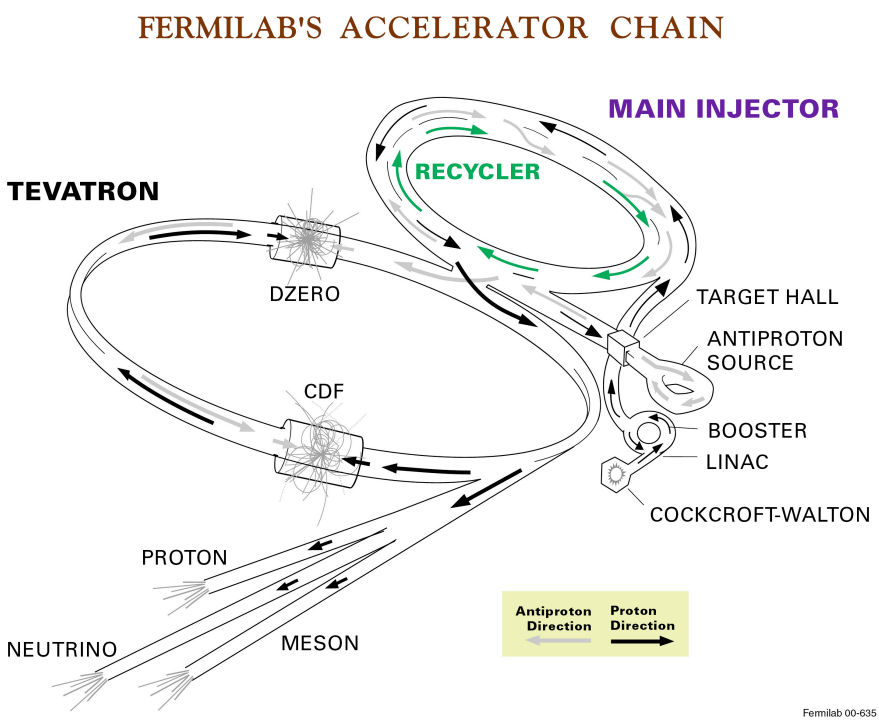
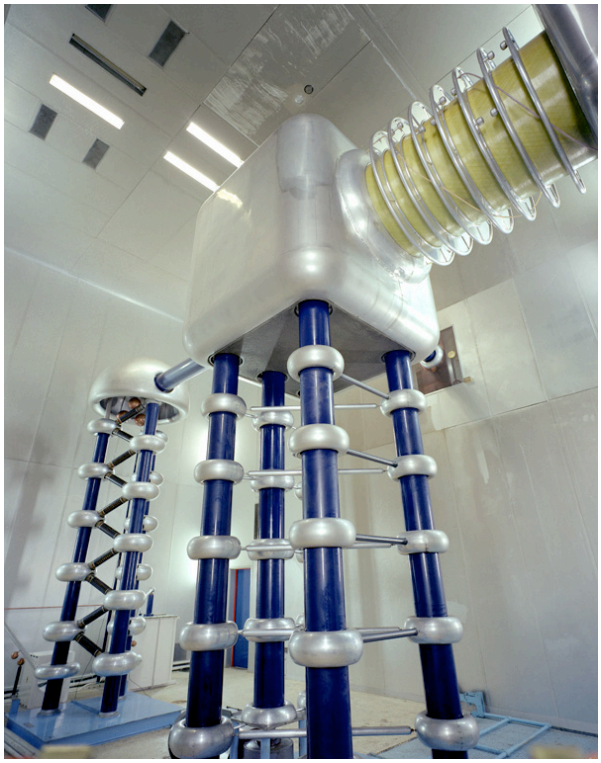


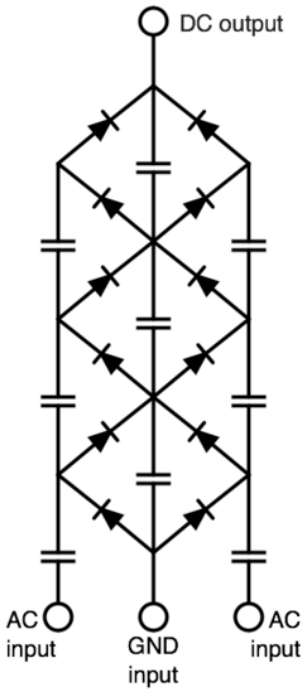
Figure 2.1: Schematic of Fermilab accelerator chain.

radio-frequency (RF) cavities that accelerate the ions to an energy of 400 MeV. The ions are then passed through a carbon foil that strips off the electrons, leaving just a proton. From the linac, protons then go to the Booster [19]. The Booster is the first synchrotron in the accelerating process. It has a radius of 75 m and uses dipole magnets to keep the protons confined to a circular path. Once the linac delivers enough protons (3×10^{12}) the Booster stops accepting and begins accelerating. The Booster accelerates the protons up to an energy of 8 GeV. This is done over the course of about 20,000 revolutions. As with all synchrotrons, the magnet strength increases as the protons are accelerated and the energy range is determined by the radius and the strength of its magnets. After the Booster comes the Main Injector [21]. The Main Injector is a synchrotron with a radius of 528 m. It accelerates protons to either 120 or 150 GeV. The 150 GeV protons are injected into the Tevatron in 36 separate bunches. The 120 GeV protons are sent to the antiproton facility.

To produce antiprotons, focused protons bombard a stationary nickle target. This causes



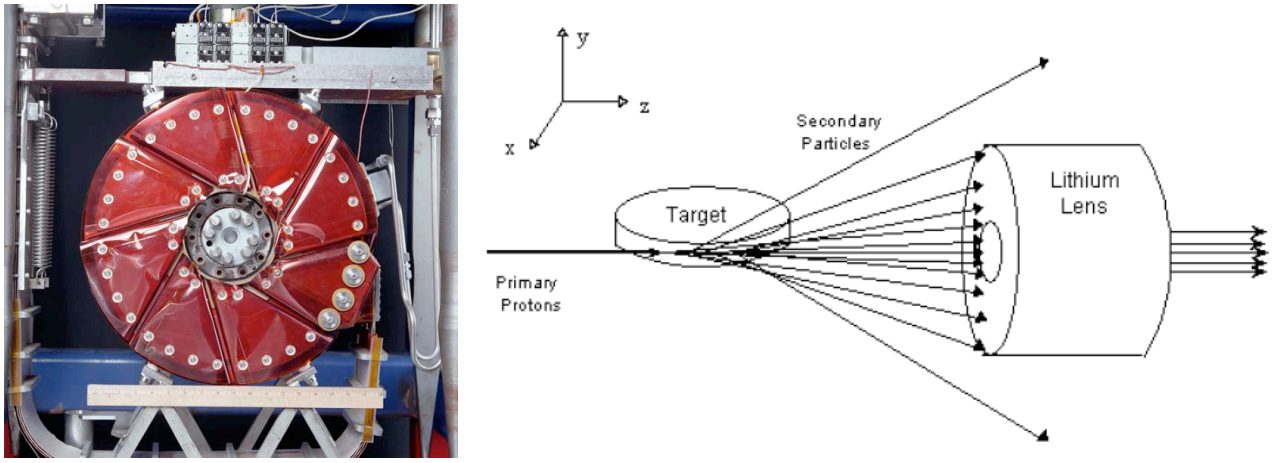
(a)



(b)

Figure 2.2: (a) Image of Cockroft-Walton pre-accelerator at Fermilab. (b) Circuit diagram of a Cockroft-Walton voltage multiplier.

a hadronic cascade that includes antiprotons at an approximate rate of one 15 antiprotons per every 10^6 protons incident on the target. Immediately downstream from the target is a lithium lens module. The lens is designed to focus most of the antiprotons of an energy around 8 GeV coming from the target. Figure 2.3 shows a schematic of the target and the particle cascade as well as a photo of the lithium lens. The lens operates at a peak current of 650 kA in order to generate the magnetic field necessary to direct the antiprotons. The antiprotons are then sent to the Debuncher.



(a) (b)

Figure 2.3: (a) Image of lithium lens (b) Schematic of target and lithium lens.

The Debuncher accepts pulses of antiprotons, due to the pulse like delivery of the protons, at a range of energies. It is a synchrotron, however, it looks more like a rounded triangle. The purpose of the Debuncher is to reduce the antiprotons momentum and transverse phase space, which allows for a more efficient transfer to the Accumulator. Each revolution the Debuncher uses an RF cavity to reduce the momentum of the antiprotons. The reduction is more severe to higher energy antiprotons. After about 100 ms, the debuncher "cools" the antiprotons creating a beam with a small spread of energy around 8 GeV. The antiprotons are now ready to be dumped into the Accumulator.

The name Accumulator is not an accident, as its job is to accumulate antiprotons. It does this by accepting cooled antiprotons from the Debuncher. The Accumulator further cools the antiprotons as it accepts more deliveries from the Debuncher. When it has collected enough antiprotons, the Accumulator sends them to the Recycler.

The antiprotons are stored and further cooled in the Recycler. When it comes time for collisions, they are then dumped into the Main Injector where they are accelerated from 8 GeV to 150 GeV. When the antiprotons reach an energy of 150 GeV, they are ready to be fed into the Tevatron.

The Tevatron is the final phase of the acceleration process. It is a synchrotron with a radius of ≈ 1 km. Unlike the other synchrotrons in the accelerating process, the Tevatron uses super-conducting magnets to produce the strong magnetic fields required to guide the beams around the ring. Both the protons and antiprotons circle the ring in 36 bunches. The proton and antiproton beams are kept in the same tunnel where they helix around each other traveling in opposite directions. The beam paths intersect at two points on the ring; one is the location of the CDF detector and the other is in the middle of the DØ detector.

2.2 DØ Detector

The DØ detector [23, 24, 25] is a multi-purpose detector suited to study high-mass (high p_T) phenomena. The detector is optimized to measure the four momentum and point of production of photons, electrons, muons, jets and neutrinos.[1] In order to accomplish this task, the detector is comprised of many different components shown in figure 2.4, starting from the collision point, the components include:

- Central Tracking System

[1] Neutrinos are not detected directly, however the missing transverse energy that they produce is detected, thus producing a signature in the detector.

This system consists of silicon and fiber tracking layers with a 2 T magnet. It is used to determine the primary vertex (PV) of the event, the paths and momenta of the particles and to help with particle identification.

- Preshower Detector

  The preshower detector is used to assist in particle identification and aids in tracking charged particles.

- Calorimeter

  The calorimeter consists of an EM and hadronic component. The EM calorimeter is used to measure the position and energy of particle showers caused by an electron or photon. The hadronic calorimeter is designed to measure the position and energy of particle showers resulting from quarks and gluons.

- Muon System

  The muon system contains drift tubes, scintillators and toroidal magnets. It is used to measure the four-momentum and position of muons.

2.2.1 DØ Coordinate System

It is now important to discuss the coordinate system used for both the DØ detector and the analysis discussed in this thesis. DØ uses a standard right-handed coordinate system with the center (0,0,0) at the center of the detector, as shown in Figure 2.5. The positive x-axis direction is pointing radially outward from the center of the Tevatron ring, the positive y-axis direction is straight up and the positive z-axis is along the direction of the protons. While DØ is roughly cylindrical in shape, a collision has a spherical rest frame symmetry. This combination is the motivation behind the cylindrical/spherical hybrid system used at DØ (z, θ, ϕ). The positive y-axis is defined by $\phi = \pi/2$.

Theta is typically replaced by another quantity called pseudorapidity, denoted as η and

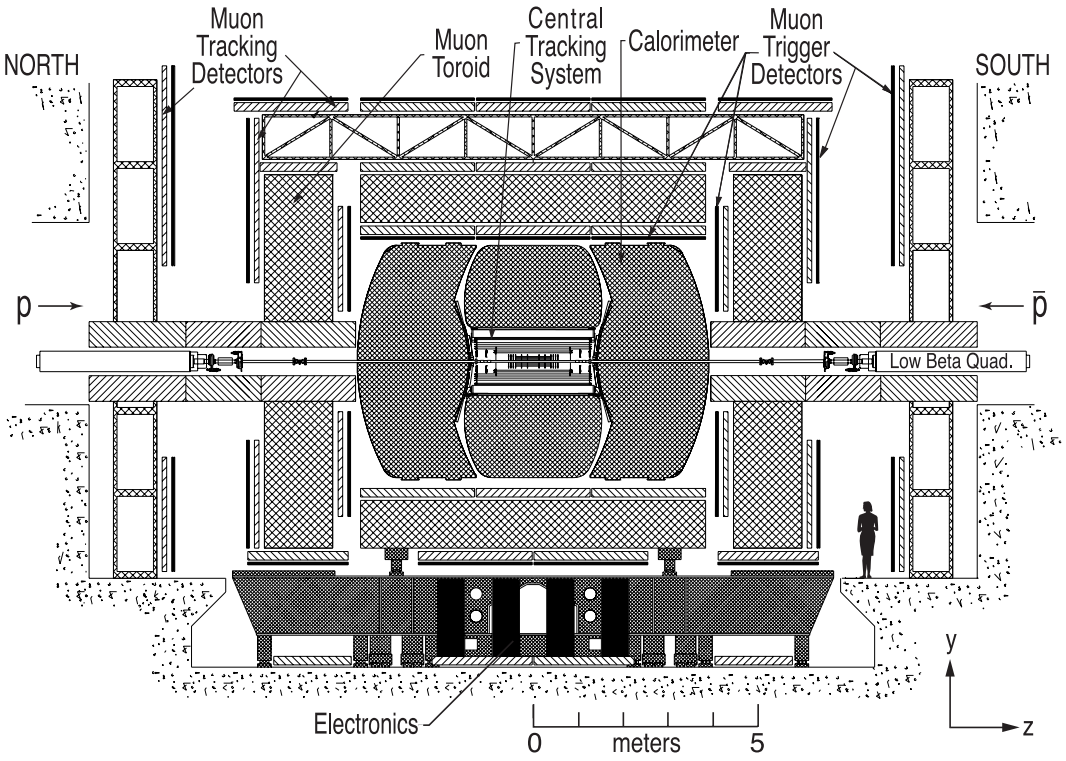


Figure 2.4: Diagram of DØ detector.

defined here:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.1}$$

In the relativistic limit ($E >> m$), η is a good approximation of the Lorentz invariant rapidity (y).

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right), \tag{2.2}$$

Collisions do not always coincide with the center of the detector. The variance from the center point in the x and y coordinates is very small ($\sim$ 40 μm) however, it can be large ($\sim$ 28 cm) in the z-direction. For this reason, it is often useful to define a "detector" η and ϕ and a "physics" η and ϕ. The detector η and ϕ are constructed by using the center of the detector as the center of coordinate system, while physics η and ϕ are constructed by using the primary vertex (PV) as the center of the coordinate system. In general, detector η and ϕ are used when referring to a location in the detector and physics η and ϕ are used when

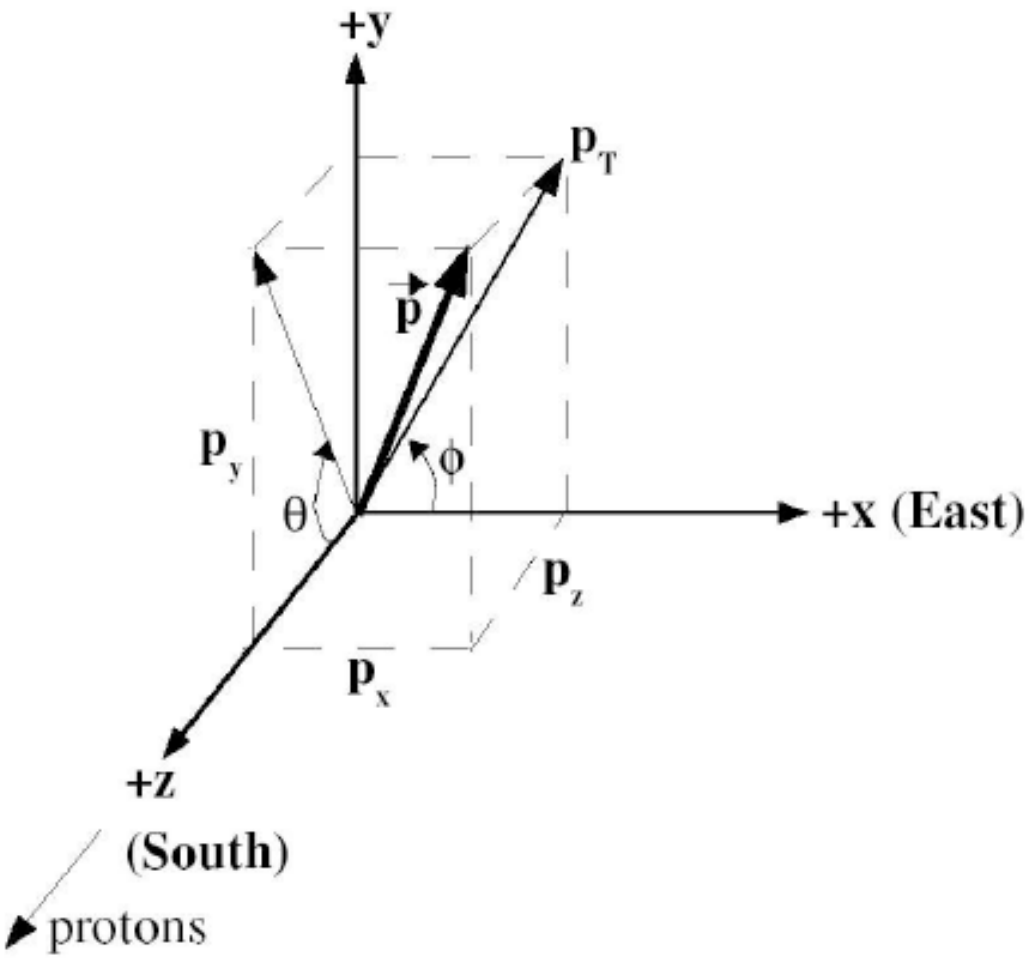


Figure 2.5: Diagram of DØ coordinates.

referring to the properties of a particle; thus no distinction between them is made, unless clarification is needed.

2.2.2 Central Tracking System

The central tracking system is comprised of the Silicon Microstrip Tracker (SMT) which is surrounded by the Central Fiber Tracker (CFT). Both of these trackers are in turn surrounded by a solenoid magnet. Figure 2.6 shows a schematic of the central tracking system.

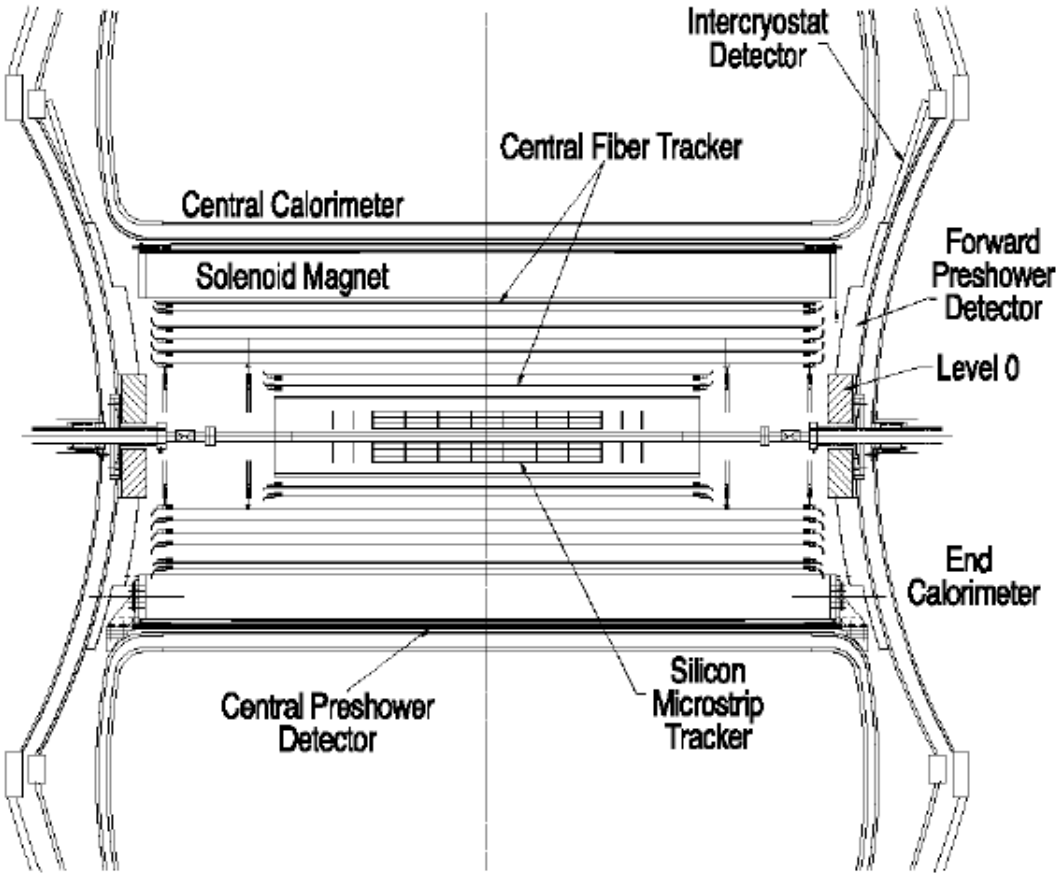


Figure 2.6: Schematic of the central tracker.

SMT

The SMT [26, 27] is composed of silicon microstrip detectors that contain 300 μm thick silicon wafers segmented into parallel strips. The microstrip detectors are arranged in 12 F-disks, four H-disks, and six barrel segments as shown in figure 2.7. This arrangement is motivated by the large z distribution for $p\bar{p}$ interactions and has a coverage of $|\eta| < 3$. Each barrel segment is 12 cm long and contains four concentric cylindrical layers of microstrip detectors, two single sided layers and two double-sided layers. The double-sided layers have one axial side and one stereo side. The axial side (on which microstrips are oriented parallel to the beamline) gives the azimuthal position of tracks. The stereo side makes an angle of 2^o or 90^o with respect to the beamline and provides a measurement of track pseudorapidity. The single-sided barrel layers are axial. The F-disks are made from 12 wedges of double-sided microstrip detectors with each side offset by an angle of $\pm$15^o from radial. The H-disks are made from two layers of 12 single sided microstrip detectors with each layer offset by $\pm$7.5^o from radial, thus forming a turbine-like structure. .

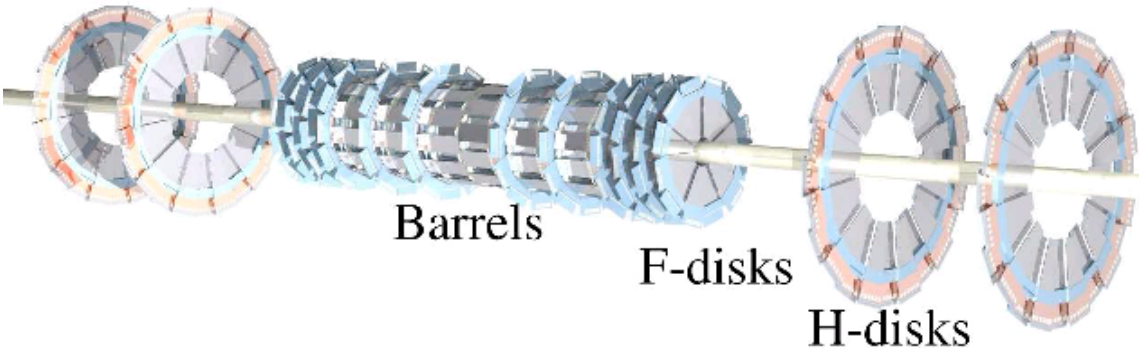


Figure 2.7: Schematic of the SMT.

2.2.3 CFT

The CFT [28] surrounds the SMT and provides tracking coverage up to $|\eta| < 2.0$. The CFT consists of eight concentric carbon fiber barrels that hold layers of scintillating fibers. Each barrel has two double-layered scintillating fibers, as shown in figure 2.8. The inner layers are

axial (parallel to beam line) and the outer layers are stereo. The stereo fibers have an offset of ± 3 °. Each double-layer is composed of 835 μm diameter scintillating fibers arranged in adjacent, parallel layers with the second layer overlapping the gaps of the first layer. The axial fibers provide a measurement of ϕ at a fixed radius, and when combined with the stereo fibers, can provide a measurement in z.

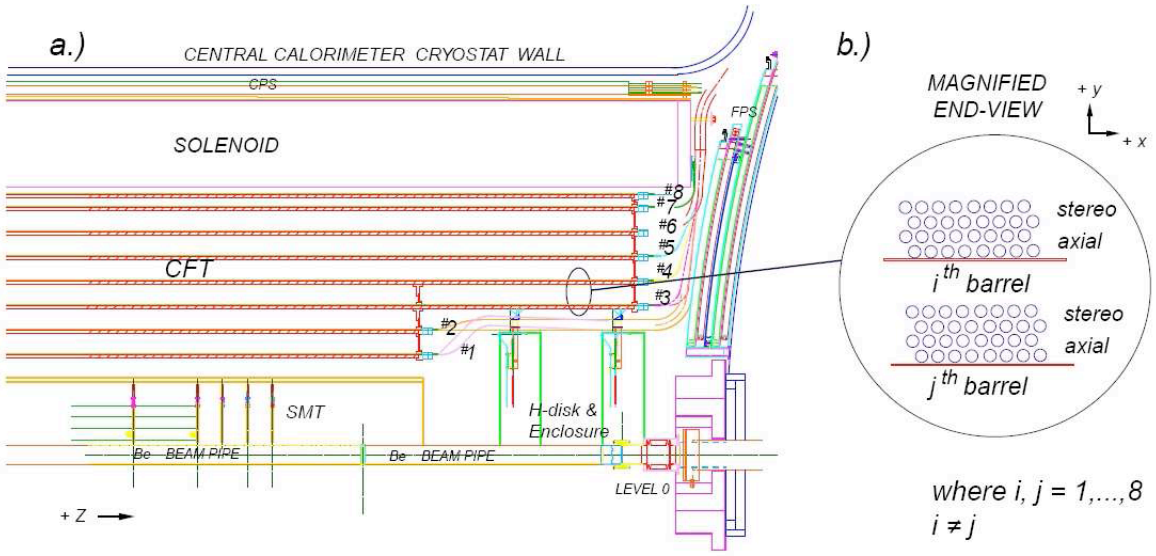


Figure 2.8: Schematic of the CFT.

2.2.4 Solenoid

The solenoid magnet [29] is made of superconducting niobium-titanium wire and is cooled to less than 4.7 K using liquid helium. It is 2.73 m in length with an outer radius of 71.0 cm. The magnet produces a magnetic field of 2 T parallel to the z-axis. This causes the trajectory of charged particles to be curved as they travel transversely from the beamline. The radius of the trajectory in meters is given by:

$$R = \frac{p_T}{0.3qB}; \tag{2.3}$$

where p_T is the transverse momentum of the particle, B is the magnetic field and q is the charge of the particle in units of electron charge, e. Thus, if a particle traveling through the

tracker can be identified, the p_T is then known. The tracker can play a part in identifying a particle as well. A charged particle will leave a track, but a neutral particle will not. For example, the only difference in signature between an electron and photon is whether or not there is a track associated to the calorimeter hit. Thus, the tracker is not only responsible for measuring the momentum of a charged particle, but it also helps with particle identification.

2.3 Preshower Detectors

There are two preshower detectors (shown in figure 2.6) located just before the calorimeters: the central preshower detector (CPS) covering $|\eta| < 1.2$ and a forward preshower detector (FPS) covering $1.4 < |\eta| < 2.5$. The preshower detectors use scintillating fibers with a triangular cross section arranged into layers as shown in figure 2.9.

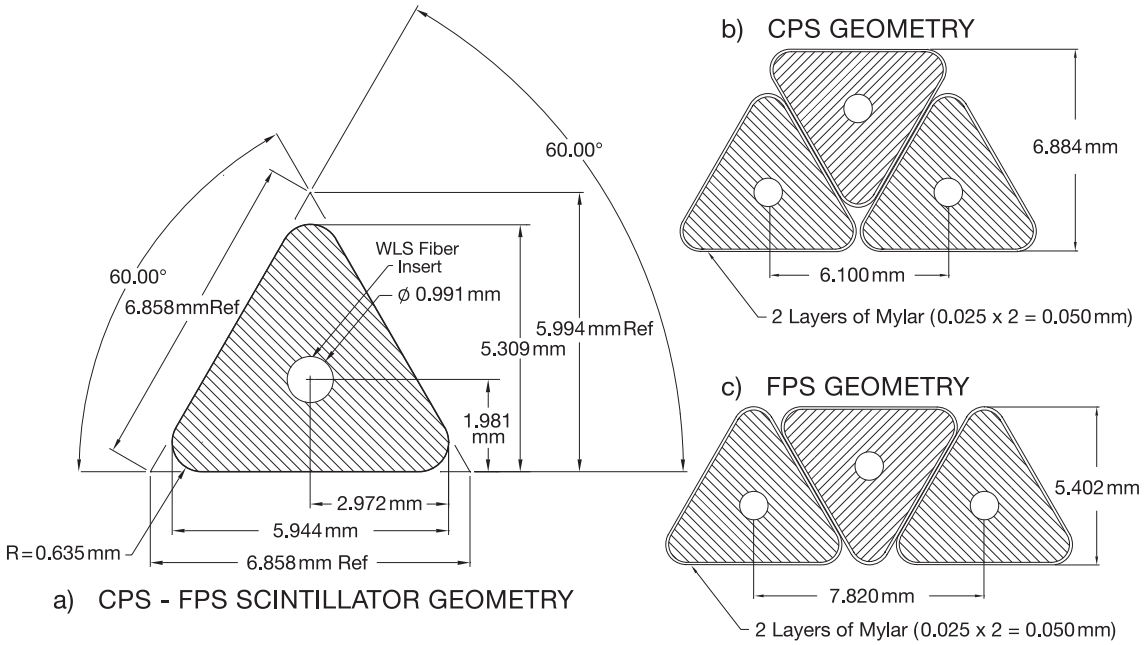


Figure 2.9: (a) Cross section diagram of a scintillating fiber used in the preshower detectors. (b) Diagram showing how fibers are stacked in each layer of the CPS. (c) Diagram showing how fibers are stacked in each layer of the FPS.

The CPS has a layer of lead followed by three layers of scintillating fibers, one axial layer, then two stereo layers with stereo angles of 24° then -24°. The layer of lead along with the

solenoid results in a thickness of approximately two radiation lengths at η=0 and increasing to around four radiation lengths near the end of the CPS at $|\eta|$=1.3. Electrons and photons will begin showering before reaching the scintillators, allowing the preshower detectors to help distinguish EM particles from hadronic particles such as pions.

Each FPS is divided into 16 wedges. Each wedge has two layers of scintillating fibers, followed by a layer of lead two radiation lengths thick, and finally two more layers of scintillating fibers. EM particles do not shower until they reach the layer of lead. Therefore, the first two layers of the FPS are used to assist the central tracking system, while the two outside layers are used to help with electron identification.

2.3.1 Calorimeter

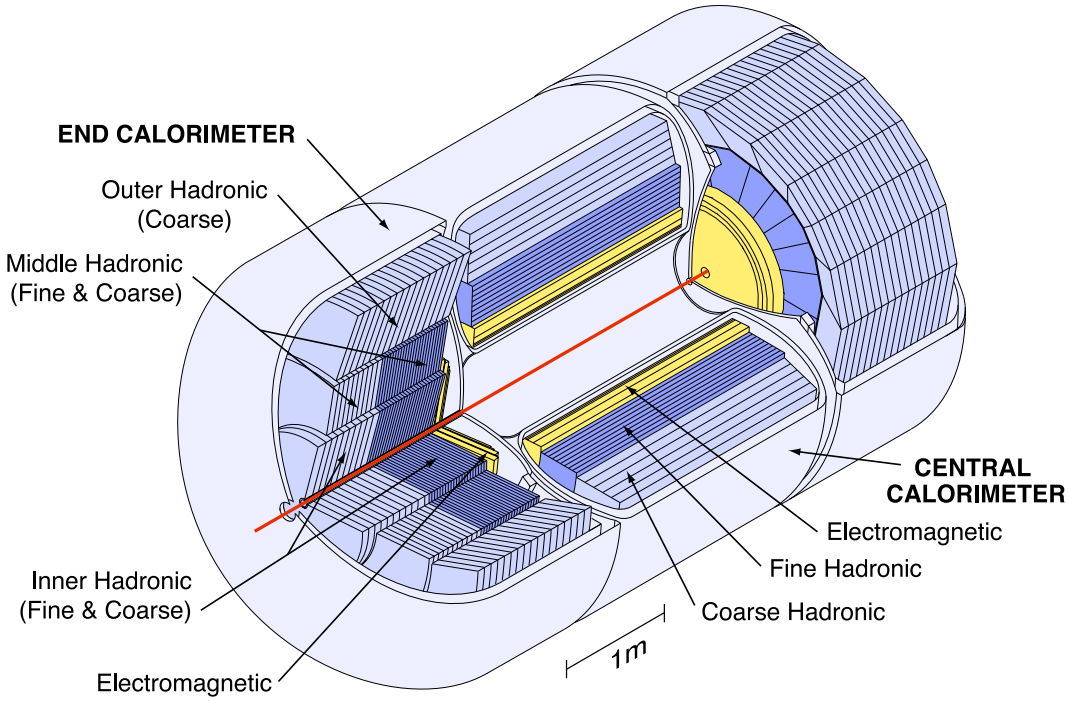


Figure 2.10: Illustration of the DØ calorimeter.

Immediately following the central tracking system and preshower detectors is the DØ calorimeter [23]. The calorimeter measures the energies of electrons, photons, and jets. The calorimeter, shown in figure 2.10, is divided into three separate pieces, a central calorimeter

(CC) covering out to $|\eta| < 1.1$ and two end calorimeters (ECs) covering from $1.1 < |\eta| < 4.0$. All three pieces are encased in a cryostat and cooled to 90 K.

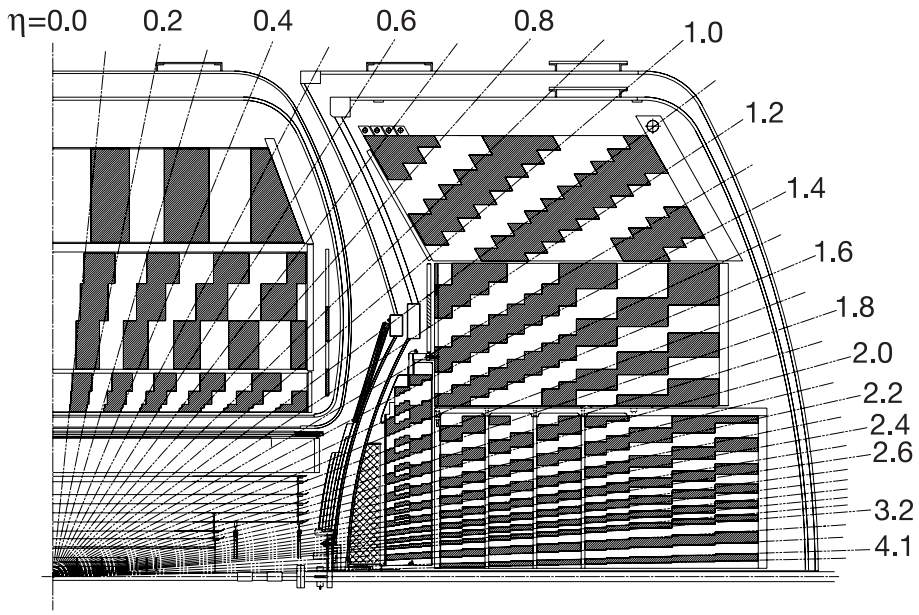


Figure 2.11: Diagram showing calorimeter towers in η.

The calorimeter is made up of many readout layers that allow energy deposits to be sampled at various depths. This provides additional information for particle identification. The inner four readout layers of the CC and ECs are the EM layers designed to measure electron and photon energies. The next three layers in the CC are fine hadronic (FH) layers and the outer most layer is the course hadronic (CH) layer. The hadronic layers in the ECs (figure 2.10) are separated into inner fine hadronic (IFH), inner course hadronic (ICH), middle fine hadronic (MFH), middle course hadronic (MCH), and outer hadronic (OH). Each readout layer is segmented into cells of approximately 0.1×0.1 in $\eta \times \phi$ space with two exceptions. These are the third EM layer, which has 0.05×0.05 segmentation for more accurate measurements of EM showers, and the ECs with $|\eta| > 3.2$, which have reduced granularity to 0.2×0.2 due to space constraints. The cells in each layer are stacked to form 0.1×0.1 towers in $\eta \times \phi$ as shown in figure 2.11.

A shower is produced in the calorimeter by an EM object through repeated bremsstrahlung

radiation ($e \rightarrow e\gamma$) and electron-positron pair production ($\gamma \rightarrow e^+e^-$). The process creates a cascade of particles until the mean energy per particle falls below a threshold where ionization becomes dominant over bremsstrahlung. The effective energy of an electron or photon in the material is the following:

$$E(x) = E_0 e^{-x/X_0} \tag{2.4}$$

where E_0 is the initial energy of the particle, x is the depth of the particle in the material and X_0 is the radiation length of the material. This formulation allows for just one variable, radiation length, to be the determining factor in how effective of an absorber a material is. For a dense material such as uranium, the radiation length is ~ 3.2 mm.

Hadrons interact with the material via the strong force. These interactions also produce secondary particles however, they are typically pions and kaons. Neutral pions (π^0) created in the shower will quickly decay to photons, resulting in some EM showering while the other hadrons will continue to shower and produce more hadrons. The hadronic showering tends be longer and span over a larger area. A good analogy for radiation length in hadronic showers is nuclear interaction length (λ_I). For uranium: $\lambda_I = 10.5$ cm, which is clearly much larger than the radiation length. Properties of the different calorimeter layers are shown in tables 2.1 and 2.2.

	EM	FH	CH
Number of readout layers	4	3	1
Signal boards per readout layer	2, 2, 7, 10	20, 16, 14	9
Absorber material	Uranium	Uranium-Niobium	Copper
Radiation lengths (X_0)	20.5	96.0	32.9
Nuclear interaction lengths (λ_I)	0.76	3.2	3.2

Table 2.1: Properties of CC layers.

Each readout cell is composed of alternating layers of grounded absorbing plates and signal boards kept at a voltage between 2.0-2.5 kV. The space between the signal boards and

	EM	IFH	ICH	MFH	MCH	OH
Number of readout layers	4	4	1	4	1	3
Signal boards per readout layer	2, 2, 6, 8	16	14	15	12	8
Absorber material	Uranium	Uranium-Niobium	Stainless Steel	Uranium-Niobium	Stainless Steel	Stainless Steel
Total radiation lengths (X_0)	20.5	121.8	32.8	115.5	37.9	65.1
Total nuclear interaction lengths (λ_I)	0.97	4.9	3.6	4.0	4.1	7.0

Table 2.2: Properties of EC layers.

the absorbing plates is filled with argon gas. A charged particle from an EM or hadronic shower will ionize the argon gas, and the free electrons drift toward the signal boards inducing a current that signals the particles presence in the detector. Figure 2.12 shows a schematic of the calorimeter readout cell.

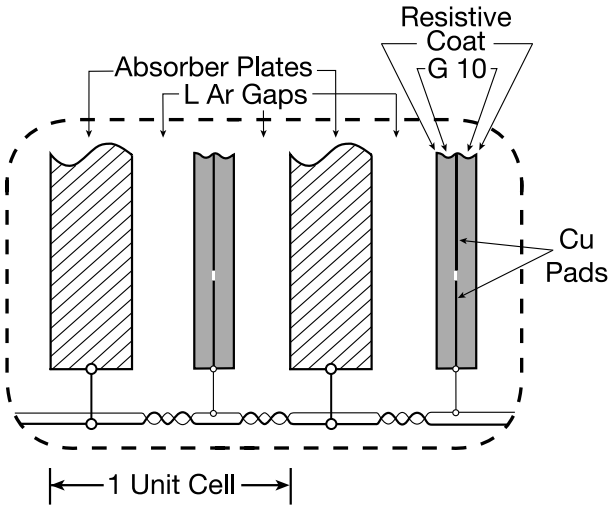


Figure 2.12: Unit cell in the calorimeter.

In total, the calorimeter has around 55,000 readout cells. With so many cells there are bound to be false signals from noise in the electronics or uranium decay. To reduce the

noise from such events, a zero-suppression algorithm [30] (called "T42") is used by removing cells that do not measure an energy significantly higher than the noise level. Specifically, cells with an energy less than four standard deviations above the noise ($E^{\text{cell}} < 4\sigma^{\text{noise}}$) are removed. In the case where an adjacent cell does measure an energy greater than four standard deviations above the noise, the cell must only measure an energy 2.5 standard deviations above in order to be kept.

2.3.2 Muon System

Because of their large mass ($\sim$ 200 times m_e), muons only leave a small fraction of their energy in the detector. Other than neutrinos, muons are the only SM particle that lives long enough to travel through the entire detector without being stopped. For this reason, the muon system is on the outside of the detector.

The muon system has a three-layered rectangular geometry, forming a cube like structure on the outside of the detector. The "A" layer is the layer closest to the beam line, followed by iron toroidal magnet, then the "B" and "C" layers, respectively. The sides of the muon system at each end (perpendicular to the beam line) form the forward angle muon system (FAMUS) and the other four sides form the wide angle muon system (WAMUS). The WAMUS covers the region $|\eta| \lesssim 1$ and the FAMUS covers the region $1 \lesssim |\eta| \lesssim 2$. The η boundaries are approximate due to the cube like structure of the system.

Figure 2.13 shows the layout of the proportional drift tubes (PDTs) in the muons system. A schematic of a PDT can be seen in figure 2.14. The tubes are 10 cm wide and 5.5 cm thick. They have copper cathodes on the upper and lower walls and a gold plated tungsten wire running down the middle. The gas filling the tubes is made up of a mixture of argon, methane and CF_4 gas. The cathode walls are held at 2.3 kV and the anode wire at 4.7 kV.

A muon that passes through a PDT will leave a trail of ions in the gas. The resultant free electrons are accelerated in the electric field, ionizing more of the gas as they travel toward the anode. The number of electrons that eventually reach the wire is a function of

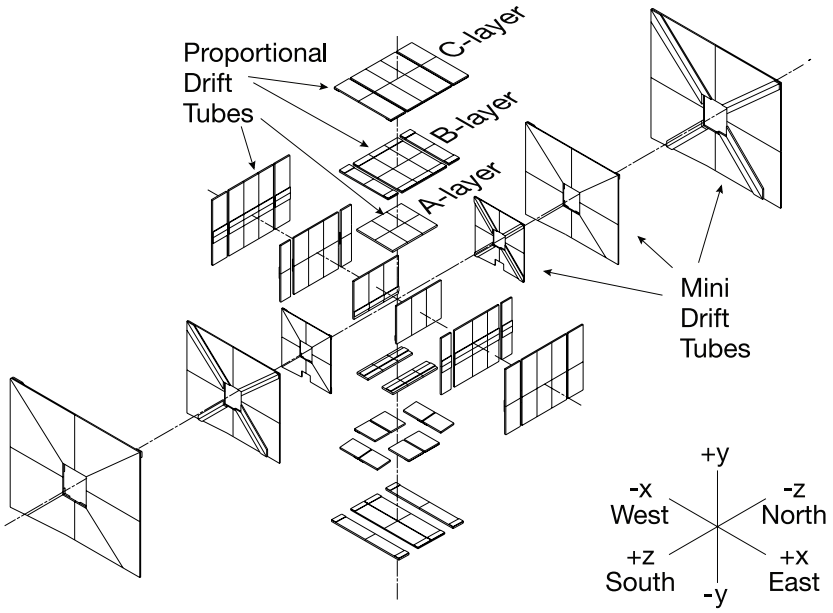


Figure 2.13: Muon System drift tubes.

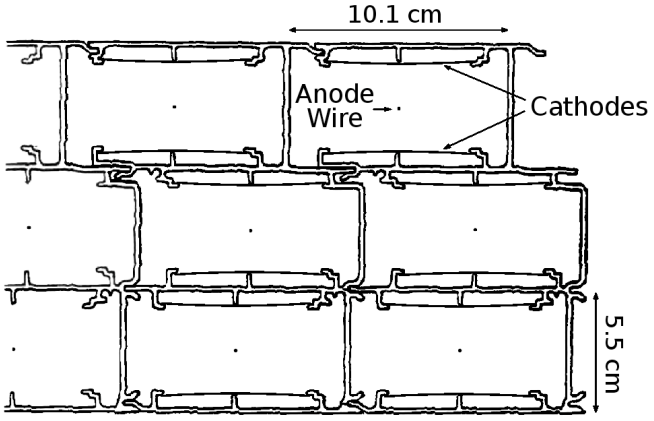


Figure 2.14: Cross section of PDTs used in WAMUS.

the distance between the wire and the muon track.

The toroidal magnet between layers A and B produces a magnetic field of 1.9 T, which causes the muons path to curve. The result is similar to that of the central tracker, a measurement of p_T. However, the momentum resolution for the muon system is not nearly as fine as the central trackers resolution. Thus, the muon system momentum measurement is typically used to help find a matching track in the central tracking system, giving a better momentum measurement. The WAMUS PDTs are oriented perpendicular to the beamline. This provides a measurement of the z-position of the tracks to a precision of 1 mm.

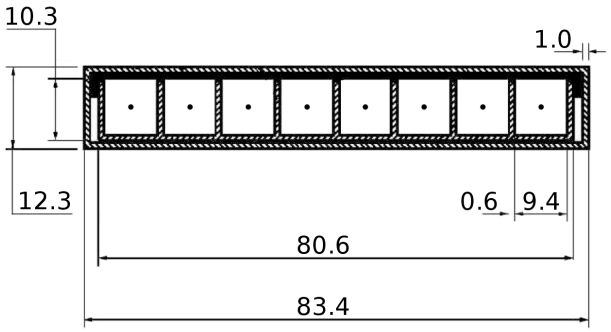


Figure 2.15: Cross section of MDTs used in FAMUS (dimensions in millimeters).

Figure 2.15 shows a cross section of the mini drift tubes (MDT) used in the FAMUS. The MDTs are arranged perpendicular to the beam line as shown in figure 2.16. The MDTs are filled with a mixture of CF_4 gas and methane. Their operation is similar to that of the PDTs with the exception that they have a much shorter response time (100 ns vs. 500 ns for the PDTs).

Figure 2.17 shows the layout of the scintillators in the muon system. As can be seen, all three layers contain scintillating panels. The "cosmic cap" consists of scintillators in the C layer on the top and sides of the WAMUS. Scintillators in the bottom of layers B and C form the "cosmic bottom". Other than the "A-ϕ" scintillators, which are segmented to provide a ϕ position measurement, the primary role of the scintillators is to provide precise timing information. The fast response time of the scintillators (1.6 ns) allows for them to be used for triggering information, matching hits to events and rejecting cosmic ray muons.

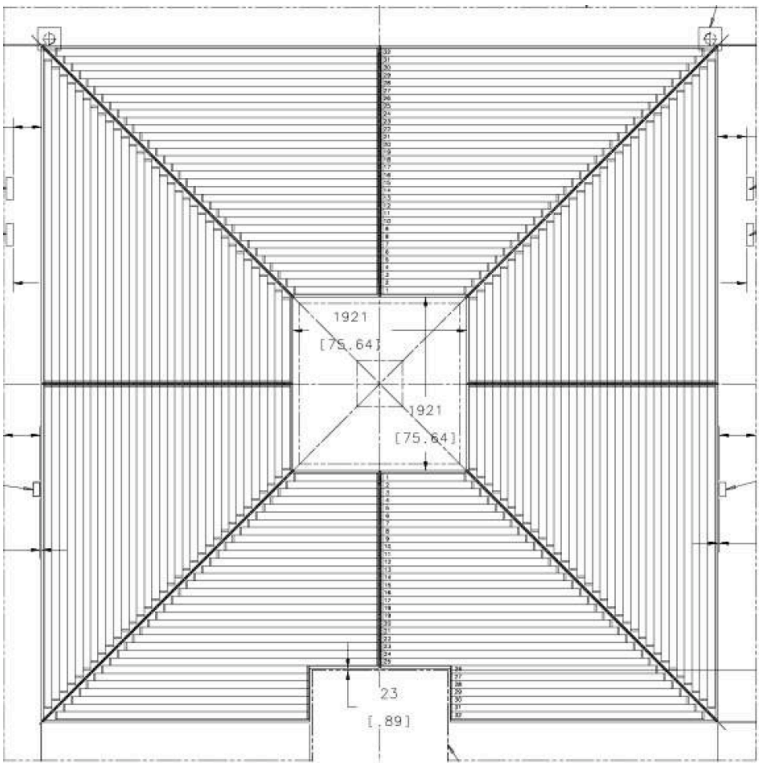


Figure 2.16: Illustration showing the orientation of MDTs in each layer of the FAMUS.

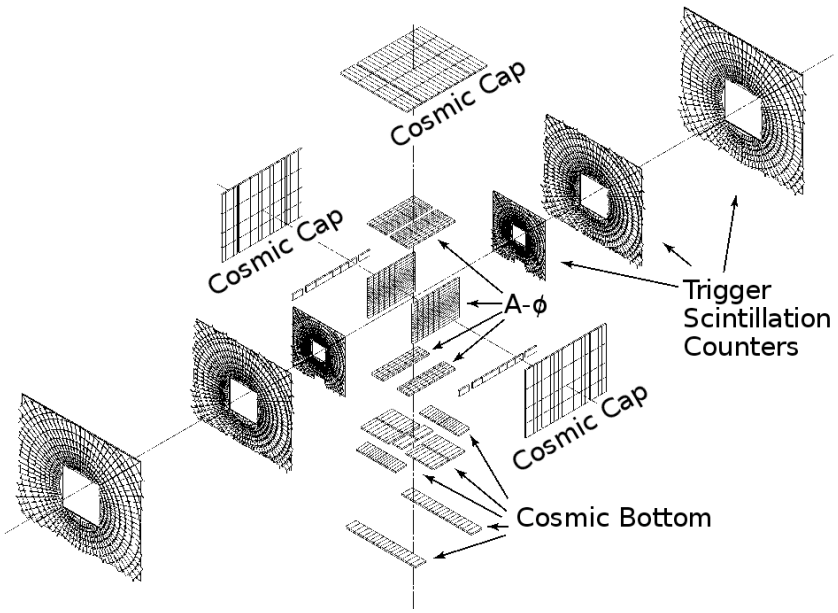


Figure 2.17: Muon System scintillators.

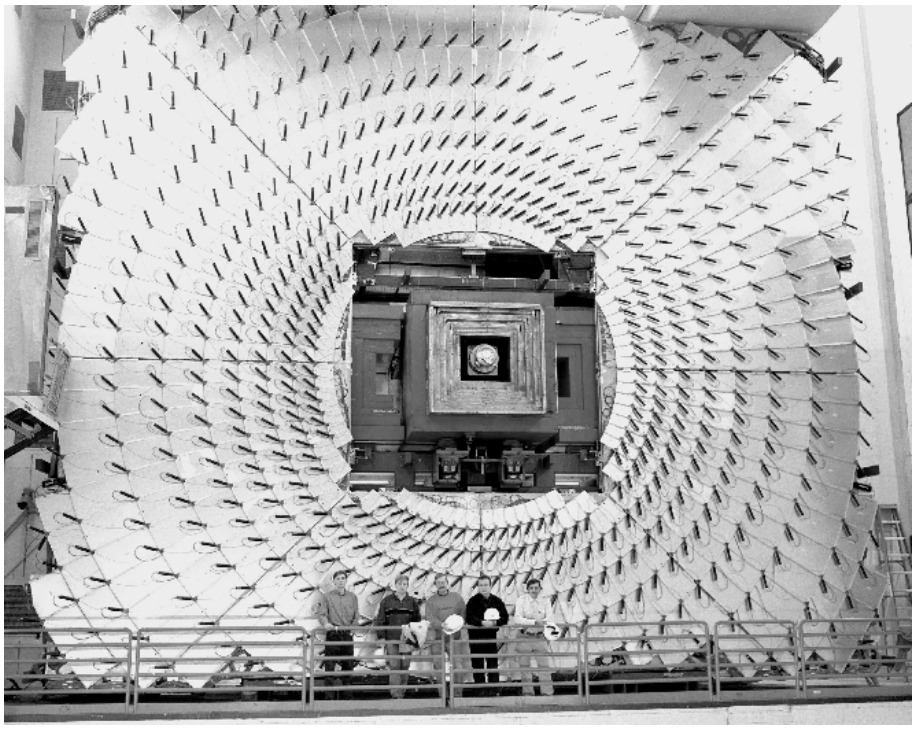


Figure 2.18: Photograph of the scintillating tiles in the C layer of the FAMUS.

The FAMUS also contains a plane of scintillators used for triggering and cosmic ray muon vetoes. The scintillating tiles form a circular matrix that has become synonymous with the image of DØ (see figure 2.18).

2.3.3 Luminosity Monitor

The luminosity monitor is used to determine the instantaneous luminosity being delivered to the DØ detector. This is done by measuring the rate of the proton-antiproton collisions. After a collision, the remnants of the proton and antiproton that were not involved in the hard interaction will hadronize in the z-direction with very small transverse momentum. To detect these remnants, a luminosity monitor is set at either end of the tracking system ($z = \pm 144$ cm) and spans a range $2.7 < |\eta| < 4.4$, as shown in figure 2.19. Each luminosity monitor in composed of 24 wedges of scintillating tiles, as shown in figure 2.20. The scintillators have a timing resolution of around 0.2 ns, allowing for a z-position resolution within 6 cm. Collisions that take place at $z < 100$ cm are used to determine the luminosity.

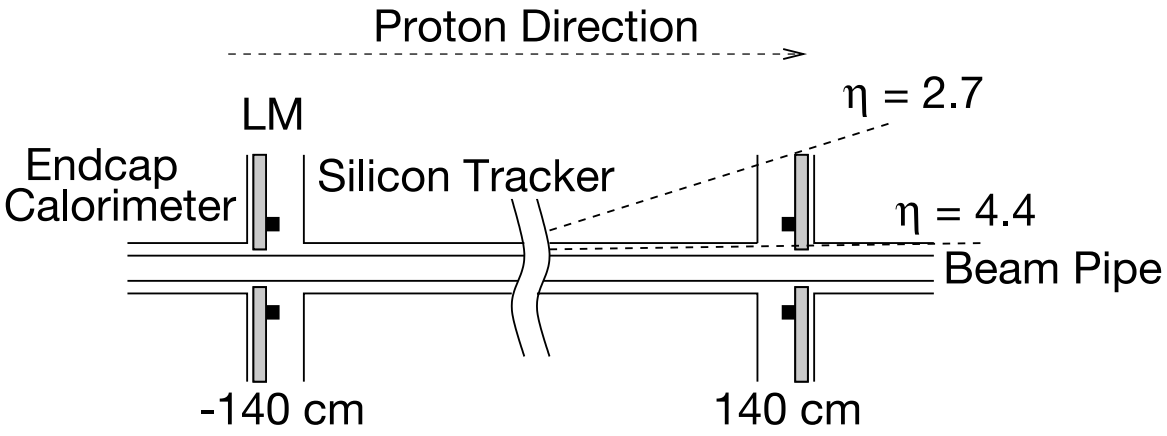


Figure 2.19: Schematic of luminosity monitor in y-z plane.

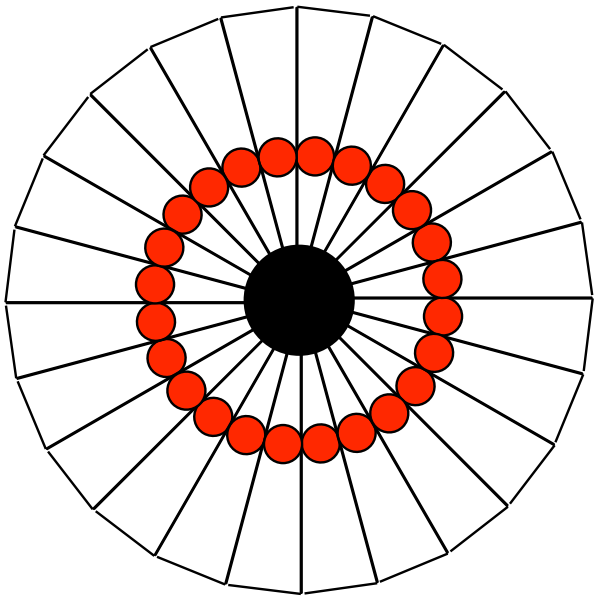


Figure 2.20: Schematic of luminosity monitor scintillators. The red spots represent the PMTs.

2.4 Trigger

With 36 bunches of protons and antiprotons traveling around the Tevatron every 21 μs, the average rate of bunch crossings at the center of the DØ detector is 1.7 MHz. A three-stage trigger system [31, 32] is used by DØ to reduce this rate to less than 100 events per second that are stored for later analysis. Of the millions of inelastic collisions happening in the DØ detector only a handful will contain physics that scientists are interested in studying. For example, only a few W bosons are produced per second and only a few top-quark pairs per hour. If every event was stored for later analysis the experiment would have to store around two terabytes of data each day, which is unnecessary. Therefore, sophisticated triggers capable of rejecting nearly 99.999% of the interactions in only a few milliseconds are used to pick out the rare events of interest.

The trigger system works by exploiting the fact that the events of interest have distinctive signatures. For example, a high p_T lepton may indicate the presence of a W or Z boson, a high p_T lepton plus two b quark jets would signal a top quark pair or Higgs boson, and leptons plus large missing energy is a signature for supersymmetry. The job of the trigger is to reject all events that do not display characteristics consistent with a signature of any interesting physics. Therefore, it is the trigger settings that ultimately determine the physics processes that can be studied.

The trigger system used by DØ is divided into three levels. An event must pass each of the three levels before it can be stored for later analysis. The Level 1 trigger is designed to find patterns of energy deposition indicating the passage of high energy particles and must do so very quickly (4.2 μs). It order to reduce the time, the level 1 trigger uses a condensed subset of the full detector readout and is implemented entirely in hardware and firmware. Each detector subsystem (calorimeter, muon, etc.) checks if the event passes preprogrammed trigger conditions (thresholds) then the results from each detector subsystem are combined to make the final Level 1 decision. After an event passes Level 1 it is sent to Level 2 where it is subjected to more refined tests that may take up to 100 μs. As well as firmware, Level 2

uses microprocessors to take advantage of more precise detector information and spatial correlations to form basic "objects" such as tracks, EM clusters, and jet clusters. Each detector subsystem has a dedicated microprocessor that reduces the data for that subsystem then sends it on to a global processor to make the final Level 2 decision. The rate at which Level 2 can pass events to Level 3 is limited to 1 kHz by the Level 2 processing time. When an event passes the Level 2 trigger the entire detector is read out and sent to the Level 3 CPU farm made up of over 100 Linux computers. The Level 3 CPUs work together to process each event in under 25 ms. During that time the entire event is reconstructed and sophisticated algorithms, similar to those discussed in the next chapter, are applied to make the final decision. The output rate for Level 3 is maintained around 50 Hz, which was chosen to keep the offline event reconstruction from piling up.

2.4.1 Single Muon and Electron Triggers

One type of trigger used in this analysis is the single muon and electron triggers. These triggers pass events that have at least one muon or electron that pass a set of requirements. A tag-and-probe method is typically used to calculate single muon or electron trigger efficiencies. This method involves using a low-background sample in data, such as $Z \rightarrow \mu\mu$, where one of the muons is chosen to be the "tag" and the other is chosen to be the "probe". The "tag" muon must fire some trigger requirement that is unbiased with respect to the "probe" muon. The "probe" muon is then tested to see at what rate it passes trigger requirements (efficiencies).

Single Muon and Electron OR Triggers

Many single muon and electron triggers exist at DØ, each with very low efficiencies. Instead of only selecting one of these triggers, it is more advantageous to take the logical OR of these trigger (*i.e.* allow an event to pass if any of a list of triggers passes.). This strategy is laid out in [33] and [34] and is followed here. The list of single muon triggers applied in the

single muon OR trigger can be found in Appendix F and the list of single electron triggers can be found in Appendix G.

Chapter 3

Event Reconstruction

Before analyzing the data, the objects in each event must be reconstructed. This process involves taking the raw data from the detector and applying a set of dedicated algorithms. This chapter describes the objects reconstructed and algorithms used to reconstruct them.

3.1 Track Reconstruction

Tracks are important for measuring momentum, object identification (ID) and b-tagging. An example where the tracker is important for object ID is EM objects. As charged particles pass through the tracking system layers, a signal is left behind (called a hit) that is then analyzed by a track finding algorithm.

A charged particle passing through the SMT and CFT will most likely interact with many channels in each layer. This results in a cluster of hits, as shown in figure 3.1. The energy-weighted center of each group of hits is then used as the location of that group and entered as an input into two track finding algorithms.

The two track finding algorithms employed by DØ attempt to find all the tracks in each event. The first method begins with the innermost tracking layers and moves outward. The algorithm finds a hit in an SMT barrel or F-disk and matches it to a hit further out with a $\Delta\phi < 0.08$ between the two. The algorithm then searches for a third hit consistent with the

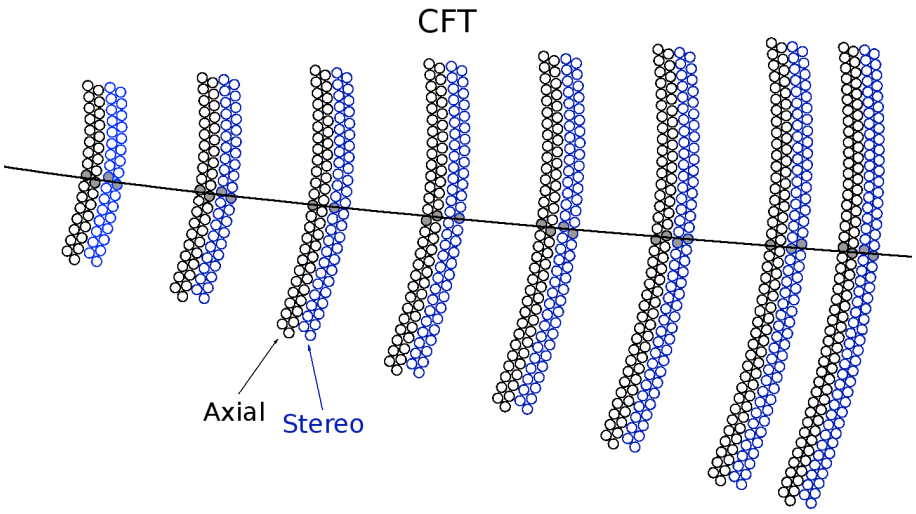


Figure 3.1: Illustration showing the clusters of hits left in each layer of the CFT from a charged particle.

first two that has a radius of curvature of $\rho > 30$ cm ($p_T > 180$ MeV), a distance of closest approach (dca) to the beamline of $dca < 2.5$ cm and a $\chi^2 < 16$ fit value. These candidates are then used to construct tracks through the rest of the tracking system with a Kalman filter [35]. Using a detailed schematic of the DØ detector, the Kalman filter extrapolates the track to the each layer of the tracking system. If a hit in a layer is found within a small window around the extrapolation, and fits the track with a $\chi^2 < 16$, then the hit is included in the track. The track is discarded if three consecutive layers are missing candidate hits, or if more than 2/3 of the hits are shared with another track candidate.

The second method takes all SMT hits and applies a Hough transformation [36] to convert the hit positions in $x \times y$ space to $\rho \times \phi_0$ space, where ρ is the curvature of a circle intersecting the origin and the hit, and ϕ_0 is the angle tangent to that circle at the origin. Hits from the same track intersect at a single point in the ρ-ϕ_0 plane. Figure 3.2 shows the process of converting from $x \times y$ space to $\rho \times \phi_0$ space. First, a 2D histogram is populated, and in the most ideal case, the hits in the most populated bin are taken as a track. Then, a 2D Kalman filter is used to more accurately calculate the track parameters and remove the tracks with large χ^2. The z-component of each track is then determined using a similar approach as the one above, then a three-dimensional Kalman filter is used to extrapolate the tracks through

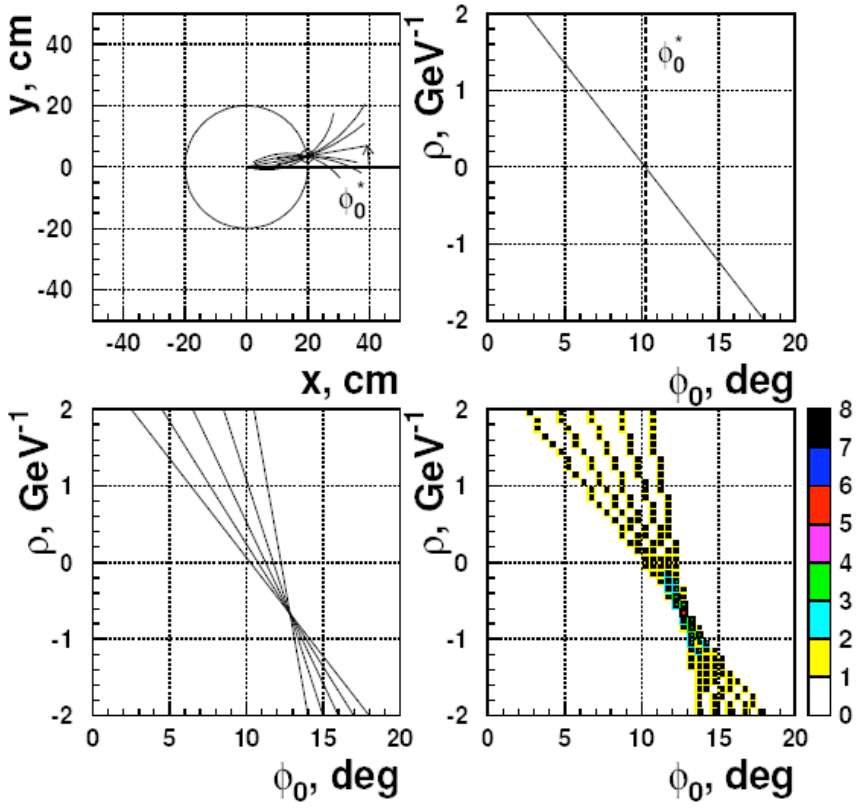


Figure 3.2: Example of track reconstruction with Hough transformation. The top left shows family of tracks that contain a given hit in x-y space. The top right shows these tracks in ρ-ϕ_0 space. The bottom left shows the intersection for each hit on a track, which indicates the parameter value for that track. The bottom right, the lines are used to fill a histogram and the intersection is identified as a maximum.

the CFT.

The final step is the repetition of the first method in the CFT. Tracks found in the CFT are extrapolated back into the SMT. The final track candidates are then fed into a Kalman filter, which determines the best candidates based on a multidimensional χ^2 minimization.

3.2 Primary Vertex Reconstruction

A vertex is the location of a proton-antiproton collision. Although the typical number of collisions in a bunch crossing is two, there can be many more. The locations of these collisions are distributed over about $\sigma_r = 40$ μm in the radial direction and about $\sigma_z = 28$ cm in the

z-direction. It is important to measure the collision point in events of interest, for instance in an event with a Z boson, so that the momentum vectors can be accurately calculated. The collision point of an event of interest is called a primary vertex. The other vertices within the same bunch crossing are referred to as minimum bias.

Reconstructing vertices is done in two steps. The first step uses tracks that have a dca to the z-axis within 100 standard deviations. These tracks are used to find the most likely location for a common vertex. If the vertex yields a $\chi^2/ndf > 10$, the track with the highest χ^2 value is removed and the vertex is recalculated. This process is iterated until the χ^2/ndf drops below 10. The process is then repeated for the removed tracks in order to find other vertices. The second step is then run on each vertex found in step one. Tracks associated to the vertex with $p_T > 0.5$ GeV, at least 2 SMT hits and a dca less than 5 standard deviations from the vertex location are used in a similar process to that of step one, where the location of the vertex is recalculated until χ^2/ndf drops below 10.

After the different vertices for the bunch crossing are found, the primary vertex needs to be determined. Tracks from a minimum bias interaction tend to have lower p_T than tracks from a hard interaction. Using the p_T of each track, the probability that it came from a minimum bias interaction is calculated. The probability of each track associated to a vertex is then multiplied together to make the probability that the vertex is a minimum bias vertex. The vertex with the smallest probability to be from a minimum bias interaction is chosen as the primary vertex.

3.3 Electron Reconstruction

The signature of an electron is an isolated track in the central tracking system followed by a narrow energy deposit in the calorimeter that ends in the first hadronic layer. To reconstruct an electron, a cone algorithm that finds clusters of energy in the EM layers of the calorimeter is used. Each identified cluster is matched to a track and is tested to see if it has certain

characteristics that are expected in an electron shower.

An EM tower is defined as the first five layers of a 0.1×0.1 ($\eta \times \phi$) tower of calorimeter cells (the four EM layers plus the first hadronic layer). All EM towers with a transverse energy (E_T) greater than 1.5 GeV go into a list of seed towers. The seed tower with the highest E_T is selected, and a cone of radius $R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ is centered on that tower. The energy-weighted center for the cluster of EM towers inside the cone is calculated and the cone is moved to the new center. This processes is repeated until a stable cluster is found.

Although the clusters have a radius of $R = 0.4$, the electron shower is actually much narrower. The energy is measured using the EM towers inside a cone of radius $R = 0.2$ (at the center of the cluster). To be considered an electron candidate, there must be a track that extrapolates to within $\Delta\eta = 0.05$ and $\Delta\phi = 0.05$ of the cluster's center. The angle of the track at the primary vertex is used to calculate the four-momentum for the electron.

Any clusters that match the above criteria must also match the following selections in order to be considered an electron candidate:

- The cluster must have a minimum E_T of 1.5 GeV
- $f_{EM} > 0.9$ where:

$$f_{EM} = \frac{E_{EM}}{E_{Tot}} \tag{3.1}$$

  That is, the energy deposited in the EM layers of the calorimeter (E_{EM}) must be greater than 90% of the entire energy of the cluster (E_{Tot}).

- The electron must have an isolation fraction (f_{iso}) < 0.2. Where f_{iso} is:

$$f_{iso} = \frac{E_{tot}^{Cal}(R < 0.4) - E_{EM}(R < 0.2)}{E_{EM}(R < 0.2)} \tag{3.2}$$

where $E_{tot}^{Cal}(R < 0.4)$ is the total energy deposited into the calorimeter within a radius of

0.4 with the exception of the course hadronic (CH) layers. In other words, only 20% more energy than that of the electron can be deposited in the calorimeter within a radius of 0.4 in order for it to be considered isolated (shown in figure 3.3).

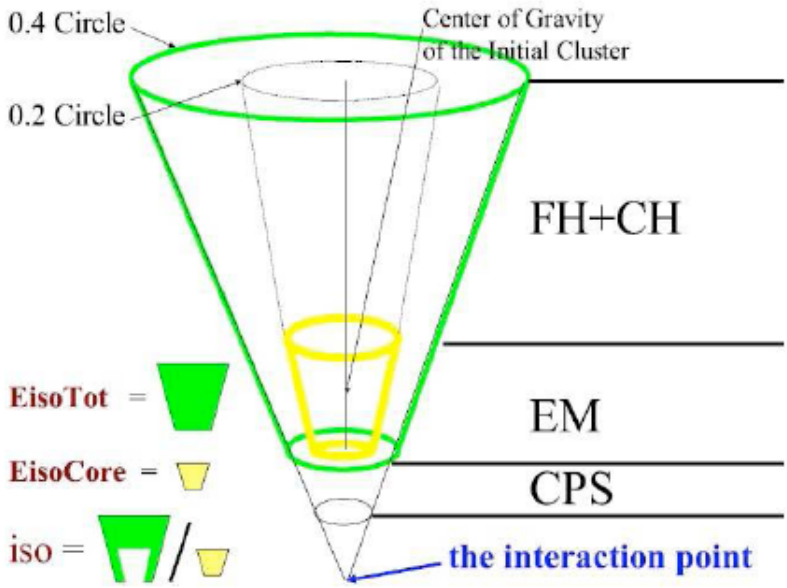


Figure 3.3: Schematic of f_{iso} for EM cluster.

3.3.1 Electron Quality

These criteria are used to select electron candidates, however the selection is quite loose. To increase the purity of electron selection, a few techniques are used.

Track Matching

A track match probability is calculated to determine how well the cluster is matched to a track. Two track matching methods are used. The spacial track-match uses the distance between the track and the center of the cluster in ϕ- and z-directions:

$$\chi^2_{\text{spacial}} = \left(\frac{\Delta\phi}{\sigma_{\Delta\phi}}\right)^2 + \left(\frac{\Delta z}{\sigma_{\Delta z}}\right)^2, \tag{3.3}$$

while the E/p track-match compares the energy, E, measured in the calorimeter to the momentum, p, of the track:

$$\chi^2_{\mathrm{E/p}} = \left(\frac{\Delta\phi}{\sigma_{\Delta\phi}}\right)^2 + \left(\frac{\Delta z}{\sigma_{\Delta z}}\right)^2 + \left(\frac{E/p - 1}{\sigma_{E/p}}\right)^2. \tag{3.4}$$

where σ_X corresponds to one standard deviation for the distribution of variable X. The corresponding χ^2 probability distribution, $P(\chi^2)$, is then used to determine the probability that the track matches the cluster. If there is more than one possible track, the track with the lowest χ^2 is used.

H-matrix Method

The H-matrix χ^2 [38] provides a comparison between the shower shape of the cluster and the expected shower shape for an electron. Seven variables are used to parameterize the shower shape:

- Fraction of energy in the first EM layer
- Fraction of energy in the second EM layer
- Fraction of energy in the third EM layer
- Fraction of energy in the fourth EM layer
- Width of shower in the third EM layer
- Logarithm of the total energy
- z coordinate of the primary vertex

A covariance matrix (M) is constructed using MC samples in order to determine the expected electron shower shape. The H-matrix is simply the inverse of the covariance matrix

$H \equiv M^{-1}$. Thus a construction of χ^2 can be carried out:

$$\chi^2_H = \sum_{i,j=1}^{7} (x_i - \bar{x}_i) H_{ij} (x_j - \bar{x}_j) \tag{3.5}$$

where x_i is the value of the ith input (listed above). A different H-matrix is used for separate η regions because the expected shower shape changes. Some H-matrix cuts are applied for electron selection in this analysis and are discussed in Chapter 6.

Electron Likelihood

The electron likelihood [39] variable is designed to discriminate between real and fake electrons. A fake electron is any electron candidate that is not the result of a real electron. Electron likelihood makes use of seven variables:

1. f_{EM}
2. Spacial track-match χ^2 probability ($P(\chi^2_{\text{spacial}})$)
3. E_T/p_T^{track}; Ratio of calorimeter energy and track momentum
4. χ^2_H
5. Track distance of closest approach to the primary vertex
6. Total p_T of other tracks inside the cone of radius $R = 0.4$
7. Number of tracks inside a cone of radius $R = 0.05$

For each variable, a probability for signal and background are determined. The overall probability that a candidate is signal (background) is then simply the product of the probabil-

ity that each of the seven variables is signal (background), as shown in equations 3.6 and 3.7:

$$P_S = \prod_{i=1}^{7} P_S^i \tag{3.6}$$

$$P_{BG} = \prod_{i=1}^{7} P_{BG}^i \tag{3.7}$$

where $P^i_{S(BG)}$ is the probability of the ith variable to be signal (background). These two probabilities are then used to calculate the electron likelihood as follows:

$$L_e = \frac{P_S}{P_S + P_{BG}} \tag{3.8}$$

3.3.2 Electron Energy Correction

The measured electron energy is corrected by selecting $Z \rightarrow ee$ events in data and comparing the dielectron mass resonance with the expectation from previous precision measurements [37]. The dielectron mass resonance obtained using uncorrected electron energies is found to be lower than expected. Thus the uncorrected electron energies are scaled up to obtain corrected energies that reproduce the expected $Z \rightarrow ee$ resonance. The true energy, E^{true}, is assumed to be related to the energy measured in the detector, E^{det}, by

$$E^{\text{det}} = \alpha E^{\text{true}} + \beta; \tag{3.9}$$

where α is the energy scale and β is an energy offset. The scale and offset corrections are determined separately for CC and EC electrons using a binned maximum likelihood method. The corrected electron energy is then calculated by inverting equation 3.9.

3.4 Jet Reconstruction

Jet reconstruction is done in three stages. The first stage is clustering, in which the four-momentum for each 0.1×0.1 ($\eta \times \phi$) tower of calorimeter cells is calculated:

$$p^{tower} = \sum_i p_i^{cell} \tag{3.10}$$

where p^{tower} is the massless four-momentum with an energy of the ith cell in the tower and a direction point from the primary vertex to the center of the cell. Towers with $p_T > 0.5$ GeV are put into a list of seeds, which are used for making clusters. The tower with the highest p_T is selected as the center, and a cone of radius 0.3 is created around the tower. Towers within the cone are summed up to form a cluster. All seeds that fall outside of this cone are put into a new list of seeds, and the process is repeated.

The second stage is the creation of protojets using the clusters found in the first stage. Clusters that contain more than one energetic tower and have a $p_T > 1.0$ GeV are put into a list of seeds for making protojets. The cluster with the highest p_T is chosen as the center, and a cone of radius $R = \sqrt{\Delta y^2 + \Delta\phi^2} = 0.5$ is formed around it. Any seeds within the cone are summed up to form a new protojet. The cone is re-centered to the center of the new protojet, and the summation process is repeated. The re-centering process is repeated until a stable protojet is found or until the protojet p_T falls below 4 GeV, at which point it is no longer considered a candidate. A new seed list is then constructed that is made up of the clusters that fall outside of a radius of 0.25 from the stable protojet, and the process is repeated.

Protojets can overlap because the size of the cone to construct them is larger than the size of the cone that clusters fall within in order to be removed from the seed list. Because of this, the final stage is the splitting or merging of overlapping protojets. If a protojet does

not overlap with another, it is made into a jet. If a protojet shares at least half of its p_T with another protojet, it is merged. The merging process is done by summing the two protojets to make a new protojet and then repeating step two until a new stable protojet is found. Splitting is done by assigning the shared towers exclusively to the closer of the two protojets. Splitting and merging is repeated until there are no more overlapping protojets.

3.4.1 Jet Quality

The following basic quality requirements are applied to the reconstructed jets to remove "fake" jets that are primarily the result of noisy cells or regions in the calorimeter.

- $0.05 < f_{EM} < 0.95$: The fraction of energy coming from the EM layers must be greater than 5% and less than 95%.
- $f_{CH} < 0.46$: The fraction of energy coming from the course hadronic calorimeter must be less than 44% in the CC and 46% in the EC.
- $E^1_{cell}/E^2_{cell} < 10$: The jet is rejected if the energy of the most energetic cell (E^1_{cell}) in the jet is more than ten times the energy of the second most energetic cell (E^2_{cell}).
- $E^1_{tower}/E_{tot} < .9$: The jet cannot have more than 90% of its total energy (E_{tot}) in a single calorimeter tower (E^1_{tower}).

Even after the above quality cuts are applied, some noise jets still exist. For this reason, an additional stipulation is required that each jet is confirmed by the Level 1 trigger (*i.e.*, verify that the jet was "seen" by the Level 1 trigger). The Level 1 trigger uses only a condensed subset of the full calorimeter data consisting of the 100 highest E_T trigger towers, where trigger towers are made up of groups of four calorimeter towers, excluding the course hadronic layer. The Level 1 confirmation requirement is:

$$\frac{E_T^{\text{trigger towers}}}{E_T^{\text{reco}} \cdot (1 - f_{\text{CH}})} > 0.5 \tag{3.11}$$

where $E_T^{\text{trigger towers}}$ is the total E_T of trigger towers inside a cone of radius $R = 0.5$ around the jet, E_T^{reco} is the E_T of the reconstructed jets, and f_{CH} is the fraction of E_T^{reco} in the course hadronic cells.

3.4.2 Jet Energy Scale

The jet energy scale (JES)[43] correction is used to correct a jet's measured energy ($E_{\text{jet}}^{\text{measured}}$) to more accurately reflect the true energy ($E_{\text{jet}}^{\text{true}}$) of all the particles in the jet before they interact with the detector. The measured energy for a jet is related to the true energy by:

$$E_{\text{jet}}^{\text{true}} = \frac{E_{\text{jet}}^{\text{measured}} - E_O}{R \cdot S} \tag{3.12}$$

where E_O is the offset to the energy resulting from uranium decay, minimum-bias interactions, pile-up, and electronics noise; R is the energy response or the fraction of a particle's energy that is actually measured, which is affected by energy lost before reaching the calorimeter, gaps/dead spots in the calorimeter, differences in EM and hadronic response, and non-linearities in energy response; and S is the showering correction due to energy that is lost outside of the cone or gained from an object leaking energy into of the cone.

By measuring the average values for each of the above quantities (as functions of energy and location in the detector), corrected jet energies can be calculated as:

$$E_{\text{jet}}^{\text{corr}} = \frac{E_{\text{jet}}^{\text{measured}} - \bar{E}_O}{\bar{R} \cdot \bar{S}} \tag{3.13}$$

where $\bar{E}$ is the measured average value for E.

The magnitude of $\bar{E}_0$ is determined in minimum bias events where the trigger is based solely on the luminosity detectors, and no other trigger requirements are set.

The calorimeter response, $\bar{R}$, is determined using the Missing E_T Projection Fraction (MPF) method [45] in data. This method takes γ+ jet events in data where the photon (γ)

and jet are back-to-back in the transverse plane. The total transverse momentum in the event is virtually 0, which means that $p_T^{\gamma} = p_T^{jet}$. The photon is required to be in the central region so that its energy can be accurately measured. Figure 3.4 shows that $\bar{R}$ is determined by measuring the missing transverse energy (MET) in the event.

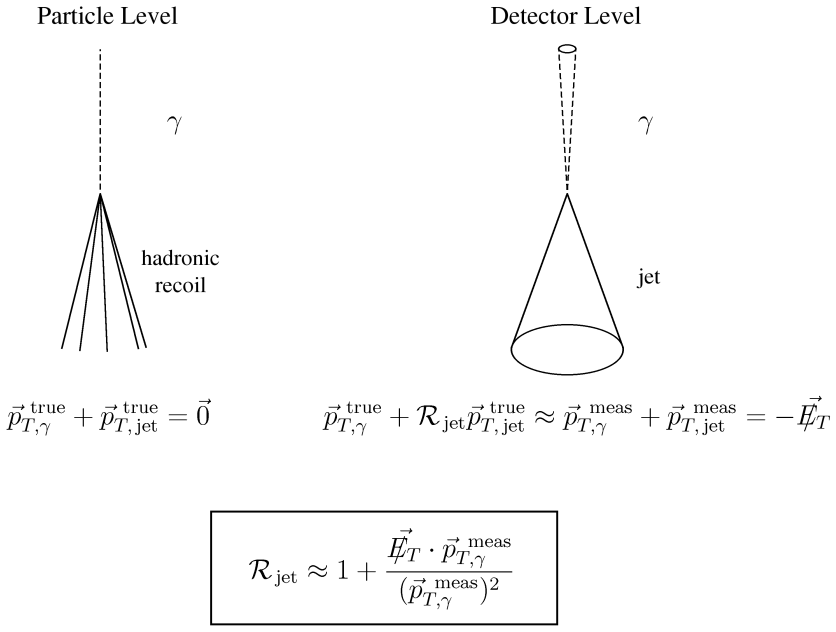


Figure 3.4: Illustration of the MPF method.

The showering correction, $\bar{S}$, is determined by measuring the shower profile (the energy contained in cones of varying radii, $0.1 \leq \tilde{R} \leq 2$, centered on the jet). MC simulation is used to create templates of the shower profiles for both particles inside and outside the jet. The templates are fit to the shower profile measured in the γ+jet data to determine the ratio of measured energy inside the cone to the true jet energy inside the cone. The JES corrections as a function of η can be seen in figure 3.5.

3.4.3 Jet Energy Resolution

The jet energy resolution (JER) is determined in dijet events where the two jets are back-to-back ($|\Delta\phi_{jets} - \pi| < \pi/36$). The jet p_T resolution is directly related to the asymmetric

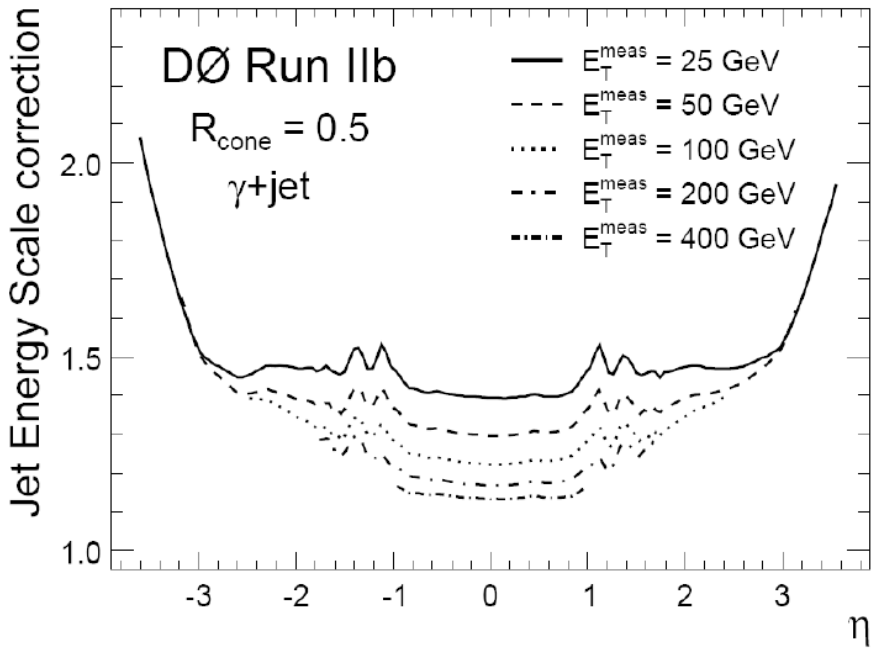


Figure 3.5: JES correction as a function of η.

resolution of the two jets:

$$|A| = \frac{|p_T^{jet_1} - p_T^{jet_2}|}{p_T^{jet_1} + p_T^{jet_2}} \tag{3.14}$$

where $p_T^{jet_{1(2)}}$ is the transverse momentum of the jets. The jet energy resolution is then given by:

$$JER = \sqrt{2}\sigma_A \tag{3.15}$$

where σ_A is the width of the A distribution obtained by a Gaussian fit.

3.5 Muon Reconstruction

The signature of a muon is a track in the muon system that can be matched to an isolated track in the tracking system. A track segment in layer A of the muon system is formed by fitting a straight line to the hits in two or more of the decks of the drift tubes. A track segment is formed in layers B and C in a similar way, however, these two layers are considered

together (called layer BC) because they fall after the toroid. When there is more than one possible fit for layers A or BC, the line with the lowest χ^2 value is used. After the segments are constructed, a fit is performed that takes into account the bending that occurs due to the toroid magnet and the energy loss from multiple scattering in the material. Because the particle identification is known, the p_T can also be known based on the radius of curvature due to the toroid's magnetic field. However, if the muon can be matched to a central track, then a more accurate p_T measurement can be made. The matching to a central track is done by extrapolating the muon track inward, while accounting for scattering in the calorimeter and matching to a central track. As usual, the track with the smallest χ^2 value is chosen as the matching central track.

3.5.1 Muon Quality

Like electrons, a muon candidate can be found, however, the quality of the candidate is determined by the following variables.

Muon System Quality

Requirements on the number of drift tube and scintillator hits in both the A and BC layers can be set in order to ensure high-quality muons and to reduce the likelihood of a fake. In addition, precise timing requirements can be placed on the scintillator hits to help reject cosmic ray muons.

Central Track Quality

The values used to quantify the quality of the central tracks are:

- Distance of closest approach to the beamline for tracks
- χ^2/ndf for the match to the central track

The distance of closest approach to the beamline is primarily used to reject muons from cosmic rays, as they are unlikely to intersect the beamline. Muons originating from pion or kaon decays inside the tracking volume will most likely have poorly reconstructed central tracks and will thus have a high χ^2/ndf value and can then be rejected.

Muon Isolation

A muon from a Z boson will result in an isolated track originating from the primary vertex. This is different than a heavy quark decay resulting in a muon (*e.g.*, a b-jet where the final state contains a muon) in which the muon would not be isolated. Thus, an isolation requirement for the muons can help distinguish between the two initial states. The isolation requirement for muons is further discussed in Chapter 6.

Two variables are used to ensure muon isolation are:

$$I_{calo} = E_T^{cal}(0.4) - E_T^{cal}(0.1) \tag{3.16}$$

$$I_{track} = p_T^{tracks}(0.5) - p_T^{\mu} \tag{3.17}$$

where $E_T^{cal}(R)$ is the E_T in the calorimeter inside a radius of R around the muon track, $p_T^{tracks}(0.5)$ is the p_T of all tracks inside a radius of 0.5 around the muon track and p_T^{μ} is the p_T of the muon.

3.6 Missing Energy

Momentum conservation requires that the total momentum in any direction at the beginning of an event be equal to the total momentum in that same direction at the end of the event. Measuring the momentum in the z-direction after an event is very difficult due to the fact that much of the event travels down the beampipe and is undetected. However, in the x and y-directions, this is not the case. The beams travel primarily in the z-direction, meaning that the momentum in the x and y directions is virtually zero. This leads to the assertion

that the total transverse momentum in any event must equal zero; $p_T^{tot} = 0$. If this is found not to be true, it indicates the presence of one or more undetected particles in the event (*e.g.* a neutrino).

3.6.1 MET Reconstruction

The first step in reconstructing the missing E_T is calculating the uncorrected missing E_T measured by the calorimeter

$$MET^{uncorr} = -\sum_i \vec{p}_i^{\,\text{cell}} \tag{3.18}$$

where $\vec{p}_i^{\,cell}$ is a three-momentum with a magnitude equal to the energy of the ith cell in the EM and fine hadronic layers of the calorimeter, that points from the primary vertex to the center of the cell. The course hadronic cells are not included because they can contain a substantial amount of noise.

Muons, which only deposit a couple GeV of energy in the calorimeter, will contribute to MET^{uncorr}. Additionally, reconstructed electrons, photons and jets have corrected energies that are more accurate than the simple sum of the calorimeter cells inside those objects. Therefore, the missing E_T is corrected by replacing the uncorrected energy with the corrected energy for each reconstructed electron, muon, photon and jet in the event. The corrected missing E_T is then given by

$$MET^{corr} = MET^{uncorr} - \sum_{objects} \left(\vec{p}^{\,object} - \sum_{i \in object} \vec{p}_i^{\,cell} \right) \tag{3.19}$$

where the outer sum is over reconstructed objects, $\vec{p}^{\,object}$ is the corrected momentum of the object, and i runs over the EM and fine hadronic calorimeter cells in the electron, photon or jet, or along the interpolated path of the muon through the calorimeter.

Chapter 4

Data

The data sample for this analysis was collected from Fermilab Tevatron $p\bar{p}$ collisions at center-of-mass energy $\sqrt{s} = 1.96$ TeV using the D0 detector between June 2006 and June 2009 (Run IIb1 and Run IIb2). The data is selected from the 2MUhighpt and 2EMhighpt skims which reject events in data that do not contain two high p_T muons or electrons, respectively.

4.1 Trigger Requirements

For the dielectron final state, a logical "OR" of the single electron triggers is used, with an efficiency of close to 100% [34]. For the dimuon final state, a logical "OR" of the single muon triggers is used, with an efficiency of 80%. The treatment of systematic uncertainty of the muon trigger is discussed in Chapter 12. The data was filtered by using a standard data quality definition which removes bad runs and luminosity blocks as defined by the SMT, CFT, Calorimeter, and Muon groups at DØ. The integrated luminosity of the data sample *after the data quality selection* was measured to be 4.2 fb^{-1} for Run IIb.

Chapter 5

MC Samples

This analysis makes use of several Monte Carlo (MC) samples for the Z+jets events selection and the estimation of background contributions. With the exception of the multijet background, all significant background processes ($t\bar{t}$, WZ, WW, and ZZ) are estimated from MC simulation. The details of these MC samples are outlined in this section.

All MC simulated events are processed by a Geant-based simulation of DØ. The events are then overlayed with a data driven minimum bias. This produces a detector level event simulation that is then passed through the same event reconstruction as data.

The MC samples used in this analysis can be found in Appendix H. The Z+jets MC samples are broken up by dilepton invariant mass. The tables list the cross section times branching ratio and total number of events for each MC sample. The WW, WZ and ZZ processes are generated using Pythia [47], while the $t\bar{t}$ and Z+jets processes are generated with Alpgen [48] interfaced with Pythia. The Z+jets MC samples are generated with Alpgen separately for each number of additional partons. The initial and final state parton radiation is then added to the event using Pythia, applying a matching scheme [50] to avoid double counting. To simulate the underlying event, the so-called "DØ-Tune" (Tune

A with CTEQ6L1 PDF) has been used. The Z+jets MC is made of events with only light partons, called Z+LP, and events with heavy-flavor partons, called Z+HF. Samples were generated specifically for $Z + 2b(2c)$ jets with additional light parton jets. Some of these samples overlap with the events generated for Z + light partons (nlp) in the final state phase space. In order to avoid double counting the overlapping events, events with b- or c-quark jets in the final state are removed from the $Z + nlp$ sample and events with c-quark jets in the final state are removed from the $Z + 2b + nlp$ samples. The cross-sections for these Heavy Flavor (HF) skimmed samples are then scaled to reflect the removal of those events.

Cross sections were taken from several different sources depending on whether the samples were produced using Pythia or Alpgen + Pythia. The WZ, WW and ZZ cross sections are taken from MCFM [68] as these samples were generated using Pythia exclusively. These cross sections are NLO and use CTEQ6.1M PDFs. Additional NLO corrections were applied to the Alpgen Z+jets cross sections and are described in the Chapter 9.

Chapter 6

Event Selection

The pre-selections on leptons, jets and the reconstruction of Z boson in the event are outlined in the following sections. Although the reconstruction of Z is different, the jet selection and b-tagging performed are the same in the $Z \to \mu\mu$ and $Z \to ee$ channels.

6.1 Primary Vertex

A primary vertex (PV) is required with at least three associated tracks and reconstructed z-position within 60 cm of the center of the detector ($z = 0$).

6.2 Muons

Events are required to have at least two muons with:

- $p_T > 15$ GeV
- Loose muon ID
- A matched central track
- A distance of closest approach of < 0.2 cm for track with no SMT hits

- $|\eta_{det}| < 2.0$

The muon p_T in data and MC is corrected using the primary vertex information for each event if the muon track has no associated SMT hits.

6.2.1 Reconstructed $Z \to \mu\mu$ boson

A good Z candidate is required in each event and is reconstructed from a pair of selected muons. The Z selection is as follows:

- 70 GeV $< M_{\mu\mu} <$ 110 GeV
- Pseudo-acolinearity > 0.05 (anti-cosmic)[1]
- Muons with opposite sign charge
- *Product scaled isolation* < 0.03 (for dimuon only)

The product scaled isolation variable is the product of the isolation variables (equation 6.1) for each of the two muons that form the Z candidate. This allows for one non-isolated muon if the other muon is sufficiently isolated. This Z selection is chosen in order to coincide with the $ZH \to llbb$ analysis [10].

$$I(\mu_i) = \frac{E_T^{cal}(0.1 - 0.4) + p_T^{trk}(0.5)}{p_T^i} \tag{6.1}$$

where $E_T^{cal}(0.1 - 0.4)$ is the calorimter E_T inside a hollow cone of $0.1 < \Delta R < 0.4$ around the muon; $p_T^{trk}(0.5)$ is the vector sum of the transverse momenta of all tracks in a cone $\Delta R < 0.5$ aorund the muon; and p_T^i is the transverse momentum of the muon.

[1] The pseudo-acolinearity between two directions is calculated as $(\pi - \Delta(\phi 1, \phi 2)) + |(\pi - (\theta 1 + \theta 2))|$. By construction it is small when directions are collinear.

6.3 Electrons

The dielectron channel contains events where both electrons are reconstructed either in the central cryostat (CC: $|\eta_{det}| < 1.1$) or in the endcap cryostat (EC: $1.5 < |\eta_{det}| < 2.5$). At least two electron candidates with p_T >15 GeV are required.

The following criteria maintain the highest efficiency for signal for a manageable multijet background:

- Isolated EM cluster $f_{iso} < 0.1$
- Energy fraction in EM calorimeter $f_{EM} > 0.95$
- Shower shape cut: HMx7 < 35 in CC, or HMx8 < 20 in EC
- IsoHC4< 3 in CC, or IsoHC4 < 2 in EC
- NNout7> 0.2 in CC, or NNout3 > 0.4 in EC
- TrkMatch $\chi^2 > 0.001$ or EMHits_e_f> 0.4 in CC

The $HMx < N >$ are H-matrix calorimeter shower shape variables. The IsoHC4 is a track isolation variable. The $NNout < N >$ are outputs of neural-networks with $< N >$ input variables including the energy deposited and number of hit cells in the first EM layer, the track isolation and energy deposited in the central preshower detector.

6.3.1 Reconstructed $Z \rightarrow ee$ boson

A good Z candidate is required in each event and is reconstructed from a pair of selected electrons, with invariant mass 70 $GeV < M_{ee} <$ 110 GeV

6.4 Jets

Jets reconstructed using a cone algorithm [69] with ΔR < 0.5 are called JCCB jets. At least one 'good' JCCB jet is required in each event, satisfying the following requirements:

- Leading jet *corrected* p_T > 20 GeV
- *Corrected* p_T > 15 GeV for all non-leading jets
- $|\eta| < 2.5$
- Vertex confirmation

Jet energies are corrected using the standard jet energy scale (JES) corrections [55]. In order to suppress additional jets originating from minimum-bias secondary interactions, jets are required to originate from the primary vertex. Vertex confirmation requires that at least two tracks associated with the jet be matched to the primary vertex.

6.5 Tagged Sample

In order to suppress events with light and c-quark jets, the Neural Network b-tagging algorithm discussed in Chapter 8 is applied.

For the tagged sample, at least one of the jets is required to have an NN value greater than 0.5.

6.6 MET Cut

One last requirement for event selection is a cut on MET. This requirement suppresses $t\bar{t}$ events while keeping most other events. MET is a typical signature of a neutrino. A high MET is a signature of one or more high p_T neutrinos being present in an event. The large mass of a top quark and its high probability of decaying with a neutrino in the final state make a high MET cut a perfect candidate for suppressing $t\bar{t}$ events. Figure 6.1 shows the MET plot for both data and the expected physics contributions. As the figure shows, events

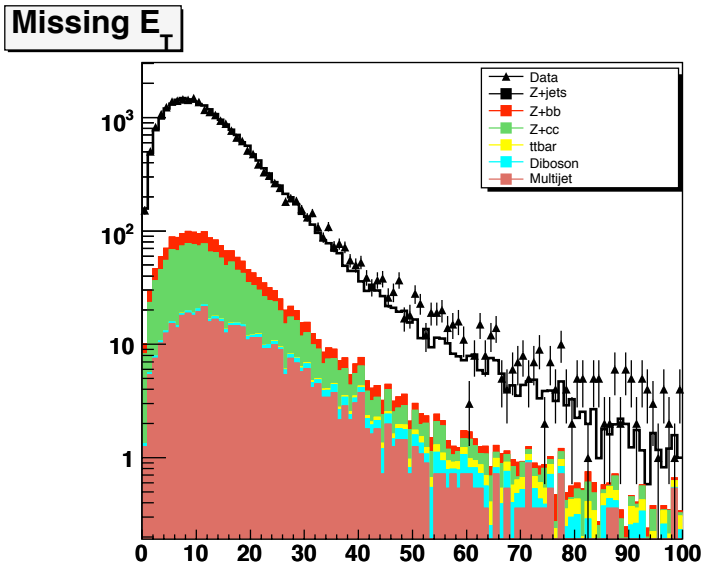


Figure 6.1: MET plot for data and expected physics contributions.

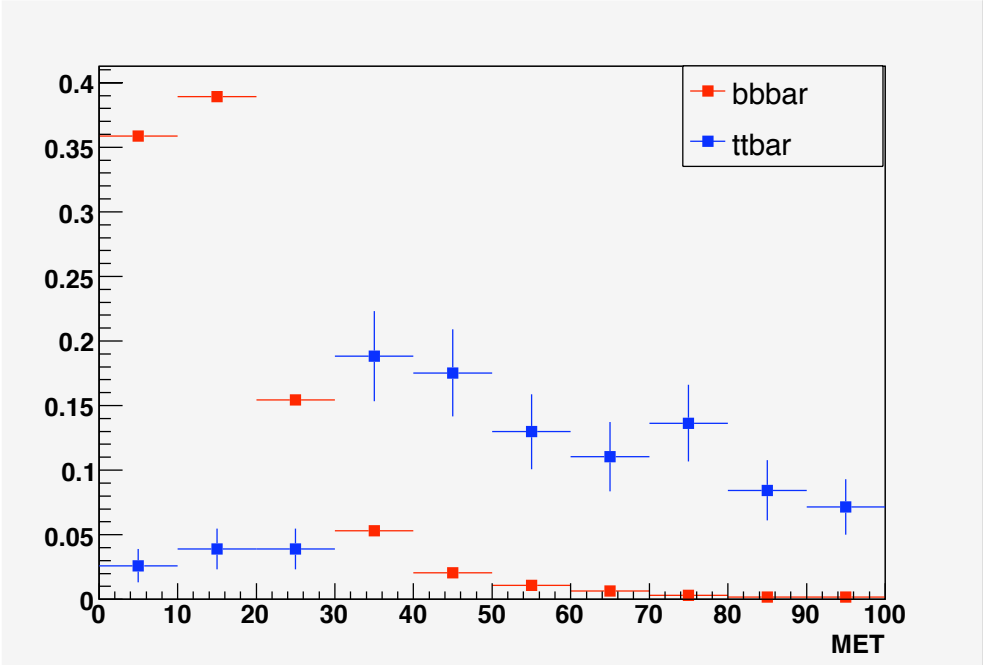


Figure 6.2: MET plot for b and t-jet events.

with high MET (> 60 GeV) are scarce ($<< 1$ %). Figure 6.2 shows a plot of MET for both $Z + b$-jet and $t\bar{t}$ events. Approximately, 99% of $Z + b$-jet events pass an MET > 60 GeV cut while the efficiency for $t\bar{t}$ events is 60% showing that an MET > 60 GeV cut will drastically reduce the $t\bar{t}$ background while preserving most of the signal.

Chapter 7

Multijet Background

The Z+jets event selection faces background contribution from multijet events. Some of these events are reconstructed with jets misidentified as leptons or with leptons from semi-leptonic decays of heavy flavor quarks appearing as prompt leptons. This instrumental multijet background is not well modeled by simulation, and so a representative sample is obtained from data instead. The conventional approach is to reverse the cuts used to reject multijet background, but the details vary depending on the channel. The selection and normalization of the multijet sample for the dielectron and dimuon sample is described below.

For the dimuon channel, the sample of multijet events is selected by reversing the muon isolation criteria, i.e. requiring that the product-scaled isolation of the two muons forming the Z candidate is > 0.03. With the exception of the isolation requirement, the multijet event must contain a Z candidate that passes all other event selection requirement. Thus, no isolated Z candidate is allowed in the event.

Similarly, for the dielectron channel, the multijet events are selected by reversing the electron shower shape requirements. The multijet selection criteria were chosen such that the jet topology is close to the selected electrons in the signal sample while still providing

sufficient multijet statistics. Accordingly, a multijet event has at least two EM objects with Isolation< 0.2, EMfraction> 0.9, HMx7> 35 in CC or HMx8> 20 in EC. All other selection criteria are applied with the exception of those using the NNout<N> variable. In addition, events with two or more electrons of type "Loose" have been discarded. The trigger prefers EM objects with low HMx values, biasing the the multijet sample. This bias is corrected by reweighting the p_T and η distributions in the sample to match those of an unbiased reference sample, as proposed in [70].

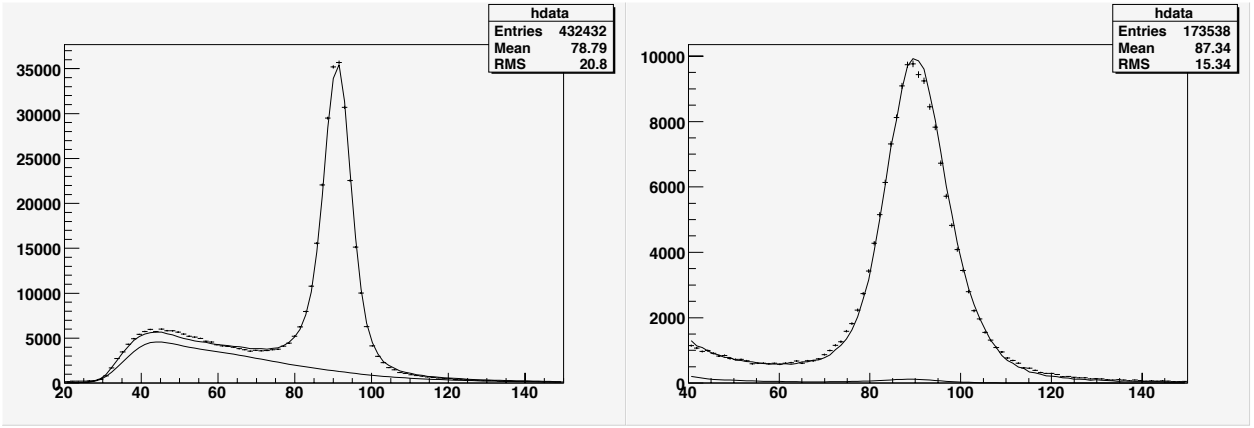
7.1 Normalizing to Data

The normalizations of the physics and multijet backgrounds are adjusted by scale factors in order to match the number of events in the data. To obtain these scale factors, the dilepton invariant mass $40 < M_{ll} < 200$ GeV distribution is used for the 0-jet, 1-jet and 2-jet inclusive bins. For each jet bin and channel, a value for α (the QCD scale factor) and β (the MC scale factor) are found that minimizes equation 7.1. Table 7.1 shows the results for each channel and jet bin, and figures 7.1-7.3 show the M_{ll} distributions for both data and αQCD + βMC

$$\chi^2 = \sum_{i=1}^{nbins} (\alpha S_i^{\mathrm{QCD}} + \beta(S_i^{\mathrm{MC}}) - D_i)^2 / D_i \tag{7.1}$$

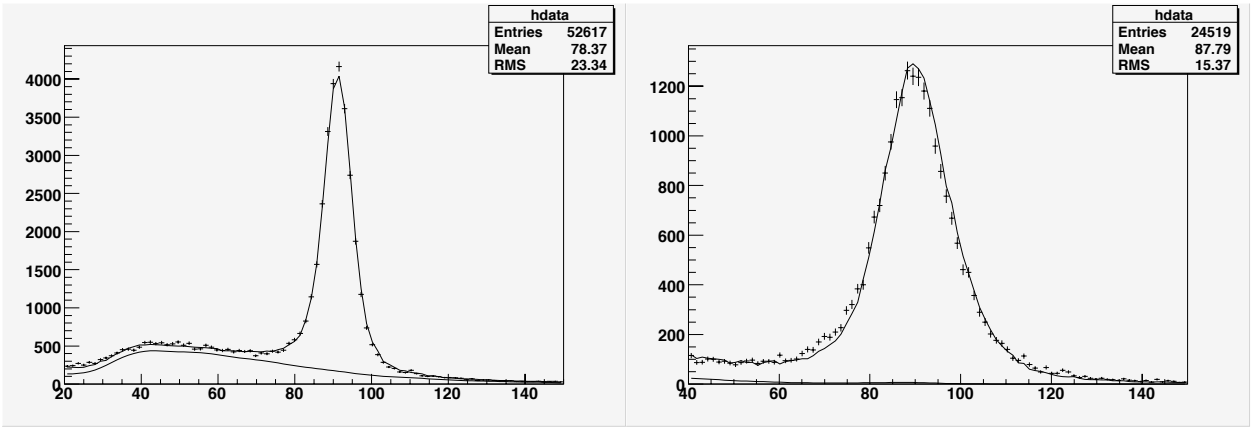
Table 7.1: Results of the multijet and background normalization in the dimuon sample.

Bin	Muon		Electron	
	α	β	α	β
0 Jet	1.2 ± 0.008	0.78 ± 0.002	0.59 ± 0.002	1.01 ± 0.002
1 Jet	0.178 ± 0.008	0.69 ± 0.002	0.59 ± 0.006	0.87 ± 0.006
2 Jet	0.016 ± 0.004	0.79 ± 0.009	0.51 ± 0.008	1.00 ± 0.012



(a) 0jet electron (b) 0jet muon

Figure 7.1: 0 jet data and αQCD + βMC distributions

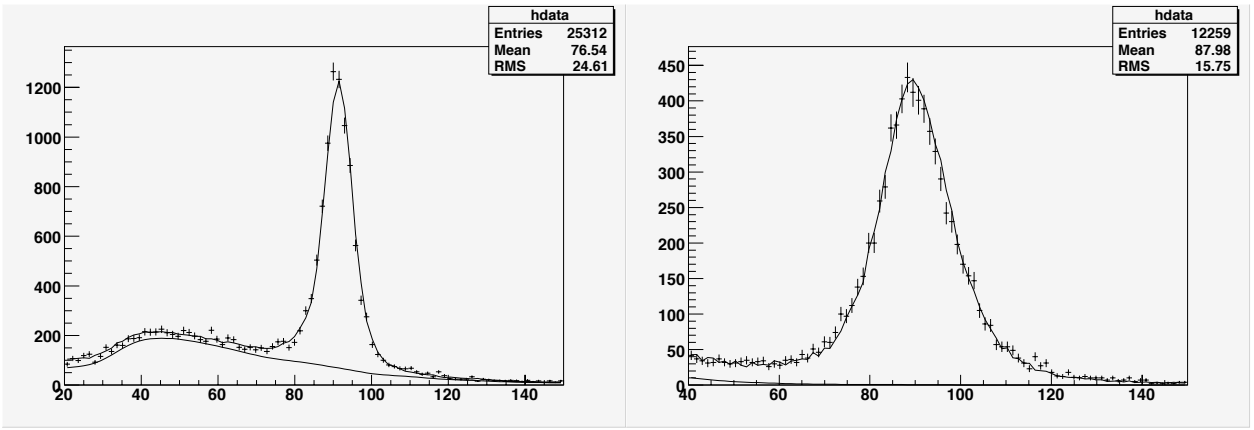


(a) 1jet electron (b) 1jet muon

Figure 7.2: 1 jet data and αQCD + βMC distributions

7.1.1 Multijet Background Contributions for Different Topologies in the ee Channel

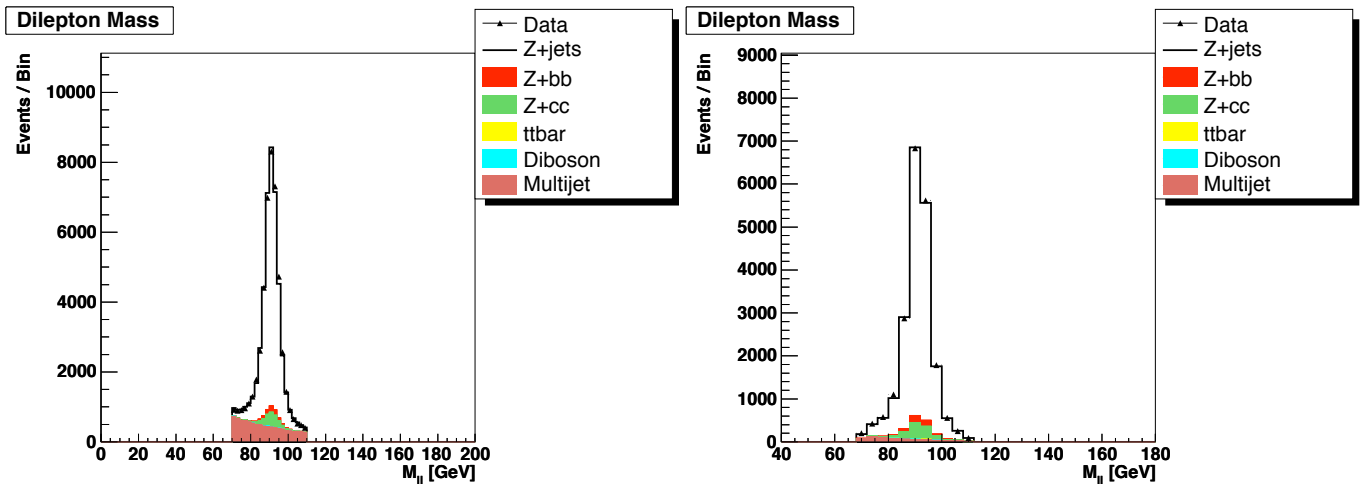
The events selected in the dielectron channel can be divided into three types based on the topology of the two electrons. Electrons can be reconstructed either in the central (CC) or endcap (EC) calorimeter and the different event topologies are: (i) CC-CC, with both CC electrons; (ii) CC-EC, with one CC and one EC electron; (iii) EC-EC, with both EC electrons. The multijet background fraction in the preselected sample of dielectron candidates with at least one jet is about 18%. However, the multijet background composition is different in the



(a) 2jet electron (b) 2jet muon

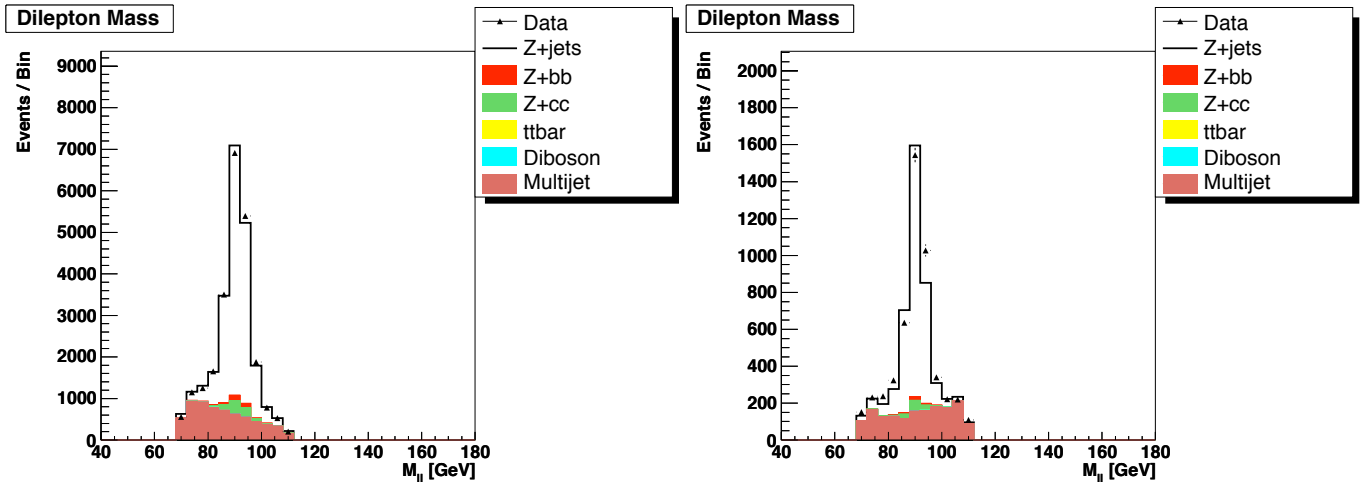
Figure 7.3: 2 jet data and αQCD + βMC distributions

three event topologies. The CC-CC events, which constitute about 41.5% of the total events, are much cleaner with a small background fraction of about 3.05% from multijets. The CC-EC events, which constitute about 48.5% of the total events, have a higher background fraction of about 26% from muljijets and the EC-EC events, which constitute the remaining 10% of the total events have the largest background contribution (about 32%) from muljitjets. Figures 7.4-7.5 show the dielectron invariant mass distributions in the complete Z+jets sample for each of the different topologies. One can clearly see that the multijet background is reasonably well modeled. The contributions from SM physics backgrounds is quite small with about 0.5% in the Z+jets sample.



(a) dielectron (b) dielectron (CCCC)

Figure 7.4: The dielectron invariant mass distributions in the complete Z+jets sample (a) and for the CC-CC (b) topology.



(a) dielectron (CCEC) (b) dielectron (ECEC)

Figure 7.5: The dielectron invariant mass distributions in the complete Z+jets sample for CC-EC(a) and for the EC-EC (b) topologies.

Chapter 8

B-tagging

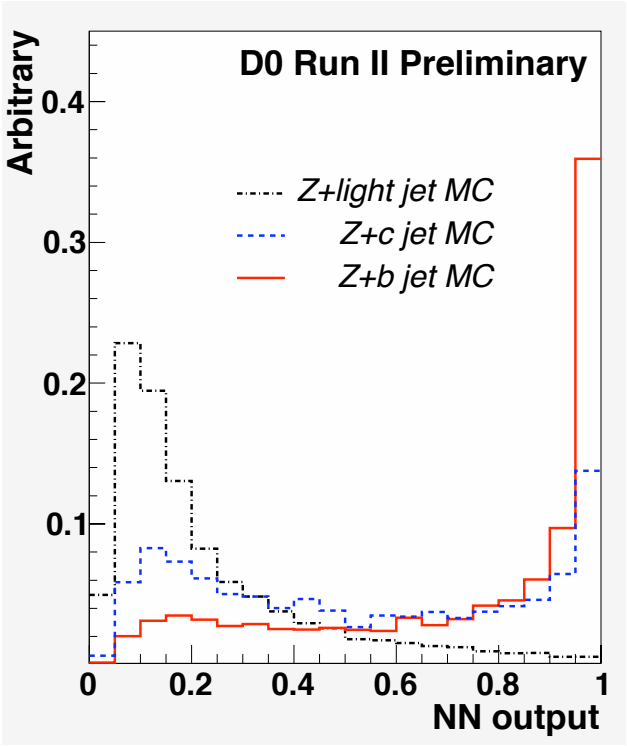
B-tagging is the process of selecting the jets that most likely came from b-quarks. This is done by exploiting the different characteristics that b-quarks have from the other quarks. B-hadrons typically take longer to decay than hadrons comprised of light quarks. This feature provides a signature for b-like jets. Thus, allowing a collection of generic jets to be separated into b-like and light-like jets.

8.1 The Neural Network Tagger

Neural Networks (NN) are designed to recognize correlations between different input variables. The NN tagger [57] takes a series of inputs, each aimed at separating b-jets from light-jets, and provides a continuous output with signal (b-like jets) being close to 1 and background (light-like jets) being close to 0. Figure 8.1 shows the output of the NN tagger for b-, c- and light-jets.

The NN tagger has nine inputs. Two of these inputs will be discussed in further detail, as they are of special importance to this analysis. The inputs are:

- S_{xy}: Decay length significance in the transverse plane with respect to the primary vertex



(a) NN output

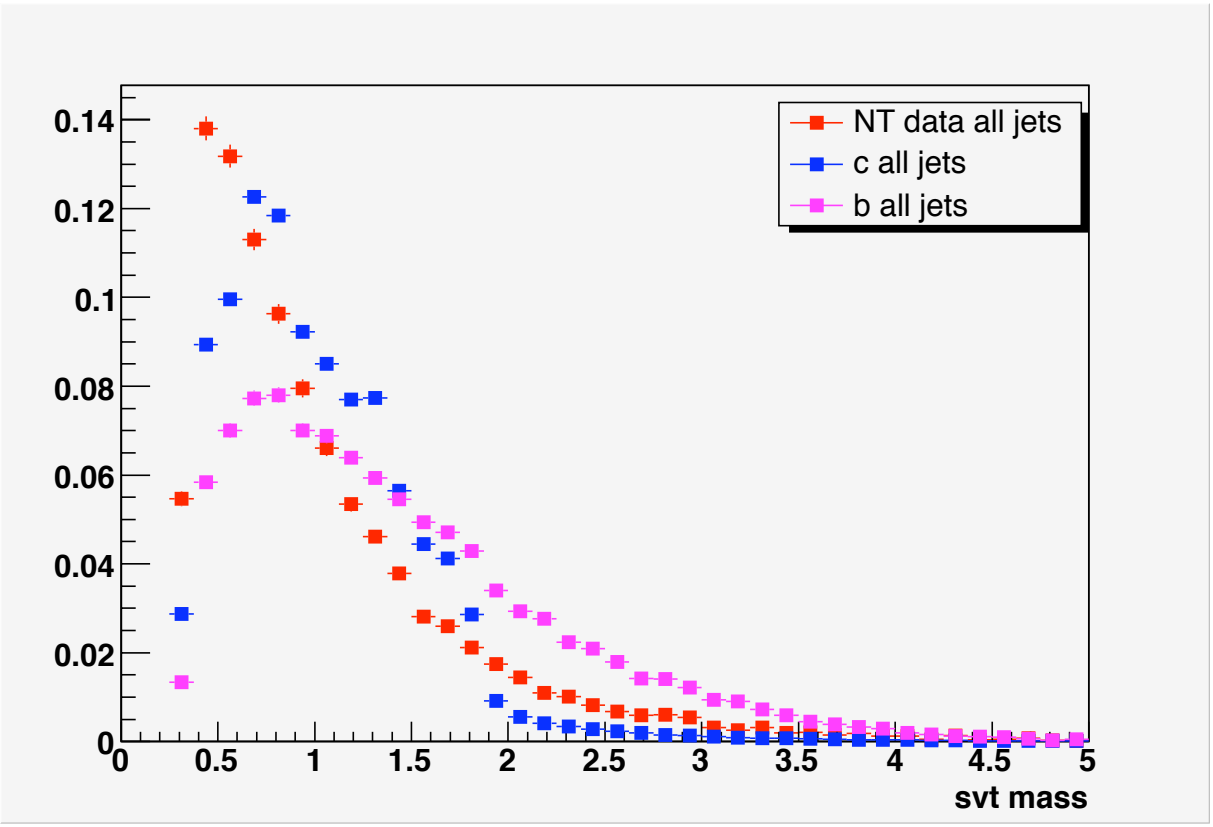
Figure 8.1: Distributions of b-jet Neural Net outputs for the b-, c- and light-jets.

- χ^2/ndf: χ^2 per number of degrees of freedom in the secondary vertex fit
- N_{trk}: The number of tracks used to reconstruct the secondary vertex
- M_{svt}: The invariant mass of the secondary vertex (M_{svt}) (This input will be discussed in further detail later in this chapter.)
- N_{svt}: Number of secondary vertices reconstructed
- ΔR: The difference in (η, ϕ) space between the jet axis and the resultant vector from the difference in the primary and secondary vertex.
- JLIP: The jet lifetime probability output (JLIP) (This input will also be discussed in greater detail later in this chapter.)
- rJLIP: The reduced jet lifetime probability output. This recalculates JLIP after removing the track that is least likely to have come from the primary vertex.

- N_{CSIP}: A combined variable based on the number of tracks with an impact parameter significance greater than an optimized value.

8.2 Secondary Vertex Mass

The secondary vertex mass (M_{svt}) is the invariant mass of all of the tracks associated to the most significant secondary vertex. This variable is dependent on the kinematics of the event. B-jets tend to have a higher M_{svt} value. Figure 8.2 shows the M_{svt} distributions for b-, c- and light-jets. M_{svt} proves to be a discriminant on its own, however, it does not utilize the geometry of an event.



(a) M_{svt}

Figure 8.2: M_{svt} distributions for b-, c- and light-jets.

8.3 Jet Lifetime Probability Tagger

The jet lifetime probability tagger [58] (JLIP) uses the impact parameter of the tracks associated to a calorimeter jet to compute the probability that that jet originated at the primary vertex. Figure 8.3 shows how the sign of the impact parameter is determined. A track within a jet is considered to have a positive IP if the extrapolation of that track crosses the jet axis before the PV. The sign is considered negative if the extrapolation of that track crosses the jet axis after passing the PV. All tracks within a jet with a positive impact parameter are assigned a probability, P, that it originated at the PV. JLIP is then calculated for the jet using equation 8.1. Figure 8.4 shows the $-ln(\text{JLIP})$ distributions for b-, c- and light-jets.

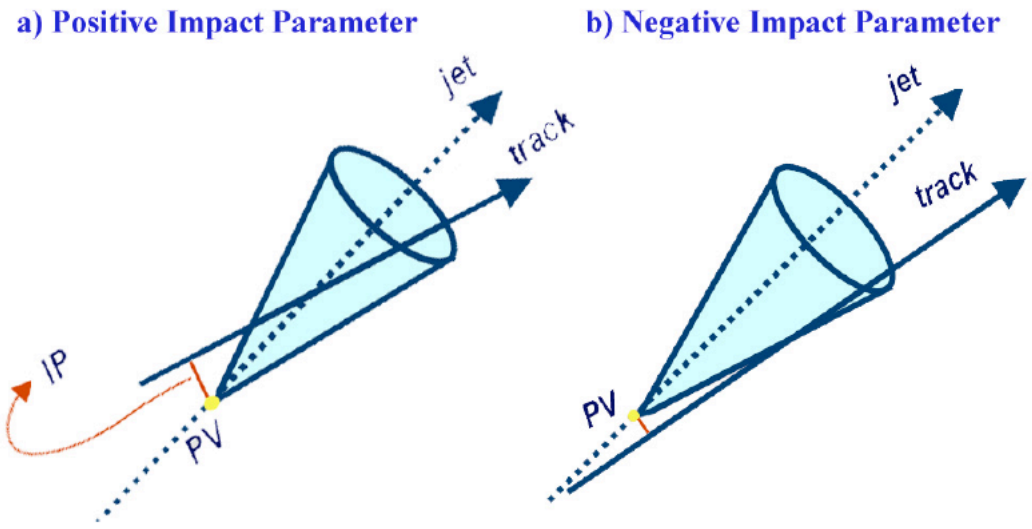


Figure 8.3: Determination of sign of IP.

$$JLIP = \prod_{i=0}^{N=Tracks} P^i \tag{8.1}$$

Much like the M_{svt} variable, JLIP provides some discriminating power on its own. However, the JLIP variable is only dependent on the event geometry. A combination of both the M_{svt} and JLIP taggers would provide a discriminant that utilized both the event geometry

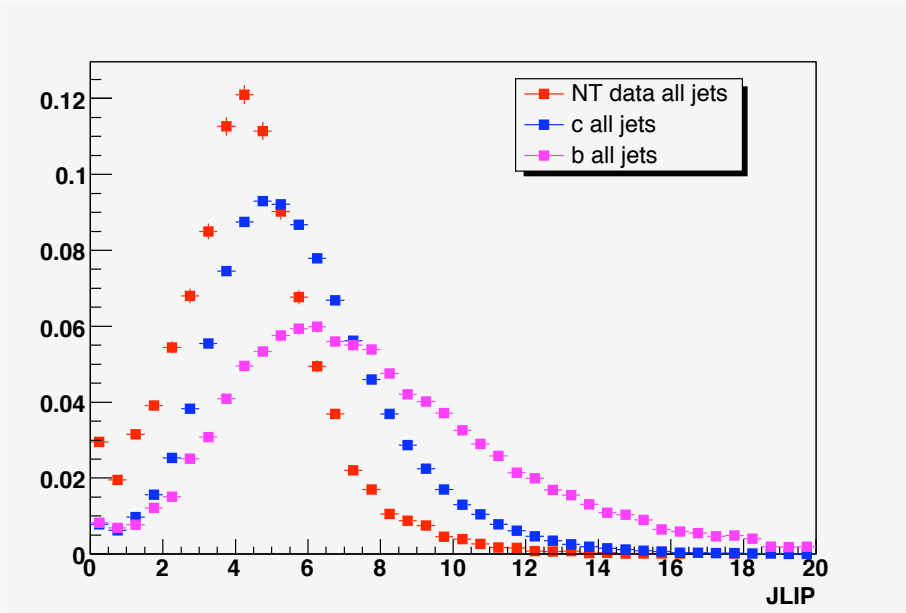


Figure 8.4: JLIP distributions for b-, c- and light-jets.

and kinematics. This method of combining both taggers into a new tagger is discussed in Chapter 11.

8.4 Negative Tagged Jets

A negative tagged jet[57] is a collection of tracks within a jet that have negative values for the NN inputs (those that can be negative). Figure 8.3 shows how a track can have a negative IP, and figure 8.5 shows how a track can have a negative decay length. These negative values are often due to resolution effects. The negative tagged jets themselves typically act in a similar fashion to light jets. Thus, it is often advantageous to use negative tagged jets in data in place of MC light jets. This replacement is further discussed in Chapter 11 as well as in Appendix E.

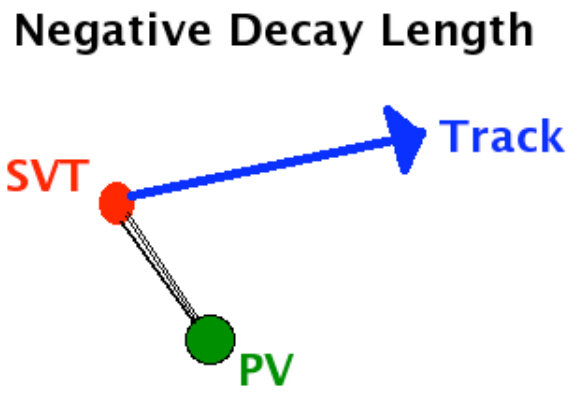


Figure 8.5: Negative decay length.

8.5 Determination of NN b-Tagging Efficiency and Data to MC Scale Factor

To find the NN b-tagging efficiency in data, a system of eight equations with eight unknowns is used. This system uses two different tags on two different samples to formulate the eight equations. The solutions to the equations include the efficiencies to both taggers. The two data samples used were a muon-in-jet sample and a subset of the muon-in-jet sample, where an away jet passes an NN cut, thus enriching the sample in b-jets. The two taggers used were the NN tagger and an SLT tagger defined as a muonic jet, where the muon-in-jet passes a $p_{TRel} > 0.5$ GeV cut with respect to the muon+jet axis. That is, the momentum of the muon perpendicular to the muon+jet axis must be greater than 0.5 GeV.

In the system of equations below, α, β, κ_b, κ_{udsc} and P_{TRel} are all coefficients related to the correlation of the two taggers. These coefficients were measured in MC and assumed to be the same in data. The system of eight equations is as follows:

$$n = n_b + n_{udsgc} \tag{8.2}$$

$$p = p_b + p_{udsgc} \tag{8.3}$$

$$n^{SLT} = \epsilon_b^{SLT} n_b + \epsilon_{udsgc}^{SLT} n_{udsgc} \tag{8.4}$$

$$p^{SLT} = \epsilon_b^{SLT} p_b + \epsilon_{udsgc}^{SLT} p_{udsgc} \tag{8.5}$$

$$n^{NN} = \epsilon_b^{NN} n_b + \epsilon_{udsgc}^{NN} n_{udsgc} \tag{8.6}$$

$$p^{NN} = \beta\epsilon_b^{NN} p_b + \alpha\epsilon_{udsgc}^{NN} p_{udsgc} \tag{8.7}$$

$$n^{SLT,NN} = \kappa_b \epsilon_b^{SLT} \epsilon_b^{NN} n_b + \kappa_{udsgc} \epsilon_{udsgc}^{SLT} \epsilon_{udsgc}^{NN} n_{udsgc} \tag{8.8}$$

$$p^{SLT,NN} = \kappa_b \beta \epsilon_b^{SLT} \epsilon_b^{NN} p_b + \kappa_{udsgc} \alpha \epsilon_{udsgc}^{SLT} \epsilon_{udsgc}^{NN} p_{udsgc} \tag{8.9}$$

where n^{Tag} is the number of events in the muon-in-jet sample after the "Tag" is applied and p^{Tag} is the number of events in the away tagged muon-in-jet sample after the Tag is applied. The eight unknowns are: the number of b-jets in the muon-in-jet sample n_b, the number of light jets in the muon-in-jet sample n_{udsgc}, the number of b-jets in the tagged muon-in-jet sample p_b, the number of light jets in the tagged muon-in-jet sample p_{udsgc}, the b-efficiency of the NN tagger ϵ_b^{NN} the mistag rate of the NN tagger ϵ_{udsgc}^{NN}, the b-efficiency of the SLT tagger ϵ_b^{SLT} and the mistag rate of the SLT tagger ϵ_{udsgc}^{SLT}.

The efficiencies are jet η and p_T dependent and vary for MC and data. To determine data to MC scale factors, a ratio of the efficiency in data over the efficiency in MC in both jet η and p_T is taken. In order to find the efficiency in MC, system 8 is run over a collection of MC samples ($Z \to b\bar{b}$, $Z \to c\bar{c}$, $Z \to q\bar{q}$ and $t\bar{t}$). The data to MC scale factors are applied to an MC jet in order to get the correct tagging efficiency for a jet in data with the same η and p_T. The application of these scale factors is discussed in chapter 11.

Chapter 9

Corrections to Monte Carlo

In an ideal world, simulations of SM events and detector performance would reproduce data perfectly. However, the complexity of nature does not allow for such agreement. In order to account for the discrepancies between data and simulation, the following corrections are applied.

9.1 Z+jets Cross Sections

The inclusive Z cross-sections determined by ALPGEN are leading log (LL) calculations and have been scaled to the next to next to leading order (NNLO) inclusive Z calculations [71]. Because this scale factor is not the ratio of a Leading Order to NNLO cross section, it is not a true k-factor. Instead, it is referred to as a k'-factor. The calculated ratio between the NNLO inclusive Z cross section to ALPGEN LL inclusive Z cross section is

$$k' = 1.30 \tag{9.1}$$

This factor is then used to scale all of the ALPGEN Z + light jets samples and an error of 10% is quoted due to variations of factorization scale, PDFs and generator cuts. Using MCFM [68], a k-factor (NLO/LO) for $Z + b\bar{b}$ and $Z + c\bar{c}$ can be calculated. Taking the ratio of the

MCFM k-factor for Z+heavy flavor versus the MCFM k-factor for Z + light jets gives an HF-factor. The ALPGEN Z+heavy flavor cross-sections are scaled by this additional factor for a total scaling of

$$k' * HF_{b\bar{b}} = 1.30 * 1.52 = 1.96 \tag{9.2}$$

$$k' * HF_{c\bar{c}} = 1.30 * 1.67 = 2.15 \tag{9.3}$$

The uncertainty on the K' factor for Z+light jet is about 6.3% and 20% for $Z + HF$ jet.

9.2 Cross Sections for background Processes

The diboson and $t\bar{t}$ cross sections are corrected by k-factors correcting LO PYTHIA and ALPGEN cross sections to MCFM [68] NLO cross sections. The calculations use CTEQ6.1M [46] PDFs. The resulting k-factors are as follows:

$$k(ZZ) = 1.030 \tag{9.4}$$

$$k(WZ) = 1.062 \tag{9.5}$$

$$k(WW) = 1.005 \tag{9.6}$$

$$k(t\bar{t}) = 1.434 \tag{9.7}$$

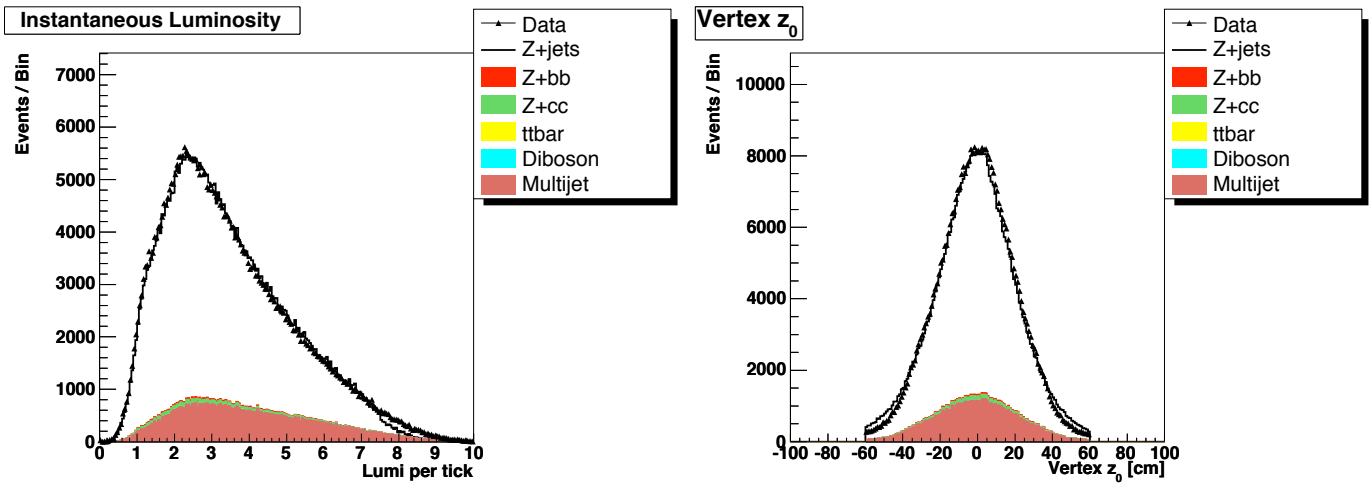
9.3 Luminosity Reweighting

In order to provide a realistic simulation of the detector's response to beam conditions, an actual data event collected using minimum-bias triggers is used to define the baseline detector response for each Monte Carlo event. The instantaneous luminosity for the minimum-bias overlay does not match the luminosity profile of the data sample, so a DØ standard luminosity correction is applied. Figure 9.1a shows the data overlayed on the reweighted instantaneous

luminosity in the MC.

9.4 Primary Vertex Reweighting

A standard reweighting for the primary vertex z-position is also applied. This is to account for the fact that primary vertex z-distributions in MC are Gaussian, while they are not in data. For this correction, the MC PV z-distribution is reweighted to match the data. The primary vertex z-distribution in data and reweighted MC is shown in figure 9.1b.



(a) Instantaneous Luminosity

(b) Z Vertex

Figure 9.1: (a) The instantaneous luminosity distributions of data and MC samples for Z inclusive sample is shown in the left plot. The MC has been reweighted so that the minimum bias overlay in the generated samples mirrors the instantaneous luminosity of the data sample. (b) The primary vertex Z distribution for data and MC samples.

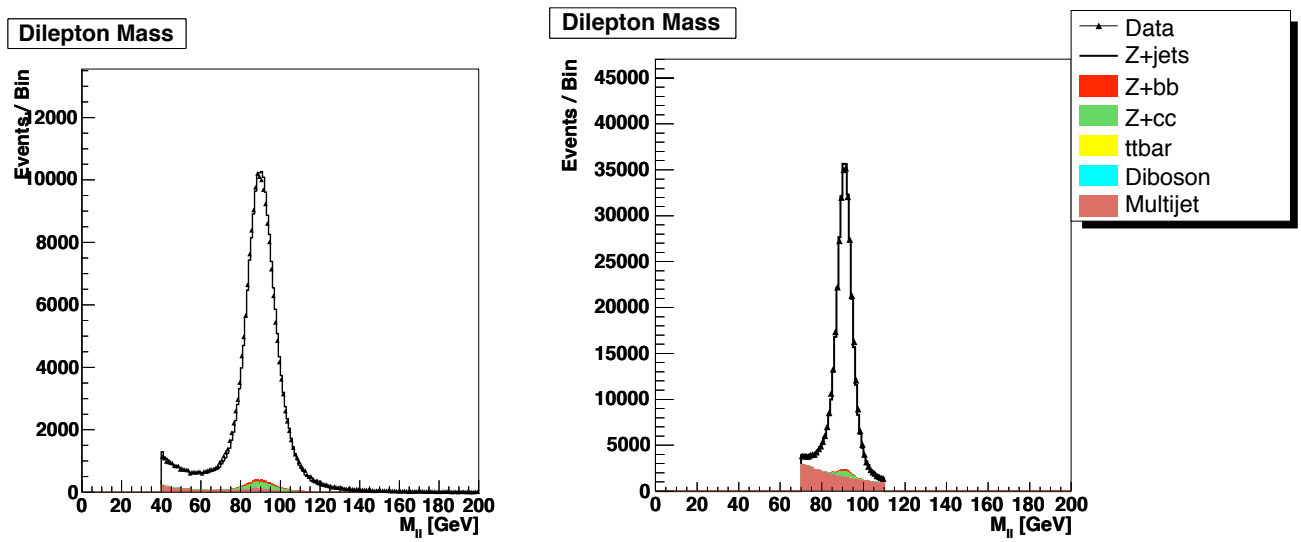
9.5 Trigger Corrections

The inclusive trigger has been found to be nearly 100% efficient for the dielectron channel, so no trigger correction is applied. For the muon channel, an "OR" of the single muon trigger is applied, leading to an efficiency of $\sim$ 80%. This efficiency is applied to the muon MC events as an event weight.

9.6 Lepton Energy and Identification Corrections

Monte Carlo simulations produce leptons with slightly better energy resolutions and identification efficiencies than exist in the data. To account for the energy resolution, DØ uses $Z \to ee$ and $Z \to \mu\mu$ samples to figure out how much to smear the MC lepton energies in order to match data. For the identification efficiency correction, a tag-and-probe method is used to measure the efficiency in both data and MC simulations. From these efficiencies a data-to-MC scale factor can be determined. This scale factor is then applied to all MC leptons. The advantage of using the data-to-MC scale factor is that it has the same effect as randomly moving events at a certain rate, however it preserves the statistics of those events.

Figure 9.2 shows the Z boson mass in the two dilepton channels.



(a) dimuon mass

(b) dielectron mass

Figure 9.2: Reconstruction of Z mass peak in (a) muon and (b) electron channels.

9.7 Z p_T Reweighting

The Z p_T distribution is poorly modeled by both the PYTHIA and ALPGEN MC generators for events with Z p_T<$\sim$100 GeV. The discrepancy between data and simulation is corrected

in MC by reweighting the Z-p_T distribution to match the data. The correction is derived from the p_T distribution at the generator level and the observed spectrum in the unfolded data [72]. In this analysis, the parametrized correction for the Z-p_T derived from ALPGEN MC is used. In addition, a reweighting parametrization is used for events with 0, 1 and 2 or more jets separately for a jet E_{T} threshold of 15 GeV. After the reweighting, there is good agreement between data and MC in the 0-jet, 1-jet and 2-jet multiplicities in the dielectron and dimuon samples. The transverse momentum of the Z candidates in the jet multiplicity samples are shown in figures 9.3–9.6.

9.8 Jet Shifting Smearing and Removal (JSSR)

The modeling of hadronization and detector response and readout are not trivial, so it is not surprising that simulated jets do not precisely match jets in data. After reconstruction and the application of the JES correction (discussed in Section 3.4), simulated jets have slightly higher energies, better energy resolution and more efficient reconstruction and identification than those in data. The differences between data and MC jets are measured by the DØ jet identification group using γ+jet events [61]. For various ranges of photon p_T (p_T^γ), histograms are filled with the fractional difference in p_T between the photon and the jet:

$$\Delta S = \frac{p_T^{\mathrm{jet}} - p_T^\gamma}{p_T^\gamma}. \tag{9.8}$$

For high p_T^γ bins (and therefore high p_T^{jet}), the distribution of ΔS is Gaussian with a mean around zero. For low p_T ($\lesssim$ 40 GeV) the distribution would be Gaussian except that the low tail that is truncated due to inefficiencies in reconstructing and identifying low p_T jets. The truncation is described by an error function as shown in figure 9.7 for the range 18 GeV $< p_T^\gamma <$ 23 GeV. The ΔS distributions are fit to a Gaussian times an error function. The difference between the data and MC Gaussian component of the ΔS distributions gives the relative data-MC JES, the difference in widths of the Gaussian component gives the

resolution difference, and the difference in the error function component gives the relative data-MC jet ID efficiency. The MC samples are then corrected by shifting and smearing the jet energies according to the relative data-MC JES and jet resolution and by randomly removing jets at a rate proportional to the ratio of the jet ID efficiency in data versus simulation.

9.9 Vertex Confirmation for Data and MC

The efficiency of vertex confirmation is different for data and MC. The CALGO group at DØ produced certified scale factors for the vertex confirmed jet cut. The scale factors were derived in both γ+ jet and Z+jet samples. A dijet sample was used for cross checks and for measuring systematic errors. This reweighting is measured in η bins of 0.4 for $|\eta| < 2.6$. The scale factor is parameterized in terms of $z_{PV} \times Sign(\eta_{jet})$ in 20 cm bins. For jets in the most central η bin ($0 < |\eta| < 0.4$) a dependence on jet p_T is found for jets with p_T < 40 GeV , while the scale factor is flat for $p_T > 40$ GeV. Therefore, the scale factor for jets in this central region with jet $p_T < 40$ GeV is also parameterized in terms of jet p_T. This reweighting is applied to all MC samples in this analysis and is processed using random removal of jets from the vertex confirmed jet branch.

A study of the fraction of reconstructed jets matching a corresponding particle jet within a cone size of $\Delta R = 0.5$ for b- and light jets was conducted and found that about 96% of the reconstructed b-jets and about 95% of the reconstructed light jets matched particle jets. The question of whether this fraction in b- and light jets has some dependence on the primary vertex multiplicity in the event was also investigated and the results are listed in table 9.1. One can see that, the difference between b- and light jets is small and that the relative change between b- and light jets is even smaller. This shows that the relative effect is not a cause for concern.

Number of PV.	b jets	light jets
1	0.9618	0.9602
2	0.9596	0.9524
3	0.9585	0.9471
> 3	0.9529	0.9375

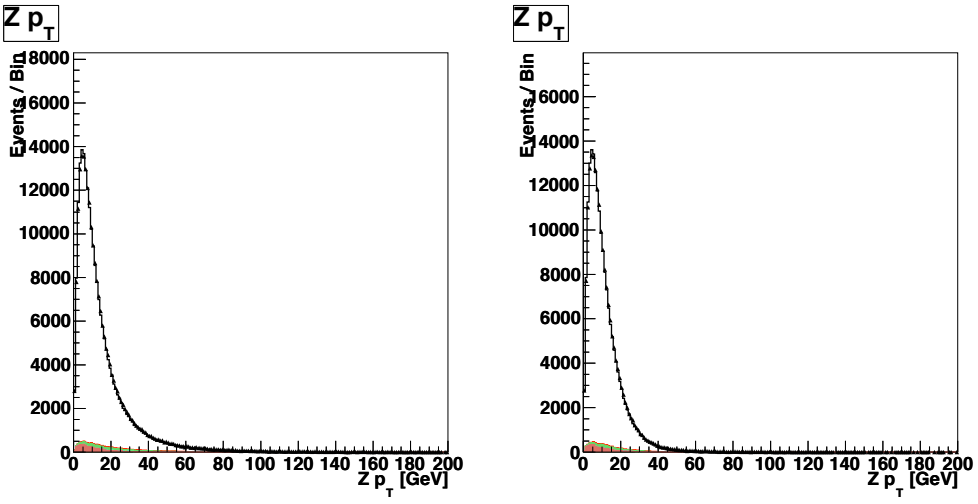
Table 9.1: The reconstructed jet matching efficiency with particle jets in $Z + b$ jets and Z+light jets as a function of the number of primary vertices in the event.

9.10 Scale Factors for b-tagging

The scale factors provided by the b-id at DØ group are calculated with a vertex cut in the z-direction of < 60 cm. For this reason, a z vertex cut cannot be applied. Figure 9.8 displays the jet p_T acceptance of a z vertex cut with respect to p_T for both b- and light jets as calculated using equation 9.9 (with all other event cuts applied).

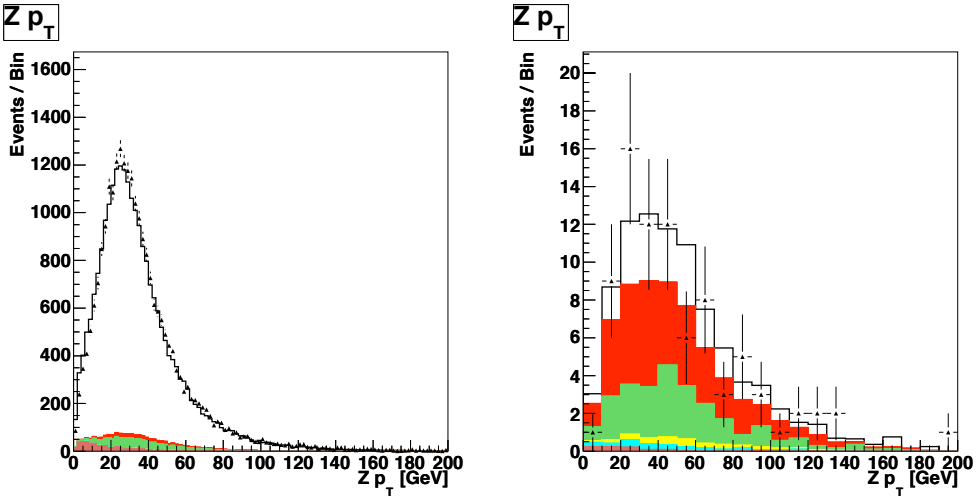
$$A = \frac{p_T^{jet,|zvtx|<40}}{p_T^{jet}} \tag{9.9}$$

Figure 9.8 shows that there is no p_T difference between jets that pass a z vertex cut of < 40 cm and jets that do not. This means that because the weighted average in p_T for the b-tagging and MJL efficiencies is used, there is no z vertex dependence.



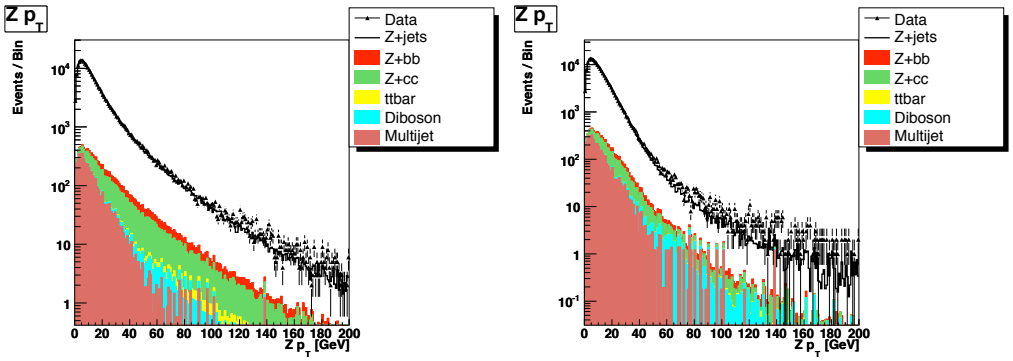
(a) Zp_T in inclusive bin (b) Zp_T in 0 jet bin

Figure 9.3: The transverse momentum of the dimuon Z candidates for (a) 0 jet inclusive bin and (b) 0 jet exclusive bin.



(a) Zp_T in 1 jet exclusive bin (b) Zp_T in 2 jet inclusive bin

Figure 9.4: The transverse momentum of dimuon Z candidates for the (a) 1 jet bin and (b) 2 jet.



(a) Zp_T in inclusive bin (b) Zp_T in 0 jet bin

Figure 9.5: The log scale transverse momentum of the dimuon Z candidates for (a) 0 jet inclusive bin and (b) 0 jet exclusive bin.

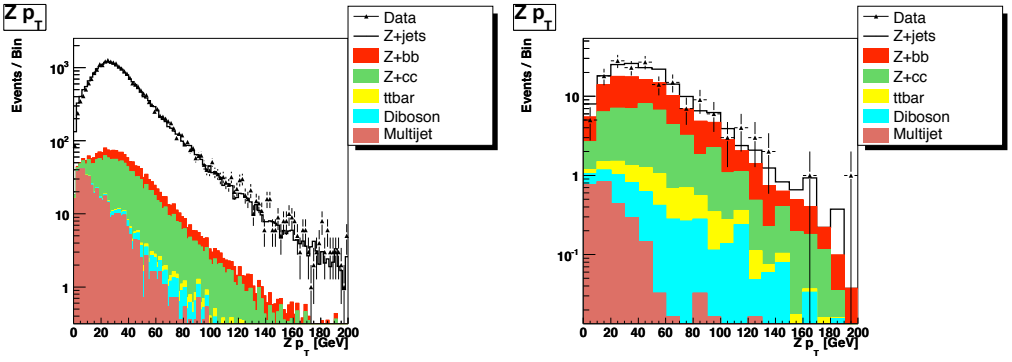


Figure 9.6: The log scale transverse momentum of dimuon Z candidates for the (a) 1 jet bin and (b) 2 jet bin.

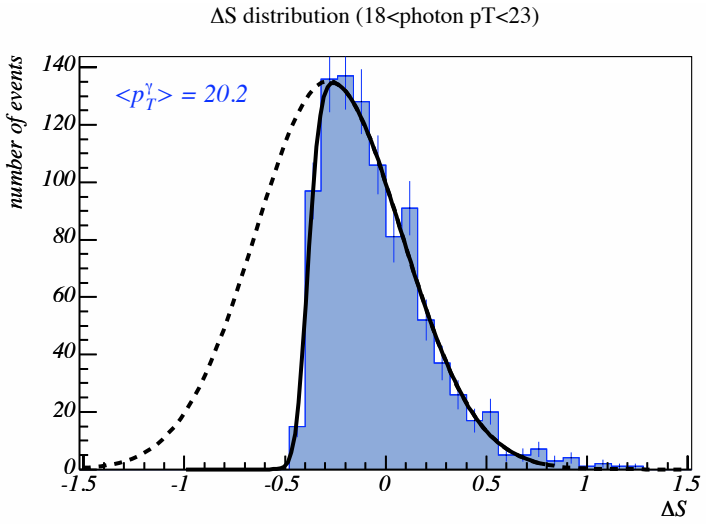


Figure 9.7: Distribution of ΔS for CC jets in data and 18 GeV $< p_T^\gamma <$ 23 GeV, fit to a Gaussian times an error function.

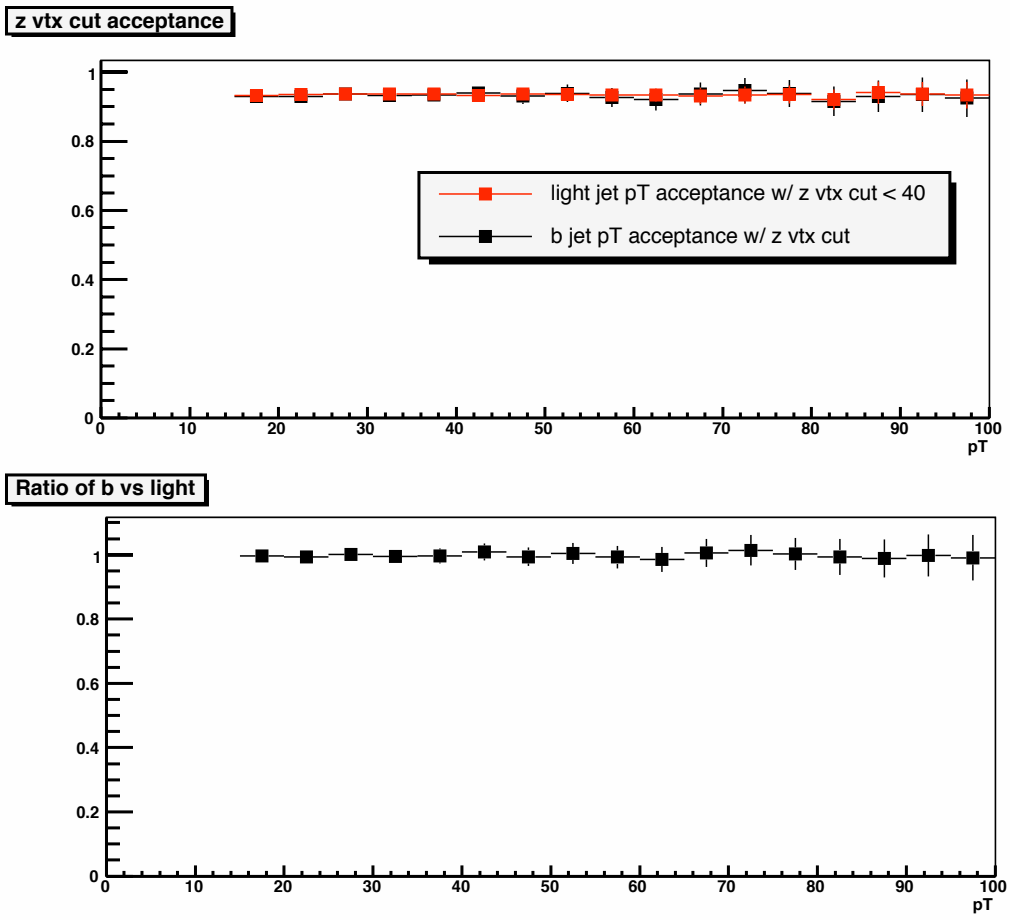


Figure 9.8: Top: b and light z vertex acceptance with respect to p_T. Bottom: ratio of b and light acceptances

Chapter 10

Z+jet Selection

After determining the amount of MC and multijet events, and applying corrections to MC, a comparison can be made to the data. Table 10.1 shows the number of events from the data, the various background components and the expected Z+jet events (events with a Z boson associated with at least one jet) at the preselection level for both dimuon and dielectron channels. The errors are due to the statistics in the MC samples and are not at all correlated to the uncertainty of their cross sections.

Figures 10.1 - 10.14 show the preselection plots for various variables in both the dimuon and dielectron channels. Good agreement is seen in these figures.

10.1 Dimuon Channel

Kinematic distributions for the dimuon Z candidate + $\geq$ 1-jet selection before any b-tagging are shown in figures 10.1-10.7.

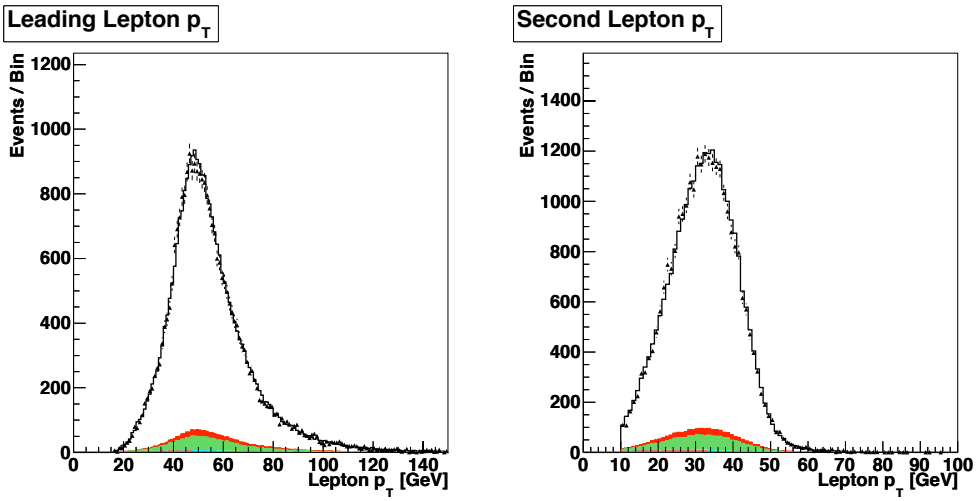


Figure 10.1: The leading and second electron p_T spectrum in $Z+ \geq 1$ jets sample before any b-tagging is applied.

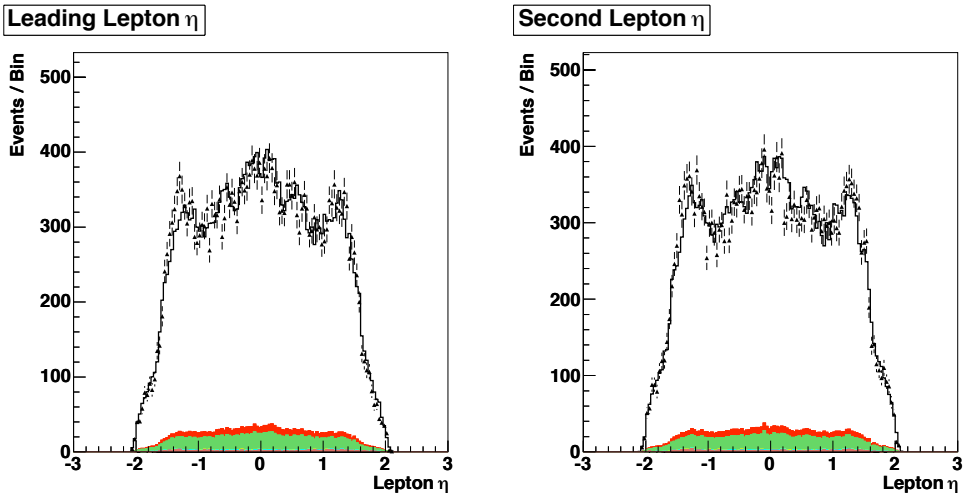


Figure 10.2: The leading and second electron η spectrum in $Z+ \geq 1$ jets sample before any b-tagging is applied.

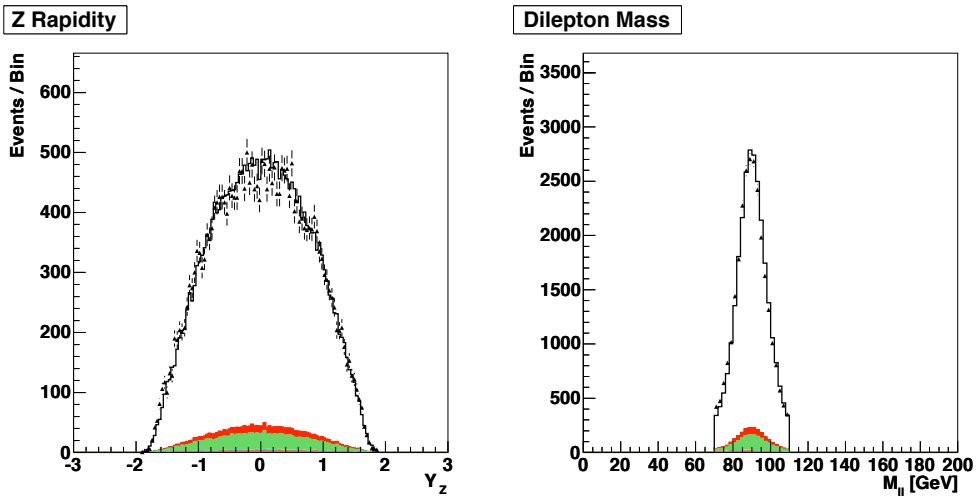


Figure 10.3: The rapidity and mass of the Z candidate in the $Z+ \geq 1$ jets sample before any b-tagging is applied.

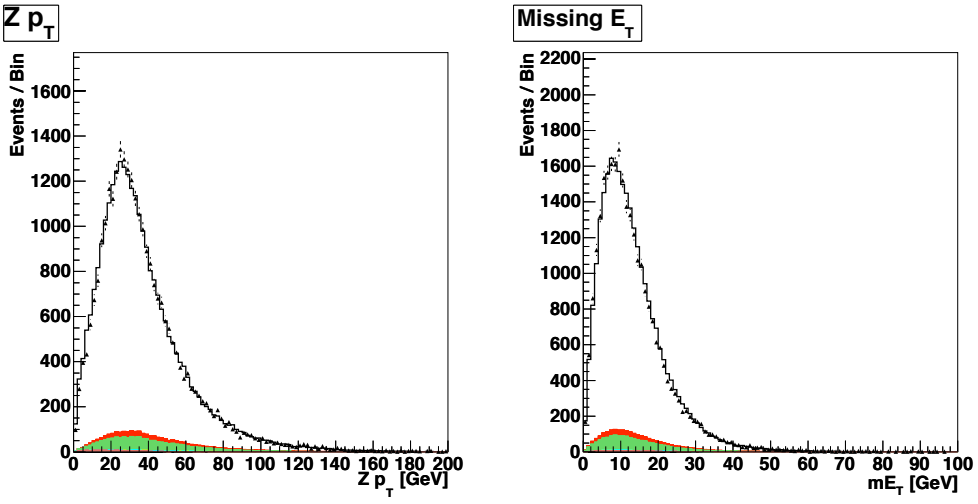


Figure 10.4: The p_T Z candidate and missing transverse energy in the $Z+ \geq 1$ jets sample before any b-tagging is applied.

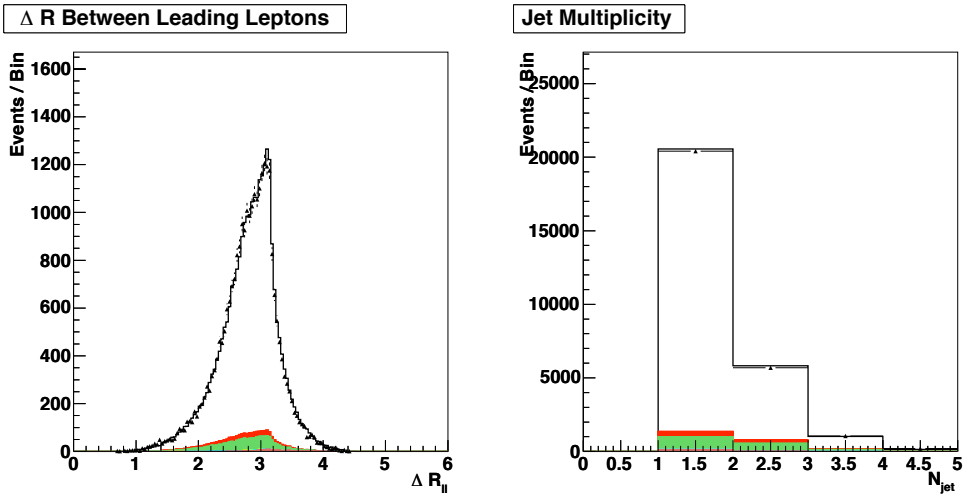
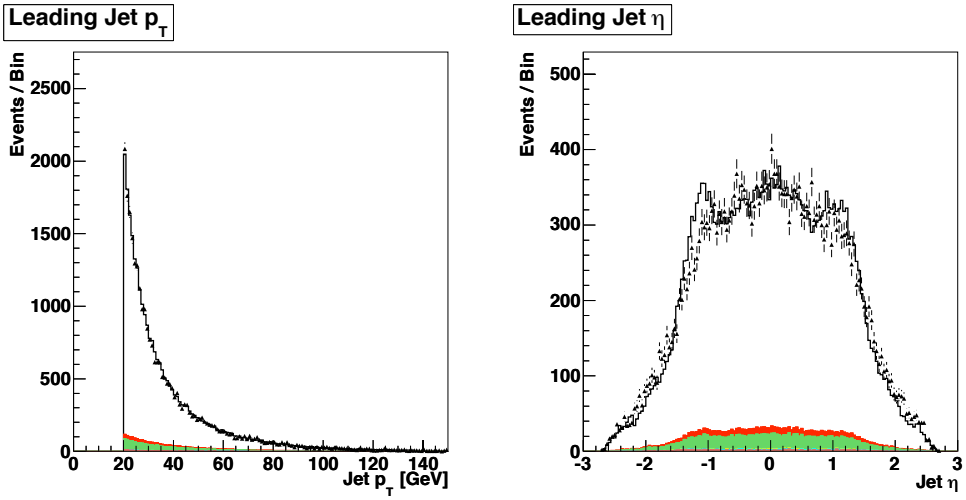


Figure 10.5: The $\Delta R(\mu\mu)$ for the Z candidate and jet multiplicity in the $Z+ \geq 1$ jets sample before any b-tagging is applied.



(a) Leading jet p_T (b) Leading jet η

Figure 10.6: The leading jet (a) p_T and (b) η for data and background in the $Z+ \geq 1$ jets sample before any b-tagging is applied.

RunIIb dimuon

sample	Dimuon	Dielectron
data	24,450	48956
Expected	$24,924.1 \pm 67.0$	49005.2 ± 153.5
$Z + j$	$22,005 \pm 67.5$	37063 ± 131.4
$Z + b$	580.9 ± 3.7	886.9 ± 6.2
$Z + c$	1315.8 ± 7.7	2054.0 ± 13.1
Multijet	53.1 ± 1.6	8748.9 ± 78.0
ZZ	$60. \pm 0.8$	96.9 ± 0.7
WZ	64.3 ± 1.1	126.3 ± 1.2
WW	6.1 ± 0.5	13.2 ± 0.7
$t\bar{t}$	11.1 ± 0.1	16.0 ± 0.2

Table 10.1: The data, backgrounds, and expected Z+jets events are listed in the preselected event sample.

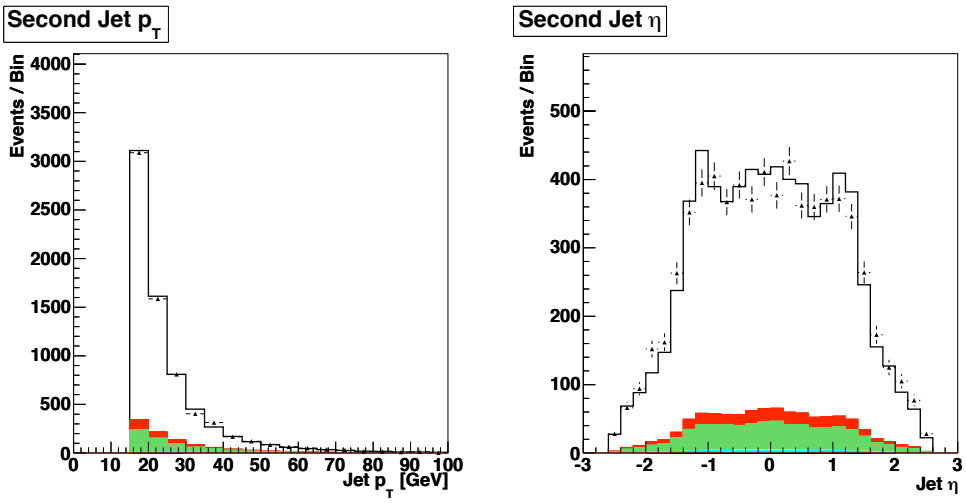
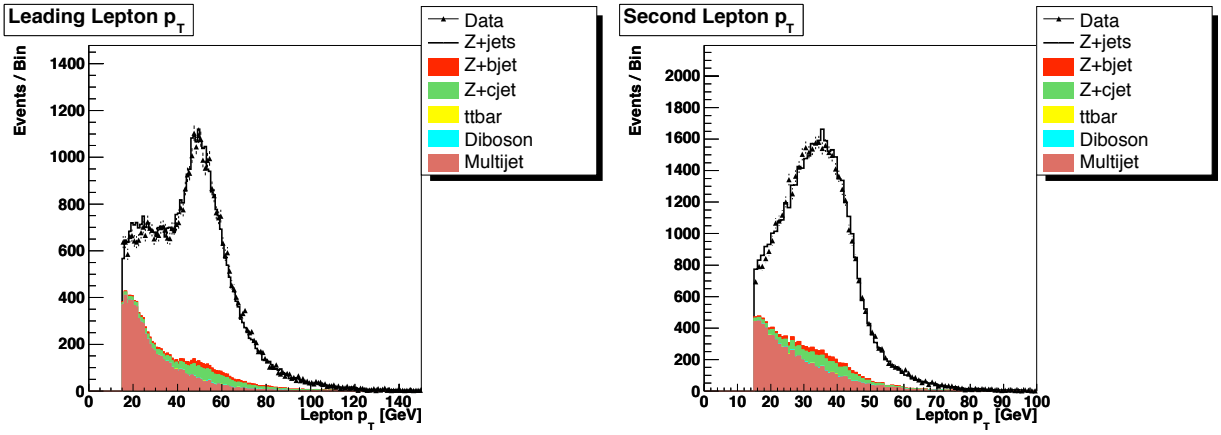


Figure 10.7: The second leading jet (a) p_T and (b) η for data and background in the $Z+ \geq 1$ jets sample before any b-tagging is applied.

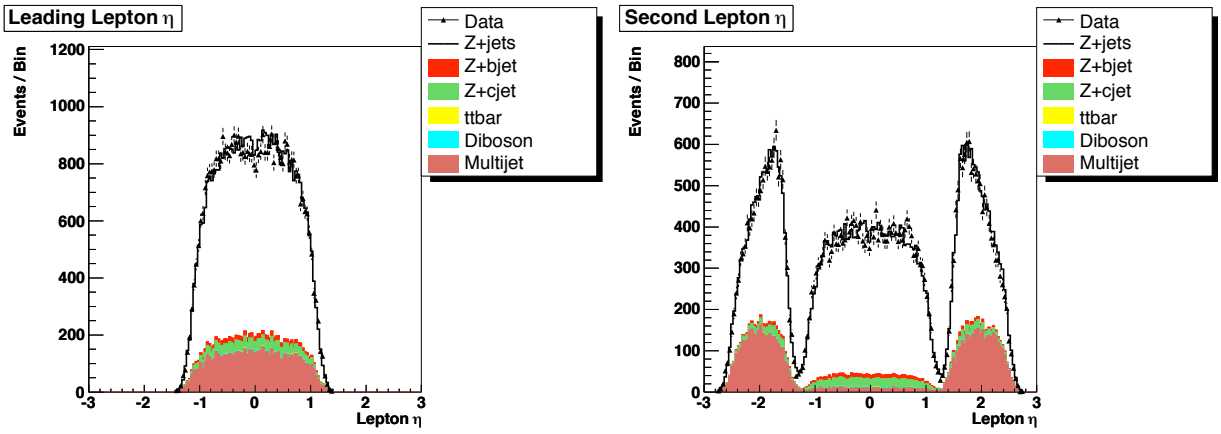
10.2 Dielectron Channel

Kinematic distributions for the dielectron Z candidate + $\geq$ 1-jet selection before any b-tagging are shown here (figures 10.8 - 10.14).



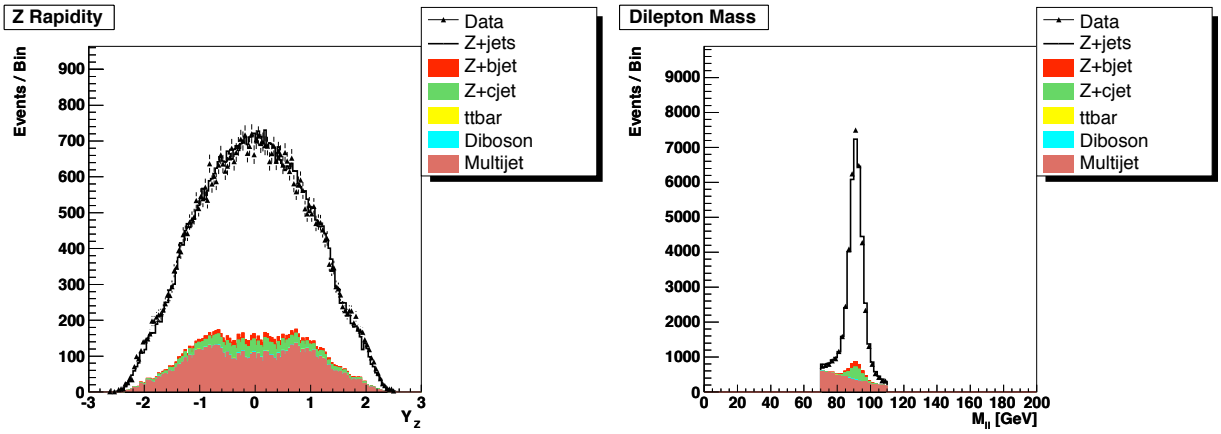
(a) RunIIb dielectron

Figure 10.8: The leading and second electron p_T spectrum in $Z+ \geq 1$ jets sample before any b-tagging is applied.



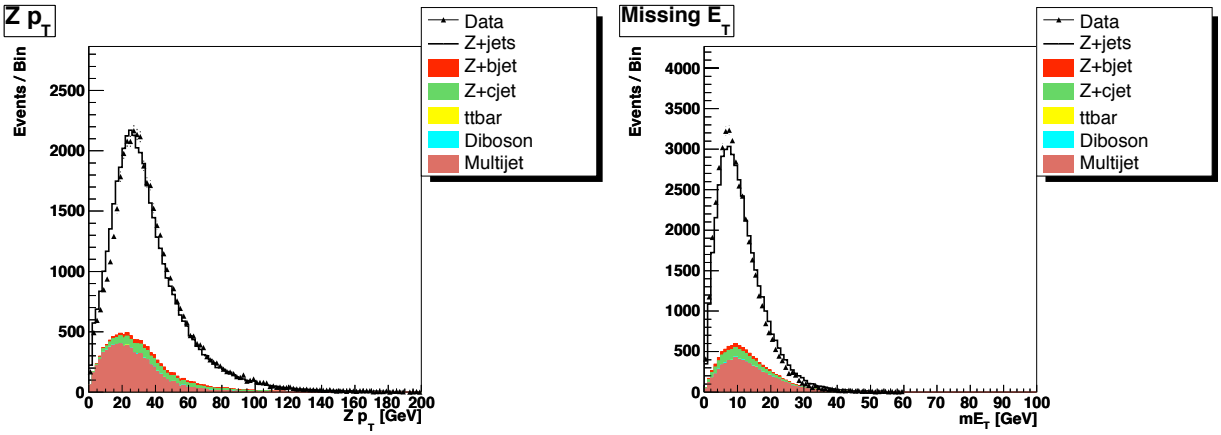
(a) RunIIb dielectron

Figure 10.9: The leading and second electron η spectrum in $Z+ \geq 1$ jets sample before any b-tagging is applied.



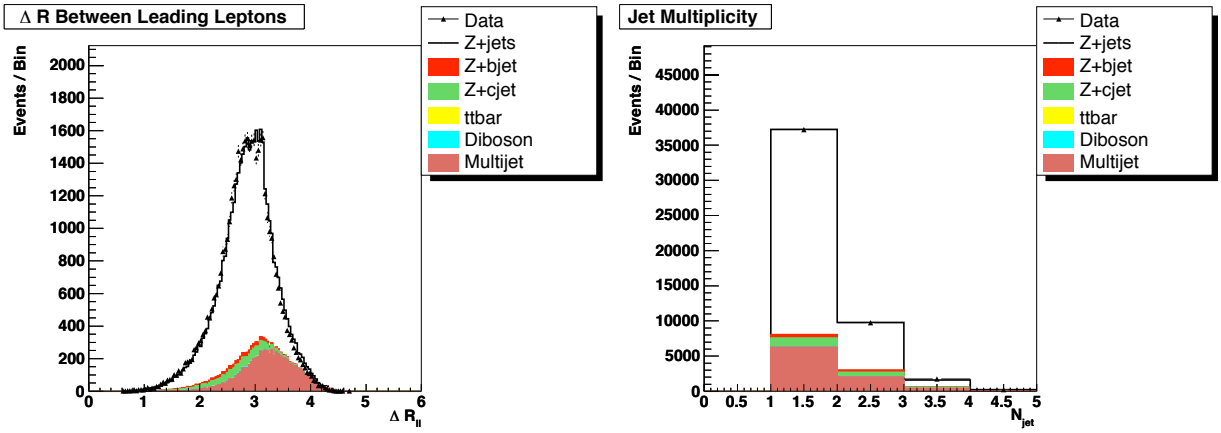
(a) RunIIb dielectron

Figure 10.10: The rapidity and mass of the Z candidate in the $Z+ \geq 1$ jets sample before any b-tagging is applied.



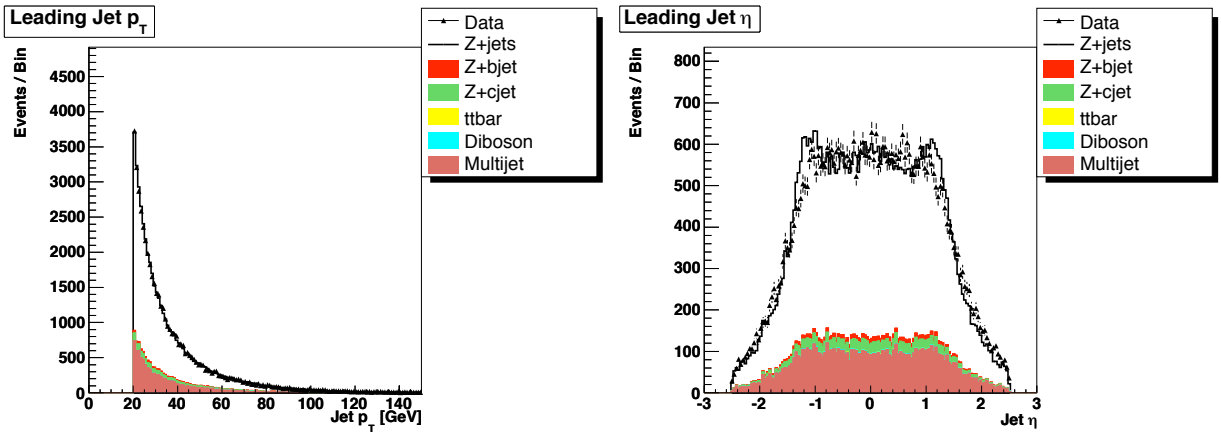
(a) RunIIb dielectron

Figure 10.11: The p_T Z candidate and missing transverse energy in the $Z+ \geq 1$ jets sample before any b-tagging is applied.



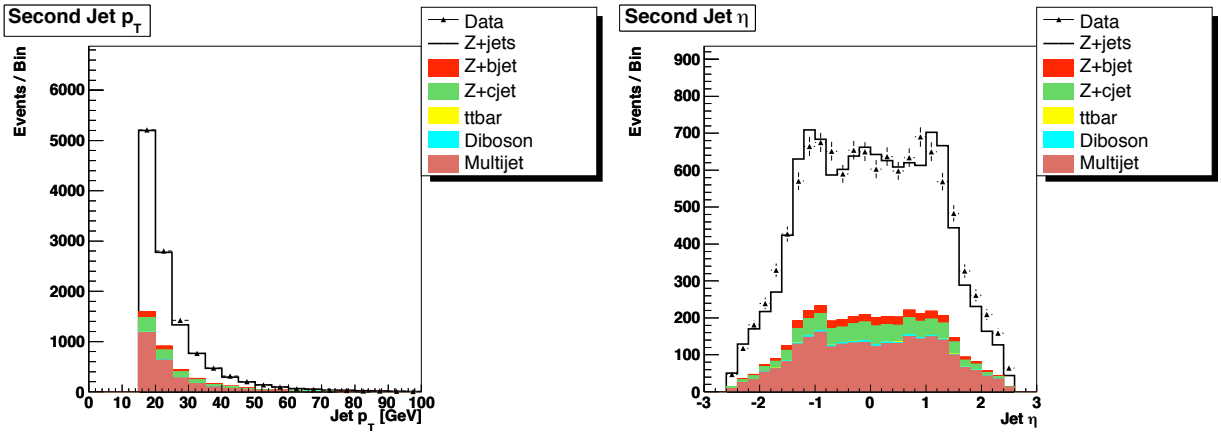
(a) RunIIb dielectron

Figure 10.12: The $\Delta R(ee)$ for the Z candidate and jet multiplicity in the $Z+ \geq 1$ jets sample before any b-tagging is applied.



(a) Leading jet p_T (b) Leading jet η

Figure 10.13: The leading jet (a) p_T and (b) η for data and background in the $Z+ \geq 1$ jets sample before any b-tagging is applied.



(a) Second Leading jet p_T

(b) Second Leading jet η

Figure 10.14: The leading jet (a) p_T and (b) η for data and background in the $Z+ \geq 2$ jets sample before any b-tagging is applied.

Chapter 11

Measurement

This analysis selects 48,956 $Z+ \geq 1$-jet candidate events in the dielectron channel and 24,450 events in the dimuon channel. The background fraction in the dielectron channel is about 16% and is dominated by the multijet production where two jets mimic isolated electrons. The dimuon channel is much cleaner, with the background contribution of 0.6%. A total of 2202 events remain in the dielectron sample and 1015 in the dimuon sample after the requirement that there be at least one b-tagged jet passing the NN output cut of 0.5.

11.1 Identifying Heavy Flavor Jets

Identification of heavy flavor jets and separation of light, c and b-jets is carried out in two steps. First, the Z+jet preselected sample is enriched with heavy flavor jets. This is done by means of the b-jet NN tagger, discussed in Chapter 8. An NN cut value of $NN > 0.5$ ("oldLoose") has been chosen for this analysis. Figure 11.1 shows the distribution of b-tagging efficiency as a function of jet p_T obtained for $Z + b$ MC for this NN operating point. In the second step, the M_{svt} + JLIP (MJL) tagger is employed via a maximum likelihood fitter in order to facilitate the separation between b-, c- and light-jets. Maximum likelihood fitters are further discussed in Appendix A

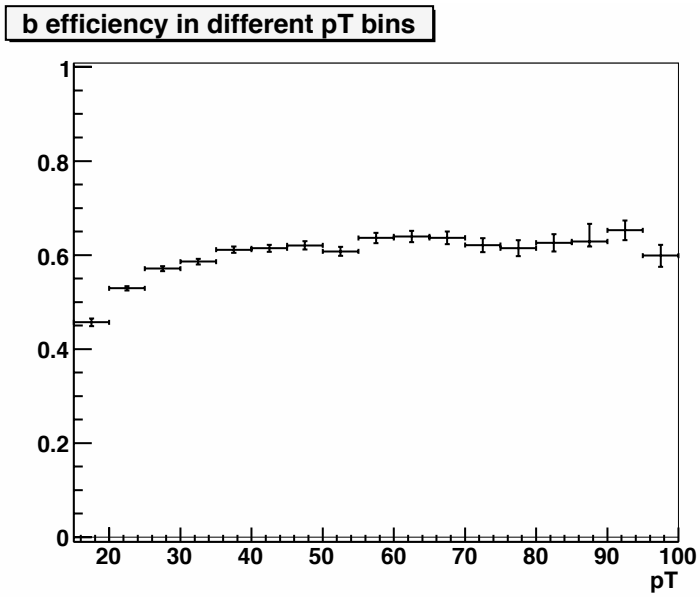


Figure 11.1: The spectrum of b-tagging efficiency as a function of jet transverse momentum measured in $Z + b$ MC for the NN operating point: oldLOOSE (NN>0.5)

11.2 Efficiencies

All efficiencies are calculated after applying the event cuts discussed in Chapter 6. This means that the following cuts are applied to both the numerator and denominator: MET < 60 GeV, $p_T^{jet_1} > 20$ GeV, 70 GeV $< Z_m <$ 110 GeV, |jet physics η|< 2.5. Table 11.1 shows the values for different efficiencies used for the ratio measurement.

Efficiency	Value	Stat. Uncertainty
ϵ^b_{MJL}	0.68	0.03
ϵ^b_{NN}	0.572	0.02
ϵ^b_{tagg}	0.866	0.02
$\epsilon^l_{reco}/\epsilon^b_{reco}$	0.985	0.02

Table 11.1: Efficiencies used in ratio calculation.

11.3 Jet Taggability Efficiency

A jet is required to be taggable before the NN tagger is applied to it. For a jet to be taggable, it must have at least two associated tracks with $p_T > 0.5$ GeV, the leading track must have a p_T of at least 1.0 GeV and both tracks must have at least one hit in the SMT. These criteria ensure that the jet has sufficient information to be classified as a heavy flavor candidate.

Figure 11.2 illustrates the taggability efficiency of jets in the $Z + b$ MC as a function of jet p_T, showing that the taggability for b-jets is nearly a constant. The weighted average yields a taggability of 0.893 ± 0.007. Because the taggability rate is greater in MC than in data, scale factors are applied to correct the MC. To estimate the difference in taggability between data and MC, a taggability requirement was applied to a distribution that showed agreement between MC and data beforehand and then measured the disagreement afterwords. Taggability efficiency in data was approximately 97% of that in MC. This difference is applied to the b-jet MC taggability.

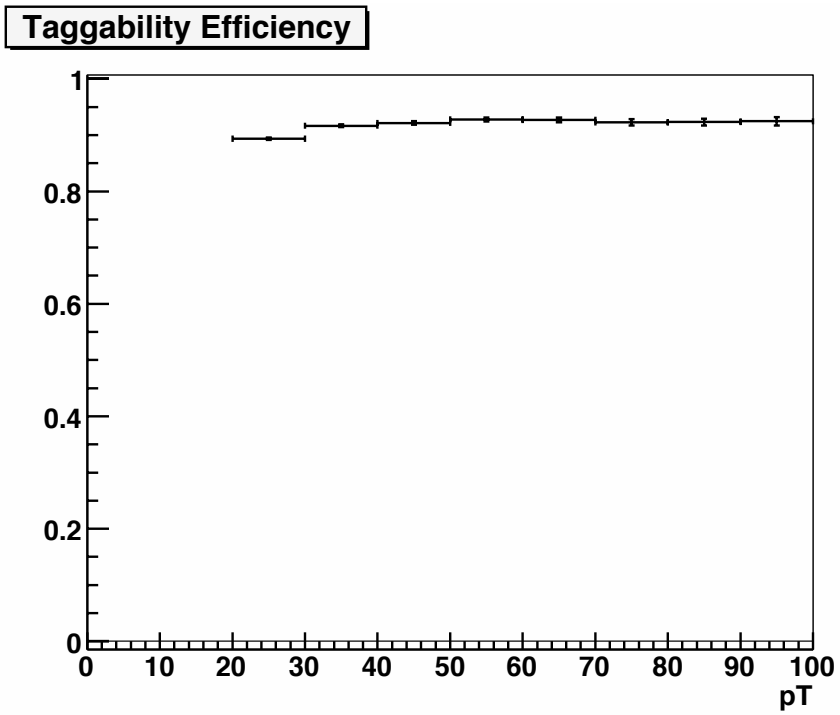


Figure 11.2: b-jet taggability efficiency as a function of jet p_T .

11.4 MJL Tagger

To further separate the b-jets from c- and light-jets in the b-jet enriched sample, a discriminant (MJL) is constructed from two variables that exploits the properties of tracks associated with the b-tagged jet. To construct the MJL discriminant, a combination of the $-ln(JLIP)$ and M_{svt} variables was taken, however a simple addition of the two was not an option. This is because the $-ln(JLIP)$ variable typically ranges between 0 and 20, while the M_{svt} only typically ranges between 0 and 5 GeV. To make the MJL discriminant, both the $-ln(JLIP)$ and M_{svt} variables were scaled down by a factor of ten. The next step was to vary the relative weight between the $-ln(JLIP)$ and M_{svt} variables , as shown in equation 11.1, and then calculate the χ^2 between the b- and c-templates:

$$MJL[i] = \frac{1}{10}\left(M_{svt} - \frac{ln(JLIP)}{i}\right) \tag{11.1}$$

where i ranges from 1 to 10. Figure 11.3 shows the b-, c- and light-jet templates for each of the ten cases from equation 11.1, normalized to unity.

The two highest χ^2 values came from the $i = 4$ and $i = 5$ templates. The $i = 4$ template was then chosen because the construction was physically intuitive. The end result is:

$$MJL = \frac{1}{2}\left(\frac{M_{svt}}{5} - \frac{ln(JLIP)}{20}\right) \tag{11.2}$$

The physical intuitiveness of equation 11.2 is based on the fact that the construction is simply the addition of each variable divided by its range. Thus, limiting the range of each variable to be between 0 and 1. The factor of 1/2 is purely aesthetic, as it confines the range of the MJL discriminant to be between 0 and 1 as well. Both M_{SVT}, which takes into account the kinematics of the event, and JLIP, which takes into account the geometry, were discussed in Chapter 8. Because MJL is formed from the combination of the M_{SVT} and JLIP variables, it has an improved discriminating power for b-, c- and light-jets over either of the

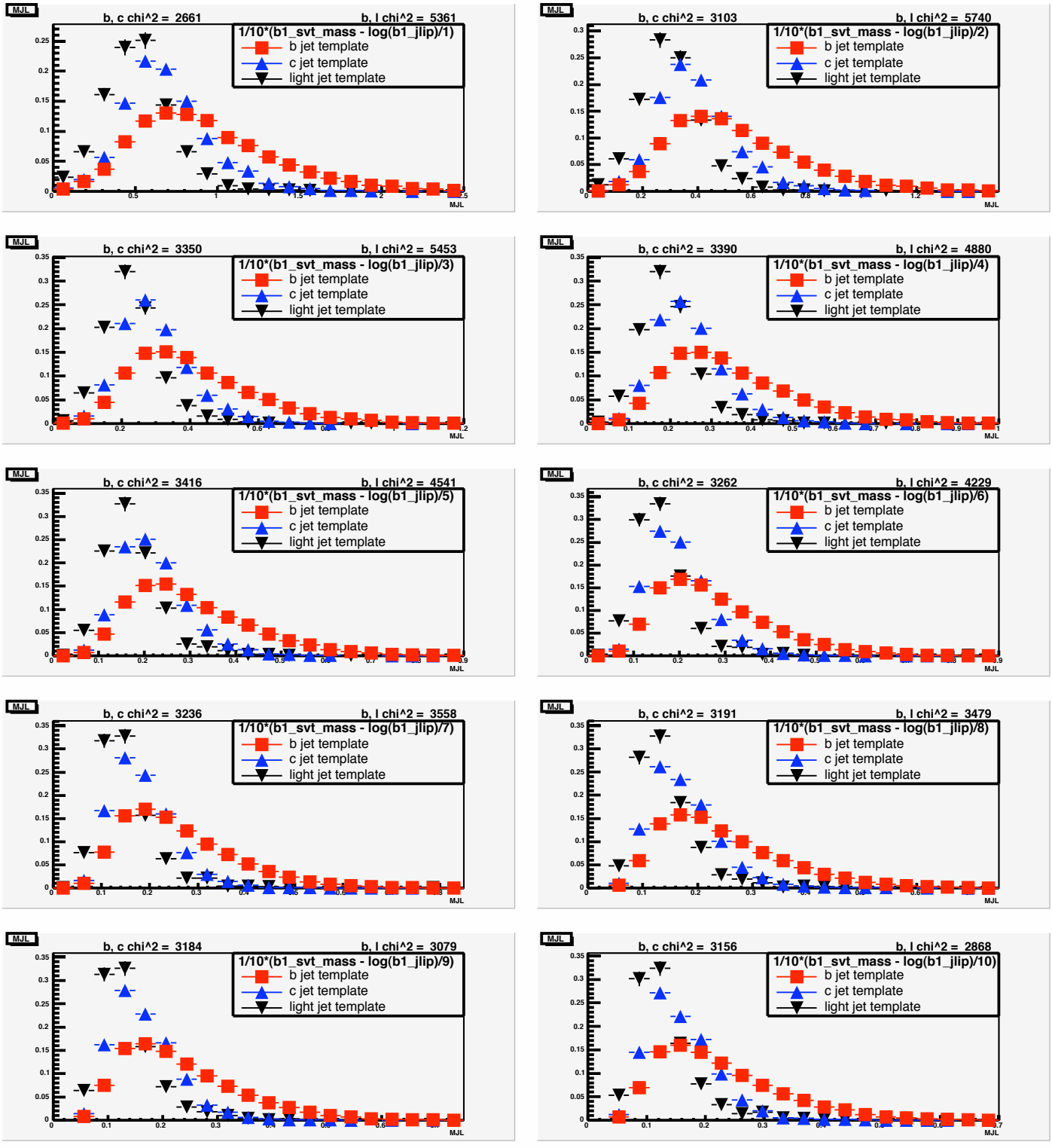
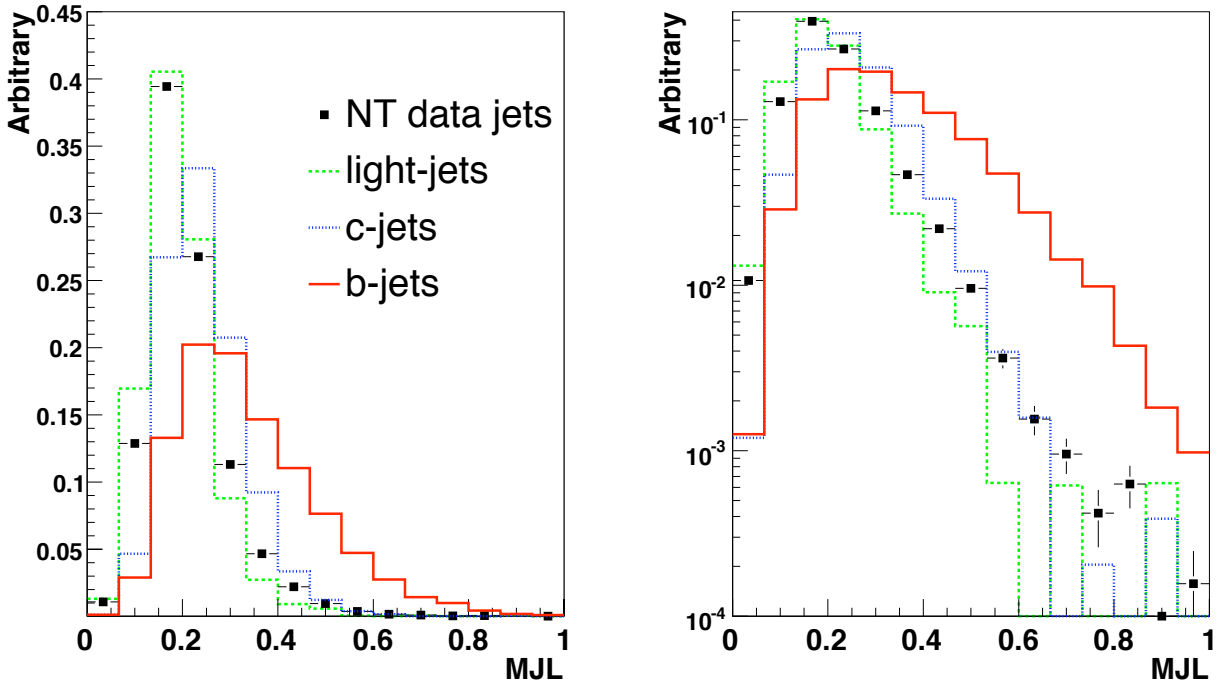


Figure 11.3: Templates for equation 11.1. i = odd on the left and even on the right both increasing on the way down.

variables individually. The M_{SVT} tagger has an efficiency of $\sim 80\%$ for b-jets in MC, while the JLIP tagger is close to 100%, leading to an MJL efficiency of 80% for b-jets in MC.

Figure 11.4 shows the MJL distributions, normalized to unity, for Z+light, Z+c, Z+b MC samples and the NT jets sample. This figure clearly shows the separation power of the new tagger. With these templates, the fraction of Z+light, Z+c and Z+b in data can be determined using a maximum likelihood fitter.



(a) MJL

(b) MJL log

Figure 11.4: MJL template for b-jets, light-jets and c-jets derived from Z+ jets MC samples and negative tagged jets from data for NN > 0.5 (a) linear and (b) log scale

11.4.1 MJL η and p_T Dependence

A major reason for using the MJL tagger is that it does not have a strong dependence in η or p_T. Figure 11.5 shows the MJL templates in different η regions and figure 11.6 shows the MJL templates for different p_T regions. The template shape in the $|\eta| < 1.1$ and $|\eta| < 2.5$

regions is very similar, which is an important feature for the tagger. This allows for the measurement to be carried out all at once, rather than in different slices of η. Although the p_T shows a slight dependence, it is not drastic. A test was run to assure that any p_T and η dependence wouldn't bias the result. This test used data from the central η (low p_T) region and the templates from the entire η (high p_T) region. This result was then compared to the result using central (low p_T) templates. The two results were found to be within the uncertainties given by the maximum likelihood fitter.

11.4.2 MJL Efficiency

Because only the fraction of tagged b-jets are used from the maximum likelihood fitter, only the MJL efficiency for b-jets is needed. Figure 11.7 shows the b-jet MJL efficiency as a function of jet p_T for the NN > 0.5 operating point. The p_T weighted average value is found to be 0.8 in MC jets. To find the value in data, a scale factor is applied to account for the difference between data and MC. This scale factor is measured similarly to that discussed in section 8.5 and has an average value of 0.85, leading to an efficiency in data of 0.68.

11.5 Increased Template Statistics

One of the major sources of uncertainty in this measurement is the statistics in each of the three MJL templates. Because of the low efficiency of the NN tagger for c- and light-jets, the statistics for both samples get greatly reduced. For this reason, steps for increasing the statistics in both the c-jet and light-jet templates were taken.

11.5.1 Negative Tagged Jets

The MJL template for light-jets is determined with data using negative tagged jets (Section 8.4), in order to increase statistics. A large sample of the Summer09 EMInclusive data set has been used for this purpose. Due to b-contamination, the NT jets template is broader

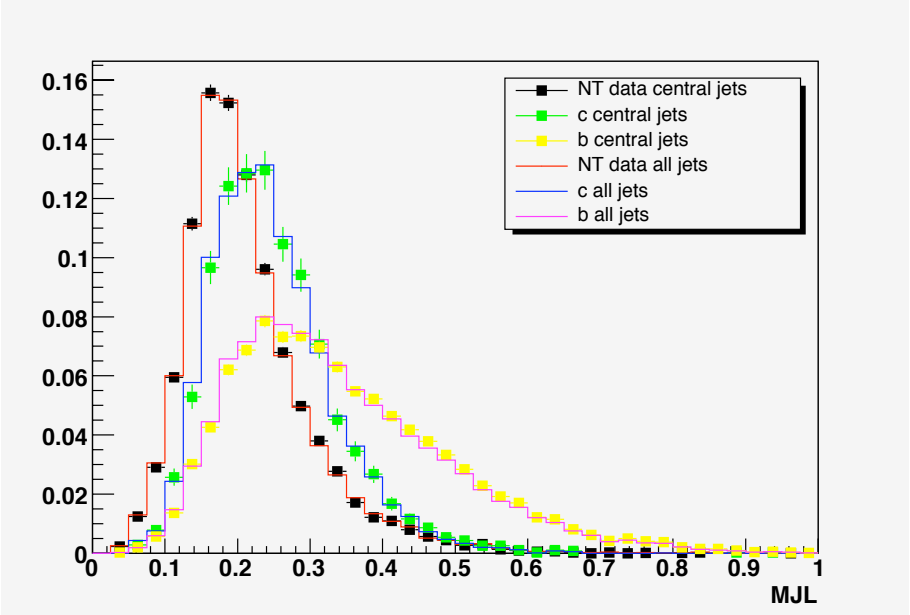


Figure 11.5: The MJL templates in different η regions.

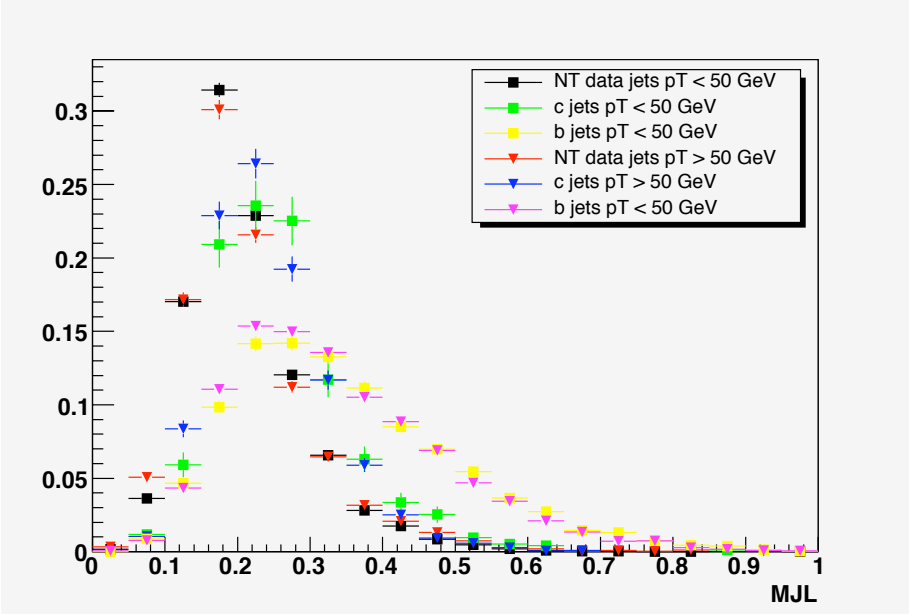


Figure 11.6: The MJL templates in different p_T regions.

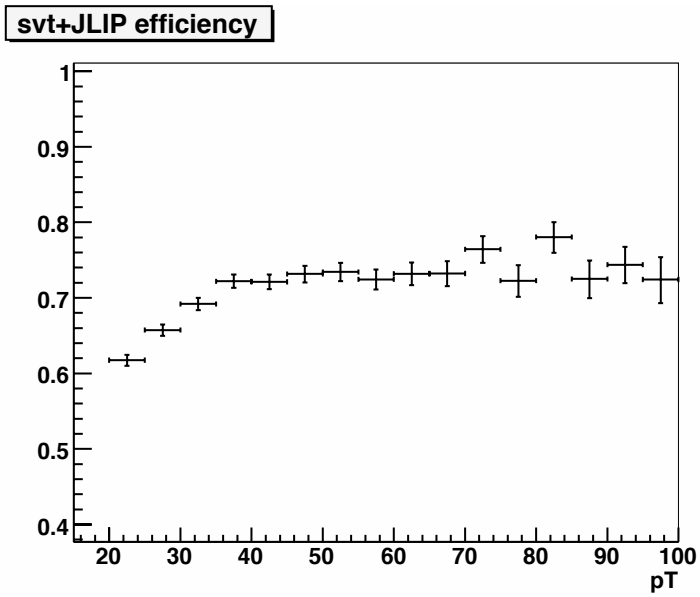


Figure 11.7: b-jet MJL efficiency with respect to p_T

than the MC light-jet template. In order to remove this contamination, a maximum likelihood fitter is used to extract the fraction of NT b-jets in the NT jets template. Once the fraction is measured, the NT b-jet fraction is subtracted from the NT jets template. Figure 11.8 shows the NT jets MJL template with the b-contamination, without the b-contamination and the MC light-jet MJL template for comparison. Without the b-contamination, the NT jets MJL template is in much better agreement with the MC light-jet template. In order to increase statistics in the light-jet template, the NT Data is used with the b-contamination subtracted out. The b-contamination found by the maximum likelihood fitter, as well as the distribution from the result can be seen in Appendix E.

11.5.2 Increased Statistics for c-jet Template

In order to increase the c-jet statistics, a Pythia c-jet sample was used. Figure 11.9 shows the comparison of the $Z + c$-jet template to the Pythia c-jet sample and the two shapes are almost identical.

The leading jet p_T distribution of the two samples is different, however (shown in fig-

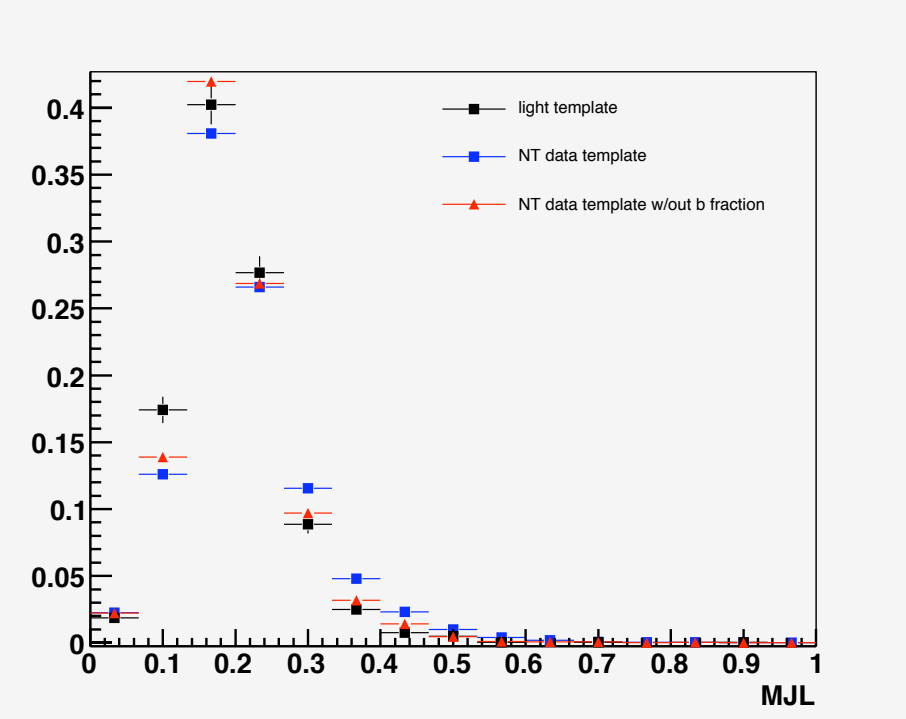


Figure 11.8: MJL templates for negative tagged jets in data and light-jets in MC

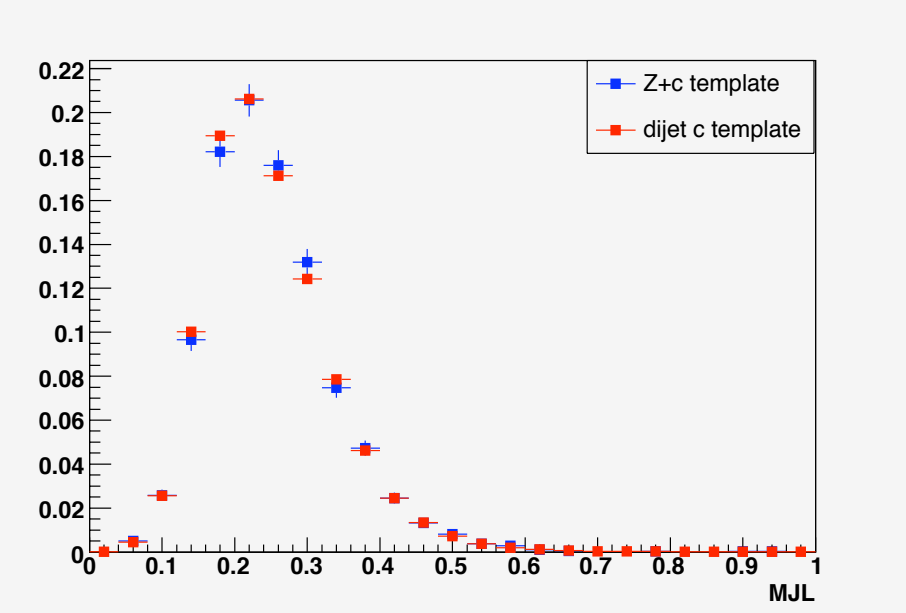


Figure 11.9: $Z + c$-jet template compared to PYTHIA c-jet template.

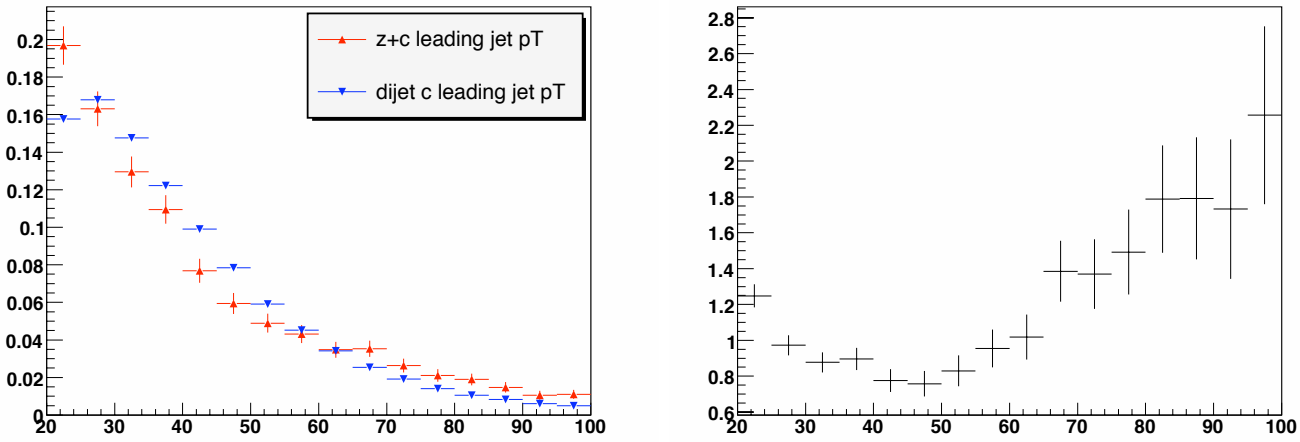
ure 11.10a). To ensure that the agreement between the two template shapes is not coincidental, the leading jet p_T of the PYTHIA c-jet sample was reweighted to match that of the $Z + c$ leading jet p_T (figures 11.10b and 11.11b). Figure 11.11a shows the $Z + c$-jet template plotted with the PYTHIA c-jet MJL template both before and after reweighting. The reweighting has virtually no effect on the template shape, showing that the different jet p_T spectra have no effect on the template shapes.

11.5.3 Uncertainty Due to Template Statistics

Limited template statistics can cause a large systematic uncertainty in the maximum likelihood fitter. This is because the templates are varied within the bin-by-bin template statistics to create a new template. Figure 11.12 shows the $Z+c$-jet template along with an even (odd) template. The even (odd) template adds the statistical uncertainty to the $Z+c$-jet template value for even (odd) bins, and subtracts the statistical uncertainty from the $Z + c$-jet template for the odd (even) bins. The even and odd templates are equally feasible, yet visibly they are different. This is due to the lack of template statistics in the $Z + c$-jet template. Figure 11.13 shows the same distributions for the PYTHIA c-jet sample and the templates are virtually indistinguishable. This greatly reduces the systematic uncertainty due to the c-template statistics and is further justification for the use of the PYTHIA c-jet (referred to as c-jet from now on) and NT jets samples.

11.6 Background Subtraction

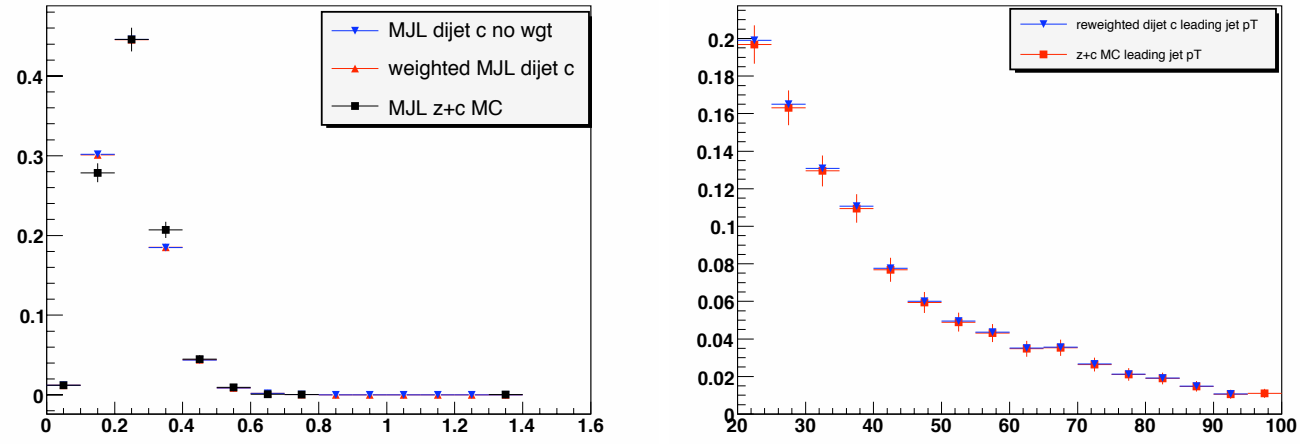
In order to fit the data with the three MJL templates, the background (everything except Z+jets) must first be subtracted out. The sources of the background are diboson, $t\bar{t}$ and QCD multijet. For the diboson and $t\bar{t}$, the MC samples discussed in Chapter 5 are used. By using the cross sections associated to these samples, the expected number of pretagged and tagged events can be determined. For QCD multijet (Chapter 7), the event weights are kept



(a) Leading c-jet p_T

(b) Ratio of leading jet p_T

Figure 11.10: a) The leading jet p_T for both the $Z + c$-jet and PYTHIA c-jet samples. b) The ratio of the two samples



(a) MJL templates

(b) Leading jet p_T after reweight

Figure 11.11: a) MJL templates for PYTHIA c-jet, reweighted PYTHIA c-jet and $Z + c$-jet samples b) p_T spectra after reweight

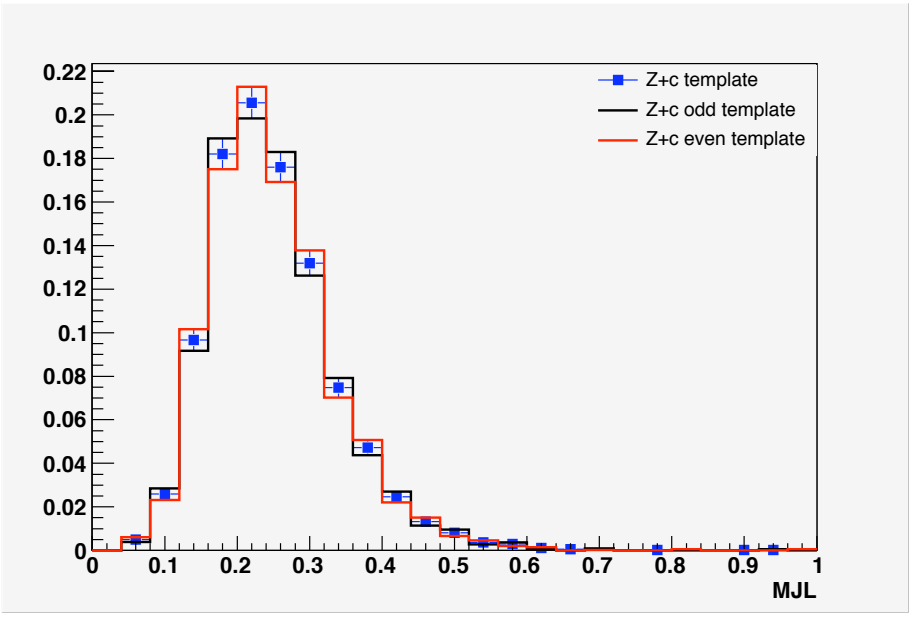


Figure 11.12: $Z + c$ template along with even and odd templates. The even (odd) template adds the $Z + c$-jet bin value to the bin uncertainty for the even (odd) bins and subtracts the bin uncertainty from the $Z + c$-jet bin value for odd (even) bins.

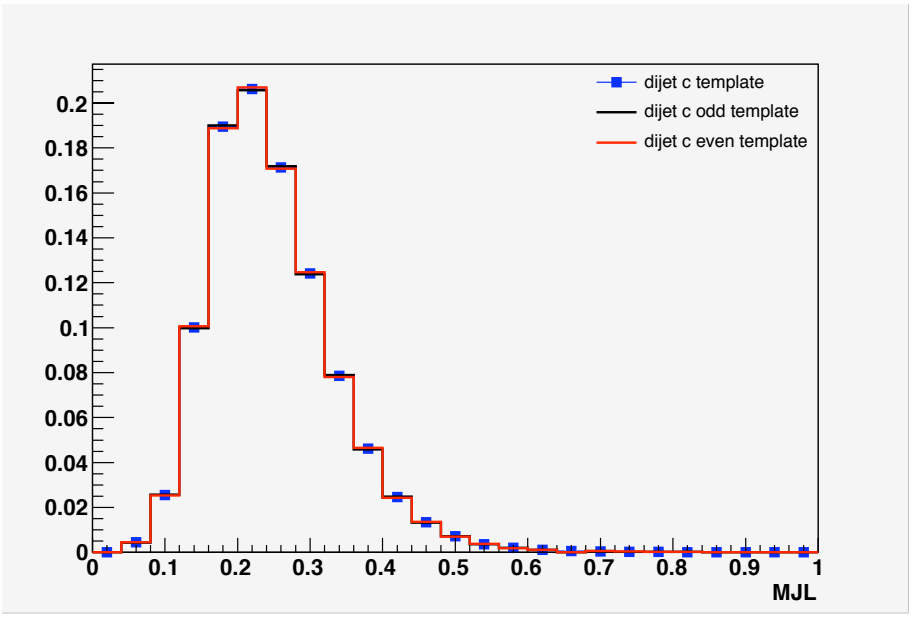


Figure 11.13: c-jet template along with even and odd templates. The even (odd) template adds the c-jet bin value to the bin uncertainty for the even (odd) bins and subtracts the bin uncertainty from the c-jet bin value for odd (even) bins.

through the tagging process and determine how many tagged events are expected. Simply subtracting the number of expected background events from the data is not a possibility due to the background being dominated by b-jets, and therefore having a b-like MJL template (shown in Figure 11.14). In order to subtract the background, the MJL shape for each background is normalized to the expected number of events, and then that distribution is subtracted from the data.

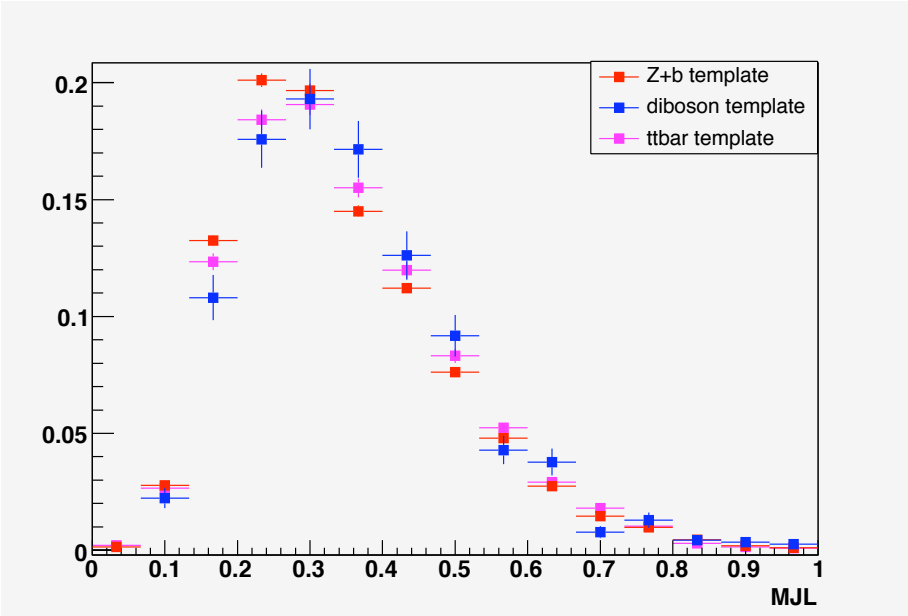


Figure 11.14: MJL templates for different backgrounds

11.7 Jet Reconstruction Efficiency, b-jet vs. light-jet

Because the ratio of cross sections is what is measured, the uncertainty in jet reconstruction efficiency should cancel out to first order. A possible bias, however, can arise from the difference between b- and other-jets. This is estimated by investigating the relative jet reconstruction efficiencies in $Z + b$-jet and $Z+$ light-jet MC.

Figure 11.15 shows the reconstruction efficiencies for b and light-jets and their ratio as a function of jet p_T. The weighted average yields a value of 0.985 ± 0.007. The scale factor $\epsilon^b_{reco}/\epsilon^l_{reco}$ is determined and applied to the ratio measurement to account for the difference

between the light and b-jet reconstruction efficiencies.

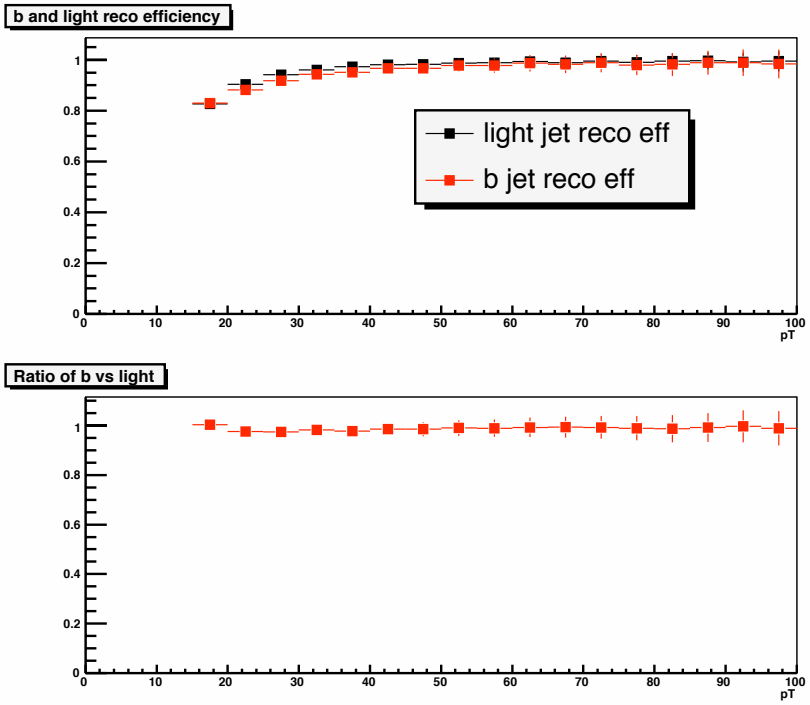


Figure 11.15: Ratio of b-jet reconstruction efficiency to light-jet reconstruction efficiency vs. p_T .

11.8 Determination of Jet Flavor Fractions

In order to measure the $\sigma(Z+b\text{-jet})/\sigma(Z+jet)$ ratio, both the number of Z+jet events and the number of $Z+b$-jet events must be determined. To find the number of Z+jet events, the number of events that match the criteria discussed above is counted. To determine the number of $Z+b$-jet events the following prescription is used. First, the NN tagger is applied to the leading jet in the data, as well as the $Z+b$-jet and c-jet MC samples and the NT jets sample. If the leading jet does not pass the NN requirement, the NN tagger is applied to next-to-leading jets with p_T >15 GeV. If either jet passes the NN cut, the event is considered "tagged," and the MJL value of the tagged jet is stored. The contribution from the next-to-leading jets is approximately 10%. The number of tagged jets after background

subtraction and MJL selection is 970 in the electron channel and 646 in the muon channel. The fraction of b-jets in data can be determined by using the MJL templates of the passed jets. The $Z + b$-jet, c-jet and NT data MJL templates are fed into a maximum likelihood fitter to extract the fractions of each component from the data. These fractions are then used to determine the $\sigma(Z + b - jet)/\sigma(Z + jet)$ ratio, as follows:

$$N = N_l + N_c + N_b \tag{11.3}$$

$$P = P_l + P_c + P_b \tag{11.4}$$

$$P_l = \epsilon_l \times N_l \tag{11.5}$$

$$P_c = \epsilon_c \times N_c \tag{11.6}$$

$$P_b = \epsilon_b \times N_b \tag{11.7}$$

where P is the number of tagged events; P_b, P_c and P_l are the number of $Z + b$, $Z + c$ and Z+light jet tagged events given by the fitter; ϵ_l, ϵ_c and ϵ_b are the corresponding (mistagging/tagging) efficiencies of the NN tagger; and N is the number of Z+jet events. The cross section ratio can then be calculated as:

$$\frac{\sigma(Z + b)}{\sigma(Z + jet)} = \frac{P_b \epsilon^{light}_{reco}}{N \epsilon^b_{NN} \epsilon^b_{MJL} \epsilon^b_{Tagg} \epsilon^b_{reco}} \tag{11.8}$$

where $\epsilon^{b(light)}_{reco}$ is the reconstruction efficiency of b (light) jets, ϵ^b_{MJL} is the MJL efficiency of b-jets and ϵ^b_{Tagg} is the taggability efficiency of b-jets. The different jet flavor fractions obtained with this method are listed in Tables 11.2 and 11.4. Results of the maximum likelihood fit for both channels are shown in figures 11.19 and 11.21; the total MC rate in these figures is normalized to the data. For the combined results, the data after subtraction for both channels is added together, and then the fractions from the templates are recalculated using

the maximum likelihood fitter. Figure 11.16 shows the data MJL shape after background subtraction for both the dimuon and dielectron channels. They are in good agreement. The similar shape in data and consistent results across both channels provides justification for the combining of the two channels. Figures 11.22 and 11.23 show the combined results for the $|\eta| < 2.5$ and $|\eta| < 1.1$ regions, respectively. The figures show both the data after background subtraction along with the $Z+b$-, $Z+c$- and Z+light-jet contributions measured by the maximum likelihood fitter.

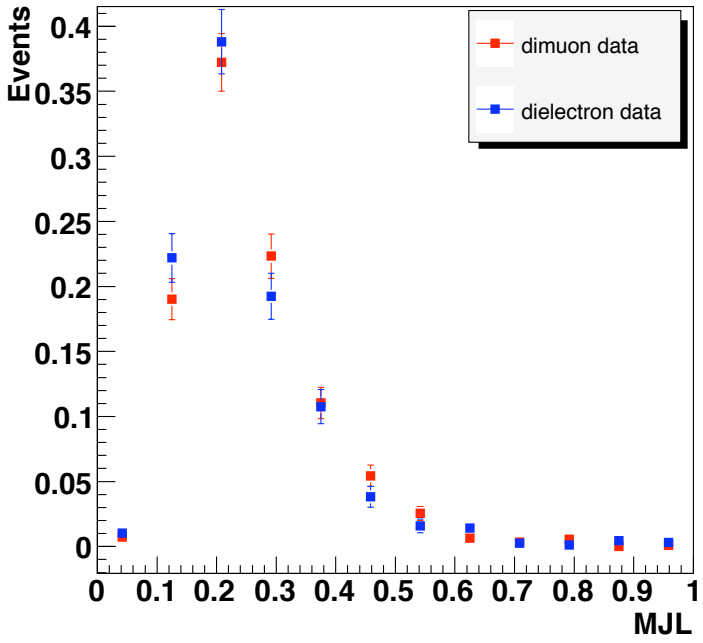


Figure 11.16: Dimuon and dielectron data after background subtraction, normalized to unity.

11.8.1 Additional Cross Check

As a cross check, an alternative method to determine the number of events with tagged b-jets, P_b is employed. This "matrix method" is based on the calculation of b-jet sample purity for different MJL tagging points [73]. The prescription is shown in equations 11.9-11.12.

$$N = N_l + N_c + N_b \tag{11.9}$$

$$\epsilon_D^1 N = \epsilon_l^1 N_l + \epsilon_c^1 N_c + \epsilon_b^1 N_b \tag{11.10}$$

$$\epsilon_D^2 N = \epsilon_l^2 N_l + \epsilon_c^2 N_c + \epsilon_b^2 N_b \tag{11.11}$$

$$N_b = \frac{(\epsilon_D^1 - \epsilon_l^1) - (\epsilon_D^2 - \epsilon_l^2)\,\dfrac{\epsilon_c^1 - \epsilon_l^1}{\epsilon_c^2 - \epsilon_l^2}}{(\epsilon_b^1 - \epsilon_l^1) - (\epsilon_b^2 - \epsilon_l^2)} \tag{11.12}$$

where N is the total number of Z+jet events, N_l, N_b, N_c is the number of light-, $Z + b$- and $Z + c$-jet events respectively, $\epsilon^1_{D(lbc)}$ is the efficiency of an MJL > 0.33 cut on data (Z+light, $Z + b$, $Z + c$), and $\epsilon^2_{D(lbc)}$ is the efficiency of an MJL > 0.66 cut on data (Z+light, $Z + b$, $Z + c$). The efficiencies for each cut are then measured in data, b-jets, c-jets and NT jets. This cross check gave a b-purity value of $33 \pm 6.2\%$ for the NN > 0.5 operating point in the combined channel and agrees with the value given by the maximum likelihood fitter: $27.2 \pm 2.8\%$.

Jet Flavor	Fraction ($\mu\mu$)	Fraction (ee)
Region	$\lvert\eta\rvert < 1.1$	$NN > 0.5$
Events	433	669
Z+b	0.274±0.051	0.283±0.042
Z+c	0.238±0.089	0.402±0.072
Z+l	0.488±0.070	0.315±0.054
Region	$\lvert\eta\rvert < 2.5$	$NN > 0.5$
Events	629	970
Z+b	0.248±0.042	0.270±0.036
Z+c	0.253±0.073	0.402±0.059
Z+l	0.500±0.058	0.315±0.046

Table 11.2: Jet flavor fractions in the tagged Z+jets sample.

Fixing c/light Ratio

An equally valid approach to measuring the $Z+b$ fraction in data is to use $Z+b$, the c/light ratio and c+light as the three parameters for the fit. This approach should yield the same result, serving as a good cross check. It also will demonstrate any sensitivity that the $Z+b$ fraction has to the c/light ratio. To obtain a result for this approach, the c/light ratio was fixed and the $Z+b$ and $Z+c$+Z+light fractions were obtained. This was repeated for many different values of the c/light ratio, increasing the value by 0.1 each time. Figure 11.17 shows the log likelihood for each c/light value for both the muon and electron channels. The figure also shows the $Z+b$ fraction divided by the $Z+b$ fraction of the most likely fit for the c/light values that fall within one sigma of the most likely c/light value. The results for this method are consistent with the results from the three template fits, shown in Table 11.3. Figure 11.17 also demonstrates the lack of sensitivity the $Z+b$ fraction has to the c/light ratio. For the electron channel, the c/light ratio can be increased by 60% from the most ideal case, and the $Z+b$ fraction still falls within one sigma of the most likely result.

Jet Flavor	Fraction ($\mu\mu$)	Fraction (ee)
Region	NN > 0.5	$\eta < 2.5$
$Z+b$	0.238±0.036	0.270±0.033
c/light	$0.6^{+0.3}_{-0.2}$	$1.1^{+0.5}_{-0.3}$
c+light	0.762±0.046	0.730±0.037

Table 11.3: Results for alternative fitting option in $\mu\mu$ and ee channels.

11.9 Final Result

The final results for the $\sigma(z+b)/\sigma(z+jet)$ ratios for both the $\mu\mu$ and ee channels and for the $|\eta| < 1.1$ and $|\eta| < 2.5$ regions are listed in table 11.5 and are found to consistent. Table 11.6 lists the $\sigma(z+b)/\sigma(z+jet)$ ratios obtained for the combined $\mu\mu$ and ee channels for the different η regions.

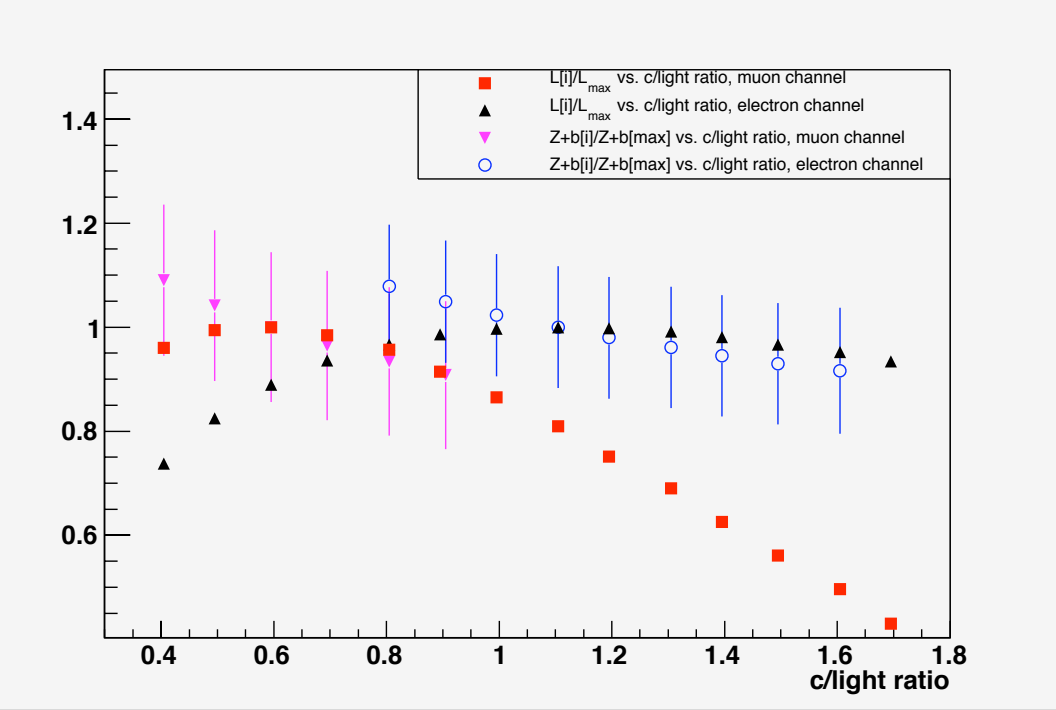
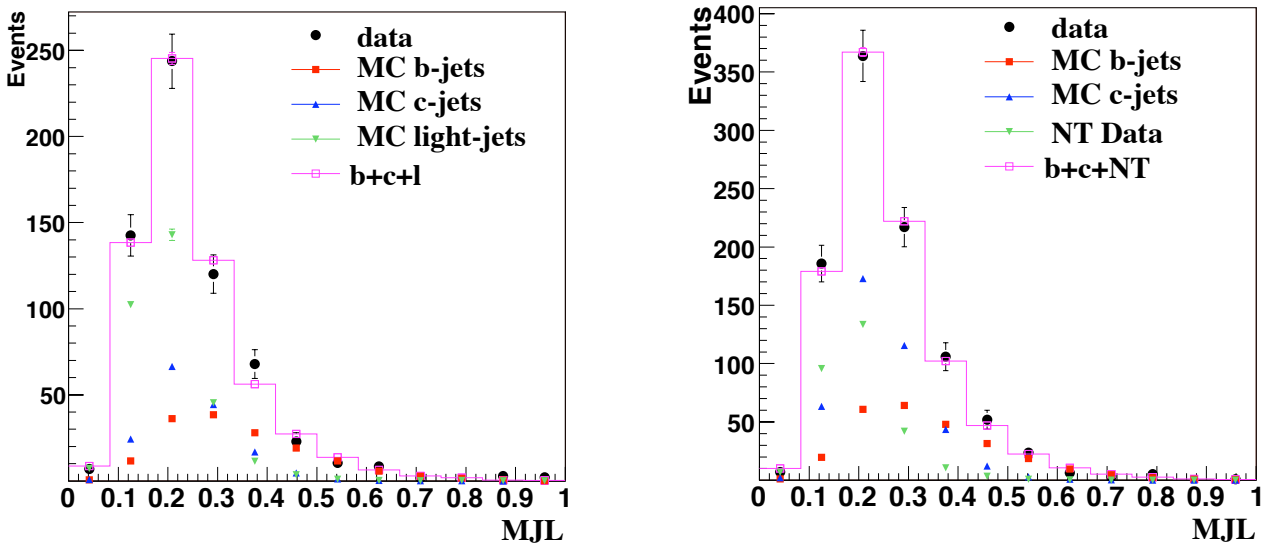


Figure 11.17: Both likelihood and $Z + b$ fraction results for different c/light values.

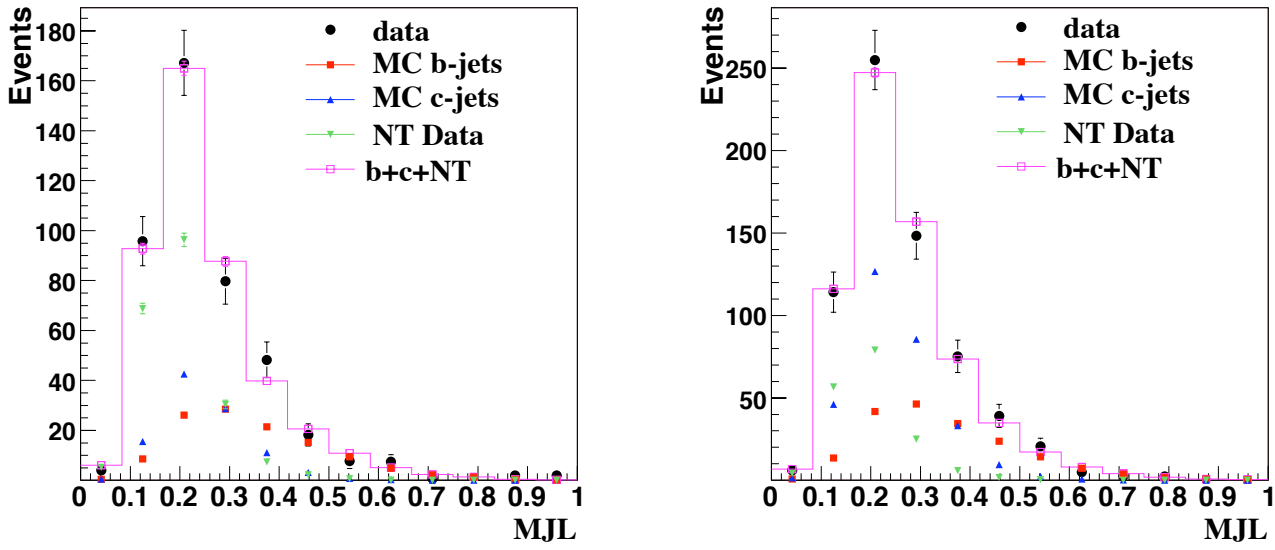
Jet Flavor	Fraction ($\mu\mu$ + ee combined)
Region	$\lvert\eta\rvert < 1.1$, $NN > 0.5$
Events	1101
Z+b	0.270±0.033
Z+c	0.347±0.058
Z+l	0.355±0.069
Region	$\lvert\eta\rvert < 2.5$, $NN > 0.5$
Events	1600
Z+b	0.259±0.028
Z+c	0.359±0.049
Z+l	0.382±0.038

Table 11.4: Jet flavor fractions $\mu\mu + ee$ combined channel.



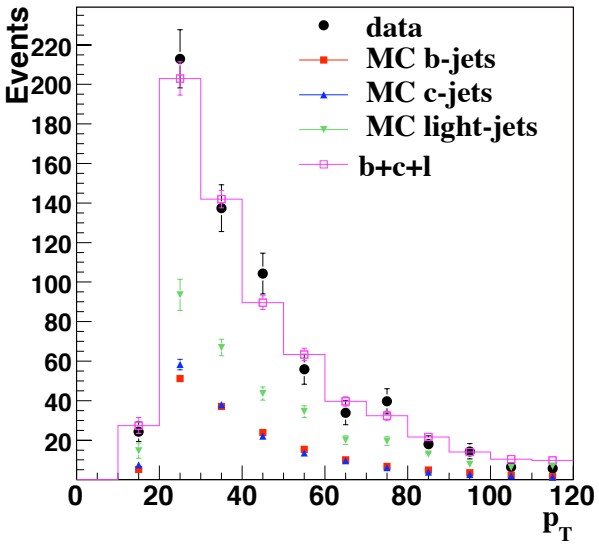
(a) Dimuon (b) Dielectron

Figure 11.18: Distribution of observed events for MJL discriminant with NN $>$ 0.5 and $|\eta| < 2.5$ cuts. The distributions for the b-, c- and light-jet templates are shown normalized to their fitted fraction. Error on the templates represent combined uncertainties from statistics of MC and the fitted jet flavor fractions.

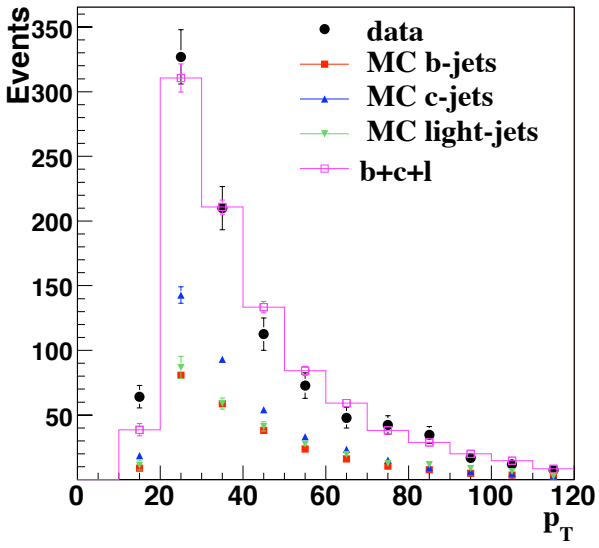


(a) Dimuon (b) Dielectron

Figure 11.19: Distribution of observed events for MJL discriminant with NN $>$ 0.5 and $|\eta| < 1.1$ cuts. .

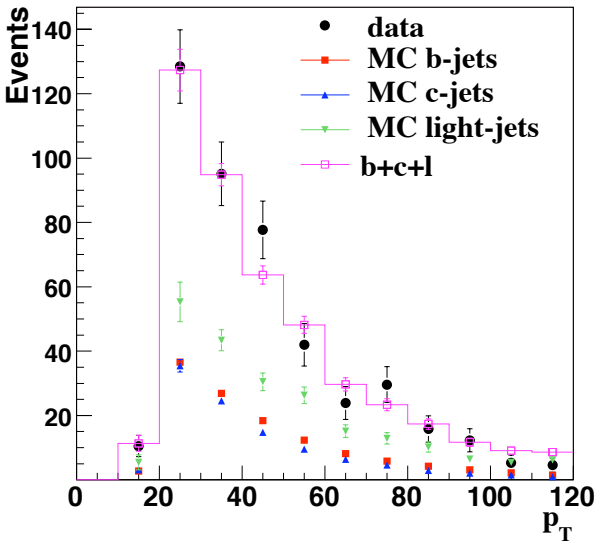


(a) Dimuon

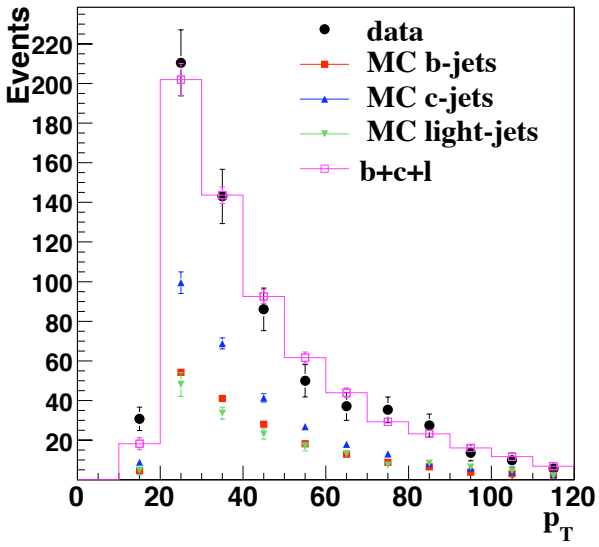


(b) Dielectron

Figure 11.20: The p_T distribution of b-tagged jets with NN > 0.5 and $|\eta| < 2.5$ cuts in each channel.

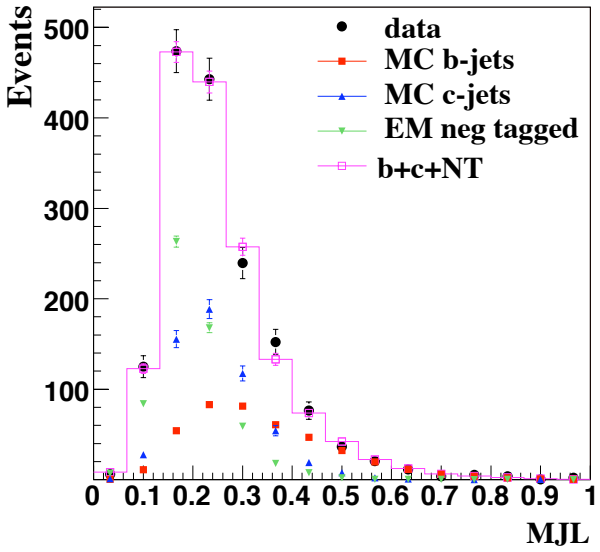


(a) Dimuon

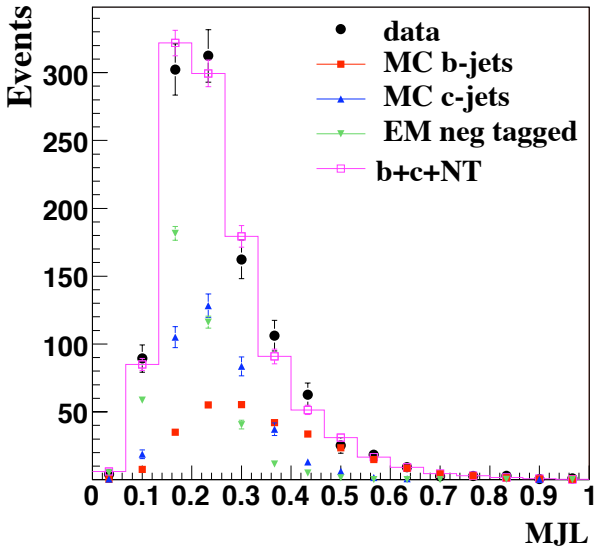


(b) Dielectron

Figure 11.21: The p_T distribution of b-tagged jets with NN > 0.5 and $|\eta| < 1.1$ cuts in each channel.

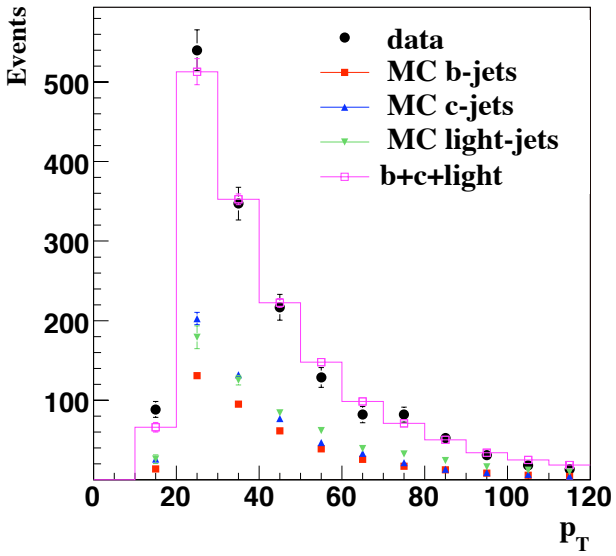


(a) $|\eta| < 2.5$

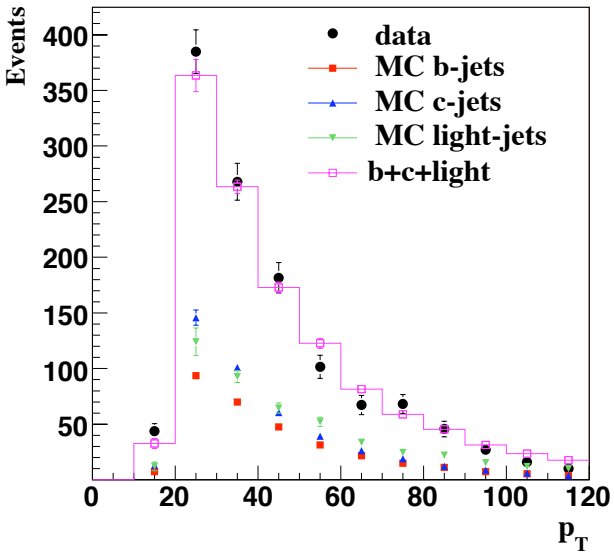


(b) $|\eta| < 1.1$

Figure 11.22: Distribution of observed events in combined sample for MJL discriminant with NN > 0.5 in (a) $|\eta| < 2.5$ and (b) $|\eta| < 1.1$ regions.



(a) $|\eta| < 2.5$



(b) $|\eta| < 1.1$

Figure 11.23: The combined p_T distributions of b-tagged jets with NN > 0.5 in (a) $|\eta| < 2.5$ cuts and (b) $|\eta| < 1.1$ regions.

The flavor fraction uncertainties obtained from the template fitting to the data are mostly statistical. The systematic component was determined by artificially increasing the statistics of the data and rerunning the maximum likelihood fitter. The result was a fit without the error due to data statistics, leaving only the error due to template statistics. The uncertainty in the $Z + b$-jet fraction from this fit was taken as the systematic uncertainty associated to template statistics.

The result for the $\sigma(z + b)/\sigma(z + jet)$ ratio in combined $\mu\mu$ and ee channels and for the $NN > 0.5$ tagging operating point is 0.0193 ± 0.0023 (stat.)$\pm$0.0004 (uncertainty from template statistics).

11.9.1 Comparison to Previous Version of Analysis

This version of the analysis is the second iteration. The first was done with a different discriminant. The original discriminant was called reduced JLIP (rJLIP) and it was similar to the JLIP variable discussed above. The variable rJLIP recalculates the JLIP value for a jet

Channel	$\mu\mu$ (Ratio$\pm$ stat.)	ee (Ratio$\pm$ stat.)
Region	$\|\eta\| < 1.1$	$NN > 0.5$
Ratio	0.0201$\pm$0.0042$\pm$0.0005	0.0198$\pm$0.0033$\pm$0.0005
Region	$\|\eta\| < 2.5$	$NN > 0.5$
Ratio	0.0191$\pm$0.0034$\pm$0.0005	0.0191$\pm$0.0028$\pm$0.0004

Table 11.5: $\frac{\sigma(z+b)}{\sigma(z+jet)}$ ratio in both $\mu\mu$ and ee channels along with uncertainty due to template fitting.

Channel	$\mu\mu + ee$ (Ratio$\pm$ stat.)
Region	$\|\eta\| < 1.1$, $NN > 0.5$
Ratio	0.0193$\pm$0.0027$\pm$0.0004
Region	$\|\eta\| < 2.5$, $NN > 0.5$
Ratio	0.0193$\pm$0.0022$\pm$0.0004

Table 11.6: $\frac{\sigma(z+b)}{\sigma(z+jet)}$ ratio in both $\mu\mu$ and ee channels combined along with uncertainty due to template fitting..

after removing the track least likely to have originated at the primary vertex. Figure 11.24 shows the rJLIP templates for light, b-, c- and NT jets and figure 11.25 shows the resultant rJLIP and jet p_T distributions with the ratios measured by the maximum likelihood fitter, using the rJLIP templates.

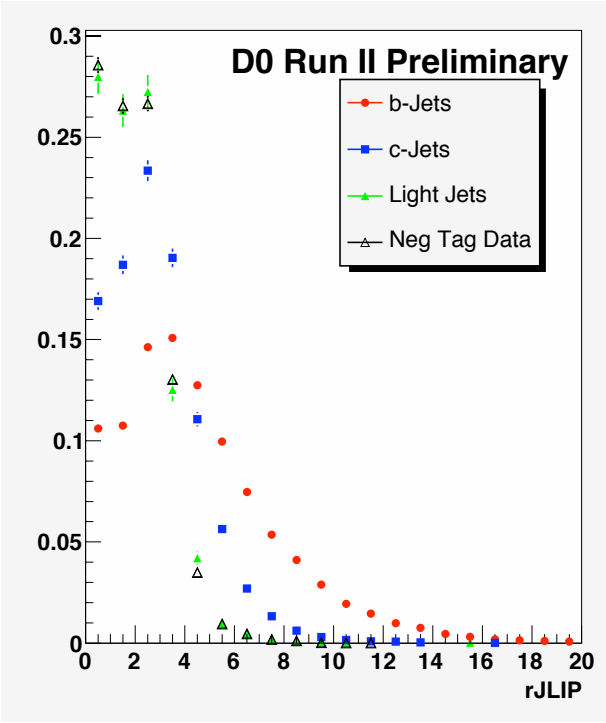
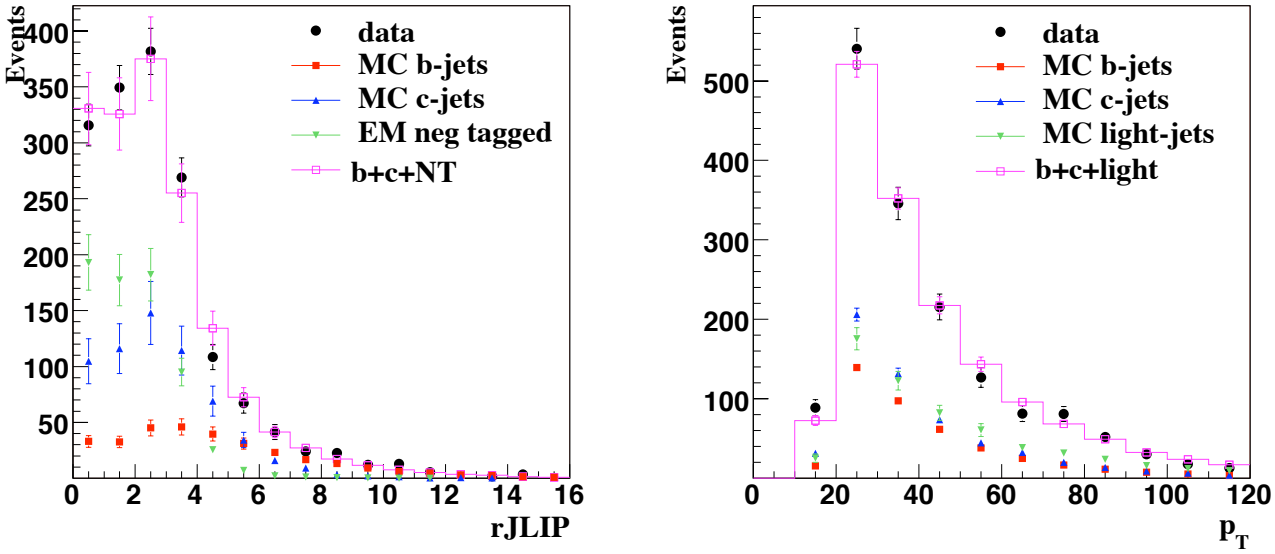


Figure 11.24: rJLIP distribution for light, b-, c- and NT jets.

The separation power of the rJLIP variable is limited to the central region ($|\eta_{jet}| < 1.1$). This was the reason for abandoning the rJLIP variable for the latest iteration of the analysis. Before changing discriminants, a value was found for the $Z + b$-jet/Z+jet ratio in the central region, thus a comparison between the two ratios can further confirm the validity of the measurement. The Z boson selection is almost identical for both analysis, however the jet selection is slightly different. For the previous iteration, only jets with $\eta_{det} < 1.1$ were considered. This is not directly comparable to the new iteration in which a physics η cut is applied. Figure 11.26 shows the physics η acceptance when a $\eta_{det} < 1.1$ cut is applied for both b- and light-jets. The figure shows that the jet acceptance is not identical between physics and detector η cuts however, the effect is expected to be small. Table 11.6 provides



(a) tagged jet rJLIP

(b) tagged jet p_T

Figure 11.25: The combined results for (a) rJLIP distribution of tagged jets and (b) jet p_T of tagged jets measured with the rJLIP templates for jets with $|\eta_{det}| < 1.1$.

a measurement in the central ($|\eta_{phys}| < 1.1$) region that should be in close proximity to the rJLIP version value in the central ($|\eta_{det}| < 1.1$) region. The rJLIP-based analysis found the $Z + b$-jet/Z+jet ratio to be 0.0176±0.0024±0.0023. This matches the value found in the latest iteration within uncertainty.

11.9.2 Cross Check with Alternative NN Operating Point

OP	$\sigma(z+b)/\sigma(z+jet)$ Ratio
NN > 0.5	0.0193±0.0022±0.0004
NN > 0.65	0.0195±0.0023±0.0004

Table 11.7: $\sigma(z+b)/\sigma(z+jet)$ Ratio for $|\eta| < 2.5$ region and NN > 0.5 and NN > 0.65 operating points.

The result should not be biased by the choice of NN operating point (OP). In order to

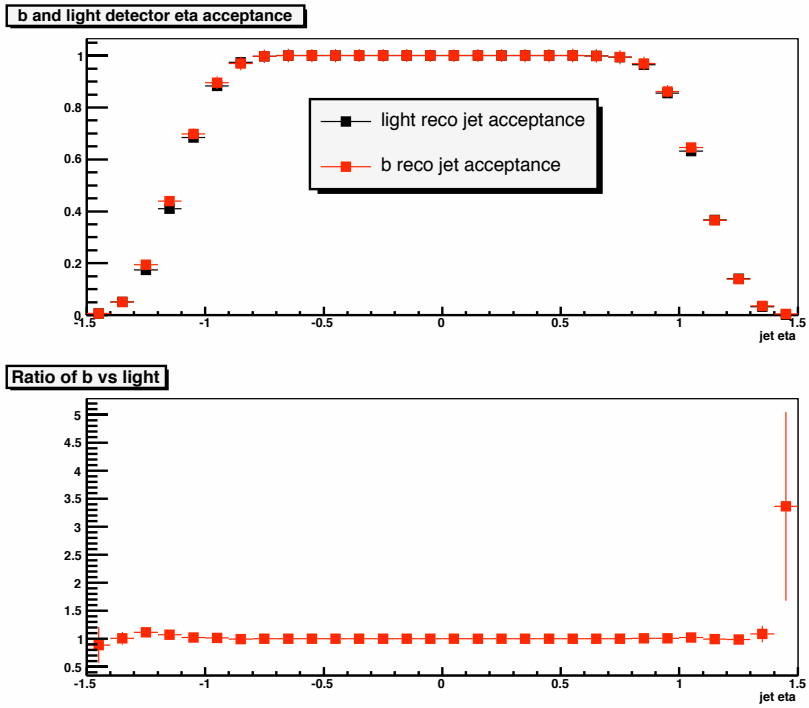


Figure 11.26: Physics η acceptance when a $\eta_{det} < 1.1$ cut is applied

show the stability of the result, a cross check was conducted that remeasured the ratio after applying an NN > 0.65 cut. The values for the $\sigma(z+b)/\sigma(z+jet)$ ratio for the two NN operating points are shown in Table 11.7 and are in good agreement. The distributions for the medium operating point as well as the result from the maximum likelihood fitter are shown in Appendix B.

Chapter 12

Systematic Uncertainties

- B-tagging Efficiency:

  B-tagging efficiency and uncertainty are determined by the DØ b-ID group [57]. In order to calculate the systematic error associated with this value, the efficiency was varied both up and down one σ and the ratio was remeasured. The difference in the central value of the ratio was taken as the uncertainty.

- Jet Energy Resolution:

  For the systematic uncertainty in jet energy resolution, the standard resolution was varied both up and down one σ, and the ratio was remeasured. The difference in the central value of the ratio was taken as the uncertainty.

- Jet Energy Scale:

  For the systematic uncertainty in jet energy scale, the standard correction was varied both up and down one σ and the ratio was remeasured, similar to the two above.

- Jet Energy Scale – b-jet vs. light-jet:

  The jet energy scale (JES) may be different for b- and light-jets, however, real data does not exist to study. Thus a standard JES correction is applied to all jets. The result of having a standard correction is the under correction of b-jets. The difference in MC

events was investigated using $Z+b$-jet and Z+light-jet samples. Figure 12.2 shows the residual for b-jets and light jets defined as the difference in p_T between reconstructed and particle jet divided by the particle jet's p_T ($\Delta p_T = \frac{p_{T\_reco} - p_{T\_particle}}{p_{T\_particle}}$). The under correction of b-jets was estimated to be approximately 6.5%, the maximum effect observed in the low p_T region. To compensate for this under correction, the processor was run with an extra 6.5% JES correction on b-jets. Figure 12.3 shows the effect of additional JES correction on the MJL template. The final $\sigma(Z+b)/\sigma(Z+j)$ ratio was calculated using the b-jets with the extra JES correction. The systematic uncertainty due to the b-JES correction is taken as half of the difference of the final ratio between the under-correct and extra-corrected b-jets.

- Jet Reconstruction Efficiencies for Heavy and Light Quarks:

  While a scale factor is applied to the final ratio to compensate for the difference in jet reconstruction, a small problem remains. $\frac{\epsilon^l_{reco}}{\epsilon^b_{reco}}$ assumes that the Z+jet sample is made up entirely of light-jets. For a systematic, the scale factor is varied by 2% and half of the difference in the ratio is taken as the systematic uncertainty.

- Background Estimations:

  In order to subtract the background from the data, Monte Carlo generators are relied on to give the number of events. To measure the systematic uncertainty associated to this method, the effect that varying the number of background events up and down by 10% had on the ratio was taken.

- Shape uncertainty – Data versus MC:

  To estimate the systematic uncertainty due to template differences in data and MC, the templates in both the negative tagged data as well as the Z+light jet MC were used. The maximum likelihood fitter was applied with both cases and for the systematic uncertainty the difference of the central values is taken; that amounts to 2.2%.

- Uncertainty Due to b-quark Fragmentation

By default, the events have been reweighted from the default PYTHIA b-fragmentation to a Bowler scheme that has been tuned to LEP data. To evaluate the systematics, the events are further reweighted to account for the difference between SLD and LEP data, as described here [74]. Figure 12.4 shows the MJL distribution for both b-fragmentation schemes. The fractions are remeasured with the new fragmentation scheme. The difference in the b-fraction between the two schemes is taken as a systematic uncertainty.

- Uncertainty Due to Collapsed b- and c-quark jets:

  In the case where two b- or c-quarks are produced close together, both quarks can wind up within the same reconstructed jet. These jets have a slightly different MJL template shape than jets with one b- or c-quark, shown in figure 12.1. An uncertainty arises from the fact that the percentage of jets with collapsed b- and c-quarks in data is not known. To measure the uncertainty in the final ratio due to this, both the c-jet and b-jet MJL templates were reweighted with a five-times larger contribution coming from the collapsed jet templates. The difference in the central value of the ratio was taken as the uncertainty and was measured to be 2.8%.

- MJL Data-to-MC Scale Factor Uncertainty:

  To estimate the systematic uncertainty from the MJL scale factor, the central value of the data to MC scale factor was varied by half of the RMS. The value for the uncertainty was taken as the difference in the central value of the ratio as a result of changing the scale factor.

- Trigger Corrections:

  A trigger is applied in the muon channel and has a corresponding efficiency. To estimate the uncertainty of the trigger efficiency, the scale factor α, discussed in Chapter 7, was reweighted to 1, and the ratio was remeasured. The difference in the central value of the ratio was taken as the systematic uncertainty.

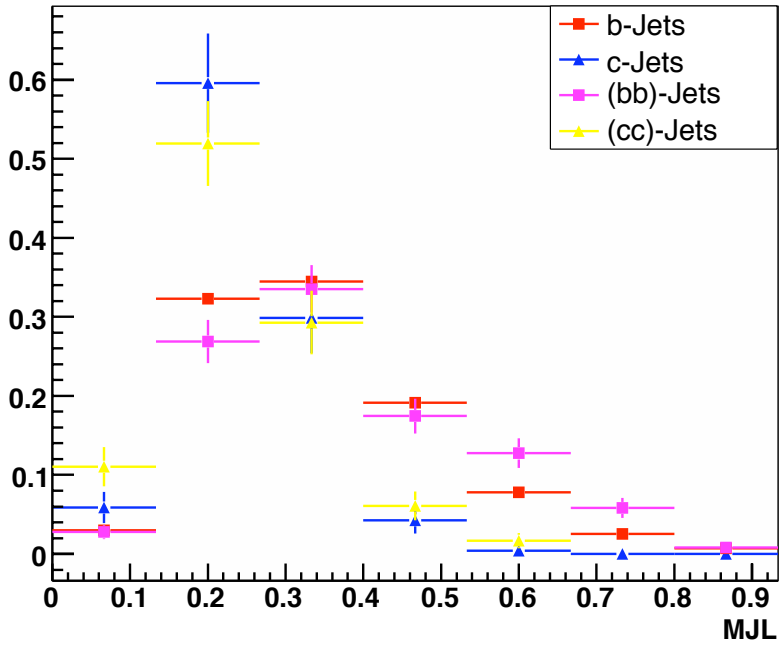


Figure 12.1: MJL template of c- and b-jets as well as collapsed c- and b-quark jets.

Systematic Uncertainty	Percentage
B-tagging	2.4%
JES	2%
JER	2.5%
JES: b vs. l	0.4%
Bkgd. Estimate	1.5%
Jet Reco: b vs. light	2%
Template Shape	2.2%
B-fragmentation	1.8%
Collapsed Quark Jet	2.8%
MJL Efficiency	3.7%
Trigger Corrections ($\mu\mu$)	0.4%
Sum	$\pm$ 7.2%

Table 12.1: Systematic uncertainties and their contribution to the ratio

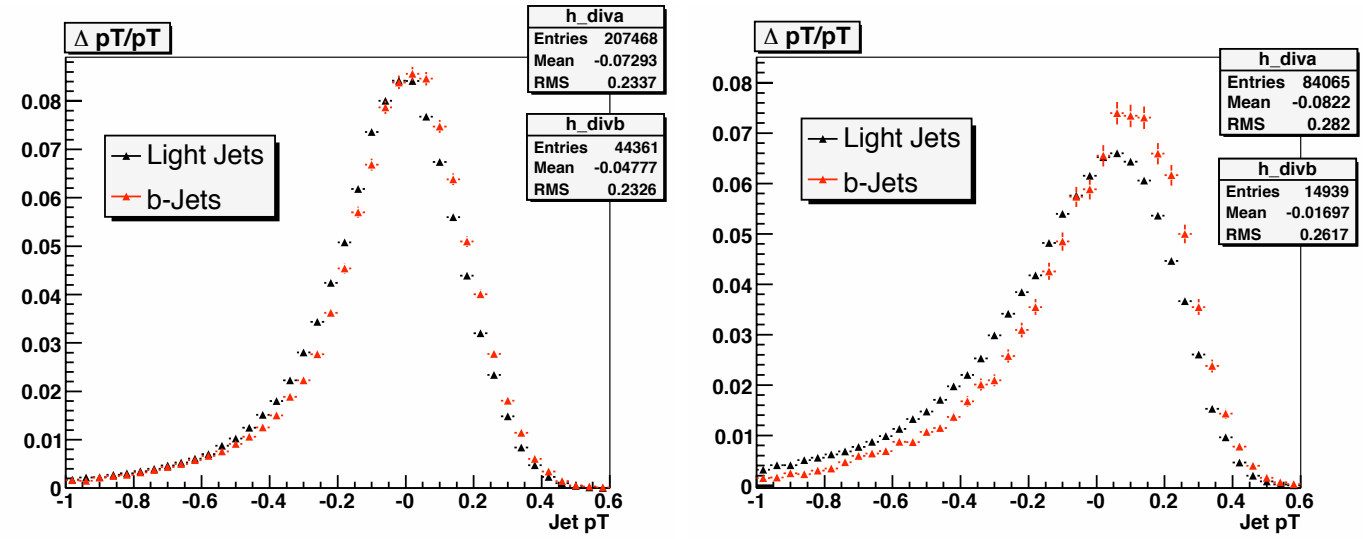


Figure 12.2: The distributions of $\frac{p_{T}\_reco - p_{T}\_particle}{p_{T}\_particle}$ for both light jets and b-jets before b-jet JES correction in the full p_T range (left) and in the low p_T range (right) of 20-30 GeV.

Taking into account above systematic uncertainties, the result for the $\sigma(z+b)/\sigma(z+jet)$ ratio in combined $\mu\mu$ and ee channels is 0.0193 ± 0.0022 (stat.) ± 0.0004 (template statistics) ± 0.0014 (syst.).

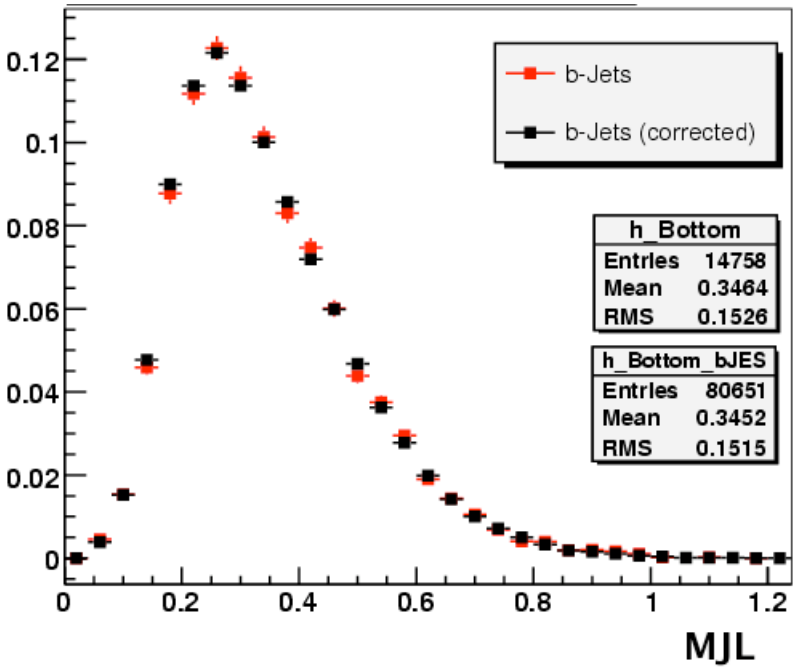


Figure 12.3: The effect of additional JES correction on the MJL distributions for b-jets.

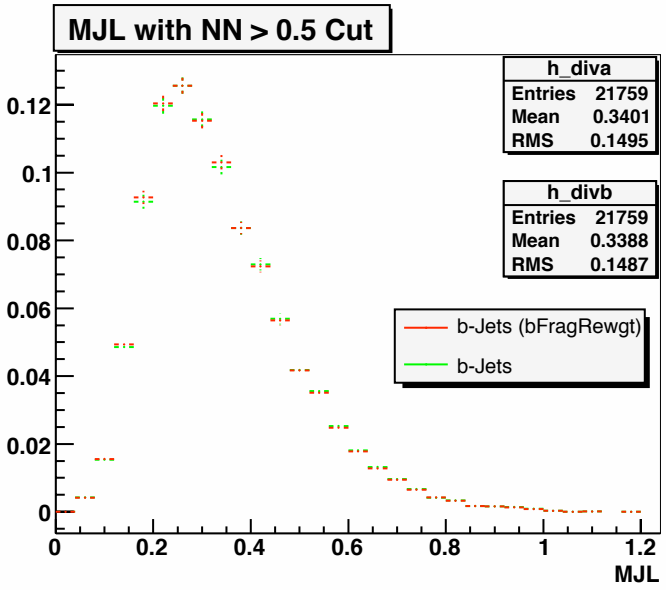


Figure 12.4: The MJL distributions for two different b-fragmentation schemes.

Chapter 13

MCFM

MCFM [68] is a Monte Carlo simulation program that calculates the theoretical cross sections for various femtobarn-level processes at hadron-hadron colliders.

13.1 Comparison with Theoretical Predictions

MCFM provides next-to-leading order (NLO) calculations [1, 68] for the inclusive cross sections of $Z + b$-jet, $Z + c$-jet and Z+light-jet production. Tables 13.1 and 13.2 show the values of expected cross sections of the different subprocesses obtained for the two set of jet kinematic cuts. These MCFM predictions are based on the scale $Q^2 = m_Z^2$ and using MRST2008 PDF. The statistical errors on the cross sections are very small ($< 0.1\%$). The theoretical systematic uncertainties from the choice of scale and PDF are expected to be around 11% for the $Z + b/c - jet$ and 4.5% for Z+light-jet production cross sections. Based on the MCFM predictions, the ratio $\sigma(z + b)\sigma(z + jet)$ is expected to be 0.018 ± 0.004 for jet $|\eta| < 2.5$. The total uncertainty on the prediction includes uncertainties arising from renormalization scale (0.0015), factorization scale (0.0011) and parton distribution functions (0.0011). The measured ratio is reasonably consistent with the theoretical prediction.

A one-to-one comparison of the measured ratio to the MCFM expected ratio is not completely accurate. This is because the measurement is carried out on reconstructed jets in

Cross sections (fb)	Tevatron (p_T >20 GeV and $\|\eta\|$ <2.5)				
	ZQ	$Z(Q\bar{Q})$	ZQj	$ZQ\bar{Q}$	$ZQ-$inclusi
$gb \to Zb$	217.6	22.2	54.7	12.3	306.8
$q\bar{q} \to Zb\bar{b}$	70.0	142.1		37.0	249.1
$Z+b-$inclusive					556 fb
$gc \to Zc$	335.6	17.9	84.1	9.8	447.4
$q\bar{q} \to Zc\bar{c}$	100.9	300.8		35.8	437.5
$Z+c-$inclusive					885 fb
	Zj	$Zjj)$			$Z+j-$inclu
$q\bar{q} \to Zg$,$gq \to Zq$	24125.9	3413.7			27540 fb
$\sigma(z+b-jet)/\sigma(z+jet)$	$0.0192\pm^{+0.0035}_{-0.0025}$				

Table 13.1: The cross section (fb) evaluated by MCFM for Z+heavy flavor and light jet production at the Tevatron. A jet lies in the range p_T >20 GeV and $|\eta|$ <2.5 and cone size of ΔR =0.5 has been used. The cross sections have been obtained for the scale $Q^2 = m_Z^2$ and using MSTW2008 PDF. The labels in the columns have the following meaning :$ZQ \equiv$ exactly one jet, which contains a heavy quark;$Z(Q\bar{Q}) \equiv$ exactly one jet which contains two heavy quarks;$ZQj \equiv$ exactly two jets, one of which contains a heavy quark; $ZQ\bar{Q} \equiv$exactly two jets, both of which contain a heavy quark. For last set of processes, the labels mean: $Zj \equiv$ exactly one jet, which does not contain a heavy quark; $Zjj \equiv$ exactly two jets, none of which contain a heavy quark. The uncertainties on the ratio are from the variation of renormalization scale, factorization scale and the PDFs.

Cross sections (fb)	Tevatron	(p_T >20 GeV,	$\|\eta\|$ <2.5)		
	ZQ	$Z(Q\bar{Q})$	ZQj	$ZQ\bar{Q}$	ZQ-inclusive
$gb \to Zb$	248.2	22.2	54.7	11.8	336.9
$q\bar{q} \to Zb\bar{b}$	70.0	142.1		37.0	249.1
$Z + b-$inclusive					586±64 fb
$gc \to Zc$	386.6	17.9	84.1	9.8	498.4
$q\bar{q} \to Zc\bar{c}$	100.9	300.8		35.8	437.5
$Z + c-$inclusive					936±103 fb
	Zj	Zjj			$Z + j$-inclusive
$q\bar{q} \to Zg$,$gq \to Zq$	25937.3	4262			30200±1359 fb

Table 13.2: The cross section (fb) evaluated by MCFM for Z+heavy flavor and light-jet production at the Tevatron. A jet lies in the range p_T >20 GeV and $|\eta|$ <2.5, and cone size of $\Delta R = 0.5$ has been used. The cross sections have been obtained for the scale $Q^2 = m_Z^2$ and using MRST2008 PDF. The labels in the columns have the following meaning: $ZQ \equiv$ exactly one jet, which contains a heavy quark; $Z(Q\bar{Q}) \equiv$ exactly one jet which contains two heavy quarks; $ZQj \equiv$ exactly two jets, one of which contains a heavy quark; $ZQ\bar{Q} \equiv$exactly two jets, both of which contain a heavy quark. For last set of processes, the labels mean: $Zj \equiv$ exactly one jet, which does not contain a heavy quark; $Zjj \equiv$ exactly two jets, none of which contain a heavy quark. The uncertainties are from the variation of renormalization scale, factorization scale and the PDFs.

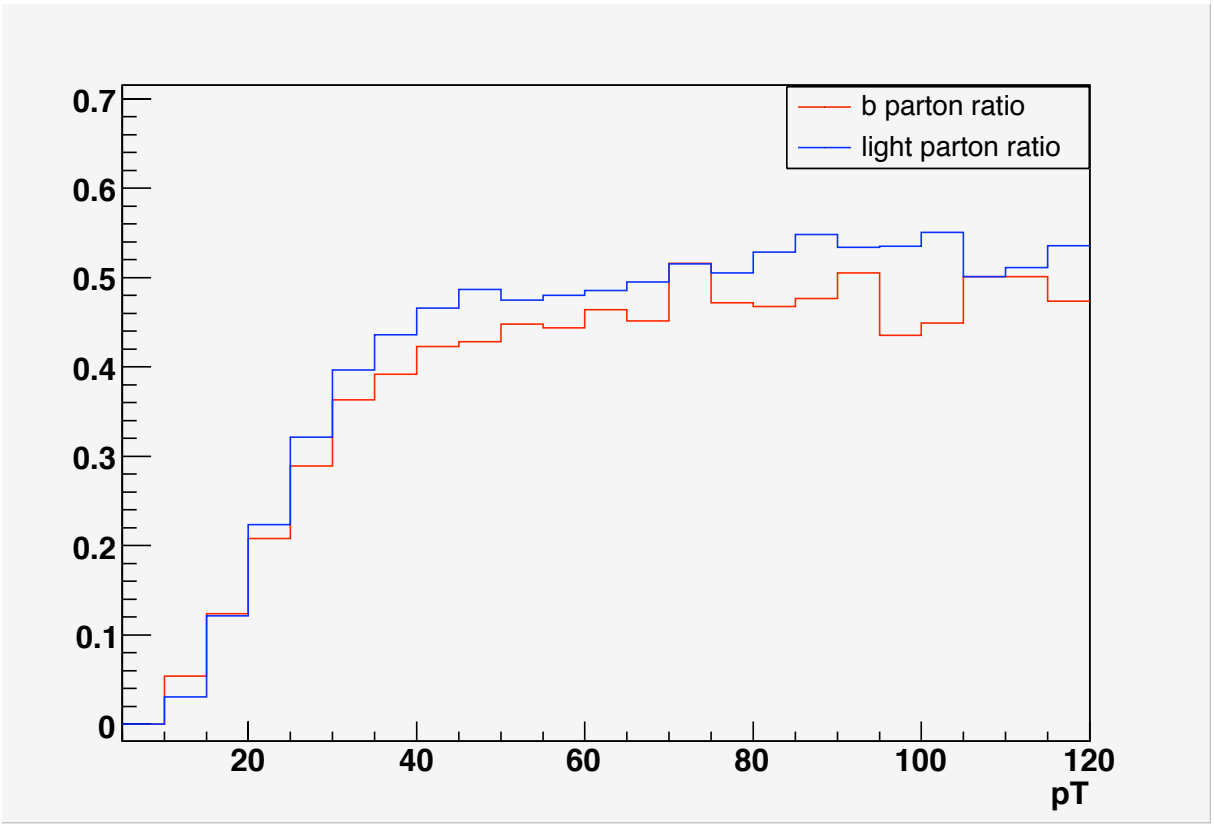


Figure 13.1: Ratio of b/light spectra for ALPGEN MC partons and ALPGEN MC partons matched to reconstructed jets.

the detector, while the MCFM prediction is for partons with no detector simulation. To get a better estimate of what the MCFM prediction would look like in the detector, the following procedure was used. First, the b and light spectra for the ALPGEN MC were created for both the MC parton level and the MC parton level where the partons were matched to a reconstructed jet that passes all cuts required in the analysis. The ratio of matched spectra versus parton p_T spectra was taken for both light and b partons with respect to parton p_T in order to demonstrate the difference between the two spectra (Figure 13.1).

Using the ratio of the two ALPGEN spectra, one can now see the expected effect on the MCFM spectrum. Equation 13.1 shows the method for getting from the MCFM parton level distribution to the expected MCFM distribution with detector effects folded in:

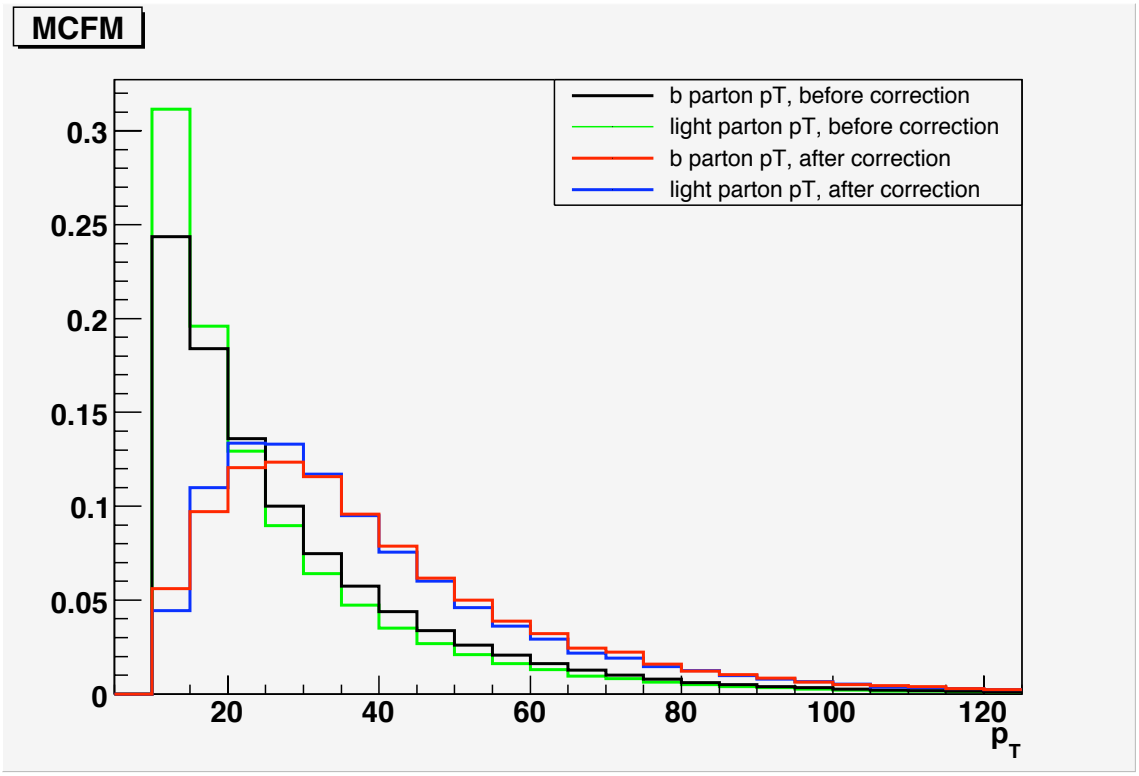


Figure 13.2: Folded and unfolded MCFM light and b parton p_T

$$\frac{MCFM^{b}_{folded}}{MCFM^{l}_{folded}} = \frac{MCFM^{b}_{parton}}{MCFM^{l}_{parton}} \frac{r_b^{\text{ALPGEN}}}{r_l^{\text{ALPGEN}}} \tag{13.1}$$

where r_b^{ALPGEN} is the ratio of matched b-spectrum versus parton p_T b-spectrum in ALPGEN and r_l^{ALPGEN} is the same for the light-spectra.

Both the original and the "folded" MCFM distributions for b and light partons are shown in figure 13.1. The reco jet > 20 GeV cut causes a large drop in the parton p_T bins less than 20 GeV, as expected. The contribution from partons with $p_T < 15$ GeV is 7% of the whole. To estimate the uncertainty due to this contribution, the $Z+b/Z+q$ ratio is calculated using the "folded" distributions and half of the difference from the MCFM parton $p_T > 20$ GeV ratio is taken as the uncertainty. The difference is found to be 3.6%. The central value for the new "folded" ratio is then used to compare to the result in data.

Chapter 14

Conclusions

This thesis presented a description of the detection and measurement of the $\sigma(Z + b\text{-jet})/\sigma(Z\text{+jet})$ ratio using 4.2 fb^{-1} of data from proton-antiproton collisions recorded by the DØ experiment at Fermi National Accelerator Laboratory at a center of mass energy, $\sqrt{s}$ of 1.96 TeV. The analysis yielded a measurement of $\sigma(Z + b\text{-jet})/\sigma(Z\text{+jet}) = 0.0193 \pm 0.0022 \pm 0.0015$ and is in good agreement with the theoretical prediction of $0.0185^{+0.0036}_{-0.0026}$ [1].

This analysis has more statistics and has less uncertainty than previous measurements made by both DØ [8] and CDF [9]. It uses more data and covers a larger region in jet η than either of the previous measurements and is the first vector boson plus jet analysis completed using RunIIb data at DØ. This analysis has helped to provide a better understanding of the b-quark density within a proton. It also narrowed down a significant background for $ZH \rightarrow llbb$ production and will aid in measuring the $\sigma(Z+b)$ cross section to better precision than ever before.

Appendix A

Maximum Likelihood Fitter

The maximum likelihood fitter used in this analysis takes the MJL template shape for b-, c- and light-jets as well as the distribution in the data as inputs. For the fitting, the sum of the b-, c- and light-jet contributions are normalized to the data. In other words, the sum of the fractions total one, as shown in equations A.1-A.3.

$$D = N_b + N_c + N_l \tag{A.1}$$

$$D = f_b D + f_c D + f_l D \tag{A.2}$$

$$1 = f_b + f_c + f_l \tag{A.3}$$

where D is the number of events in the data, N_b, N_c, N_l are the number of b-, c- and light-jet events in data and f_b, f_c, and f_l are the fractions of b-, c- and light-jet events in data.

Because of the binning (or discontinuous) makeup of the MJL templates, a Poissonian approach is taken by the maximum likelihood fitter. Equation A.4 shows the Poisson distribution, which computes the probability (P) that the number of actual events is r (number of MC events) given μ (number of observed events):

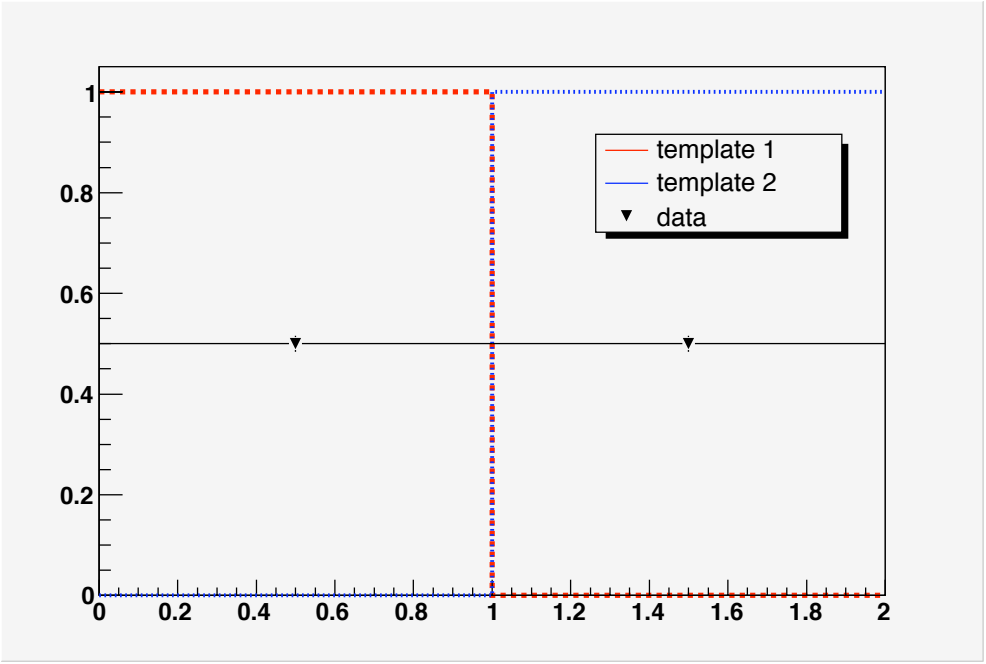


Figure A.1: Simple example of distribution and its two components, all normalized to unity.

$$P(r) = \frac{\mu^r e^{-\mu}}{r!} \tag{A.4}$$

where r is the number of MC events, which is normalized to data. The number r, however is also a combination of the f_b, f_c and f_l fractions. The fractions are then varied within the normalization. For each combination, a maximum likelihood value is calculated by maximizing a log-likelihood value over the number of bins (N) in the data, as shown in equation A.5:

$$L(\mu) = ln \prod_{i=1}^{N} \frac{\mu_i^{r_i} e^{-\mu_i}}{r_i!} \tag{A.5}$$

The value for r that maximizes equation A.5 is then returned as the combination of fractions that best describes the data.

A simplified example is shown in figure A.1. In this case, the data has an equal number of entries in the two bins that cover its range. One template has only entries in the first

bin and the other only has entries in the second bin. The stipulation that $D = N_1 + N_2$ is set, where N_1 is the number of events from template one, N_2 is the number of events from template two and D is the number of events in data. The two extreme cases where either N_1 or N_2 equal the number of events in data should give the same result, as shown in equation A.7. Whereas the case where $N_1 = N_2 = D/2$ gives a different result, as shown in equation A.8. The only difference between equations A.7 and A.8 is $D!$ vs. $(D/2)!^2$. For all values of D in this case, $(D/2)!^2 < D!$, giving the expected result that the $N_1 = N_2 = D/2$ case gives a higher log-likelihood than the $N_1 = D$, $N_2 = 0$ and the $N_2 = D$, $N_1 = 0$ cases.

$$L(\mu) = -\sum_{i=1}^{N} \mu_i + \sum_{i=0}^{N} r_i ln(\mu_i) - \sum_{i=0}^{N} ln(r_i!) \tag{A.6}$$

$$L(\mu) = -D + Dln(D/2) - ln(D!) = -D + Dln(D/2) - ln(D!) \tag{A.7}$$

$$L(\mu) = -D + 2D/2ln(D/2) - 2ln((D/2)!) \tag{A.8}$$

To find extrema of $L(\mu)$, set the first derivative to zero (equation A.9). As expected, the $N_1 = N_2 = D/2$ formulation gives an extremum ($r_i = \mu_i$).

$$\frac{dL(\mu)}{d\mu} = -N + \sum_{i=0}^{N} \frac{r_i}{\mu_i} = 0 \tag{A.9}$$

And, finally, check that the second derivative is negative for the above value:

$$\frac{\partial^2 L}{\partial \mu^2} = -\sum_{i=1}^{N} \frac{r_i}{\mu_i^2} \tag{A.10}$$

The second derivative is negative for all positive values of r_i, thus this extremum is a maximum.

Both the statistical uncertainty in the data and the templates contribute to the uncertainty in the fit. Each bin for the data and templates is fluctuated within its statistical uncertainty, and then the maximum likelihood is remeasured. Uncertainty due to data statistics is considered statistical, while uncertainty due to template statistics is taken as systematic uncertainty. Clearly, low statistics can cause a high uncertainty in the fitter. This is the motivation behind increasing the template statistics.

Appendix B

Medium Operating Point

This appendix shows the distributions and maximum likelihood fitter result for the operating point NN > 0.65. Figure B.1 shows the MJL template for the medium (NN > 0.65) operating point. The separation of b-jets from the others is still strong, however the separating power between c- and light-jets has diminished considerably.

Table B.1 shows the fractions for the ee, $\mu\mu$ and combined channels given by the maximum likelihood fitter, and table B.2 shows the results for the $\sigma(z + b\text{-jet})/\sigma(z\text{+jet})$ ratio for each channel. Figures B.2 - B.5 show the jet p_T and MJL distributions from the maximum likelihood fitter output for the combined, ee and $\mu\mu$ channels. All three channels are in excellent agreement with the NN > 0.5 operating point used for this analysis.

Events	788	534	1321
Channel	ee	$\mu\mu$	Combined
Z+b	0.298 ± 0.044	0.277 ± 0.047	0.290 ± 0.031
Z+c	0.505 ± 0.068	0.401 ± 0.077	0.461 ± 0.055
Z+l	0.197 ± 0.053	0.322 ± 0.064	0.249 ± 0.044

Table B.1: Fractions from maximum likelihood fitter for ee and $\mu\mu$ and combined channels for $|\eta| < 2.5$ region and NN > 0.65 operating point.

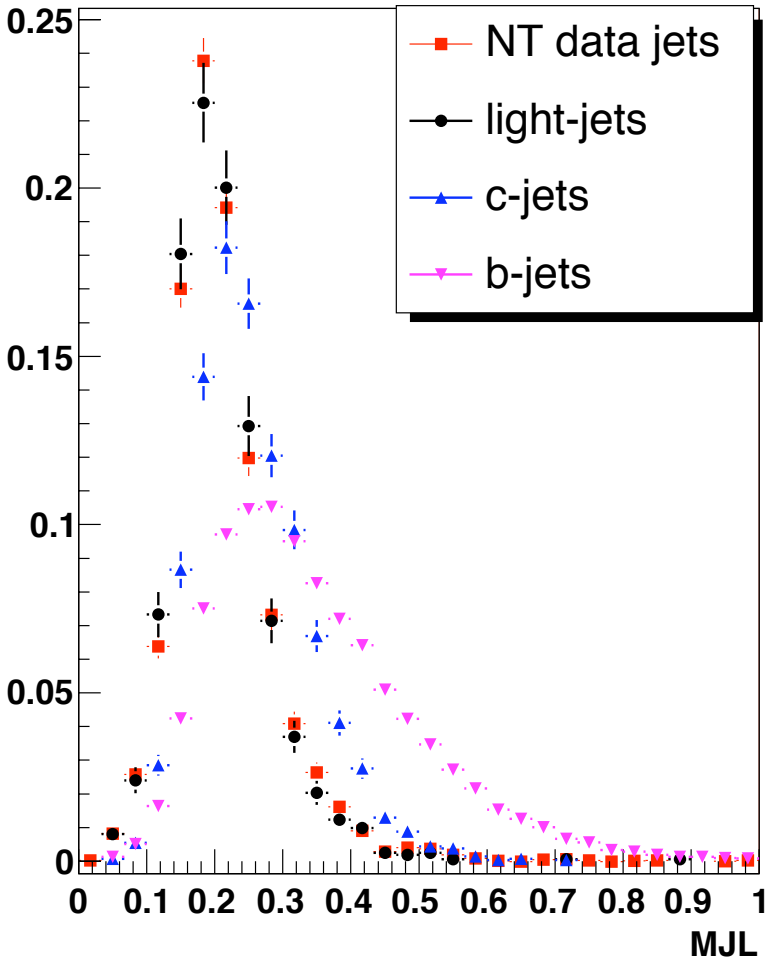


Figure B.1: MJL template for NT data, MC b-, c- and light-jets for NN $>$ .65 operating point

Channel	$\sigma(z+b)/\sigma(z+jet)$ Ratio
ee	0.0197±0.0031±0.0006
$\mu\mu$	0.0191±0.0033±0.0006
Combined	0.0195±0.0023±0.0004

Table B.2: $\sigma(z+b)/\sigma(z+jet)$ ratio for $|\eta| < 2.5$ for region and NN > 0.65 operating point.

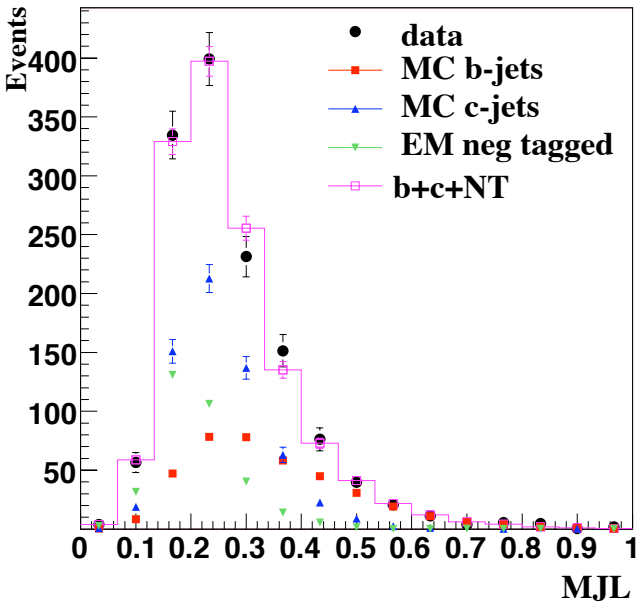


Figure B.2: MJL distribution from maximum likelihood fitter for $|\eta| < 2.5$ and NN > 0.65 region

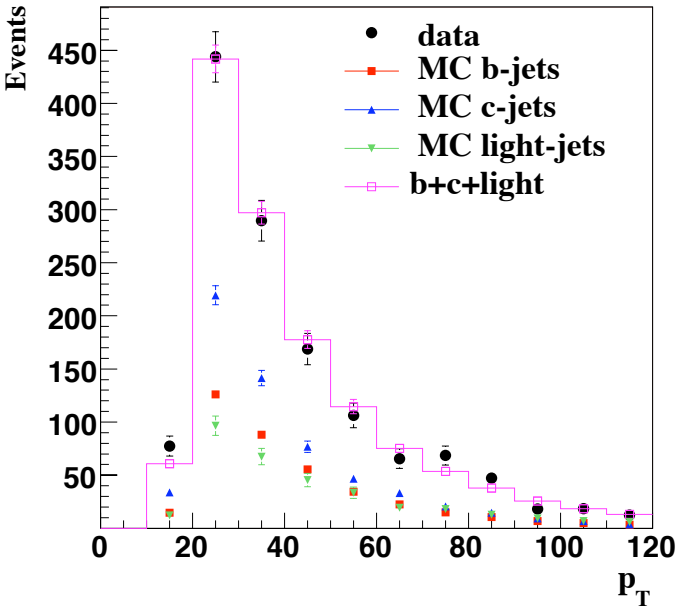
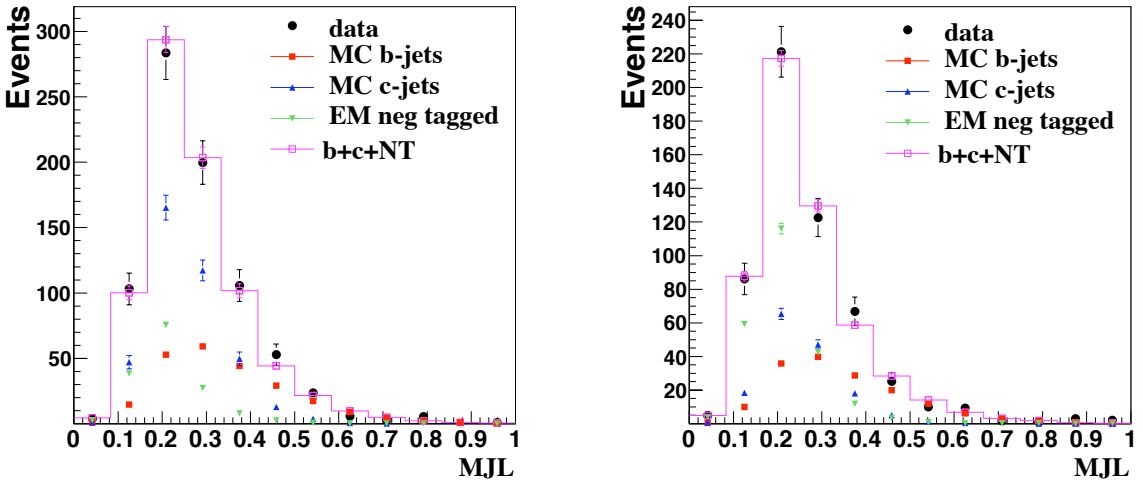


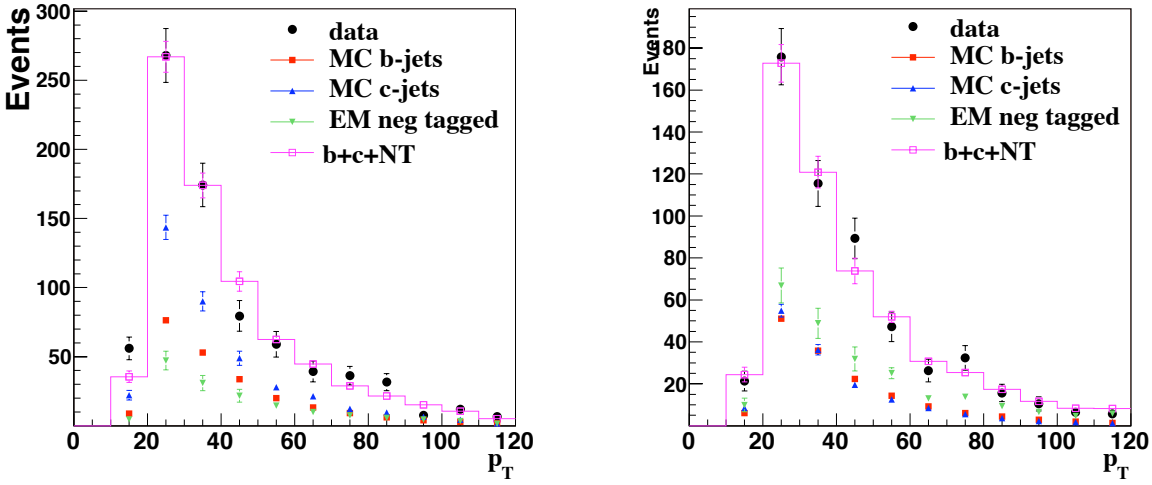
Figure B.3: Jet p_T distribution from maximum likelihood fitter for $|\eta| < 2.5$ and NN > 0.65 region



(a) ee MJL

(b) $\mu\mu$ MJL

Figure B.4: MJL distributions for ee and $\mu\mu$ channels in the $|\eta| < 2.5$ region for NN > 0.65 cut.



(a) ee jet p_T

(b) $\mu\mu$ jet p_T

Figure B.5: MJL distributions for ee and $\mu\mu$ channels in the $|\eta| < 2.5$ region for NN > 0.65 cut.

B.1 MJL Dependence on NN Scale Factors

When applying a cut on the NN tagger, a data-to-MC scale factor that depends on the cut made has to be applied. The scale factor applied for an NN cut of > 0.5 is different than the scale factor applied for an NN cut of > 0.65. The application of the scale factors can then be carried out in two ways. One way, the method used in this analysis, is to simply use the scale factor for each jet given by the NN > 0.5 operating point. The other way is to apply the scale factor of the region that the NN value of the jet falls in. In other words, if the NN value of the jet falls between $0.5 <$ NN < 0.65, then the scale factor exclusively for that region is applied. The exclusive scale factor is different than the scale factor given by the NN tagger. The scale factor given by the NN tagger is the data-to-MC correction for any jet that has an NN value greater than the operating point used, while the exclusive scale factor for a given operating point, i is calculated using equation B.1:

$$SF_{excl}[i] = \frac{SF_{NN}[i] * \epsilon^{MC}_{NN}[i] - SF_{NN}[i+1] * \epsilon^{MC}_{NN}[i+1]}{\epsilon^{MC}_{NN}[i] - \epsilon^{MC}_{NN}[i+1]} \tag{B.1}$$

where $SF_{NN}[i]$ is the scale factor given by the NN tagger for the operating point i, and $\epsilon^{MC}_{NN}[i]$ is the NN tagger efficiency of that operating point.

Figure B.6 shows the MJL template for b-jets with the NN > 0.5 scale factors applied to the jets, as well as the reweighted template with the exclusive scale factors applied to each jet. The templates are very similar, and result in a 0.4% difference in the central value of the $Z + b$ fraction.

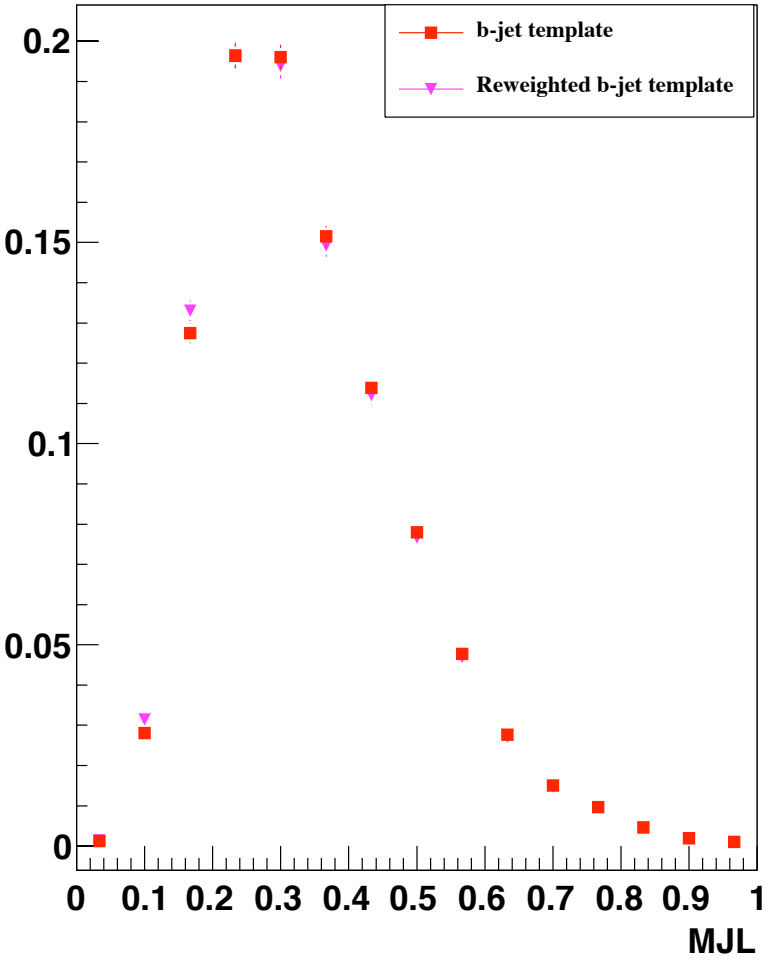


Figure B.6: MJL template for b-jets with NN > 0.5 scale factor and exclusive scale factor ("reweighed") applied.

Appendix C

Study of p_T Dependence of MJL Templates for b-jets

Figure C.1 shows the MJL shapes for b-jets from different sources and their corresponding p_T distributions. Although the p_T spectra for b-jets are different, the MJL shapes are quite stable.

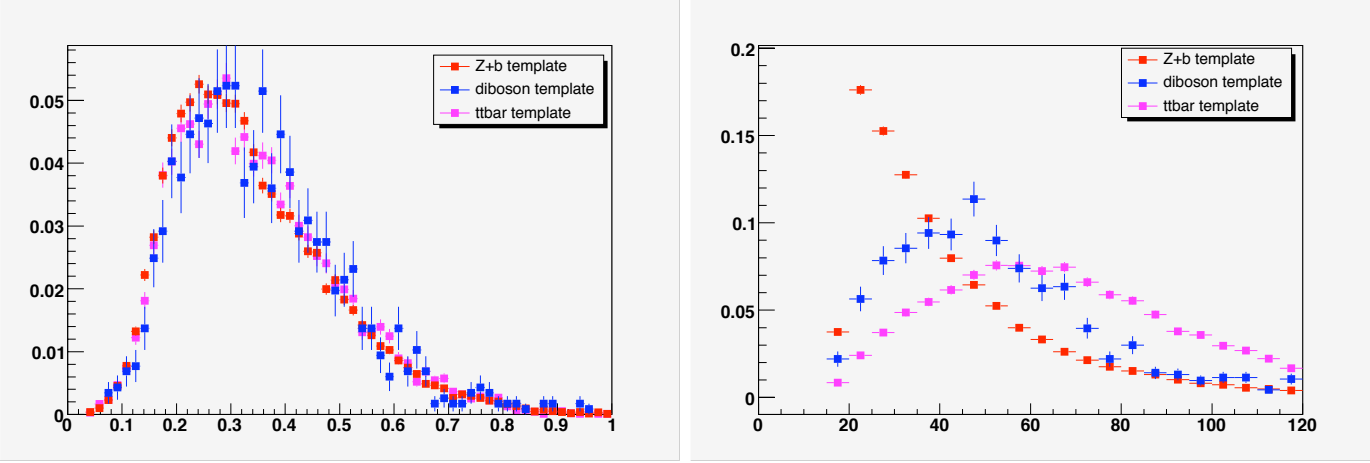


Figure C.1: a) MJL and b) p_T distributions for b-jets coming from $Z + b$, diboson and $t\bar{t}$ samples. The jets pass the $NN > 0.5$ requirement.

Figure C.2 shows the MJL templates for $Z + b$-jets with $p_T > 20$ GeV and $p_T < 35$ GeV, and figure C.3 shows the template for b-jets with $p_T < 50$ GeV. The three regions are in

good agreement and display the lack of a large p_T dependence in the MJL discriminant.

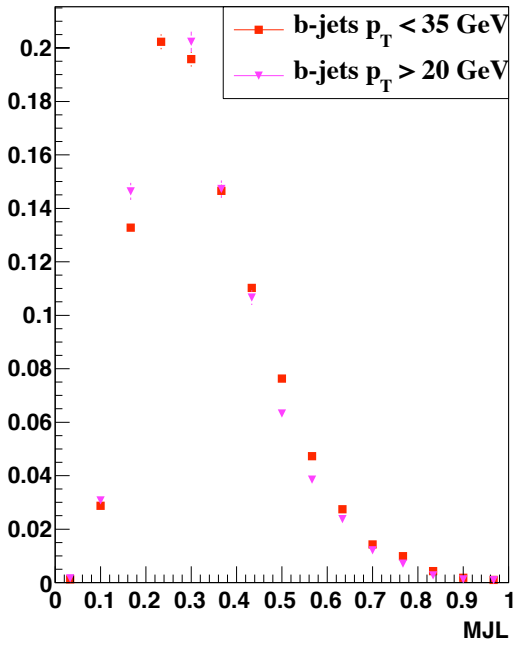


Figure C.2: MJL template for b-jets with (a) $p_T > 20$ GeV and (b) $p_T < 35$ GeV

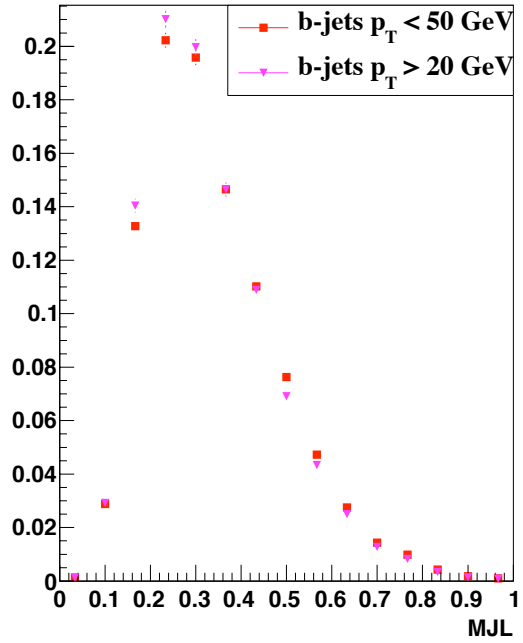


Figure C.3: MJL template for b-jets with (a) $p_T > 20$ GeV and (b) $p_T < 50$ GeV

Appendix D

Jet p_T distributions Across Channels

D.1 Distributions from Monte Carlo

Figures D.1-D.6 show the leading jet p_T spectra for b-, c-, and light-jets in the dielectron and dimuon channels before and after tagging. They are very consistent across channels, as expected.

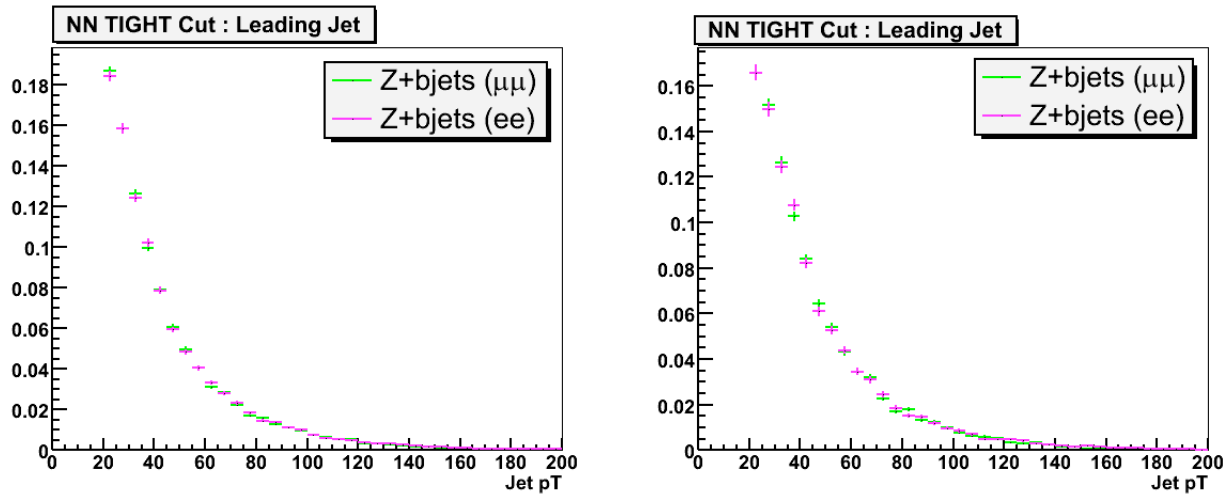


Figure D.1: The b-jet p_T distributions in the dielectron and dimuon channels before (left) and after tagging (right). For the tagging, $NN > 0.775$ requirement has been used.

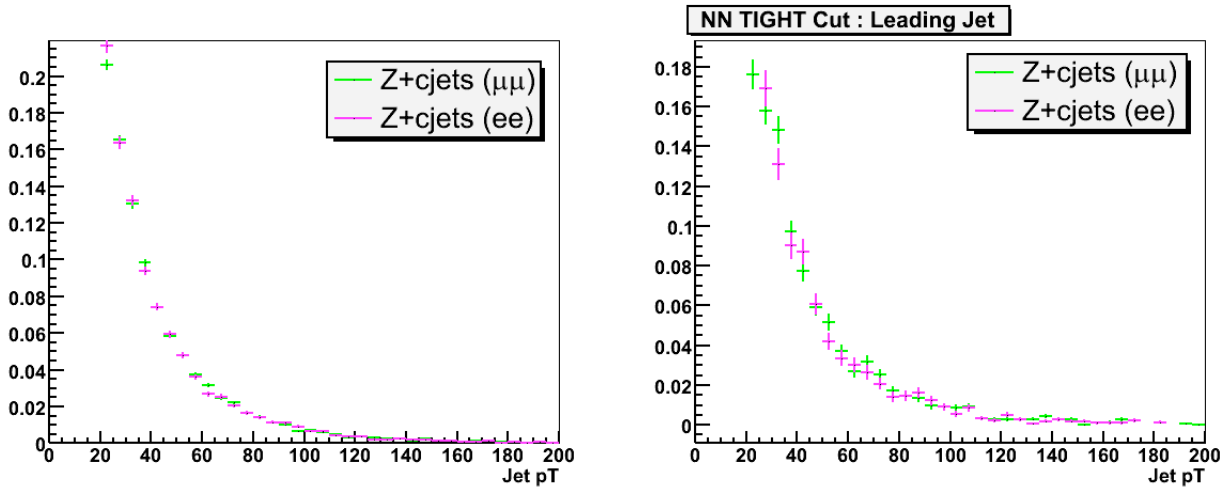


Figure D.2: The c-jet p_T distributions in the dielectron and dimuon channels before (left) and after tagging (right). For the tagging, $NN > 0.775$ requirement has been used.

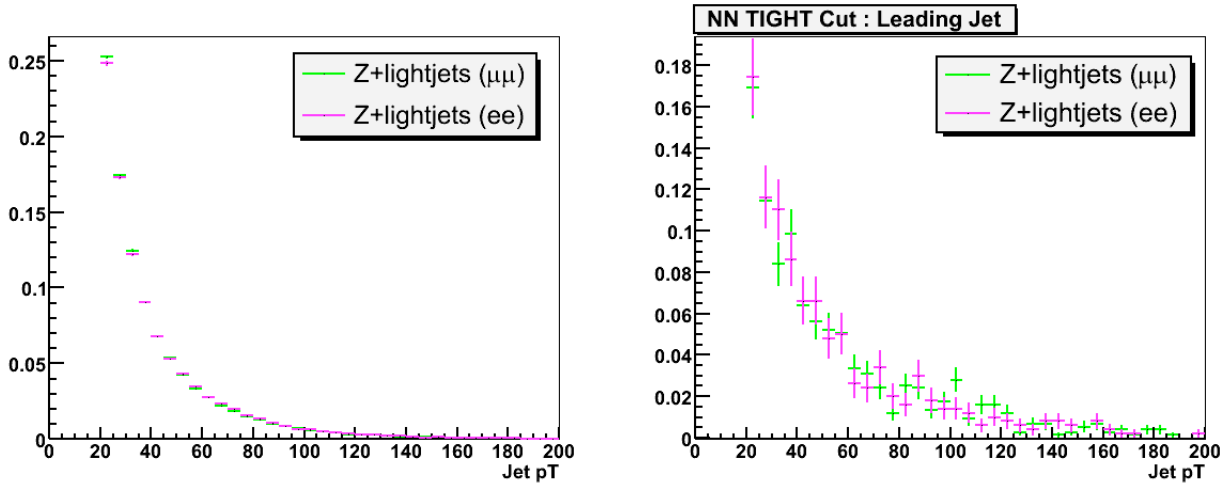


Figure D.3: The light-jet p_T distributions in the dielectron and dimuon channels before (left) and after tagging (right). For the tagging, $NN > 0.775$ requirement has been used.

D.2 Distributions from Data

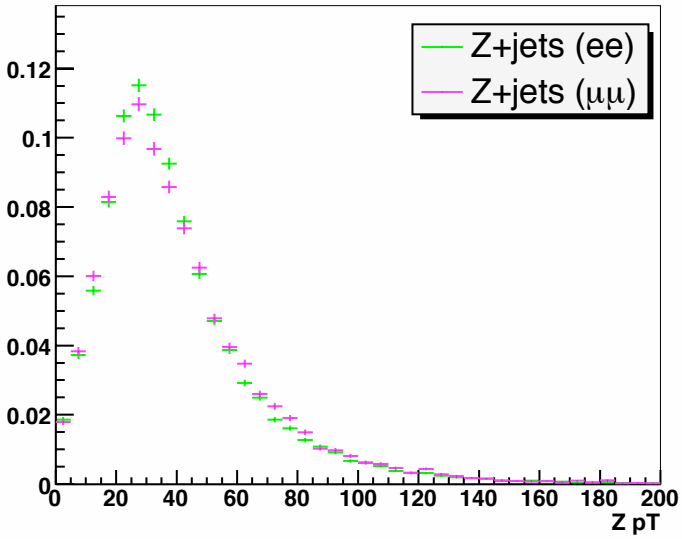


Figure D.4: The Zp_T distributions in the dielectron and dimuon channels in data before tagging.

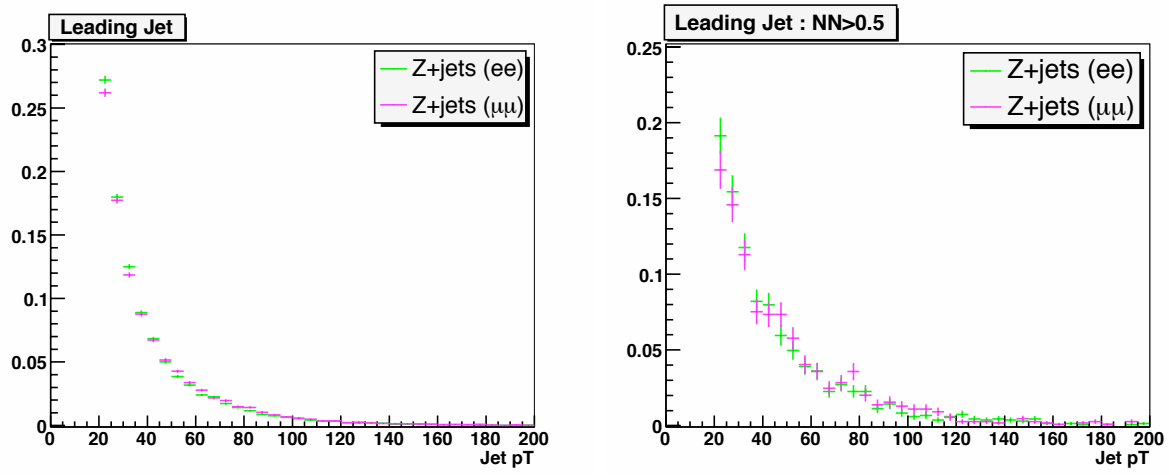


Figure D.5: The leading jet p_T distributions in the dielectron and dimuon channels in data before (left) and after tagging (right). For the tagging, $NN > 0.5$ requirement has been used.

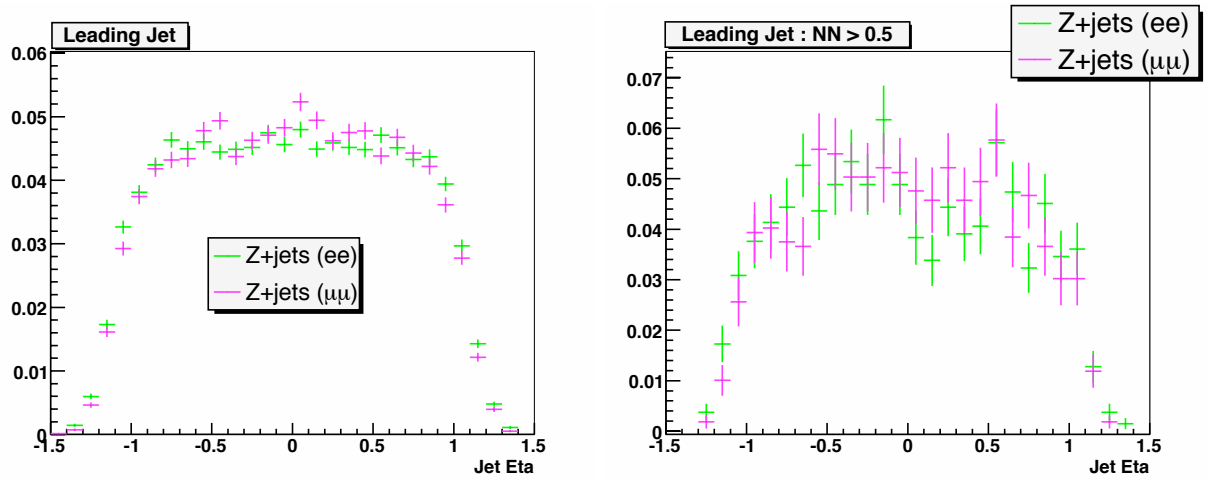


Figure D.6: The leading jet η distributions in the dielectron and dimuon channels in data before (left) and after tagging (right). For the tagging, $NN > 0.5$ requirement has been used.

Appendix E

Negative Tagged Jets

As discussed in the Chapter 11, the MJL template for NT jets from data has been used for the light-jet template. Before the NT jets sample can be used however, the b-contamination must be removed. Figure E.1 shows the MJL template shapes for NT jets in the $Z+b$, $Z+c$, and $Z+l$ MC. Using these templates, the b-contamination in the data can be measured with a maximum likelihood fitter. The fractions were found to be:

	Fraction
NT b	0.176 ± 0.017
NT c	0.218 ± 0.042
NT light	0.607 ± 0.034

Table E.1: Fractions from maximum likelihood fitter for NT b-, NT c- and NT light-jets in NT jets from data. The NT b-, NT c- and NT light-jet templates were taken from MC.

Figure E.2 shows the MJL distribution in NT jet in data along with contributions from b-, c- and light-jets given by the maximum likelihood fitter.

Using NT data with the b-contamination removed is based on the assumption that the MJL templates of NT light-jets and tagged light-jets are similar. Figure E.3 shows the distributions for both tagged Z+light-jet and NT Z+light-jet samples. The distributions are

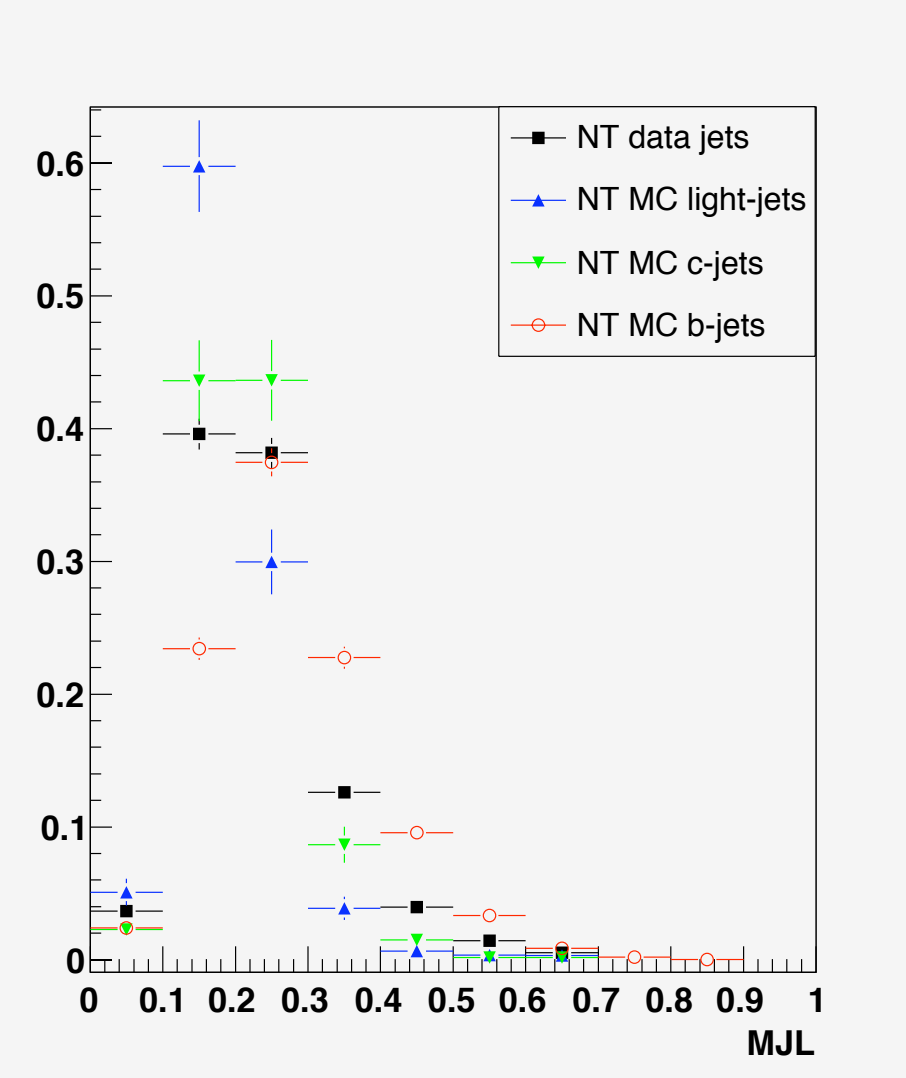


Figure E.1: The MJL distributions for b-, c- and light-jets obtained from Monte Carlo negative tagged jets along with negative tagged jets in data.

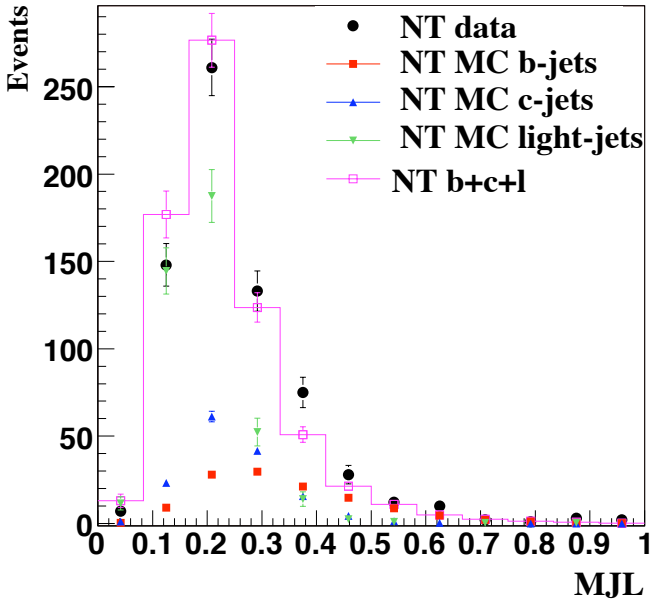


Figure E.2: The MJL distribution in NT data along with contributions from NT b-, c- and light-jets.

almost identical. For further inspection, a generic MC multijet sample was compared to that of NT jets in data (figure E.4). Both shapes show the "tail" from a large b-contamination.

Finally, the multijet sample was used to check the accuracy of the maximum likelihood fitter. Templates from negative tagged b-, c- and light-jets were used to measure the fractions of each in the negative tagged multijet sample. The maximum likelihood fitter gave a b-contamination of $36 \pm 4\%$, while MC truth gave a b-contamination of $\sim 34\%$, well within the range given by the maximum likelihood fitter.

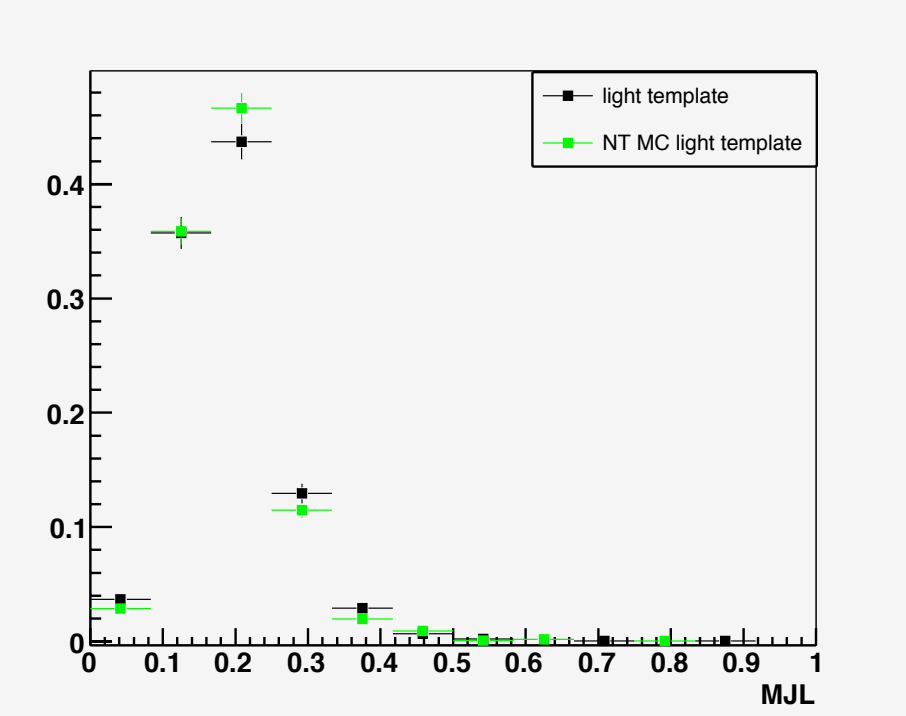


Figure E.3: The MJL distributions for NT MC light-jets and tagged MC light-jets.

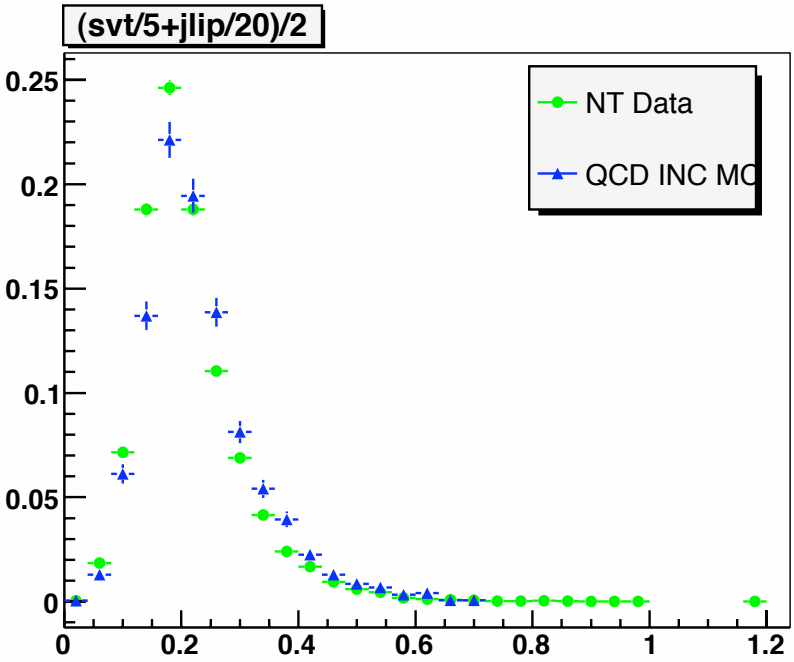


Figure E.4: The MJL distributions for NT MC multijet sample and NT jets in data.

Appendix F

Single Muon Trigger List

Information for each trigger level is embedded in each of the names in table F.1. For example, MUHI2_TMM10 requires a Level 1 track > 10 GeV.

Single Muon Trigger List
MUHI1_ITLM10_a
MUHI1_TK12_TLM12
MUHI1_ILM15
MUHI2_ITLM10_a
MUHI2_TK12_TLM12
MUHI2_ILM15
MUHI3_ITLM10
MUHI3_TK12_TLM12
MUHI3_ILM15
MUHI1_ITLM10_b
MUHI2_ITLM10_b
MUHI1_ILM10
MUHI2_ILM10
MUHI1_TLM12
MUHI2_TLM12
MUHI1_MM10
MUHI1_TMM10
MUHI2_MM10
MUHI2_TMM10

Table F.1: Single muon triggers used in analysis.

Appendix G

Single Electron Trigger List

Information for each trigger level is imbedded in each of the names in table G.1 and table G.2. For example, E1_L70 requires a Level 1 EM object with $E_T > 12$ GeV.

Single Electron	Trigger List
E1_ISH30	TE1_L80_NOLUM
E1_ISHT15_TK13	TE3_LH2ISH24
E1_ISHT22	TE3_L80_NOLUM
E1_L70	TE3_L80
E1_L80	TE3_L70
E1_LH2ISH24	TE3_ISHT22
E1_LH2ISHT17T14	TE3_ISHT15_TK13
E1_LH2L70	TE3_ISH30
E1_LH2SH27	TE2_T15SH20
E1_LH3ISH25	TE2_T14LH2SH17
E1_LH3SH27	TE2_T13SHT15
E1_SH35	TE2_SHT50
E1_SH60	TE2_SHT25
E1_SHT25	TE2_SH60
E1_SHT27	TE2_SH35
E1_SHT27_NOLUM	TE2_LH2SH27
E1_SHT50	TE2_LH2L70
E1_T13SHT15	TE2_LH2ISHT17T14
E1_T14LH2SH17	TE2_LH2ISH24
E1_T15SH20	TE2_L80
E2_ISH30	TE2_L70
E2_ISHT15_TK13	TE2_ISHT22
E2_ISHT22	TE2_ISHT15_TK13
E2_L70	TE2_ISH30
E2_L80	TE1_T15SH20
E2_LH2ISH24	TE1_T14LH2SH17_NOLUM
E2_LH2ISHT17T14	TE1_T14LH2SH17
E2_LH2L70	TE1_T13SHT15
E2_LH2SH27	TE1_SHT50_NOLUM
E2_LH3ISH25	TE1_SHT50
E2_LH3SH27	TE1_SHT25_NOLUM
E2_SH35	TE1_SHT25
E2_SH60	TE1_SH60_NOLUM
E2_SHT25	TE1_SH60
E2_SHT27	TE1_SH35
E2_SHT50	TE1_LH2SH27_NOLUM
E2_T13SHT15	TE1_LH2SH27
E2_T14LH2SH17	TE1_LH2L70_NOLUM
E2_T15SH20	TE1_LH2L70
TE1_ISH30	TE1_LH2ISHT17T14_NOLUM
TE1_ISHT15_TK13	TE1_LH2ISHT17T14
TE1_ISHT22	TE1_LH2ISH24_NOLUM
TE1_L70	TE1_LH2ISH24
TE1_L80	TE1_L80_NOLUM
TE3_LH2ISH24_NOLUM	TE3_LH2ISHT17T14
TE3_LH2ISHT17T14_NOLUM	TE3_LH2L70

Table G.1: Single electron triggers used in analysis.

TE3_LH2L70_NOLUM	TE3_LH2SH27
TE3_LH2SH27_NOLUM	TE3_SH35
TE3_SH60	TE3_SHT25
TE3_SHT25_NOLUM	TE3_SHT50
TE3_SHT50_NOLUM	TE3_T13SHT15
TE3_T14LH2SH17	TE3_T14LH2SH17_NOLUM
TE3_T15SH20	TE4_ISH30
TE4_ISHT15_TK13	TE4_ISHT22
TE4_L70	TE4_L80
TE4_L80_NOLUM	TE4_LH2ISH24
TE4_LH2ISH24_NOLUM	TE4_LH2ISHT17T14
TE4_LH2ISHT17T14_NOLUM	TE4_LH2L70
TE4_LH2L70_NOLUM	TE4_LH2SH27
TE4_LH2SH27_NOLUM	TE4_SH35
TE4_SH60	TE4_SHT25
TE4_SHT25_NOLUM	TE4_SHT50
TE4_SHT50_NOLUM	TE4_T13SHT15
TE4_T14LH2SH17	TE4_T14LH2SH17_NOLUM
TE4_T15SH20	TE5_ISH30
TE5_ISHT15_TK13	TE5_ISHT22
TE5_L70	TE5_L80
TE5_LH2ISH24	TE5_LH2ISHT17T14
TE5_LH2L70	TE5_LH2SH27
TE5_SH35	TE5_SH60
TE5_SHT25	TE5_SHT50
TE5_T13SHT15	TE5_T14LH2SH17
TE5_T15SH20	

Table G.2: Single electron triggers used in analysis, continued.

Appendix H

MC Samples

Tables H.1 - H.2 show the MC samples used in this analysis.

Process and Mass Range (GeV)	$\sigma\times$ BR (pb)	Generator	Events
$Z/\gamma^* + 0lp \to \mu\mu$ $15 < M(Z) < 75$	338.363	Alpgen+ Pythia	1805548
$Z/\gamma^* + 0lp \to \mu\mu$ $75 < M(Z) < 130$	133.103	Alpgen+ Pythia	1036678
$Z/\gamma^* + 0lp \to \mu\mu$ $130 < M(Z) < 250$	1.335	Alpgen+ Pythia	271249
$Z/\gamma^* + 0lp \to \mu\mu$ $M(Z) > 250$	0.072	Alpgen+ Pythia	541752
$Z/\gamma^* + 1lp \to \mu\mu$ $15 < M(Z) < 75$	39.942	Alpgen+ Pythia	336634
$Z/\gamma^* + 1lp \to \mu\mu$ $75 < M(Z) < 130$	40.677	Alpgen+ Pythia	616917
$Z/\gamma^* + 1lp \to \mu\mu$ $130 < M(Z) < 250$	0.931	Alpgen+ Pythia	173777
$Z/\gamma^* + 1lp \to \mu\mu$ $M(Z) > 250$	0.036	Alpgen+ Pythia	450281
$Z/\gamma^* + 2lp \to \mu\mu$ $15 < M(Z) < 75$	9.870	Alpgen+ Pythia	197528
$Z/\gamma^* + 2lp \to \mu\mu$ $75 < M(Z) < 130$	9.7514	Alpgen+ Pythia	432267
$Z/\gamma^* + 2lp \to \mu\mu$ $130 < M(Z) < 250$	0.480	Alpgen+ Pythia	173118
$Z/\gamma^* + 2lp \to \mu\mu$ $M(Z) > 250$	0.011	Alpgen+ Pythia	354830
$Z/\gamma^* + 3lp \to \mu\mu$ $15 < M(Z) < 75$	2.739	Alpgen+ Pythia	317512
$Z/\gamma^* + 3lp \to \mu\mu$ $75 < M(Z) < 130$	3.1148	Alpgen+ Pythia	166902
$Z/\gamma^* + 3lp \to \mu\mu$ $130 < M(Z) < 250$	0.053	Alpgen+ Pythia	164316
$Z/\gamma^* + 3lp \to \mu\mu$ $M(Z) > 250$	0.0066	Alpgen+ Pythia	354415
$Z/\gamma^* + 2b + 0lp \to \mu\mu$ $15 < M(Z) < 75$	0.5087	Alpgen+ Pythia	188546
$Z/\gamma^* + 2b + 0lp \to \mu\mu$ $75 < M(Z) < 130$	0.4175	Alpgen+ Pythia	179558
$Z/\gamma^* + 2b + 0lp \to \mu\mu$ $130 < M(Z) < 250$	0.0034	Alpgen+ Pythia	90624
$Z/\gamma^* + 2b + 0lp \to \mu\mu$ $M(Z) > 250$	0.00034	Alpgen+ Pythia	178546
$Z/\gamma^* + 2b + 1lp \to \mu\mu$ $15 < M(Z) < 75$	0.1988	Alpgen+ Pythia	92018
$Z/\gamma^* + 2b + 1lp \to \mu\mu$ $75 < M(Z) < 130$	0.1904	Alpgen+ Pythia	96941
$Z/\gamma^* + 2b + 1lp \to \mu\mu$ $130 < M(Z) < 250$	0.0018	Alpgen+ Pythia	46222
$Z/\gamma^* + 2b + 1lp \to \mu\mu$ $M(Z) > 250$	0.00018	Alpgen+ Pythia	180821
$Z/\gamma^* + 2b + 2lp \to \mu\mu$ $15 < M(Z) < 75$	0.01835	Alpgen+ Pythia	89922
$Z/\gamma^* + 2b + 2lp \to \mu\mu$ $75 < M(Z) < 130$	0.09859	Alpgen+ Pythia	47925
$Z/\gamma^* + 2b + 2lp \to \mu\mu$ $130 < M(Z) < 250$	0.00088	Alpgen+ Pythia	45903
$Z/\gamma^* + 2b + 2lp \to \mu\mu$ $M(Z) > 250$	0.0001	Alpgen+ Pythia	187820
$Z/\gamma^* + 2c + 0lp \to \mu\mu$ $15 < M(Z) < 75$	4.144	Alpgen+ Pythia	180580
$Z/\gamma^* + 2c + 0lp \to \mu\mu$ $75 < M(Z) < 130$	0.9322	Alpgen+ Pythia	187636
$Z/\gamma^* + 2c + 0lp \to \mu\mu$ $130 < M(Z) < 250$	0.0076	Alpgen+ Pythia	79493
$Z/\gamma^* + 2c + 0lp \to \mu\mu$ $M(Z) > 250$	0.00062	Alpgen+ Pythia	191017
$Z/\gamma^* + 2c + 1lp \to \mu\mu$ $15 < M(Z) < 75$	0.9530	Alpgen+ Pythia	93093
$Z/\gamma^* + 2c + 1lp \to \mu\mu$ $75 < M(Z) < 130$	0.5481	Alpgen+ Pythia	187636
$Z/\gamma^* + 2c + 1lp \to \mu\mu$ $130 < M(Z) < 250$	0.00439	Alpgen+ Pythia	45857
$Z/\gamma^* + 2c + 1lp \to \mu\mu$ $M(Z) > 250$	0.00044	Alpgen+ Pythia	170822
$Z/\gamma^* + 2c + 2lp \to \mu\mu$ $15 < M(Z) < 75$	0.3431	Alpgen+ Pythia	95436
$Z/\gamma^* + 2c + 2lp \to \mu\mu$ $75 < M(Z) < 130$	0.0985	Alpgen+ Pythia	49608
$Z/\gamma^* + 2c + 2lp \to \mu\mu$ $130 < M(Z) < 250$	0.00283	Alpgen+ Pythia	47946
$Z/\gamma^* + 2c + 2lp \to \mu\mu$ $M(Z) > 250$	0.00026	Alpgen+ Pythia	181158
$ZZ \to 2j2\ell$	0.226	Pythia	105325
$WZ \to 2j2\ell$	0.275	Pythia	273344
$WW \to incl$	12.35	Pythia	675814
$tt \to 2b2\ell2\nu + 0lp$	0.33	Alpgen+Pythia	749642
$tt \to 2b2\ell2\nu + 1lp$	0.14	Alpgen+Pythia	452177
$tt \to 2b2\ell2\nu + 2lp$	0.07	Alpgen+Pythia	281453

Table H.1: The MC samples used and their cross sections (before any scaling corrections) and number of events generated after removing events failing DQ cuts. Here, ℓ stands for any of the charged leptons: e or μ. "lp" stands for "light-parton" in Alpgen and indicates the number of hard jets at the parton level for each Alpgen+Pythia sub-sample. "excl." means that the sub-sample is exclusive and requires exactly the number of light partons indicated, whereas "incl." denotes that all higher parton multiplicity states were also allowed. Some samples used have been processed through a HF skimming in order to avoid double counting of b- and c-quark production.

sample	cross-section(xBR)	num. events
Z+0lp excl. (M_Z=15-75 GeV)	337 pb	1580k
Z+0lp excl. (M_Z=75-130 GeV)	132 pb	1010k
Z+0lp excl. (M_Z=130-250 GeV)	0.891 pb	350k
Z+1lp excl. (M_Z=15-75 GeV)	40.0 pb	1580k
Z+1lp excl. (M_Z=75-130 GeV)	40.1 pb	570k
Z+1lp excl. (M_Z=130-250 GeV)	0.376 pb	180k
Z+2lp excl. (M_Z=15-75 GeV)	9.95 pb	550k
Z+2lp excl. (M_Z=75-130 GeV)	9.98 pb	270k
Z+2lp excl. (M_Z=130-250 GeV)	102 fb	160k
Z+3lp incl. (M_Z=15-75 GeV)	2.77 pb	540k
Z+3lp incl. (M_Z=75-130 GeV)	3.30 pb	130k
Z+3lp incl. (M_Z=130-250 GeV)	33.1 fb	300k
Z+2b+0lp excl. (M_Z=15-75 GeV)	0.518 pb	170k
Z+2b+0lp excl. (M_Z=75-130 GeV)	0.401 pb	200k
Z+2b+0lp excl. (M_Z=130-250 GeV)	3.40 fb	90k
Z+2b+1lp excl. (M_Z=15-75 GeV)	0.207 pb	85k
Z+2b+1lp excl. (M_Z=75-130 GeV)	0.173 pb	93k
Z+2b+1lp excl. (M_Z=130-250 GeV)	1.67 fb	45k
Z+2b+2lp incl. (M_Z=15-75 GeV)	78.3 fb	82k
Z+2b+2lp incl. (M_Z=75-130 GeV)	107 fb	44k
Z+2b+2lp incl. (M_Z=130-250 GeV)	0.938 fb	44k
Z+2c+0lp excl. (M_Z=15-75 GeV)	4.09 pb	180k
Z+2c+0lp excl. (M_Z=75-130 GeV)	0.901 pb	180k
Z+2c+0lp excl. (M_Z=130-250 GeV)	7.50 fb	180k
Z+2c+1lp excl. (M_Z=15-75 GeV)	1.03 pb	180k
Z+2c+1lp excl. (M_Z=75-130 GeV)	0.506 pb	90k
Z+2c+1lp excl. (M_Z=130-250 GeV)	4.33 fb	90k
Z+2c+2lp incl. (M_Z=15-75 GeV)	0.382 pb	170k
Z+2c+2lp incl. (M_Z=75-130 GeV)	0.286 pb	50k
Z+2c+2lp incl. (M_Z=130-250 GeV)	2.67 fb	90k

Table H.2: The MC samples used, their cross sections (before any scaling corrections) and the number of events generated after removing events failing DQ cuts. Here, ℓ stands for any of the charged leptons: e or μ. "lp" stands for "light-parton" in ALPGEN and indicates the number of hard jets at the parton level for each ALPGEN+PYTHIA sub-sample. "excl." means that the sub-sample is exclusive and requires exactly the number of light partons indicated, whereas "incl." denotes that all higher parton multiplicity states were also allowed. Some samples used have been processed through a HF skimming in order to avoid double counting of b and c quark production.

Bibliography

[1] J.M. Campbell, R.K. Ellis, F. Maltoni, and S. Willenbrock, Phys. Rev. D **69**, 074021 (2004); Phys. Rev. D **73**, 054007 (2006).

[2] A. Djouadi, K. Kalinowski, M. Spira, Comp. Phys. Commun. 108C, 56 (1998)

[3] M.A. Aivazis, J.C. Collins, F.I. Olness and W.K. Tung, Phys. Rev. D **50**, 3102 (1994); J.C. Collins, Phys. Rev. D **58**, 094002 (1998); J. Pumplin et. al., JHEP **0207**, 012 (2002).

[4] T. Stelzer, Z. Sullivan, and S. Willenbrock, Phys. Rev. D **56**, 5919 (1997);**58**, 094021 (1998).

[5] D. Dicus, T. Stelzer, Z. Sullivan, and S. Willenbrock, Phys. Rev. D **59**, 094016 (1999); C. Balazs, H. J. He, and C. P. Yuan, Phys. Rev. D **60**, 114001 (1999); F. Maltoni, Z. Sullivan, and S. Willenbrock, Phys. Rev. D **67**, 093005 (2003).

[6] C. S. Huang and S. H. Zhu, Phys. Rev. D **60**, 075012 (1999); J. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, Phys. Rev. D **67**, 095002 (2003).

[7] F.F. Cordero, L. Reina, and D. Wackeroth, Phys. Rev. D **78**, 074014 (2008).

[8] V. Abazov et al. (D0 Collaboration), Phys. Rev. Lett. **94**, 161801 (2005).

[9] A. Abulencia et al. (CDF Collaboration), Phys. Rev. D **79**, 052008 (2009).

[10] L.Ancu *et al.*. "Search for $ZH \rightarrow llbb$ in 4.2fb^{-1}of Data Using a Random Forest in Dimuon and Dielectron Events", Internal D0 Note 6071, June 2010.

[11] V. Abazov et al. (D0 Collaboration), Phy. Lett. B. **658**, 112 (2008).

[12] L. M. Lederman, "The Tevatron", Scientific American, **264(3)**, 48 (1991).

[13] H. T. Edwards, "The Tevatron Energy Doubler: A Superconducting Accelerator", Ann. Rev. Nucl. Part. Sci., **35**, 605 (1985).

[14] TeVI Group, "Design Report Tevatron I projects", Fermilab Technical Report FERMILAB-DESIGN-1984-01 (1984).

[15] F. T. Cole *et al.*, "A Report on the Design of the Fermi National Accelerator Laboratory Superconducting Accelerator", Beams Document 1888-v1 (1979).

[16] J. Thompson, "Introduction to Colliding Beams at Fermilab", Fermilab Technical Report FermiLab TM-1909 (1994).

[17] C. W. Schmidt and C. D. Curtis, "A 50 mA Negative Hydrogen-Ion Source", IEEE Transactions on Nuclear Science, NS-26, 4120 (1979).

[18] C. D. Curtis *et al.*, "Linac H-Beam Operation and Uses at FermiLab", IEEE Transactions on Nuclear Science, NS-26, 3760 (1979).

[19] E. L. Hubbard *et al.*, "Booster synchrotron", Fermilab Technical Report Fermilab TM-405 (1973).

[20] Jim Morgan, *et al*, "Antiproton Source Rookie Book", $http : //www - bdnew.fnal.gov/operations/rookie_books/Pbar_v2.pdf$

[21] C. S. Mishra, "The Fermilab Main Injector", Fermilab Technical Report Fermilab Conf-92/372 (1992).

[22] S. Nagaitsev *et al.*, "Antiproton Cooling in the Fermilab Recycler Ring", Fermilab Technical Report Fermilab-Conf-05-550-AD (2005).

[23] S. Abachi *et al.* (D0 Collaboration), "The D0 Detector", Nucl. Inst. Methods **A338**, 185 (1994).

[24] S. Abachi *et al.* (D0 Collaboration), "The D0 Upgrade: The Detector and its Physics", Fermilab Technical Report Fermilab-Pub-96/357-E (1996).

[25] V. M. Abazov *et al.* (D0 Collaboration), "The Upgraded D0 Detector", Nucl. Inst. Methods **A565**, 463 (2006).

[26] M. Roco, "The Silicon Microstrip Tracker for the D0 Upgrade", Internal D0 Note 3553 (1998).

[27] D0 Upgrade Collaboration, "D0 Silicon Tracker Technical Design Report" (1994).

[28] F. Borcherding *et al.* (D0 Collaboration), "Technical Design Report for the Upgrade L1/L2 Tracking Trigger - TDR for the CTT", Internal D0 Note 3551 (1998).

[29] J. Brzezniak (D0 Collaboration), "Conceptual Design of a 2 Tesla Superconducting Solenoid for the Fermilab D0 Detector Upgrade", Fermilab Technical Report FERMILAB-TM-1886 (1994).

[30] G. Bernardi *et al.*, "Improvements from the T42 Algorithm on Calorimeter Objects Reconstruction", Internal D0 Note 4335 (2004).

[31] M. Abolins *et al.*, "The Run IIb Trigger Upgrade for the D0 Experiment" IEEE Trans. Nucl. Sci. **51**, 340344 (2004).

[32] V. M. Abazov *et al.* (D0 Collaboration), "Run IIb Upgrade Technical Design Report", Fermilab Pub-02-327-E (2002).

[33] P. Calfayan, "ORing single muon triggers in p17 data", Internal D0 Note 5329 (2007).

[34] Marc Hohlfield, "Measurement of Single Electron Trigger Efficiencies for RunIIB", D0 Note 5783 (2008).

[35] R. Fruhwirth, "Application of Kalman Filtering to Track and Vertex Fitting", Nucl. Inst. Methods **A262**, 444 (1987).

[36] P. V. C. Hough, "Machine Analysis of Bubble Chamber Pictures", Proc. Int. Conf. High Energy Accelerators and Instrumentation (1959).

[37] J. Zhu, "Determination of Electron Energy Scale and Energy Resolution using P14 $Z^0 \to e^+e^-$ data", Internal D0 Note 4323 (2003).

[38] M. S. Anzelc and J. Hays, "Comparison of em_cert H-Matrix Cuts and Efficiencies in Data and Monte Carlo", Internal D0 Note 5075 (2006).

[39] C. Schwanenberger, "Electron Likelihood Efficiency in p17", Internal D0 Note 5114 (2006).

[40] P. Calfayan *et al.*, "Muon Identification Certification for p17 data", Internal D0 Note 5157 (2007).

[41] G. C. Blazey *et al.*, "Run II Jet Physics", arXiv:hep-ex/0005012 (2000).

[42] E. Busato and B. Andrieu, "Jet Algorithms in the D0 Run II Software: Description and User's Guide", Internal D0 Note 4457 (2004).

[43] D0 JES Group, "Jet Energy Scale Determination at D0 Run II (final p17 version)", Internal D0 Note 5382 (2008).

[44] M. Voutilainen and C. Royon, "Jet p_T resolution using v7.1 JES", D0 Note 5381 (2007).

[45] D. Gillberg *et al.*, "Measuring Jet Response Using the Missing E_T Projection Fraction Method in γ + jet Events", Internal D0 Note 4571 (2004).

[46] J. Pumplin, *et al.*, JHEP **0207**, 012 (2002).

[47] T. Sjöstrand, *et al.*, "High-Energy-Physics Event Generation with PYTHIA 6.1", Computer Phys. Commun. **135** 238 (2001) [arXiv:hep-ph/0010017].

[48] M. L. Mangano *et al.*, "ALPGEN, a generator for hard multiparton processes in hadronic collisions", JHEP **0307**, 001 (2003) [arXiv:hep-ph/0206293].

[49] G. Corcella, *et al.*, JHEP 0101 (2001) 010 [hep-ph/0011363]; hep-ph/0210213.

[50] MLM matching procedure, http://www-d0.fnal.gov/computing/MonteCarlo/generators/common_alp

[51] S. Höche *et al.*, "Matching Parton Showers and Matrix Elements", (2006) [arXiv:hep-ph/0602031v1].

[52] D. Gillberg, "Heavy Flavour Removal and Determination of Weighting Factors for ALPGEN W+jet Monte Carlo", Internal D0 Note 5129 (2006).

[53] Z. Sullivan, "Understanding single-top-quark production and jets at hadron colliders", Phys. Rev. D **70**, 114012 (2004) [arXiv:hep-ph/0408049].

[54] Kidonakis *et al.*, "Top quark production at the Tevatron at NNLO", Eur. Phys. J. **C33**, s466-s468 (2004).

[55] K. DeVaughan et. al., D0 Note 5801, Sept 2009.

[56] F. James, "MINUIT Function Minimization and Error Analysis, Reference Manual", http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html.

[57] T. Gadfort, A. Haas, D. Johnston, D. Lincoln, T. Scanlon, S. Schlobohm, " Performance of the Dzero NN b-tagging Tool on p20 Data", D0 Note 5554, December 19, 2007.

[58] D. Bloch et al., "Update of the JLIP b-tagger Performance in p14/pass2 with JES 5.3" D0 Note 4824, D. Bloch, et al., "Performance of the JLIP b-tagger in p14" D0 Note 4348.

[59] M. Owen and M. Hohlfeld, "Trigger Efficiencies for the OR of Single Electron Triggers in p17 Data", Internal D0 Note 5409 (2007).

[60] M. Voutilainen, "Jet trigger efficiencies for Run IIa", Internal D0 Note 5549 (2008).

[61] N. Makovec and J.-F. Grivaz, "Shifting, Smearing and Removing Simulated Jets", Internal D0 Note 4914 (2005).

[62] M. Sanders, "Multi-Parton Interaction Studies in p17.09 W+jets", D0 Collaboration Joint MC Generators and MC Verification Meeting (2007).
http://www-d0.hef.kun.nl//fullAgenda.php?ida=a072093&fid=109

[63] H. Nilsen, "Alpgen multiple parton reweighting", D0 Collaboration Joint MC Generators and MC Simulation Meeting (2007).
http://www-d0.hef.kun.nl//fullAgenda.php?ida=a071637&fid=47

[64] H. Schellman, "The longitudinal shape of the luminous region at D0", Internal D0 Note 5142 (2006).

[65] J. Alwall *et al.*, "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", arxiv:0706.2569v1 [hep-ph] (2007).

[66] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", JHEP **0206** 029 (2002) [hep-ph/0204244]; S. Frixione and B. R. Webber, "The MC@NLO 3.3 Event Generator", Cavendish-HEP-06/28, GEF-TH-19/2006 (2006) [arXiv:hep-ph/0612272v1]

[67] CDF Collaboration: D. Acosta *et al.*, Phys. Rev. Lett. **94**, 211801 (2005); A. Abulencia *et al.*, Phys. Rev. Lett. **98**, 161801 (2007); T. Aaltonen *et al.*, Phys. Rev. Lett. **100**, 201801 (2008).

[68] J. Campbell, R. K. Ellis, Phys Rev D, Vol 60, 113006 (1999). http://mcfm.fnal.gov/.

[69] G. C. Blazey *et al.*, in *Proceedings of the Workshop: "QCD and Weak Boson Physics in Run II,"* edited by U. Baur, R. K. Ellis, and D. Zeppenfeld, (Fermilab, Batavia, IL, 2000) p. 47; see Sec. 3.5 for details.

[70] S. Calvent, A. Duperrin, J.-F. Grivaz, D0 Note 5367.

[71] J. Hobbs, T. Nunnemann, and R. Van Kooten, "Study of the Z to ee and Z to mumu Yields", D0 Note 5268, Oct 27, 2006.

[72] M. Shamim and T. Bolton, D0 Note 5565.

[73] Dmitry Bandurin, Michael Begel, Jochen Cammin, and Dan Duggan, "Measurement of the Triple Differential Photon Plus Heavy-Flavor Jet Cross Section in p anti-p Collisions at sqrt(s)=1.96 TeV in D0", D0 Note 5702, June 2008.

[74] Y. Peters, M. Begel, K. Hamacher and D. Wicke, "Reweighting of the fragmenation functions for the D0 Monte carlo", D0 Note 5325, Jan 2007.