1-1-2010

A Search For The Sm Higgs Boson In The Process Zhllbb In 4.1/fb Of Cdf Ii Data

Shalhout Zaki Shalhout
Wayne State University,

A SEARCH FOR THE STANDARD MODEL HIGGS BOSON IN THE PROCESS $ZH \to \ell^+\ell^- b\bar{b}$ IN 4.1 fb^{-1} OF CDF II DATA

by

SHALHOUT Z. SHALHOUT

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2010

MAJOR: PHYSICS

Approved by:

_____

Advisor Date

_____

_____

_____

Table of Contents

List of Tables vi

List of Figures ix

Chapter 1 Introduction **1**

Chapter 2 The Standard Model and the Higgs Boson **3**

2.1 The Standard Model 3

2.1.0.1 The Strong Force 4

2.1.0.2 The Electroweak Force 5

2.1.1 Experimental Verification and Challenges 7

2.1.2 Gauge Invariance 7

2.1.3 Spontaneous Symmetry Breaking and the Higgs Mechanism 9

2.2 The SM Higgs boson 14

2.2.1 Production Mechanisms 14

2.2.2 Decay Modes 17

2.2.3 The SM Higgs Mass 17

Chapter 3 Experimental Searches for the SM Higgs **22**

3.1 LEP Searches 22

3.2 Tevatron Searches 23

3.3 Tevatron Search for $ZH \to \ell^+\ell^- b\bar{b}$ 27

Chapter 4 The Tevatron and CDF II **30**

4.1 The Tevatron and the FNAL Accelerator Complex 30

4.2 Tevatron Luminosity 32

4.3 The CDF II Detector 32

4.3.1 The CDF Coordinate System 35

4.3.2 The Cherenkov Luminosity Counter . . . 37
4.3.3 The Silicon Detectors . . . 38
4.3.4 The Central Outer Tracker . . . 41
4.3.5 The Calorimeters . . . 41
4.3.6 The Muon Detectors . . . 43
4.3.7 The Trigger System . . . 45

Chapter 5 Object Reconstruction 48

5.1 Track Reconstruction . . . 48
5.2 Electrons . . . 50
5.3 Muons . . . 52
5.4 Jets . . . 54
5.5 Missing Transverse Energy . . . 56

Chapter 6 Data Sample and Model 58

6.1 Data Samples . . . 58
6.2 Data Model . . . 59
6.2.1 Simulated Processes . . . 59
6.2.2 Data-Derived Backgrounds . . . 63

Chapter 7 Event Selection 65

7.1 Trigger Selection . . . 65
7.1.1 Trigger Efficiencies . . . 66
7.2 Muon Identification . . . 69
7.3 Electron Identification . . . 69
7.3.1 Central Electron Identification . . . 71
7.3.2 Plug Electron Identification . . . 74
7.3.3 Electron Reconstruction Scale Factors . . . 76
7.3.4 Fake Rates . . . 78

7.4 Z Reconstruction and Classification . . . 85
7.4.1 Additional Requirements on very Loose Z's . . . 85
7.4.2 EM energy corrections . . . 88
7.5 Jet Selection . . . 92
7.6 b-tagging . . . 93
7.7 Final Analysis Channels . . . 107

Chapter 8 Multvariate Techniques 108

8.1 Artificial Neural Networks . . . 108
8.2 Matrix Elements . . . 110
8.3 NN Jet Energy Corrections . . . 112
8.4 Karlsruhe Flavor Separator . . . 113
8.5 Matrix Element Probabilities . . . 117
8.6 Two Dimensional NNs for S/B Discrimination . . . 117

Chapter 9 Systematic Uncertainties 132

9.1 Rate Uncertainties . . . 134
9.2 Uncertainties Affecting Normalization and NN Output Shapes . . . 135

Chapter 10 Results 142

10.1 Results . . . 142
10.2 Observed and Expected Limits . . . 150

Chapter 11 Conclusions 171

11.1 Sensitivity Gain . . . 171
11.2 Combination With Other CDF II Higgs Searches . . . 172
11.3 Tevatron Combination . . . 172
11.4 Future Prospects . . . 175
References . . . 176

Abstract 186
Autobiographical Statement 187

List of Tables

Table 2.1 : The Properties of the Quarks. 3

Table 2.2 : The Properties of the Leptons. 4

Table 2.3 : The Properties of the Gauge bosons. 4

Table 4.1 : Average values for Tevatron beam parameters in Run II. 33

Table 4.2 : Location and energy resolution of the CDF calorimeter systems. The quoted resolutions are from [52] and references therein. 43

Table 6.1 : Integrated luminosity and calendar range by data period. Numbers are from [70]. The integrated luminosity is shown for run numbers with operational silicon, muon and calorimeter systems and active ELECTRON_CENTRAL_18 and Z_NOTRACK triggers. 60

Table 6.2 : Luminosity by trigger. The silicon systems are required to be operational. The CMUP_18 luminosity is reduced (in comparison to the ELECTRON_CENTRAL_18/Z_NOTRACK luminosity) due to a rejection of runs with an improper CMUP_18 trigger configuration in period 18. The lower CMX_18 trigger luminosity is the result of the CMX trigger not being fully implemented for run numbers before 15014. 60

Table 6.3 : Signal MC samples. Numbers are taken from [77] where the cross-sections (σ_{ZH}) are computed using the method outlined in Ref. [78] and the branching ratios are computed with the HDECAY code discussed in Ref. [79]. A 5% uncertainty is assumed on all cross-sections. 62

Table 6.4 : Backgrounds modeled with MC. The $\sigma \times BR$ for $Z+$ lf (light flavor) jets, $c\bar{c}$ and $b\bar{b}$ are from [73], and vary according to $\ell^+\ell^-$ mass and the jet multiplicity. These samples are scaled by an additional factor of 1.4 to account for the difference between leading and next to leading order predictions. WW, WZ and ZZ cross-sections are from [74]. The $t\bar{t}$ cross-section is from [75]. 63

Table 6.5 : Variables entering the mistag probability calculation 64

Table 7.1 : ELECTRON_CENTRAL_18 trigger path 67

Table 7.2 : Z_NOTRACK trigger path 68

Table 7.3 : MUON_CMX_18 trigger path 68

Table 7.4 : MUON_CMUP_18 trigger path 69

Table 7.5 : CMX muon selection. 70

Table 7.6 : CMUP muon selection. 70

Table 7.7 : The loose muon selection criteria for CMIO muons. 70

Table 7.8 : The electron types we consider in this analysis. 71

Table 7.9 : Tight central (Tc) electron selection. 72
Table 7.10 : Loose central 1 (Lc1) electron selection. 72
Table 7.11 : Loose central 2 (Lc2) electron selection. 73
Table 7.12 : Crack Track (CrkTrk) electron selection. 73
Table 7.13 : Tight plug (Tp) electron selection. 74
Table 7.14 : Loose plug phoenix(LpPhx) electron selection. 75
Table 7.15 : Loose plug non-phoenix(LpNphx25) electron selection. 75
Table 7.16 : Loose plug non-phoenix(LpNphx33) electron selection. 75
Table 7.17 : Loose plug (Lp) electron selection. 76
Table 7.18 : Probe (CEM) electron selection. 77
Table 7.19 : Probe (PEM) electron selection. 77
Table 7.20 : Electron scale factors in different mass regions. 77
Table 7.21 : For each dilepton category, we list the percentage of Z's in data, as well as the expected fake percentage. 79
Table 7.22 : Allowed dilepton types considered in this analysis. 85
Table 7.23 : Event totals for $Z \rightarrow ee$ selection. The systematic uncertainty on the background is about 20% 86
Table 7.24 : Event totals for $Z \rightarrow \mu\mu$ selection. The systematic uncertainty on the background is about 20% 87
Table 7.25 : Summary of additional requirements on electrons forming loose Z's. . . . 89
Table 7.26 : Scale factors for central and plug electron energies. 91
Table 7.27 : Z type and $\frac{S}{\sqrt{B}}$ category. 93
Table 7.28 : Preselection event totals (high S/B). 94
Table 7.29 : Preselection event totals (low S/B). 95
Table 7.30 : Summary of jet/b-tag selection. 106
Table 7.31 : Analysis channels. 107
Table 8.1 : Variables used to correct jet energies to parton level. 113
Table 9.1 : Summary of systematic uncertainties in terms of fractional acceptance change on samples. 133
Table 10.1 : Comparison of observed and predicted event totals for tag level high S/B selection. 142
Table 10.2 : Comparison of observed and predicted event totals for tag level low S/B selection. Blank entries denote negligible contributions. 149

Table 10.3 : Expected and observed limits divided by the SM value for $\sigma_{ZH} \times BR(H \to b\bar{b})$ for individual channels at $M_H = 115\,\mathrm{GeV}/c^2$. The total number of observed data events, predicted total background events, and the $S/\sqrt{B}$ are also shown. 166

Table 10.4 : Expected and observed 95% CL upper limits on $\sigma_{ZH} \times BR(H \to b\bar{b})$ divided by the SM values. 167

Table 10.5 : Expected and observed 95% CL upper limits on $\sigma_{ZH} \times BR(H \to b\bar{b})$. The SM values are included for comparison. 168

Table 11.1 : Comparison of observed 95% CL upper limits on $\sigma_{ZH}^{SM} \times BR(H \to b\bar{b})$ for CDF II searches @ $M_H = 115\,\mathrm{GeV}/c^2$. The absolute improvement is calculated by dividing previous limits by the current limit (in $4.1\,\mathrm{fb}^{-1}$). The projected limits are calculated under the assumption that sensitivity improves with increased integrated luminosity ($\mathcal{L}$) as $1/\sqrt{\mathcal{L}}$. The relative improvement is the ratio of the projected limits to the current. 171

List of Figures

Figure 2.1 : These are the fundamental strong interaction vertices. 5

Figure 2.2 : This is the fundamental vertex of the electromagnetic interaction. Here, f represents a fermion. 5

Figure 2.3 : These are the fundamental leptonic weak interaction vertices. In the diagram on the left, ℓ^- represents a charged lepton, while in the diagram on the right ℓ denotes a charged lepton or a neutrino. 6

Figure 2.4 : The four primary Higgs production mechanisms. They are (a) gluon-gluon fusion, (b) associated production with a W, (c) associated production with a $t\bar{t}$ pair, and (d) vector boson fusion. 15

Figure 2.5 : Higgs production cross-sections in picobarns (a) and decay branching ratios (b). Cross-sections and branching ratios are shown for various Higgs boson masses. Taken from [24]. 16

Figure 2.6 : Upper and lower bounds on the Higgs mass. Bounds are plotted as a function of energy scale Λ. Taken from [24]. 19

Figure 2.7 : Comparison of indirect (dashed contour) and direct (solid contour) constraints on M_W and M_t. The SM relationship for the masses as a function of the Higgs mass is also shown, with the $\Delta\alpha$ arrow representing the shift in this relationship as strong coupling is varied by one standard deviation. Figure taken from [7]. 20

Figure 2.8 : $\Delta\chi^2$ vs. M_H curve resulting from a fit to electroweak data. The yellow bands show the LEP and Tevatron 95% CL direct search exclusion limits on M_H. Figure taken from [7]. 21

Figure 3.1 : Behavior of $-2\ln(Q)$ for combined LEP data. The green and yellow bands represent the 68% and 95% probability bands about the median background expectation. Taken from [9]. 24

Figure 3.2 : Expected (dashed line) and observed (solid line) Tevatron 95% CL upper limits on SM Higgs production cross-sections $\times$ branching ratios for Higgs masses between 100 and 200 $GeV/\,\mathrm{c}^2$. The green and yellow bands represent the 68% and 95% probability bands about the expected limits. The pink shaded regions show the LEP and Tevatron 95% CL exclusions. Taken from [34]. 26

Figure 3.3 : Leading order Feynman diagrams for $ZH \to \ell^+\ell^- b\bar{b}$ [a], $Z + b\bar{b}$ [b], $t\bar{t}$ [c], and Diboson ZZ production [d]. $Z + b\bar{b}$ is the dominant background to Tevatron $ZH \to \ell^+\ell^- b\bar{b}$ searches. 28

Figure 3.4 : Comparison of the reconstructed $b\bar{b}$ mass in $ZH \to \ell^+\ell^- b\bar{b}$ (red) and $Z + b\bar{b}$ (black) simulated events. The simulation uses a Higgs mass of $120\,\mathrm{GeV}/c^2$. 29

Figure 3.5 : Comparison of the magnitude of the missing transverse energy in $ZH \rightarrow \ell^+\ell^- b\bar{b}$ (red) and $t\bar{t}$ (black) simulated events. The simulation uses a Higgs mass of $120\,\mathrm{GeV}/c^2$. 29

Figure 4.1 : Aerial view of the FNAL accelerator complex. The Tevatron (yellow highlight) is the large circular (radius of 1 km) object in the center of the photo. The main injector (orange highlight) is also shown. 33

Figure 4.2 : The Tevatron accelerator chain. Figure from [39]. 34

Figure 4.3 : Isometric view of the CDF II Detector with silicon detectors, Central Outer Tracker (COT), solenoid, calorimeter and muon detector components labeled. Image from [42] with modification. 35

Figure 4.4 : The CDF Coordinate System. 37

Figure 4.5 : Schematic view of the CDF CLC. Figure taken from [46]. 39

Figure 4.6 : Axial (left) and $r-z$ (right) views of the CDFII Silicon Detectors. Figure taken from [50]. 40

Figure 4.7 : End view of one sector of the COT. Dimensions are in cm. Figure taken from [51]. 42

Figure 4.8 : Elevation view of one half of the CDF II detector with the calorimeter systems labeled. Figure taken from [56]. 44

Figure 4.9 : Depiction of the flow of CDF II data through the 3-level trigger system to mass storage. 47

Figure 5.1 : ϕ vs. η of tracks extrapolated to non-fiducial regions of the CDF calorimeter in CDF data. The dots outline the location of unistrumented calorimeter space. 53

Figure 5.2 : Comparison of corrected (red) to uncorrected (black) jet E_T for jets with a corrected jet $E_T > 15\,\mathrm{GeV}$ and $|\eta| < 2.0$ in CDF data. 56

Figure 7.1 : Distribution of leading and next-to-leading lepton (left) and jet (right) transverse energies in simulated $ZH \rightarrow \ell\ell b\bar{b}$ events. The simulation has a Higgs mass of $120\,\mathrm{GeV}/c^2$. 66

Figure 7.2 : Fake rates for Tc, Lc1 and Lc2 type electrons. The mean fake rate from jet20 (black) , jet50 (red), jet70 (green) and jet100 (blue) data is applied to the high p_T electron data. The 50 % error band is shown in grey. . . . 80

Figure 7.3 : Fake rates for Tp, Lp and LpPhx type electrons. 81

Figure 7.4 : Fake rates for LpNphx25 and LpNphx33 type electrons. 82

Figure 7.5 : Ratio of Em object E_t to jet E_t in jet triggered data. 83

Figure 7.6 : Fake (rates measured in jet triggered events) component (yellow) of total data (black) for $Z \rightarrow ee$. 84

Figure 7.7 : Fake (from like-sign events) component (yellow) of total data (black) for $Z \to \mu\mu$ and $Z \to Tc + CrkTrk$ 84

Figure 7.8 : Dilepton mass in Z candidates reconstructed from two electron candidates. 88

Figure 7.9 : Dilepton mass in Z candidates reconstructed from two muon candidates. 88

Figure 7.10 : Distributions of the Z mass, dijet Mass and Number of tight jets for Lc1Lc1, LpLp, TpLp, and Lc1Lp Z's before and after the additional cuts listed in Table 7.25 are applied. 90

Figure 7.11 : Z mass in electron triggered data before and after corrections are applied. 92

Figure 7.12 : Distributions of the number of tight jets, Z p_T and $\not\!\!E_T$ in preTag events in the high S/B Z category. 96

Figure 7.13 : Distributions of the number of tight jets, $Z + 2$ jet Mass and jet E_T's in preTag events in the high S/B Z category. 97

Figure 7.14 : Jet 1, Jet 2, and Z ηs and $\not\!\!E_T$ ϕ in preTag events in the high S/B Z category. 98

Figure 7.15 : Distributions of the number of tight jets, Z p_T and $\not\!\!E_T$ in preTag event in the low S/B Z category. 99

Figure 7.16 : Distributions of the number of tight jets, $Z + 2$ jet Mass and jet E_T's in preTag events in the low S/B Z category. 100

Figure 7.17 : Jet 1, Jet 2, and Z ηs and $\not\!\!E_T$ ϕ in preTag events in the low S/B Z category. 101

Figure 7.18 : Depiction of a b jet with secondary vertex. d0 is the impact parameter of a displaced track. For simplicity, prompt tracks are not shown. Image from [85]. 102

Figure 7.19 : Tagging efficiency for b jets in $t\bar{t}$ MC as a function of jet E_T (top left) and η (top right). Mistag rates for jets in jet data as a function of jet E_T (bottom left) and η (bottom right). Figures from [87]. 103

Figure 7.20 : The sign of the impact parameter of a track. The impact parameter is assigned a negative (positive) sign if the angle ϕ is greater (less) than $\pi/2$. Figure from [88]. 104

Figure 7.21 : Distribution of values returned by the JP algorithm for simulated b (red), c (blue) and light flavor jets (green). Figure from [88]. 104

Figure 7.22 : (left) Efficiency of the JP tagging algorithm on b jets in $t\bar{t}$ MC as a function of jet E_T. The efficiency is shown for $P_J < 0.01$ (red) and $P_J < 0.05$ (blue) selection. (right) The likelihood of mistagging light flavor jets in jet data as a function of jet E_T. The mistag rate is shown for $P_J < 0.01$ (red) and $P_J < 0.05$ (blue) selection. We impose the $P_J < 0.05$ b tag cutoff (as opposed to 0.01) due to the significantly increased b tag efficiency. Figures from [88]. 105

Figure 8.1 : Structure of a NN with 3 nodes in the input layer (yellow band), 2 nodes in a single hidden layer (pink band) and 1 node in the output layer (purple band). Connections between nodes are represented by solid black lines with weights W_{ij} and M_{jk} indicated. The values X_1, X_2, X_3 represent experimentally observed quantities. Equations 8.1 and 8.3 show expressions for the values Y_j, while the final NN value Z_1 is given in Eq. 8.6. 111

Figure 8.2 : Effect of jet energy NN corrections on signal. The blue or green line shows the signal dijet mass before corrections. The red line shows dijet mass after corrections. The dijet mass is shown for the data and background model after corrections are applied. 114

Figure 8.3 : Effect of jet energy NN corrections on backgrounds. The teal line shows the background dijet mass before corrections. The teal line does not include the mistag background in the tag level plots. 115

Figure 8.4 : The magnitude of missing $\vec{E_T}$ without (left) and with (right) Neural Network corrections. Missing E_T is shown for ZH(brown), $t\bar{t}$ (black) and Z+jets (red) MC. 115

Figure 8.5 : Distribution of KNN flavor separator output for single b-tag events (top row) and for the lead E_T jet in double b-tag events (bottom row). . . . 116

Figure 8.6 : Distribution of matrix element probabilities for events in the TT tag high S/B sample. 118

Figure 8.7 : Distribution of matrix element probabilities for events in the L+JP tag high S/B sample. 119

Figure 8.8 : Distribution of matrix element probabilities for events in the T tag high S/B sample. 120

Figure 8.9 : Distribution of matrix element probabilities for events in the TT tag low S/B sample. 121

Figure 8.10 : Distribution of matrix element probabilities for events in the L+JP tag low S/B sample. 122

Figure 8.11 : Distribution of matrix element probabilities for events in the T tag low S/B sample. 123

Figure 8.12 : The selected NN variables, in order of selection from left to right, and the average testing error produced by their addition. 125

Figure 8.13 : Pre-Tag high S/B NN inputs 126

Figure 8.14 : Pre-Tag high S/B NN inputs. Due to computing time required for ME calculations, a random 10% of each MC background is shown (scaled by 10). 127

Figure 8.15 : Pre-Tag low S/B NN inputs 128

Figure 8.16 : Pre-Tag low S/B NN inputs. Due to computing time required for ME calculations, a random 10% of each MC background is shown (scaled by 10). 129

Figure 8.17 : Two dimensional NN output projections for high S/B pretag data. A cut on NNy<0.1 is made to highlight the signal region. Due to computing time required for ME calculations, a random 10% of each MC background is shown (scaled by 10). 130

Figure 8.18 : Two dimensional NN output projections for low S/B pretag data. A cut on NNy<0.1 is made to highlight the signal region. Due to computing time required for ME calculations, a random 10% of each MC background is shown (scaled by 10). 131

Figure 9.1 : Projections of templates affected by shape uncertainties used in the calculation of limits for the single tag high S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts. 136

Figure 9.2 : Projections of templates affected by shape uncertainties used in the calculation of limits for the single tag low S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts. 137

Figure 9.3 : Projections of templates affected by shape uncertainties used in the calculation of limits for the double tag high S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts. 138

Figure 9.4 : Projections of templates affected by shape uncertainties used in the calculation of limits for the double tag low S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts. 139

Figure 9.5 : Projections of templates affected by shape uncertainties used in the calculation of limits for the L+JP tag high S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts. 140

Figure 9.6 : Projections of templates affected by shape uncertainties used in the calculation of limits for the L+JP tag low S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts. 141

Figure 10.1 : NN input distributions for the single tag high S/B channel. 143

Figure 10.2 : NN input distributions for the single tag low S/B channel. 144

Figure 10.3 : NN input distributions for the L+JP tag high S/B channel. 145

Figure 10.4 : NN input distributions for the L+JP tag low S/B channel. 146

Figure 10.5 : NN input distributions for the double tag high S/B channel. 147

Figure 10.6 : NN input distributions for the double tag low S/B channel. 148

Figure 10.7 : NN output for high S/B single tag (T) channel data, ZH ($M_H = 120\text{GeV}/c^2$), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 151

Figure 10.8 : NN output for high S/B single tag (T) channel $Z + b\bar{b}$, $Z + b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 152

Figure 10.9 : NN output for high S/B single tag (L+JP) channel data, ZH ($M_H = 120\text{GeV}/c^2$), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 153

Figure 10.10: NN output for high S/B single tag (L+JP) channel $Z + b\bar{b}$, $Z + b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 154

Figure 10.11: NN output for high S/B single tag (TT) channel data, ZH ($M_H = 120\text{GeV}/c^2$), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 155

Figure 10.12: NN output for high S/B single tag (TT) channel $Z + b\bar{b}$, $Z + b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 156

Figure 10.13: NN output for low S/B single tag (T) channel data, ZH ($M_H = 120\text{GeV}/c^2$), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 157

Figure 10.14: NN output for low S/B single tag (T) channel $Z + b\bar{b}$, $Z + b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 158

Figure 10.15: NN output for low S/B single tag (L+JP) channel data, ZH ($M_H = 120\text{GeV}/c^2$), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 159

Figure 10.16: NN output for low S/B single tag (L+JP) channel $Z + b\bar{b}$, $Z + b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 160

Figure 10.17: NN output for low S/B single tag (TT) channel data, ZH ($M_H = 120\text{GeV}/c^2$), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 161

Figure 10.18: NN output for low S/B single tag (TT) channel $Z + b\bar{b}$, $Z + b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text. 162

Figure 10.19: NN output projections for the final high (S/B) analysis channels onto the Z+jets – ZH (x) axis of the 2D output. A cut of NNy<0.1 is made to emphasize the signal region. 163

Figure 10.20: NN output projections for the final low (S/B) analysis channels onto the Z+jets – ZH (x) axis of the 2D output. A cut of NNy<0.1 is made to emphasize the signal region. 164

Figure 10.21: Distribution of 95% CL upper limits from 1000 pseudo-experiements at $M_H = 115\,\mathrm{GeV}/c^2$ with 1σ (green) and 2σ (yellow) bands indicated. The expected limit (black arrow) is the median of the pseudo-experiement limit distribution. The observed limit (red arrow) is shown for comparison. . 166

Figure 10.22: Expected and observed limits with ± 1 and 2 σ bands. 169

Figure 10.23: Expected and observed limits with ± 1 and 2 σ bands. The SM values of $\sigma_{ZH} \times BR(H \to b\bar{b})$ are indicated by the red line. 170

Figure 11.1 : CDF Higgs Working Group upper limits on SM Higgs production as a function of Higgs mass. The combined upper limits are shown as solid (observed) and dashed (expected) lines in dark red. The limits combine the eight Higgs modes listed in the legend. The $ZH \to \ell^+\ell^- b\bar{b}$ limits are indicated by the solid (observed) and dashed (expected) blue lines. . . . 173

Figure 11.2 : Tevatron New Phenomena and Higgs Working Group upper limits on SM Higgs production as a function of Higgs mass. The combined upper limits are shown as solid (observed) and dashed (expected) lines, with $\pm 1\sigma$ (green) and $\pm 2\sigma$ (yellow) bands indicated. The pink shaded regions show the LEP and Tevatron 95% CL exclusions. 174

Chapter 1: Introduction

The standard model of particle physics provides a detailed description of a universe in which all matter is composed of a small number of fundamental particles, which interact through the exchange of force–carrying gauge bosons (the photon, $W^{\pm}$, Z and gluons). The organization of the matter and energy in this universe is determined by the effects of three forces; the strong, weak, and electromagnetic. The weak and electromagnetic forces are the low energy manifestations of a single electro-weak force, while the strong force binds quarks into protons and neutrons. The standard model does not include gravity, as the effect of this force on fundamental particles is negligible.

Four decades of experimental tests, spanning energies from a few electron-volts (eV) up to nearly two TeV, confirm that the universe described by the standard model is a reasonable approximation of our world. For example, experiments have confirmed the existence of the top quark, the $W^{\pm}$ and the Z bosons, as predicted by the standard model [1, 2, 3, 4, 5]. The latest experimental averages for the masses of the top quark, $W^{\pm}$ and Z are respectively $173.1 \pm 0.6(\text{stat.}) \pm 1.1(\text{syst.})$ [6], 80.399 ± 0.023 [7] and $91.1876 \pm 0.0021\,\text{GeV}/c^2$ [8].

The SM is a gauge field theory of zero mass particles. However, the SM is able to accommodate particles with non–zero mass through the introduction of a theoretical Higgs field which permeates all of space. Fermions gain mass through interactions with this field, while the longitudinal components of the massive $W^{\pm}$ and Z are the physical manifestations of the field itself. Introduction of the Higgs field, directly leads to the predicted existence of an additional particle, the Higgs boson. The Higgs boson is the only particle of the standard model that has not been observed, and is the only unconfirmed prediction of the theory. The standard model describes the properties of the Higgs boson in terms of its mass, which is a free parameter in the theory. Experimental evidence [9, 7] suggests that the Higgs mass has a value between 114.4 and 186 GeV/c^2.

Particles with a mass in this range can be produced in collisions of less massive particles

accelerated to near the speed of light. Currently, one of only a few machines capable of achieving collision energies large enough to potentially produce a standard model Higgs boson is the Tevatron proton—antiproton collider located at Fermi National Accelerator Laboratory in Batavia, Illinois.

This dissertation describes the effort to observe the standard model Higgs in Tevatron collisions recorded by the Collider Detector at Fermilab (CDF) II [10] experiment in the $ZH \to \ell^+\ell^- b\bar{b}$ production and decay channel. In this process, the Higgs is produced along with a Z boson which decays to a pair of electrons or muons ($Z \to \ell^+\ell^-$), while the Higgs decays to a bottom anti–bottom quark pair ($H \to b\bar{b}$).

A breif overview of the standard model and Higgs theory is presented in Chapter 2. Chapter 3 explores previous searches for the standard model Higgs at the Tevatron and elsewhere. The search presented in this dissertation expands upon the techniques and methods developed in previous searches. The fourth chapter contains a description of the Tevatron collider and the CDF II detector. The scope of the discussion in Chapter 4 is limited to the experimental components relevant to the current $ZH \to \ell^+\ell^- b\bar{b}$ search.

Chapter 5 presents the details of object reconstruction; the methods used to convert detector signals into potential electrons, muons or quarks. Chapter six describes the data sample studied for the presence of a $ZH \to \ell^+\ell^- b\bar{b}$ signal and details the techniques used to model the data. The model accounts for both signal and non–signal processes (backgrounds) which are expected to contribute to the observed event sample.

Chapters 7 and 8 summarize the event selection applied to isolate $ZH \to \ell^+\ell^- b\bar{b}$ candidate events from the data sample, and the advanced techniques employed to maximize the separation of the signal from background processes.

Chapters 9 and 10 present the systematic uncertainties affecting our modeling of the data sample and the results of the search. Chapter 11 presents a discussion of $ZH \to \ell^+\ell^- b\bar{b}$ in the context of the overall Tevatron efforts to observe a standard model Higgs signal.

Chapter 2: The Standard Model and the Higgs Boson

This chapter presents an outline of the standard model. The chapter centers on a discussion of the Higgs Mechanism and an overview of the Higgs production and decay modes most relevant to experimental searches.

2.1 The Standard Model

In the standard model [11], particles are divided into two categories, fermions and bosons. Fermions are particles of half integer spin, while bosons have integer spin. Naturally, the elementary particles of the model (12 leptons, 36 quarks, and 12 gauge particles), fall into these two categories. Leptons and quarks are fermions, while the gauge particles are bosons. The properties of the quarks, leptons and gauge bosons are summarized in Tables 2.1-2.3.

Table 2.1: The Properties of the Quarks.

particle	charge	spin	$\sim$ mass
u	$+\frac{2}{3}$	$\frac{1}{2}$	$1.5 - 3.3$ MeV/c^2
d	$-\frac{1}{3}$	$\frac{1}{2}$	$3.5 - 6$ MeV/c^2
c	$+\frac{2}{3}$	$\frac{1}{2}$	1.27 GeV/c^2
s	$-\frac{1}{3}$	$\frac{1}{2}$	105 MeV/c^2
t	$+\frac{2}{3}$	$\frac{1}{2}$	171.3 GeV/c^2
b	$-\frac{1}{3}$	$\frac{1}{2}$	4.2 GeV/c^2

There are four types of interactions between particles: electromagnetic, weak, strong and gravitational. The standard model describes only the electromagnetic, weak, and strong interactions, ignoring gravity, whose effects are too feeble to be significant at the

Table 2.2: The Properties of the Leptons.

particle	charge	spin	$\sim$ mass($\frac{\text{MeV}}{c^2}$)
e	-1	$\frac{1}{2}$	.511
ν_e	0	$\frac{1}{2}$	$< 15\frac{eV}{c^2}$
μ	-1	$\frac{1}{2}$	105
ν_μ	0	$\frac{1}{2}$	$< .17$
τ	-1	$\frac{1}{2}$	1777
ν_τ	0	$\frac{1}{2}$	< 24

Table 2.3: The Properties of the Gauge bosons.

particle	charge	spin	$\sim$ mass($\frac{\text{GeV}}{c^2}$)
γ	0	$+1$	0
g	0	$+1$	0
$W^\pm$	± 1	$+1$	80
Z^0	0	$+1$	91

level of elementary particle interactions. The strong interactions are described by quantum chromodynamics (QCD), and a combined description of the electromagnetic and weak interactions is contained in electroweak theory.

Standard model interactions are mediated by the exchange of spin–1 gauge bosons. For the electromagnetic interaction, the mediating particle is the massless photon (γ), while the mediators of the strong force are eight massless gluons (g). Similarly, the weak force is mediated by the Z^0, W^+ and W^- bosons.

2.1.0.1 The Strong Force

Quarks and gluons carry color charge, usually denoted r,g,b, for red, green and blue. The concept of color charge was suggested by Greenberg [12], as a way to ensure that the quark model would abide by the Pauli exclusion principle. Color serves as a distinguishing

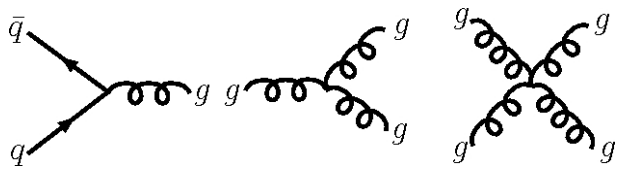


Figure 2.1: These are the fundamental strong interaction vertices.

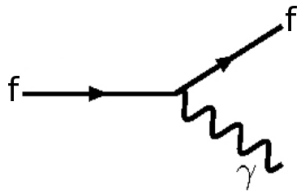


Figure 2.2: This is the fundamental vertex of the electromagnetic interaction. Here, f represents a fermion.

feature, allowing otherwise identical quarks to occupy the same state. Particles containing a $q\bar{q}$ combination are called mesons, while those consisting of a qqq combination are called baryons. Mesons, baryons and any other quark combination, must have zero net color charge (for example, blue with anti-blue, or red, green and blue). Objects carrying color charge participate in strong interactions. The fundamental strong interaction vertices are shown in Fig. 2.1.

2.1.0.2 The Electroweak Force

Electroweak theory contains a unified description of the weak and electromagnetic forces under the $SU(2)_L \bigotimes U(1)_Y$ gauge group. The electromagnetic force is experienced by all charged particles. The basic electromagnetic interaction vertex is shown in Fig. 2.2.

Leptons and quarks participate in weak interactions, with neutral current weak interactions mediated by the Z^0, and charged current interactions mediated by W^+ or W^-.

The fundamental vertices of leptonic weak interactions are shown in Fig. 2.3. It should be noted (assuming that neutrinos are massless) that leptonic weak interactions conserve

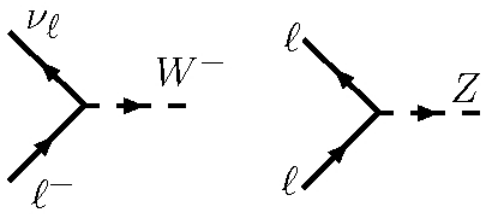


Figure 2.3: These are the fundamental leptonic weak interaction vertices. In the diagram on the left, ℓ^- represents a charged lepton, while in the diagram on the right ℓ denotes a charged lepton or a neutrino.

the lepton numbers L_e, L_μ and L_τ; meaning cross–generational leptonic weak interactions are forbidden.

Quark weak interactions also come in two forms, neutral and charged. Neutral current quark interactions are similar to leptonic weak interactions, in that they do not mix generations; they are flavor conserving. However, charged current weak interactions can mix generations. This is because the weak eigenstates (d',s',b'), are not equivalent to the mass eigenstates (d,s,b), with the weak eigenstates being linear combinations of the mass eigenstates.

The two sets of eigenstates, are connected by the CKM matrix.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.1}$$

Where the absolute value of the CKM matrix entries measures the likelihood of quark mixing. The $V-A$ [13] structure of the weak interaction allows for maximal parity violation, while the inclusion of a complex phase for some CKM matrix elements, incorporates the much smaller effect of CP violation into the standard model.

2.1.1 Experimental Verification and Challenges

To date, all experimental tests (with the possible exception of neutrino mixing) are found to be consistent with the standard model. For example, the predicted existence of the massive gauge bosons, W^+, W^- [1] and Z^0 [2], was confirmed at CERN in 1983. In addition, weak neutral current interactions were first proposed in the standard model. The discovery of this type of weak interaction at CERN in 1973 [14], was a large success for the model.

The observation of neutrino flavor oscillation [15, 16] may be the first significant evidence of a physical process contrary to the standard model. In the SM, neutrino masses are typically assumed to be zero. However, neutrino mixing implies non-zero masses. A detailed discussion of non-zero neutrino masses and the consequences for the standard model is available in [17].

The only particle predicted by the standard model which has not been observed is the Higgs boson. The Higgs boson is a by-product of spontaneous symmetry breaking under the Higgs mechanism [18].

2.1.2 Gauge Invariance

As early as 1961 [19], the derivation of particle interactions from a Lagrangian has been tried, starting from the imposition of local gauge invariance. A simple demonstration of the gauge principle can be seen in QED under the gauge group $U(1)_{em}$. The following discussion is adapted from more detailed presentations in Griffiths' "Introduction to Elementary Particles" [20] and Pich's "The Standard model of electroweak interactions" [21]. The Dirac equation for a free particle wave-function ψ, can be derived from the Lagrangian density $\mathcal{L}_f$:

$$\mathcal{L}_f = \bar{\psi}(i\gamma^\mu \partial_\mu)\psi - \bar{\psi}m\psi \tag{2.2}$$

Applying the following global $U(1)_{em}$ gauge transformation to the wavefunctions leaves $\mathcal{L}_f$ unchanged.

$$\psi \rightarrow e^{i\chi}\psi \tag{2.3}$$

If however, χ depends upon the space time coordinates x, then the transformation is a local gauge transformation and $\mathcal{L}$ is not invariant:

$$\mathcal{L}_f \rightarrow e^{-i\chi}\bar{\psi}(i\gamma^\mu\partial_\mu)e^{i\chi}\psi - \bar{\psi}m\psi$$

$$= \mathcal{L}_f - \bar{\psi}[\gamma^\mu\partial_\mu\chi(x)]\psi \tag{2.4}$$

$\mathcal{L}_f$ can be made invariant under the local gauge transformation by introducing the covariant derivative $\mathcal{D}_\mu$,

$$\mathcal{D}_\mu = \partial_\mu + ieA_\mu \tag{2.5}$$

where A_μ is a gauge field, chosen with the transformation property:

$$A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu\chi(x) \tag{2.6}$$

Then prior to applying the local transformation, the ∂_μ in Eq. 2.2 is replaced by $\mathcal{D}_\mu$ to get:

$$\mathcal{L}_f \longrightarrow \mathcal{L}'_f = \bar{\psi}[i\gamma^\mu(\partial_\mu + ieA_\mu)]\psi - \bar{\psi}m\psi$$

$$= \mathcal{L}_f - e\bar{\psi}\gamma^\mu A_\mu\psi \tag{2.7}$$

Applying the local transformation to the wavefunctions in $\mathcal{L}'_f$ and using the transformation property of the gauge field gives an invariant free Lagrangian:

$$\mathcal{L}'_f \rightarrow \mathcal{L}_f - \bar{\psi}[\gamma^\mu\partial_\mu\chi(x)]\psi - e\bar{\psi}\gamma^\mu[A_\mu - \frac{1}{e}\partial_\mu\chi(x)]\psi$$

$$= \mathcal{L}_f - e\bar{\psi}\gamma^\mu A_\mu \psi$$

$$= \mathcal{L}'_f \tag{2.8}$$

The introduction of the gauge field A_μ, requires the addition of two terms, $\mathcal{L}_g$ and $\mathcal{L}_{m_\gamma}$, to the free Lagrangian $\mathcal{L}'_f$. $\mathcal{L}_g$ is the Lagrangian for the gauge field, and has the form:

$$\mathcal{L}_g = -\frac{1}{4}\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu} \tag{2.9}$$

where, $\mathcal{F}_{\mu\nu}$ is the electromagnetic strength tensor.

$$\mathcal{F}_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.10}$$

The addition of $\mathcal{L}_g$ to the Lagrangian does not affect the invariance of $\mathcal{L}'_f$ under the local transformation, since $\mathcal{F}_{\mu\nu}$ is invariant under the transformation property of A_μ.

The gauge field also adds a mass term $\mathcal{L}_{m_\gamma}$, to the Lagrangian $\mathcal{L}'_f$, of the form:

$$\mathcal{L}_{m_\gamma} \sim (m_\gamma)^2 A^\mu A_\mu \tag{2.11}$$

which is not invariant. For gauge fields expressed as massless bosons, this term is zero; in this case the boson is the photon with $m_\gamma = 0$.

2.1.3 Spontaneous Symmetry Breaking and the Higgs Mechanism

In the standard model, unification of the weak and electromagnetic forces occurs under the gauge group $SU(2)_L \bigotimes U(1)_Y$, and the spontaneous breaking of this symmetry via the Higgs mechanism imparts mass to the W^+,W^- and Z^0 bosons.

Considering only leptons and assuming a non-zero mass for neutrinos, let

$$\psi_1 = \begin{pmatrix} \nu_L \\ \ell_L \end{pmatrix} \tag{2.12}$$

where, ℓ_L represents a left-handed lepton 4-spinor, and ν_L is the spinor for the left-handed component of the corresponding neutrino. Similarly, let

$$\psi_2 = \nu_R \quad \text{and} \quad \psi_3 = \ell_R \tag{2.13}$$

where, ν_R and ℓ_R are the right handed components of the neutrino and lepton. The free Lagrangian for the ψ_i,

$$\mathcal{L}_f = \sum_{k=1}^{3} i\bar{\psi}_k(x)\gamma^\mu\partial_\mu\psi_k(x) \tag{2.14}$$

is invariant under the global $U(1)_Y$ and $SU(2)_L$ transformations, where $SU(2)_L$ only acts on the left-handed ψ_k doublet, and Y is the hypercharge.

Local invariance of $\mathcal{L}_f$, under $SU(2)_L \bigotimes U(1)_Y$ is accomplished by introducing zero-mass gauge fields W_μ^1, W_μ^2, W_μ^3 and B_μ with appropriate transformation properties, and the covariant derivatives:

$$\mathcal{D}_\mu^1 = \partial_\mu - \frac{ig}{2}\sigma^1 W_\mu^1 - ig\prime y_1 B_\mu \tag{2.15}$$

and

$$\mathcal{D}_\rho = \partial_\mu - ig\prime y_\rho B_\mu \qquad \rho = 2, 3 \tag{2.16}$$

where, g and $g\prime$ are coupling constants, the y_ρ are hypercharge components, and the σ^i are the Pauli matrices.

Including the Lagrangian for the gauge fields, $\mathcal{L}_g$, the combined Lagrangian $\mathcal{L}$

$$\mathcal{L} = \mathcal{L}_f + \mathcal{L}_g = \sum_{k=1}^{3} i\bar{\psi}_k(x)\gamma^\mu \mathcal{D}_k \psi_k(x) - \frac{1}{4}(\mathcal{W}_{\mu\nu})^i(\mathcal{W}^{\mu\nu})^i - \frac{1}{4}\mathcal{B}_{\mu\nu}\mathcal{B}^{\mu\nu} \tag{2.17}$$

is invariant locally under $SU(2)_L \bigotimes U(1)_Y$ transformations. Although, $\mathcal{L}$ is invariant, it contains no mass terms. Inclusion of lepton masses would have mixed right and left-handed states breaking the invariance, while masses for the gauge fields would introduce invariant terms into $\mathcal{L}$ as discussed in Section 2.1.2.

Before proceeding, it should be noted that the gauge fields, W_μ^i and B_μ, are not equivalent to the physical fields $W_\mu^\pm$, Z_μ^0 and γ_μ^0. Instead, the physical fields are linear combinations of the gauge fields:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp W_\mu^2) \tag{2.18}$$

and

$$\begin{pmatrix} \gamma_\mu^0 \\ Z_\mu^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_w & \sin\theta_w \\ -\sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{2.19}$$

where θ_w is the Weinberg angle.

The gauge fields in Eq. 2.17 can acquire the appropriate masses by the process known as the Higgs mechanism. To begin, let the scalar doublet, Φ and Lagrangian, $\mathcal{L}_\Phi$ be defined as,

$$\Phi = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix} \tag{2.20}$$

and

$$\mathcal{L}_\Phi = \partial_\mu \Phi^\dagger \partial^\mu \Phi - \mu^2 \Phi^\dagger \Phi - \lambda(\Phi^\dagger \Phi)^2 \tag{2.21}$$

The Lagrangian $\mathcal{L}_\Phi$, is invariant locally under $SU(2)_L \bigotimes U(1)_Y$ when the covariant

derivative $\mathcal{D}_\mu^\Phi$ is substituted for ∂_μ,

$$\mathcal{D}_\mu^\Phi = \partial_\mu - \frac{ig}{2}\sigma^i W_\mu^i - \frac{ig\prime}{2}B_\mu \tag{2.22}$$

The potential term,

$$V = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2 \tag{2.23}$$

is chosen to allow the symmetry to be spontaneously broken from $SU(2)_L \bigotimes U(1)_Y$ to $U(1)_{em}$

Spontaneous symmetry breaking refers to the arbitrary selection of one particular minima for the potential. Expansion of the potential term gives:

$$V = \mu^2[(\varphi^+)^2 + (\varphi^0)^2] + \lambda[(\varphi^+)^2 + (\varphi^0)^2]^2 \tag{2.24}$$

which is minimized for $\mu^2 < 0$ whenever,

$$[(\varphi^+)^2 + (\varphi^0)^2] = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{2.25}$$

Choosing a particular pair of minimizing values, say $\varphi^+ = 0$ and $\varphi^0 = \frac{v}{\sqrt{2}}$ gives,

$$\Phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.26}$$

which is the vacuum expectation of the Higgs field.

The next step is to express $\mathcal{L}_\Phi$ in terms of a state, $\Phi_0^\sim$ that differs slightly from Φ_0.

$$\Phi_0^\sim = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v+h \end{pmatrix} \tag{2.27}$$

Only displaying terms relevant to the current discussion, the Lagrangian $\mathcal{L}_\Phi$ in terms of

$\Phi_0^{\sim}$ is:

$$\mathcal{L}_\Phi = \frac{1}{2}\partial_\mu(0, v+h)\partial^\mu \begin{pmatrix} 0 \\ v+h \end{pmatrix} - V(\Phi^\dagger\Phi)$$

$$= .... + \frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \frac{1}{2}(-2\mu^2)h^2 + ... \tag{2.28}$$

The term $(-2\mu^2)h^2$, is the mass term for a scalar boson with mass,

$$m_h = \sqrt{-2\mu^2} \tag{2.29}$$

which is the Higgs boson.

In order to give mass to the correct gauge bosons, the lagrangian, $\mathcal{L}_\Phi$ is expressed in terms of $\Phi_0^{\sim}$ and $\mathcal{D}_\mu^\Phi$. Then the non-physical fields are replaced by the physical fields as in Eqs. 2.18 and 2.19. This results in a lagrangian with terms:

$$\mathcal{L}_\Phi = ... + \frac{g^2v^2}{4}W_\mu^+W^{-\mu} + \frac{g^2v^2}{8(\cos^2\theta_w)}Z_\mu^0Z^{0\mu} + ... \tag{2.30}$$

where the masses of the gauge bosons $W^\pm$ and Z^0 are identified as,

$$M_W = \frac{gv}{2} \quad and \quad M_Z = \frac{gv}{2(\cos\theta_w)} \tag{2.31}$$

and no such term exists for the photon field, which remains massless. In a similar manner, fermion masses can be introduced by the addition of a term to the lagrangian which includes Yukawa couplings of the leptons with the Higgs field Φ.

Employing gauge invariance, spontaneous symmetry breaking and the Higgs mechanism, results in four mediator bosons, each with the correct mass. The byproduct of these procedures is the Higgs boson, whose existence remains unverified.

2.2 The SM Higgs boson

This subsection contains a brief description of basic properties of the Higgs boson, with the focus primarily on behavior at hadronic colliders. Behavior of the Higgs at electron-positron colliders is discussed elsewhere [22].

2.2.1 Production Mechanisms

At $p\bar{p}$ colliders such as the Tevatron, there are four primary Higgs boson production processes: gluon-gluon fusion, associated production with a $W^{\pm}$ or Z^0, associated production with $b\bar{b}$ or $t\bar{t}$ and vector boson fusion. Feynman diagrams for these processes are shown in Fig. 2.4, and cross-sections for Higgs production are shown in Fig. 2.5 (a).

Gluon-gluon fusion is the most important production mechanism. While gluons do not couple directly to the Higgs boson, gluon-gluon fusion does occur through an intermediate quark loop. Since Higgs couplings to fermions are proportional to the fermion's mass, gluon-gluon fusion usually occurs through a top quark loop as shown in Fig. 2.4(a). Despite occurring at a higher order, gluon-gluon fusion is the dominant production mechanism at the Tevatron.

Feynman diagrams for associated Higgs production are shown in Figs. 2.4(b) and 2.4(c). Despite smaller cross-sections than gluon-gluon fusion, the decay of the associated particle can result in easily distinguished final states containing, for example, charged leptons or missing energy (from neutrinos) produced in the decay of $W^{\pm}$ or Z^0. These decay products are used to distinguish events containing a Higgs boson from multi-jet QCD events.

Vector boson fusion consists of a quark and antiquark each coupling to a vector boson which then annihilate to produces a Higgs boson. The Feynman diagram in Fig. 2.4(d) is an example of vector boson fusion. Searches for Higgs production through vector boson fusion look for the characteristic presence of two forward energetic jets [23].

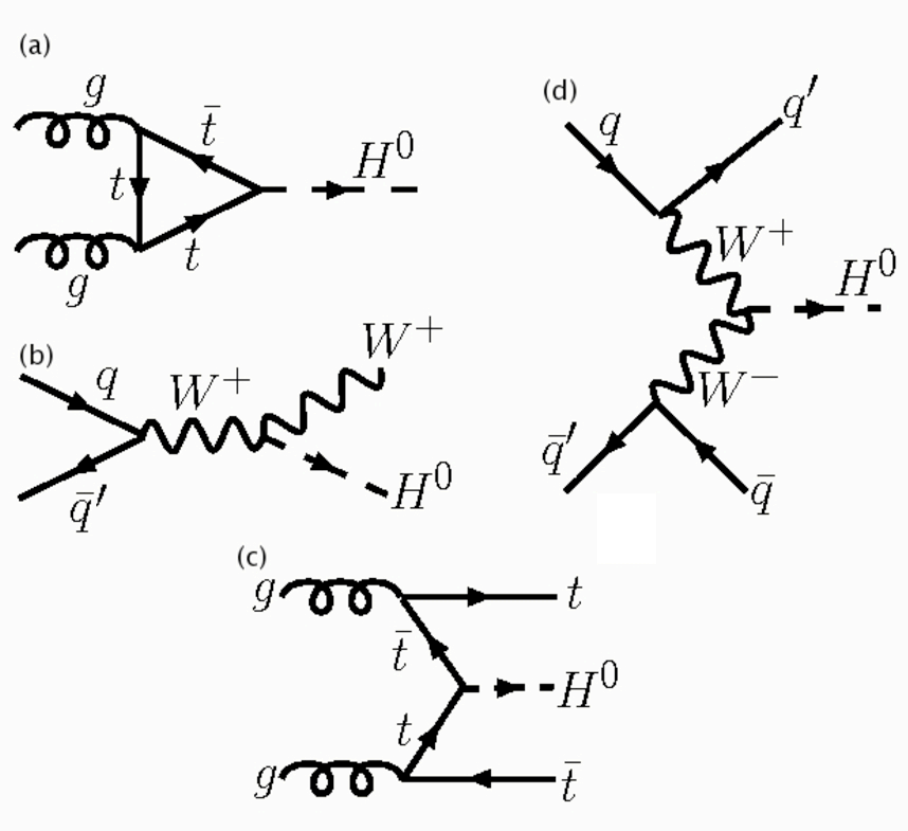


Figure 2.4: The four primary Higgs production mechanisms. They are (a) gluon-gluon fusion, (b) associated production with a W, (c) associated production with a $t\bar{t}$ pair, and (d) vector boson fusion.

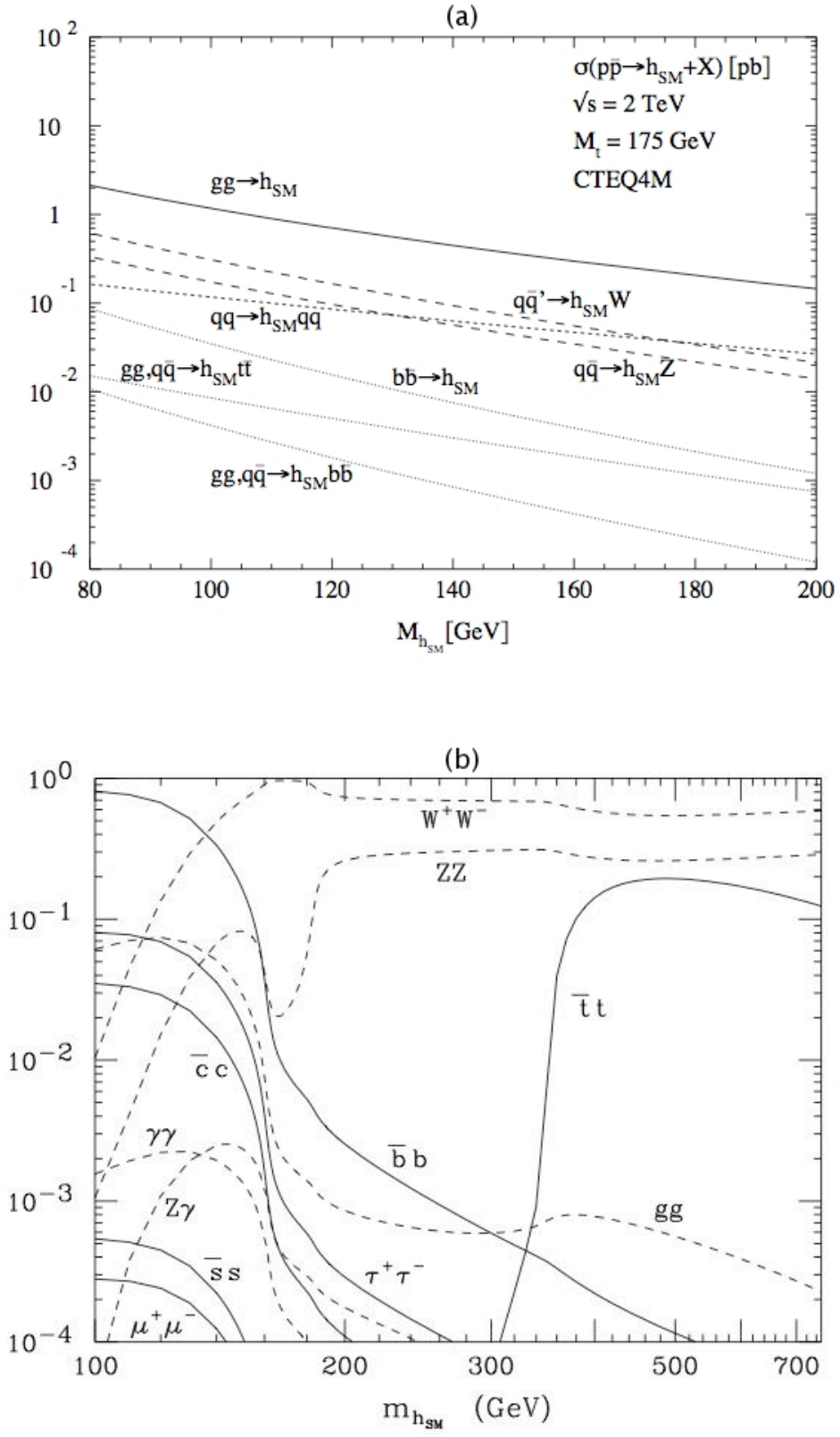


Figure 2.5: Higgs production cross-sections in picobarns (a) and decay branching ratios (b). Cross-sections and branching ratios are shown for various Higgs boson masses. Taken from [24].

2.2.2 Decay Modes

Figure 2.5(b), shows the branching ratio of the Higgs boson to various final states as a function of Higgs mass. A low mass Higgs boson, $M_h < 135\ \mathrm{GeV}/c^2$, decays predominantly to a pair of low mass fermions, $H \rightarrow f\bar{f}$. As mentioned in Sub-Section 2.2.1, the Higgs coupling to fermions is proportional to the fermion's mass. Since top quark pairs are too massive for a Higgs mass in this range, the dominant decay is $H \rightarrow b\bar{b}$, with the next most common fermion – anti-fermion mode, $H \rightarrow \tau\tau$, an order of magnitude smaller.

For a Higgs of intermediate mass, $120 < M_h < 135\ \mathrm{GeV}/c^2$, the channel $H \rightarrow gg$ becomes significant, having the third largest branching ratio for this Higgs mass range.

For a larger Higgs mass, $M_h > 135\ \mathrm{GeV}/c^2$, the primary decay mode is $H \rightarrow W^+W^-$ with a branching ratio of 0.6 to 0.9. The decay channel $H \rightarrow Z^0Z^0$ has a branching ratio of about 0.3 in this range.

2.2.3 The SM Higgs Mass

Equation 2.29 relates the mass of the Higgs boson to the factor μ, which is an arbitrary parameter whose value is unknown. Nonetheless, theoretical bounds on the mass of the Higgs have been calculated by excluding values of m_h for which the standard model becomes non-perturbative. A lower bound on m_h can be obtained by considering the instability of the Higgs potential when quantum loop corrections are included [25]. An upper bound can be calculated by considering energy scales for which the coupling factor λ tends to ∞ [26]. Figure 2.6 shows the allowed range for the Higgs mass as a function of Λ, the energy scale at which the standard model breaks down.

The LEP experiments provide a lower limit on the Higgs mass. The combined Higgs lower mass limit in direct searches by the four experiments is $M_h > 114.4\ \mathrm{GeV}/c^2$ at the 95% confidence level [27]. As of March 2009, the two Tevatron experiments CDF and D0, have (in combination) excluded the range 160 to 170 GeV/c^2 at the 95% confidence level. Details of the LEP and Tevatron exclusions will be presented in Chaper 3.

The values of SM parameters such as the masses of the W and top quark are sensitive to the mass of the Higgs boson. Precision measurements of various electroweak parameters by the LEP, SLD, CDF, and D0 experiments provide indirect experimental constraints on the value of M_H. Figure 2.7 shows the indirect (from LEP-1 and SLD) and direct (from LEP-2, CDF, and D0) 68% CL contours [7] on the masses of the W and top quark (M_W and M_t respectively), the values or which are modified by radiative corrections from Higgs loops. In the SM, knowing the masses for two of the three particles, W, H, or t, determines the mass of the third. This relationship is represented by the solid lines in Fig. 2.7. The intersection of the 68% CL contour on M_W and M_t favors a Higgs mass near the $M_h > 114.4$ GeV/c^2 LEP direct exclusion bound.

Figure 2.8 shows the constraint on M_H derived from a SM fit to experimental measurements of electroweak parameters. The minimum of the $\Delta\chi^2$ curve corresponds to the preferred value of M_H ($87^{+35}_{-26}\,\mathrm{GeV}/c^2$ at $68\%CL$). While this value does fall below the LEP exclusion bound, the experimental uncertainty of +35 and -26 GeV (at $68\%CL$) means that the two results are compatible. When the LEP direct search results and the precision electroweak measurements are combined, the 95% CL upper limit on the Higgs mass is $186\,\mathrm{GeV}/c^2$ [7].

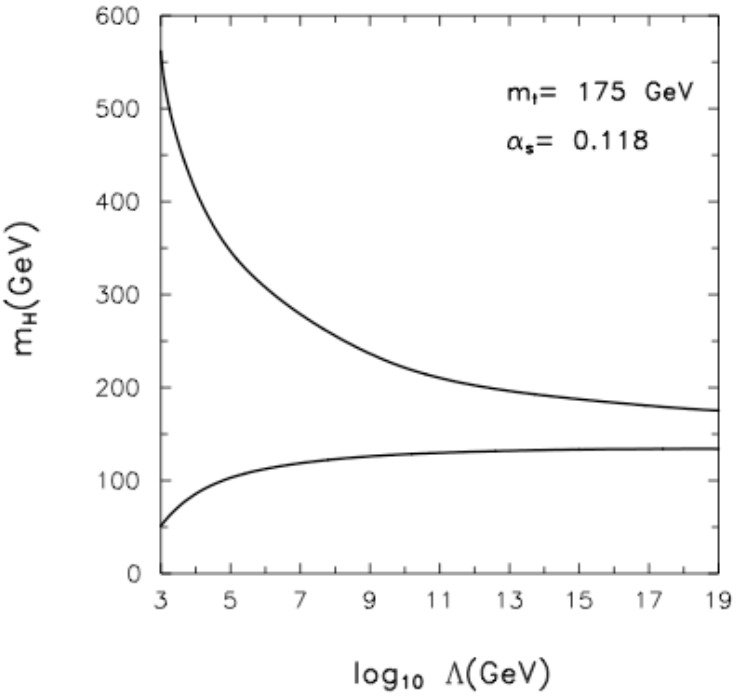


Figure 2.6: Upper and lower bounds on the Higgs mass. Bounds are plotted as a function of energy scale Λ. Taken from [24].

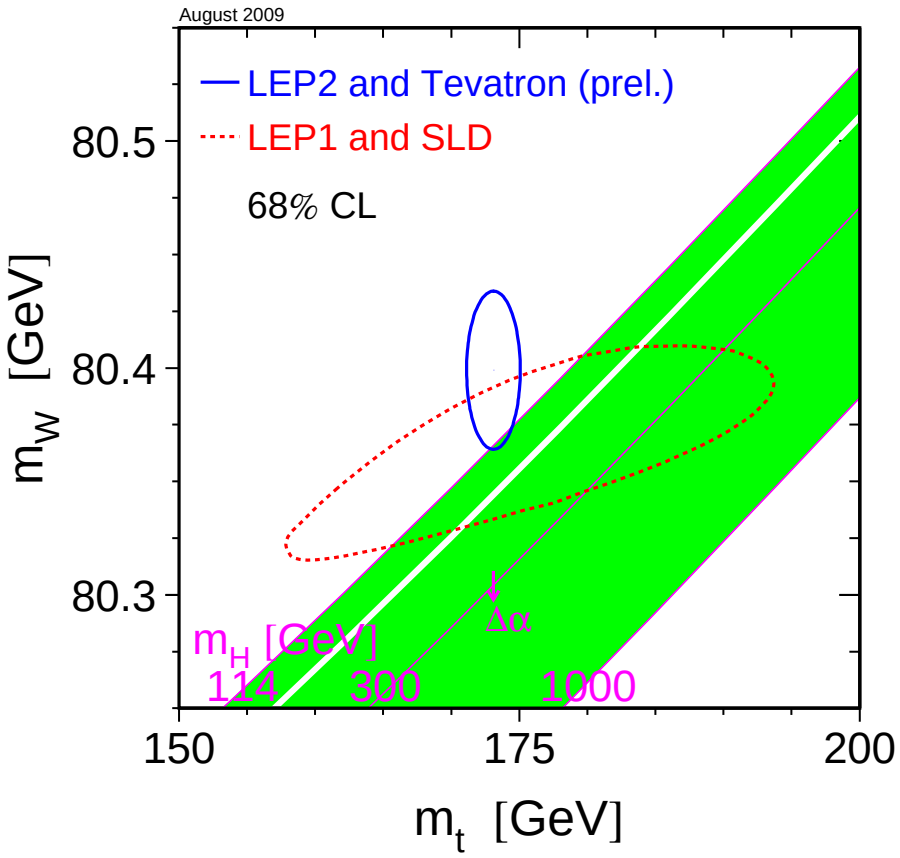


Figure 2.7: Comparison of indirect (dashed contour) and direct (solid contour) constraints on M_W and M_t. The SM relationship for the masses as a function of the Higgs mass is also shown, with the $\Delta\alpha$ arrow representing the shift in this relationship as strong coupling is varied by one standard deviation. Figure taken from [7].

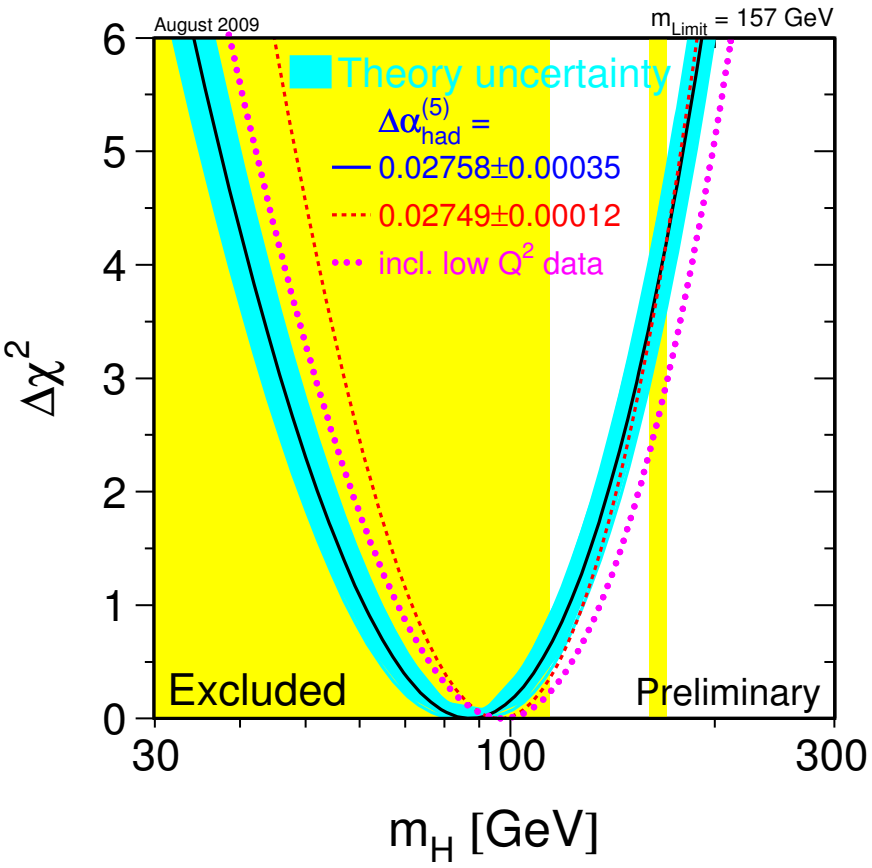


Figure 2.8: $\Delta\chi^2$ vs. M_H curve resulting from a fit to electroweak data. The yellow bands show the LEP and Tevatron 95% CL direct search exclusion limits on M_H. Figure taken from [7].

Chapter 3: Experimental Searches for the SM Higgs

The following sections outline efforts to observe the SM Higgs in searches performed by the LEP and Tevatron experiments. The chapter closes with a discussion of the $ZH \rightarrow \ell^+\ell^- b\bar{b}$ search at the Tevatron.

3.1 LEP Searches

Each of the four LEP experiments (ALEPH, DELPHI, L3 and OPAL) recorded approximately $700\,\mathrm{pb}^{-1}$ of e^+e^- collisions at center–of–mass energies ranging from 189 to 209 GeV. The primary Higgs production mechanism at LEP was $e^+e^- \rightarrow ZH$ [28]. The LEP searches focused on events with one of the following final states:

- $(H \rightarrow b\bar{b})(Z \rightarrow q\bar{q})$
- $(H \rightarrow b\bar{b})(Z \rightarrow \nu\bar{\nu})$
- $(H \rightarrow b\bar{b})(Z \rightarrow \ell^+\ell^-)$ where ℓ =electron or muon
- $(H \rightarrow \tau^+\tau^-)(Z \rightarrow q\bar{q})$.

The main background processes at LEP were two-photon exchange, and $e^+e^- \rightarrow$ fermion pairs, ZZ, or W^+W^-.

The individual data samples were examined for the presence of a SM Higgs signal [29, 30, 31, 32] with the ALEPH experiment reporting an excess in the data over the SM background expectation with a significance of approximately 3σ. By convention, a 3σ significance constitutes 'evidence' while 5σ significance represents 'discovery'. The ALEPH excess was consistent with a Higgs signal where the Higgs mass is roughly $115\,\mathrm{GeV/c^2}$. The

L3 and OPAL data samples were consistent with SM backgrounds, while DELPHI observed a slight deficit of events.

Once LEP operations ceased in 2000, the data recorded by the four experiments, totaling 2461 pb^{-1}, was simultaneously examined for the presence of a SM Higgs signal [9]. Analysis of the combined LEP dataset employed a modified frequentist approach [33], where the combined dataset was compared to two Monte-Carlo simulated sets, each corresponding to one of two hypothesis: the "(b)-hypothesis" containing only background events, and the "(s+b)-hypothesis" which contained a standard model Higgs of some assumed mass in addition to background. The quantity $-2\ln(Q)$, where Q is the ratio of the binned likelihoods for each hypothesis,

$$Q = \frac{\mathcal{L}_{s+b}}{\mathcal{L}_b} \tag{3.1}$$

was used as the test statistic. Figure 3.1 shows the test statistic plotted against the Higgs mass, where the negative values of the observed $-2\ln(Q)$ correspond to mass values where the (s+b)-hypothesis is slightly favored. Although the (s+b)-hypothesis is favored for Higgs masses in the range of 114 to 120 $GeV/\,\text{c}^2$, the departure from the background only hypothesis is statistically insignificant. The final LEP lower bound on the Higgs mass, obtained from the combined dataset, was $M_h > 114.4\ GeV/\,\text{c}^2$ at the 95% confidence level.

3.2 Tevatron Searches

Standard model Higgs searches by the CDF and D0 collaborations fall into two categories based on the decay mode of the Higgs boson: low mass searches where $M_H \leq 135\,\text{GeV}/c^2$, and high mass searches, $M_H > 135\,\text{GeV}/c^2$. For $M_H \leq 135\,\text{GeV}/c^2$, the main search modes at the Tevatron are associated production of the Higgs with a vector boson, with the Higgs decaying to $b\bar{b}$, and the vector boson decaying leptonically:

- $WH \rightarrow \ell^{\pm}\nu b\bar{b}$

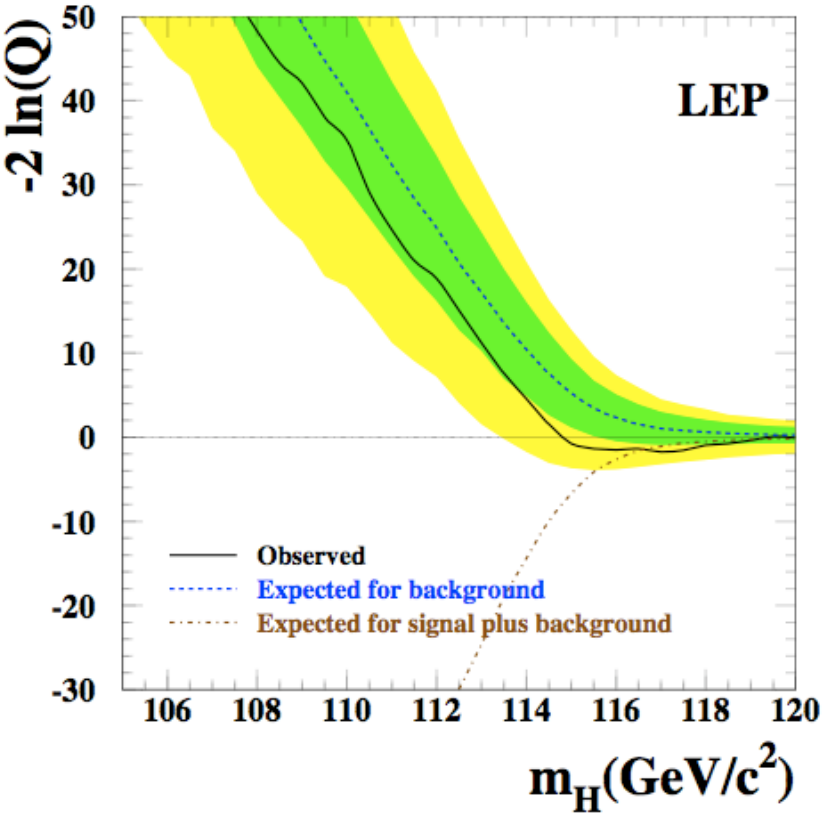


Figure 3.1: Behavior of $-2\ln(Q)$ for combined LEP data. The green and yellow bands represent the 68% and 95% probability bands about the median background expectation. Taken from [9].

- $ZH \to \nu\nu b\bar{b}$
- $ZH \to \ell^+\ell^- b\bar{b}$

Tevatron searches in the above modes attempt to identifying Higgs bosons decaying to a high transverse momentum b quark and b anti-quark pair. The hadronization of a high momentum quark produces a spray of secondary hadrons called a "jet". Tevatron low mass searches rely on algorithms designed to distinguish b quark jets from the jets produced when non-b quarks or gluons hadronize. Several of these algorithms are discussed in Chapter 7.

The low mass modes listed above are associated production modes where the Higgs is produced with a W/Z partner. Although the Higgs is more likely to be produced without a partner at the Tevatron by the gluon fusion process $gg \to H$, the detector signature of that process is difficult to distinguish from QCD multi-jet events. In associated Higgs production the decays of the W or Z to charged lepton(s) and/or neutrinos produce a final state that is distinguishable from multi-jet backgrounds.

Figure 3.2 shows the upper limits (as of March 2009) on SM Higgs production cross-sections $\times$ branching ratios obtained from a combination of searches performed by the Tevatron (CDF and D0) collaborations. The individual searches were performed in data samples of integrated luminosity between 0.9 and 4.2 fb^{-1}. The Tevatron experiments set a 95% CL upper limit of 2.5 times the standard model's values for the cross-sections $\times$ branching ratios for a Higgs with $M_H = 115\,\mathrm{GeV}/c^2$.

In the high mass ($M_H > 135\,\mathrm{GeV}/c^2$) region, the Tevatron is most sensitive to a SM Higgs in the production/decay mode $gg \to H \to WW \to \ell\nu\ell\nu$, with additional sensitivity from (Z/W) associated Higgs production where the Higgs decays to a pair of W bosons ($H \to W^+W^-$). The combination of D0 and CDF results excludes the existence of a SM Higgs boson with $160 \leq M_H \leq 170\ GeV/\,\mathrm{c}^2$ at the 95% confidence level.

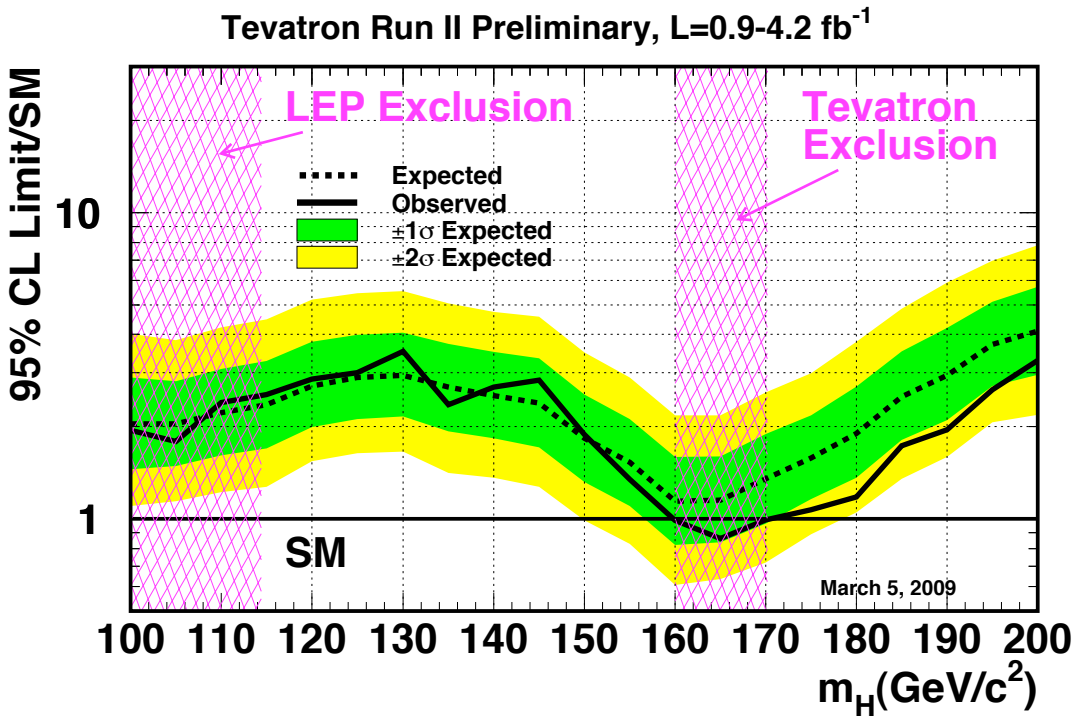


Figure 3.2: Expected (dashed line) and observed (solid line) Tevatron 95% CL upper limits on SM Higgs production cross-sections × branching ratios for Higgs masses between 100 and 200 $GeV/\,c^2$. The green and yellow bands represent the 68% and 95% probability bands about the expected limits. The pink shaded regions show the LEP and Tevatron 95% CL exclusions. Taken from [34].

3.3 Tevatron Search for $ZH \to \ell^+\ell^- b\bar{b}$

The detector signature of a $ZH \to \ell^+\ell^- b\bar{b}$ event at the Tevatron is two energetic b jets and two charged leptons whose combined mass is near that of the Z. The $ZH \to \ell^+\ell^- b\bar{b}$ Higgs search channel is distinguished from other Tevatron modes by the lack of neutrinos in the final state. The lack of neutrinos means that both the $H \to b\bar{b}$ and $Z \to \ell^+\ell^-$ decays can be reconstructed without the need to infer the presence of particles from missing energy. This feature provides a strong control on background processes that compensates for the low production cross section $\times$ $Z \to \ell^+\ell^-$ branching fraction.

Leading order Feynman diagrams for $ZH \to \ell^+\ell^- b\bar{b}$ and important background processes to the $ZH \to \ell^+\ell^- b\bar{b}$ search are shown in Fig. 3.3. The primary standard model background to $ZH \to \ell^+\ell^- b\bar{b}$ is $Z + b\bar{b}$ production. This process shares a final state that while similar to $ZH \to \ell^+\ell^- b\bar{b}$ is distinguished by the lack of a resonance in the distribution of the reconstructed dijet mass as shown in Fig. 3.4. The next largest background to Tevatron $ZH \to \ell^+\ell^- b\bar{b}$ searches is $t\bar{t}$ production where each top decays to a W and a b quark. The detector signature of a $t\bar{t}$ event is similar to that of $ZH \to \ell^+\ell^- b\bar{b}$ in that it has two b jets and two charged leptons (from $W \to \ell\nu$). However, the two neutrinos from the W decays appear as significant missing transverse energy; a feature that is not present in $ZH \to \ell^+\ell^- b\bar{b}$ events as seen in Fig. 3.5.

Previous Tevatron searches for the $ZH \to \ell^+\ell^- b\bar{b}$ process are described in [35, 36, 37]. The general strategy employed in these searches is to select a sample of events with an identified $Z \to e^+e^-$ or $Z \to \mu^+\mu^-$ decay, and two jets in the event. Background processes with non-b jets are reduced by requiring at least one of the two jets to be identified as a b quark jet. The sample of events with a Z candidate and at least two b jets is compared to the SM background for any excess consistent with a ZH signal. To date no experiment has observed an excess over background processes consistent with a Higgs signal. The most stringent 95% CL upper limit on $\sigma_{ZH} \times BR(H \to b\bar{b})$ obtained in [36] is 8.2$\times$ the standard model values for a Higgs mass of $115\,\mathrm{GeV}/c^2$.

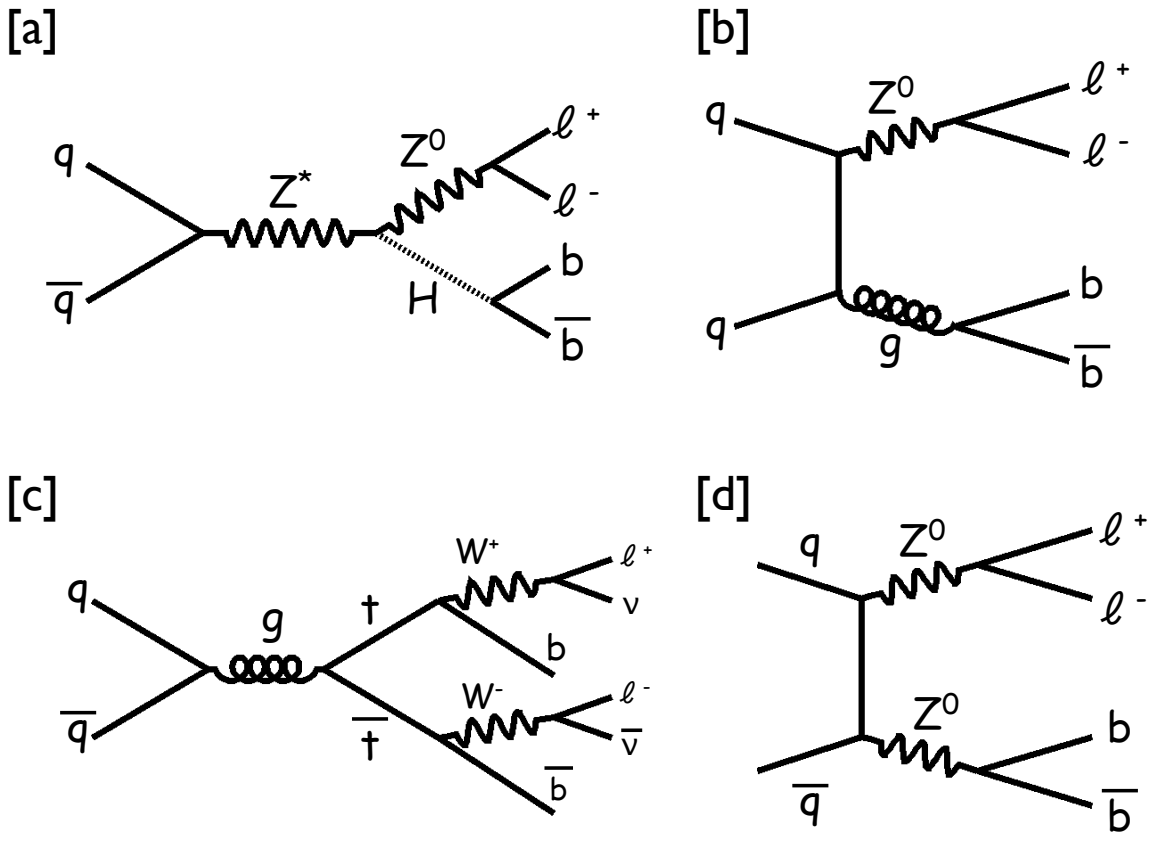


Figure 3.3: Leading order Feynman diagrams for $ZH \to \ell^+\ell^- b\bar{b}$ [a], $Z + b\bar{b}$ [b], $t\bar{t}$ [c], and Diboson ZZ production [d]. $Z + b\bar{b}$ is the dominant background to Tevatron $ZH \to \ell^+\ell^- b\bar{b}$ searches.

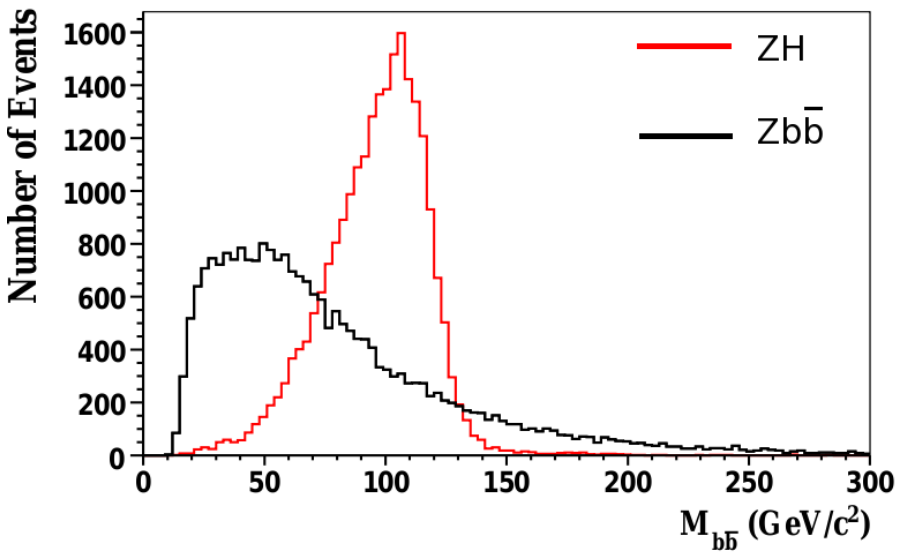


Figure 3.4: Comparison of the reconstructed $b\bar{b}$ mass in $ZH \to \ell^+\ell^- b\bar{b}$ (red) and $Z + b\bar{b}$ (black) simulated events. The simulation uses a Higgs mass of 120 GeV/c^2.

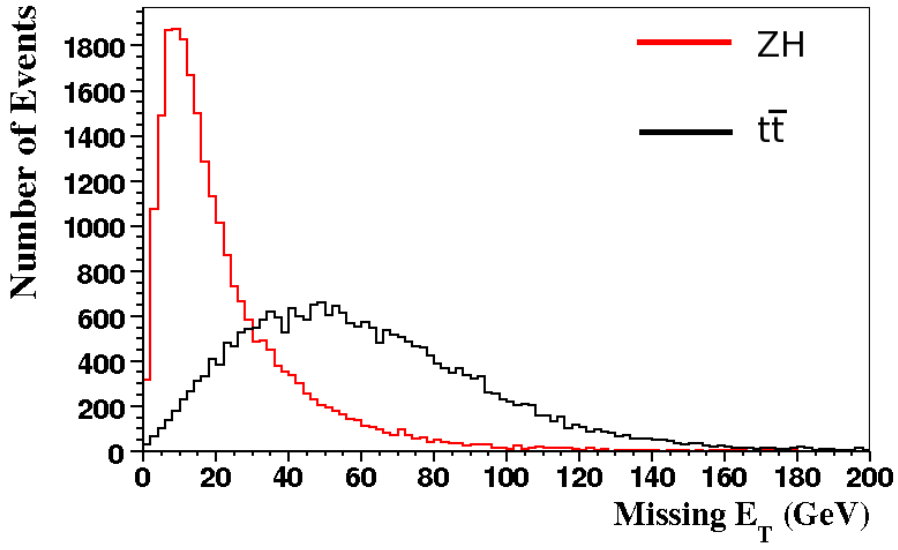


Figure 3.5: Comparison of the magnitude of the missing transverse energy in $ZH \to \ell^+\ell^- b\bar{b}$ (red) and $t\bar{t}$ (black) simulated events. The simulation uses a Higgs mass of 120 GeV/c^2.

Chapter 4: The Tevatron and CDF II

The data analyzed in this thesis was recorded by the Collider Detector at Fermilab (CDF II) and represents approximately seven years of Tevatron collisons. The following three sections (4.1,4.2,4.3) describe the main features of the Fermi National Accelerator Laboratory (FNAL) accelerator complex, the Tevatron beam, and the components of the CDF II detector most relevant to this analysis.

4.1 The Tevatron and the FNAL Accelerator Complex

The Tevatron [38] is a $p\bar{p}$ collider located at FNAL in Batavia, IL. An aerial view of the Tevatron is shown in Figure 4.1. Since 2001 (the phase of operation known as Run II), the Tevatron has operated at a center of mass energy of 1.96 TeV, making it the world's highest energy particle collider. The Tevatron held this distinction until the Large Hadron Collider (LHC) became operational in late 2009.

To reach a collision energy of 1.96 TeV, a series of accelerators is employed to produce beams of protons and antiprotons with an energy of 980 GeV for collisions at either of the Tevatron's two interaction regions (labeled B0 and D0 in Fig. 4.2). The CDF II detector is located at the B0 interaction point, while D0 (as the name implies) is situated at the D0 point.

Figure 4.2 shows the various accelerators in the Tevatron accelerator chain. The preaccelerator is the first stage in the chain, were negative hydrogen ions, (H^-), are accelerated by either of two Cockroft-Walton accelerators to 750 keV. The ions are then injected into a linear accelerator (Linac) which raises the kinetic energy of the H^- ions to 400 MeV. The energy increase in the Linac is achieved through the use of Radio Frequency Cavities (RFC) [39]. As it travels through the "Linac Buncher" the continuous beam of

H^- ions is separated into 201.24 MHz bunches.

These 400 MeV H^- ions are then directed onto a thin carbon foil. As they pass through the carbon foil, two electrons are stripped from each ion leaving bare protons. The protons enter the "booster" synchrotron. Unlike the linac where varying drift tube lengths ensure that particles are shielded from the decelerating phase of electric field, synchrotrons rely on precise synchronization of the field to the beam momentum to achieve acceleration. In the booster, the protons are accelerated to an energy of 8 GeV.

After acceleration in the booster, the protons are transferred to another synchrotron called the main injector, where they are accelerated to an energy of either 120 or 150 GeV. From the main injector, 150 GeV protons intended for collision enter the Tevatron, while others are accelerated to 120 GeV before being directed onto a nickel source to produce antiprotons [40]. Approximately one antiproton is recovered for every $50,000$ proton-nickel target collisions. Through various improvements to the FNAL accelerator chain (including the addition of the main injector and the Recycler $\bar{p}$ storage ring) the antiproton production rate has increased from 6×10^{10} to 2×10^{11} $\bar{\mathrm{p}}$/hr since 1986.

After antiprotons are produced, they enter the antiproton ring which consists of the Debuncher and the Accumulator. The Debuncher is a rounded triangular synchrotron which serves to reduce the momentum spread of the antiprotons, while at the same time "debunching" them— that is increasing their physical separation. The antiprotons leave the Debuncher and enter the Accumulator, another triangular-shaped synchrotron, where they are stored until enough are present for collisions in the Tevatron. When enough protons and antiprotons are available, (approximately $6.5 \times 10^{12} p$ and $4.3 \times 10^{11} \bar{p}$), they are injected into the Tevatron from the main injector at 150 GeV. There, the protons and antiprotons are accelerated to 980 GeV, as they travel in opposite directions around the nearly four mile circumference of the Tevatron. The protons and antiprotons are grouped into 36 bunches each, and directed towards the two collision points where the CDF and D0 detectors are located. The beams cross every 396 ns.

4.2 Tevatron Luminosity

The Tevatron is characterized by the instantaneous luminosity L which is defined [38] as:

$$L = \frac{f \times n \times N_p \times N_{\bar{p}}}{A} \tag{4.1}$$

where f is the beam revolution frequency, n is the number of proton and antiproton bunches, N_p ($N_{\bar{p}}$) is the number of protons (antiprotons) in each bunch, and A is a factor which depends on the width of the gaussian beam shapes and a form factor dependent on the bunch size. Average values for these quantities are listed in Table 4.1.

The average instantaneous luminosity at the Tevatron in Run II is approximately $1.7 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ with peak values of $3.2 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

The expected number of ZH events is given by the product of the associated production cross-section (σ_{ZH}) and the integrated luminosity $\mathcal{L} = \int L\ dt$. As of October 2009 the Tevatron has delivered an integrated luminosity of nearly $6\,\mathrm{fb}^{-1}$ to the CDF experiment, and with $4.1\,\mathrm{fb}^{-1}$ analyzed in this thesis, the expected number of (inclusive) ZH events produced in the CDF detector is between 690 and 170 events, depending on the assumed Higgs mass. With SM branching fractions for $H \to b\bar{b}$ between 0.81 and 0.18 and a branching fraction of 0.066 for $Z \to \ell^+\ell^-$ ($\ell = e, \mu$), the SM prediction for the number of $ZH \to \ell^+\ell^- b\bar{b}$ events falls between 37 and 2 events for Higgs masses between 100 and $150\,\mathrm{GeV}/c^2$.

4.3 The CDF II Detector

The Collider Detector at Fermilab II (CDFII) [41] is a general purpose $p\bar{p}$ detector consisting of tracking systems, calorimeters and muon detectors, designed for the study of $p\bar{p}$ collisions at the Tevatron. An isometric view of the CDFII detector with the major components labeled is shown in Fig. 4.3. The following subsections present a brief description of the CDFII detector components, as well as the trigger and data acquisition

Table 4.1: Average values for Tevatron beam parameters in Run II.

Quantity	Tevatron Run II Value
Revolution Frequency (f)	$\sim 47\,\mathrm{kHz}$
Number of Bunches (n)	36
Number of Protons per Bunch (N_p)	$\sim 2.7 \times 10^{11}$
Number of Antiprotons per Bunch ($N_{\bar{p}}$)	$\sim 7 \times 10^{10}$
A	$1.838 \times 10^{-4}\,\mathrm{cm}^2$
Instantaneous Luminosity (L)	$1.7 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$

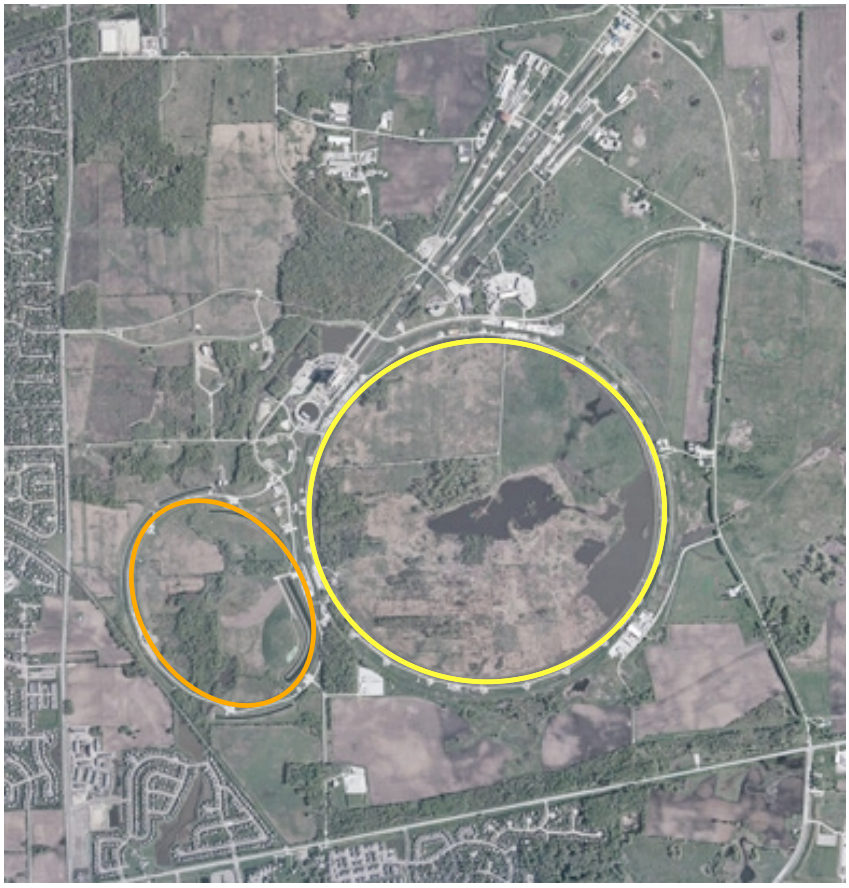


Figure 4.1: Aerial view of the FNAL accelerator complex. The Tevatron (yellow highlight) is the large circular (radius of 1 km) object in the center of the photo. The main injector (orange highlight) is also shown.

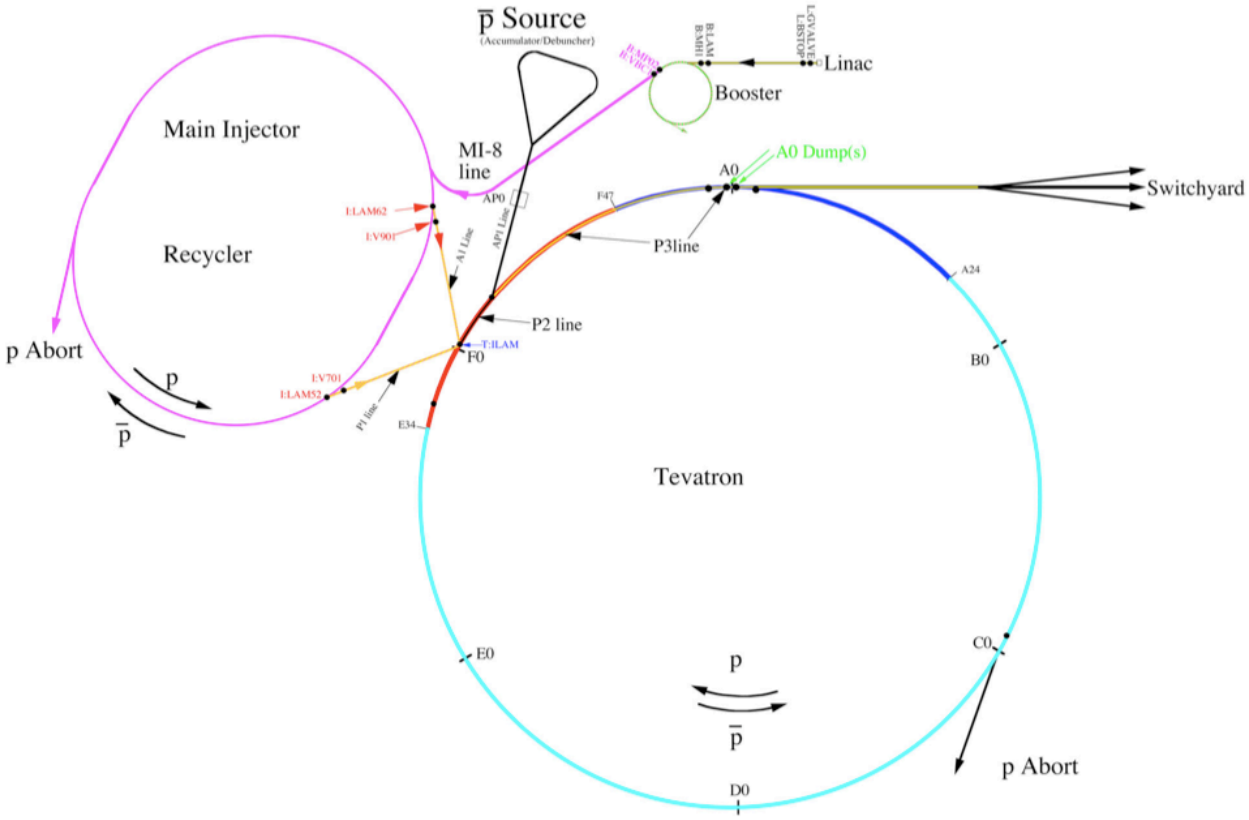


Figure 4.2: The Tevatron accelerator chain. Figure from [39].

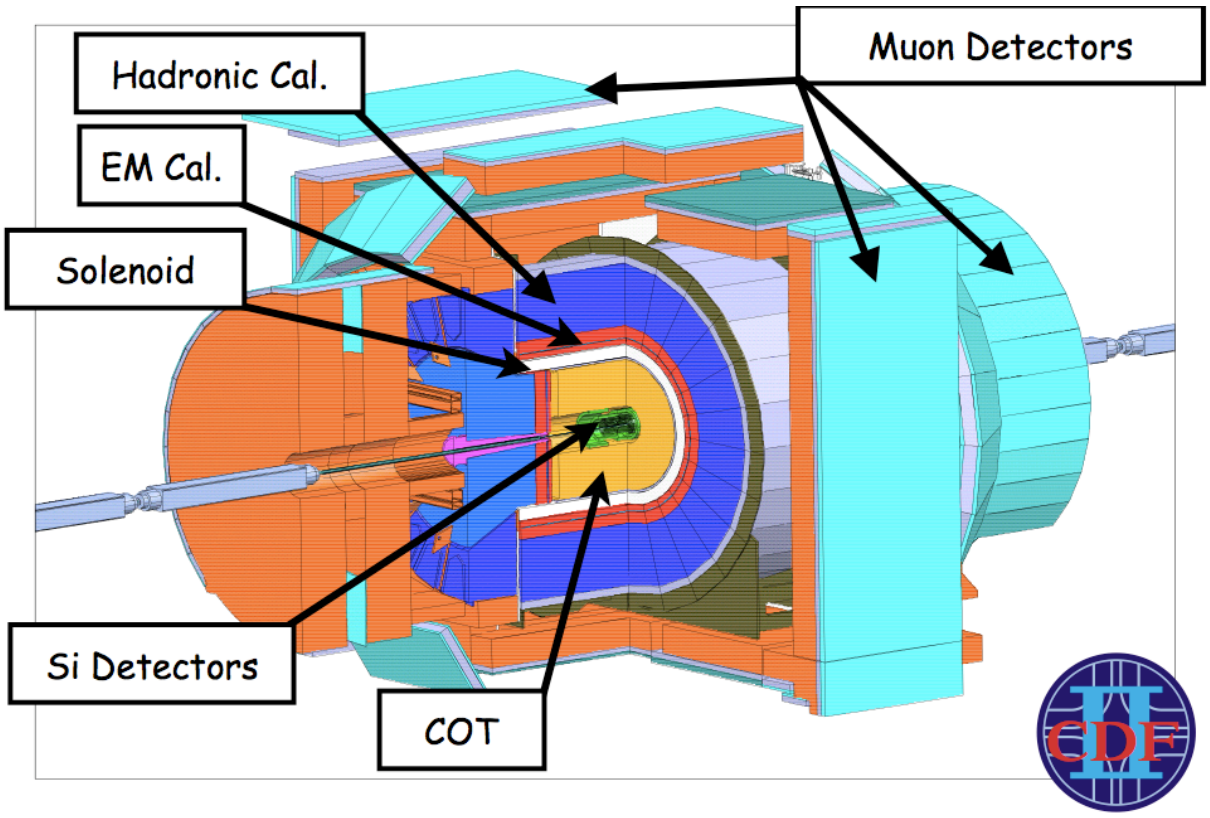


Figure 4.3: Isometric view of the CDF II Detector with silicon detectors, Central Outer Tracker (COT), solenoid, calorimeter and muon detector components labeled. Image from [42] with modification.

(DAQ) systems.

4.3.1 The CDF Coordinate System

The CDF II detector is a roughly cylindrical device built around the Tevatron beampipe, measuring approximately 40 feet in length, width, and height and weighing roughly 100 tons [43]. We use a cylindrical coordinate system, as depicted in Fig. 4.4. Protons travel eastward through CDF in the $+z$ direction while anti-protons travel westward in the $-z$ direction. Beam collisions occur near $z = 0$, at the origin of the detector coordinate system. The radial distance from the z axis is denoted by r, while the angle θ measures the polar angle from the beamline and ϕ is the azimuthal angle around the beamline. The psuedorapidity

(η) defined as :

$$\eta = -\ln\left\{\tan\frac{\theta}{2}\right\} \tag{4.2}$$

is used to describe polar angles. Objects perpendicular to the beamline have psuedorapidity zero, while psuedorapidity approaches positive (negative) infinity along the $+z$ ($-z$) direction. The *central* region of the CDF coordinate space is defined by $|\eta| \leq 1.1$ while the *forward* (also called *plug*) region covers values of $1.1 \leq |\eta| \leq 3.6$.

The separation between two points in the $\eta - \phi$ plane is given by the quantity ΔR which is defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{4.3}$$

In collisions between composite objects (p and $\bar{p}$ for example) the initial state momentum is known only in the plane transverse to the beampipe (*i.e.* zero GeV/c). The consequence of this is that conservation of momentum and energy can only be applied in the transverse plane. We define the *transverse momentum* (p_T) of an object with momentum $\vec{p}$ and polar angle θ as

$$p_T = |\vec{p}| \times \sin(\theta). \tag{4.4}$$

Similarly, the *transverse energy* (E_T) of an object with energy E and polar angle θ is given by

$$E_T = E \times \sin(\theta). \tag{4.5}$$

The energy imbalance in the transverse plane is denoted as the missing transverse energy ($\vec{\not\!\!E}_T$) with

$$\not\!\!E_T = |\vec{\not\!\!E}_T|. \tag{4.6}$$

and where $\vec{\not\!\!E}_T$ is defined as :

$$\vec{\not\!\!E}_T = -\sum_i (E_T)_i \times \hat{n}_i \tag{4.7}$$

were i denotes the calorimeter tower (discussed in Chapter 5) number and $\hat{n}_i$ is a unit vector

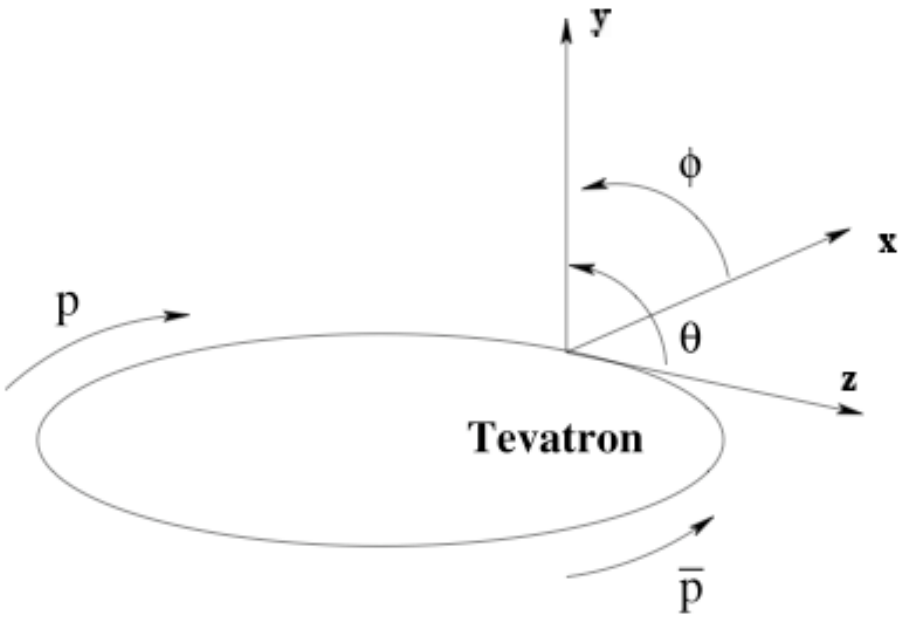


Figure 4.4: The CDF Coordinate System.

pointing out radially from the origin to the ith calorimeter tower.

4.3.2 The Cherenkov Luminosity Counter

The instantaneous luminosity (L) can be expressed as

$$L = \frac{\mu \times f_{bc}}{\sigma_{in}}, \tag{4.8}$$

where μ is the average number of $p\bar{p}$ interactions per bunch crossing, f_{bc} is the Tevatron bunch crossing frequency (1.515×10^7 Hz) and σ_{in} (approx. 60 mb [44]) is the total inelastic $p\bar{p}$ interaction cross-section. The purpose of the CDF Cherenkov Luminosity Counter (CLC) is to obtain an accurate measurement of μ.

The CLC consists of two Cherenkov light detectors located at the opposite ends ($\pm z$) of the CDF detector's plug regions. Each CLC module consists of 48 isobutane filled Cherenkov

counters surrounding the beampipe and providing coverage for the space $3.7 < |\eta| < 4.7$. A photomultiplier tube (PMT) at the end of each counter collects the Cherenkov radiation produced as charged particles travel through the isobutane gas. A schematic view of the CLC is presented in Fig. 4.5.

The μ parameter is measured under the assumption of Poisson statistics by recording the fraction of bunch crossings with no significant Cherenkov light in the CLC modules. Since the Poisson probability for such a "no hit" event is given by

$$P_{no\ hit} = e^{-\mu}, \tag{4.9}$$

the value of μ can be obtained. The value of μ for CDF in Run II is about 6.

The value μ is then combined with experimental measurements of the CLC detector efficiency [45, 46] to produce measurements of the instantaneous luminosity.

4.3.3 The Silicon Detectors

In order to provide accurate track reconstruction, and to distinguish between charged particles coming from the primary $p\bar{p}$ interaction vertex and those from the decays of secondary particles, CDF contains a high resolution silicon tracking detector [47, 48, 49] close to the beampipe. The impact of the silicon detector system on CDF low mass Higgs searches cannot be overstated; without the ability provided by the high resolution tracking of Si detectors to distinguish between particles from secondary B meson decays and particles coming from the primary vertex, backgrounds to $ZH \rightarrow \ell^+\ell^- b\bar{b}$ would be approximately 35 times larger.

A silicon detector is created from doped silicon strips, each with a bias voltage applied. A charged particle traveling through a particular strip produces a current; by combining the current readings from multiple (376 modules in the CDF Si detector) silicon strips the particle's path through the silicon detector can be reconstructed. The tracking resolution of

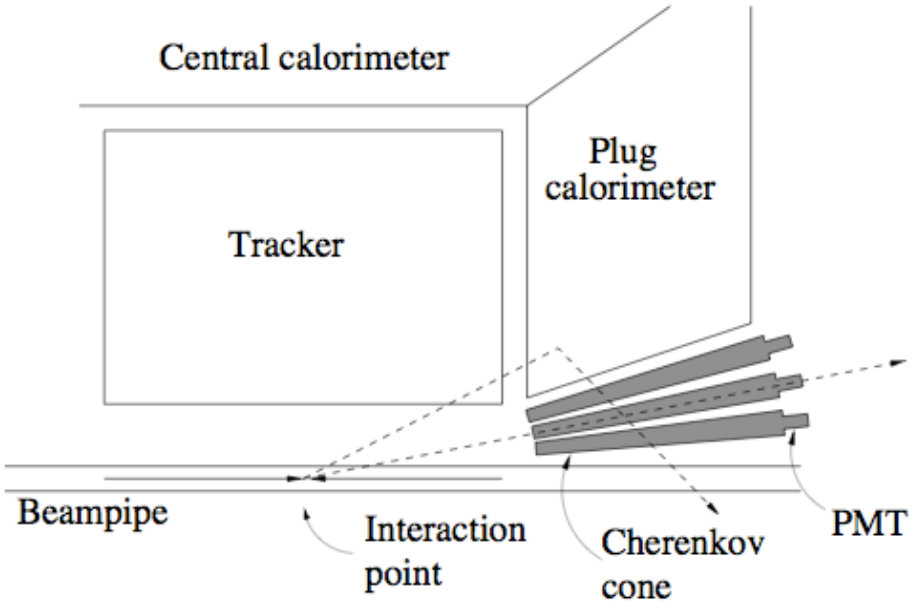


Figure 4.5: Schematic view of the CDF CLC. Figure taken from [46].

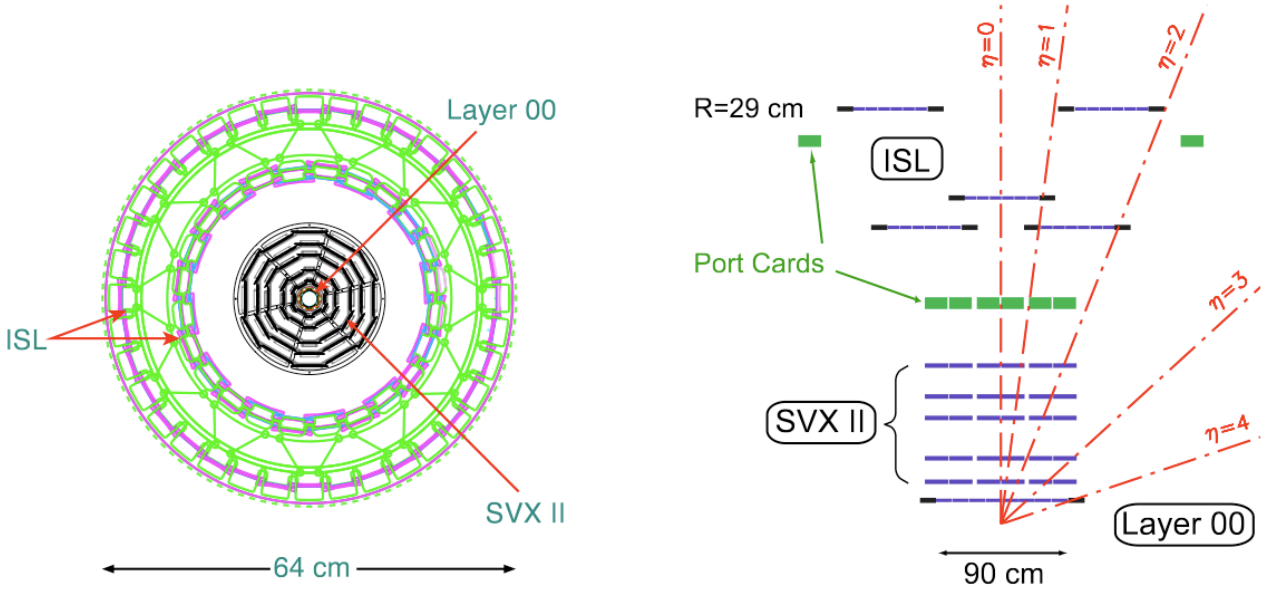


Figure 4.6: Axial (left) and $r-z$ (right) views of the CDFII Silicon Detectors. Figure taken from [50].

the CDF silicon detector is on the order of 10μm, with a track impact parameter resolution of approximately 40μm.

The silicon detector has three components, Layer 00 (L00), the Silicon Vertex Detector (SVX) and the Intermediate Silicon Layer (ISL). Axial and $r-z$ views of the silicon detector are shown in Fig. 4.6. L00 is located directly outside the beampipe, at roughly $r = 1.6\,\text{cm}$ and provides tracking coverage for $|\eta| < 4.0$. L00 is a single-sided silicon microstrip detector designed for making precision track measurements close to the beampipe. The SVX is a high-precision tracking and secondary vertex detector, consisting of five layers of double-sided silicon microstrip detectors, located from $r = 2.1$ to $17.3\,\text{cm}$ and covering $|\eta| < 2.0$. The ISL is located outside of the SVX and is a double-sided silicon microstrip detector designed to provide track linking between the SVX and the CDF Central Outer Tracker.

4.3.4 The Central Outer Tracker

The Central Outer Tracker (COT) [51], is a cylindrical open-cell drift chamber using a 50/50 mix of argon-ethane gas. The COT extends from $r = 40$ to 137 cm and provides tracking for $|\eta| < 1.0$. The COT contains $30,240$ sense wires organized into 96 layers in r, which in turn are grouped into eight "superlayers". Each superlayer is divided in ϕ forming 2520 "supercells" which contain both sense and field wires.

A voltage maintained on the field wires accelerates electrons, produced when charged particles travel through the gas, towards the sense wires. The sense wires span the length of the COT (approximately 310 cm in z), and register the current produced when electrons arrive at the wires as "hits". The r and ϕ information from multiple supercell hits can be fit to a helix reconstructing the path of a charged particle through the COT.

The COT is surrounded by a solenoidal magnet which generates a 1.4 Tesla magnetic field. The magnetic field of the NbTi superconducting solenoid deflects the path of charged particles in the COT and silicon systems, allowing for charge sign determination and momentum measurement.

The COT has a hit position resolution of roughly 140 μm in the $r - \phi$ plane and a transverse momentum resolution $(\sigma(p_T)/(p_T)^2)$ of 0.0015 $(\mathrm{GeV}/c)^{-1}$ [52]. The COT is depicted in Fig. 4.7.

4.3.5 The Calorimeters

CDF calorimeters [53, 54, 55] are constructed from alternating layers of absorption and detection (scintillator) materials. A particle interacting with the absorption material will produce a shower of secondary particles which produce light as they travel in the scintillator layer. The amount of light produced is proportional to the number of shower particles which depends on the energy of the original particle; from measurements of the light yield we can estimate the original particle's energy.

CDF features a combination of electromagnetic and hadronic calorimeter systems

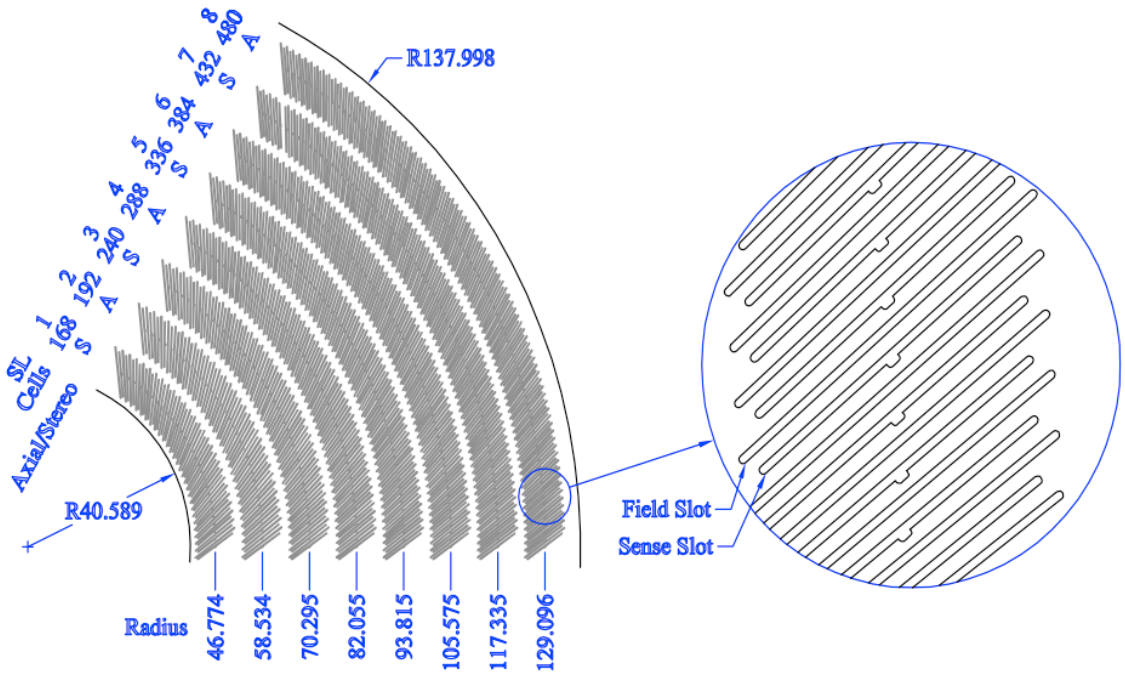


Figure 4.7: End view of one sector of the COT. Dimensions are in cm. Figure taken from [51].

designed to measure particle energies. Hadronic calorimeters, consisting of alternating layers of scintillator and iron absorption material, measure the energy of hadronic showers, while electromagnetic calorimeters, consisting of scintillators with lead as the absorption material, measure the energy of electromagnetic showers.

The main calorimeter systems are :

- central electromagnetic calorimeter (CEM)
- plug electromagnetic calorimeter (PEM)
- central hadronic calorimeter (CHA)
- plug hadronic calorimeter (PHA)
- endwall hadronic calorimeter (WHA)

Table 4.2 presents a summary of the main features of the CDF calorimeter systems which are depicted in Fig. 4.8.

Table 4.2: Location and energy resolution of the CDF calorimeter systems. The quoted resolutions are from [52] and references therein.

Calorimeter	Location	Energy Resolution
CEM	$\|\eta\| < 1.1$, outside solenoid	$13.5\%/\sqrt{E_T} \oplus 2\%$
CHA	$\|\eta\| < 0.9$, outside CEM	$75\%/\sqrt{E_T} \oplus 3\%$
PEM	$1.1 < \|\eta\| < 3.6$, outside COT end planes	$16\%/\sqrt{E} \oplus 2\%$
PHA	$1.2 < \|\eta\| < 3.6$, outside PEM	$74\%/\sqrt{E} \oplus 4\%$
WHA	$0.8 < \|\eta\| < 1.2$	$75\%/\sqrt{E} \oplus 4\%$

4.3.6 The Muon Detectors

While electrons, photons, and hadronic particles are expected to be absorbed in the calorimeter materials, muons do not lose a significant amount of energy to bremsstrahlung and can reach the outer radii of the CDF detector. In order to detect the presence of these particles, CDF has several muon detectors positioned outside of the hadronic calorimeters. The CDF muon detectors [57] are wire chambers operating in "proportional" mode; meaning that a strong electric field is maintained throughout the detector. This causes the ionization electrons produced as a muon passes through the chamber, to themselves become ionizing. This cascade of secondary particles are collected on sense wires and produces a signal whose strength is proportional to the energy of the original ionization. Combined with drift time information, the signals produced as a muon passes through multiple muon detector chambers can be combined to form a small track segment called a "stub". When matched to track information from the silicon detectors and the COT, the path of the muon from the production point to the muon detector can be reconstructed.

The primary CDF muon systems are the :

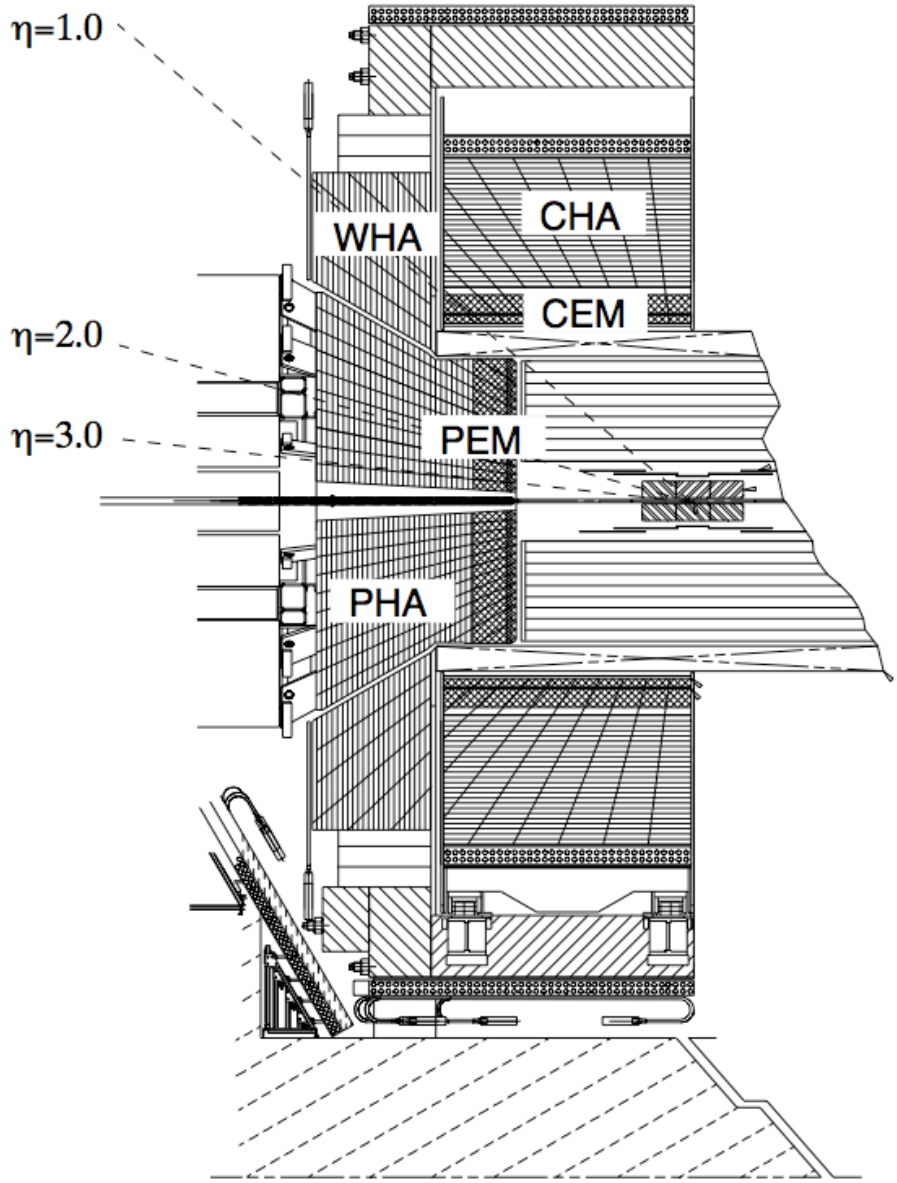


Figure 4.8: Elevation view of one half of the CDF II detector with the calorimeter systems labeled. Figure taken from [56].

- central muon chambers (CMU)
- central muon upgrade (CMP)
- central muon extension (CMX)

Muons with a transverse momentum as low as $1.4\,\mathrm{GeV}/c$ can be detected in the CMU detectors which are located behind the CHA detectors and cover the region $|\eta| < 0.6$. In order to reduce the background from surviving hadronic particles, CMU signals are matched to stubs in the CMP. The CMP is situated behind an additional layer of steel shielding, designed to limit the number of hadronic particles surviving to the CMP. Due to the extra shielding, hadronic particles are unlikely to produce CMU signals that are matched to a corresponding signal in the CMP. Objects with matching signals in the CMU and CMP are reconstructed as 'CMUP' muons. The $|\eta|$ range from 0.6 to 1.0 is covered by the CMX chambers.

4.3.7 The Trigger System

The Tevatron bunch crossing rate is approximately 1.7 MHz, while event data can be stored at a rate of about 100 Hz. Therefore the vast majority of events cannot be recorded. In order to ensure the efficient selection of events significant for physics studies (*i.e.* those with energetic particles, displaced vertices, or large $\not\!\!E_T$), CDFII employs a three-level trigger system. A schematic of the data flow is depicted in Fig. 4.9.

At level one, hardware triggers use signals from the calorimeters, COT and muon detectors, to decide weather an event should be considered further. Decision times at level one are about 5 μs with an acceptance rate near 30 kHz. Level two combines software and hardware triggers, with a typical decision time of about 20 μs. Information from additional systems is available at level two, such as data from the SVX. Level two passes events at rates on the order of a few hundred Hz. Event data passing the level two trigger is sent to data acquisition storage buffers, and then transferred to a level three decision farm node

for complete event reconstruction. Events meeting the requirements of the level three trigger are accepted at a rate near 100 Hz for storage.

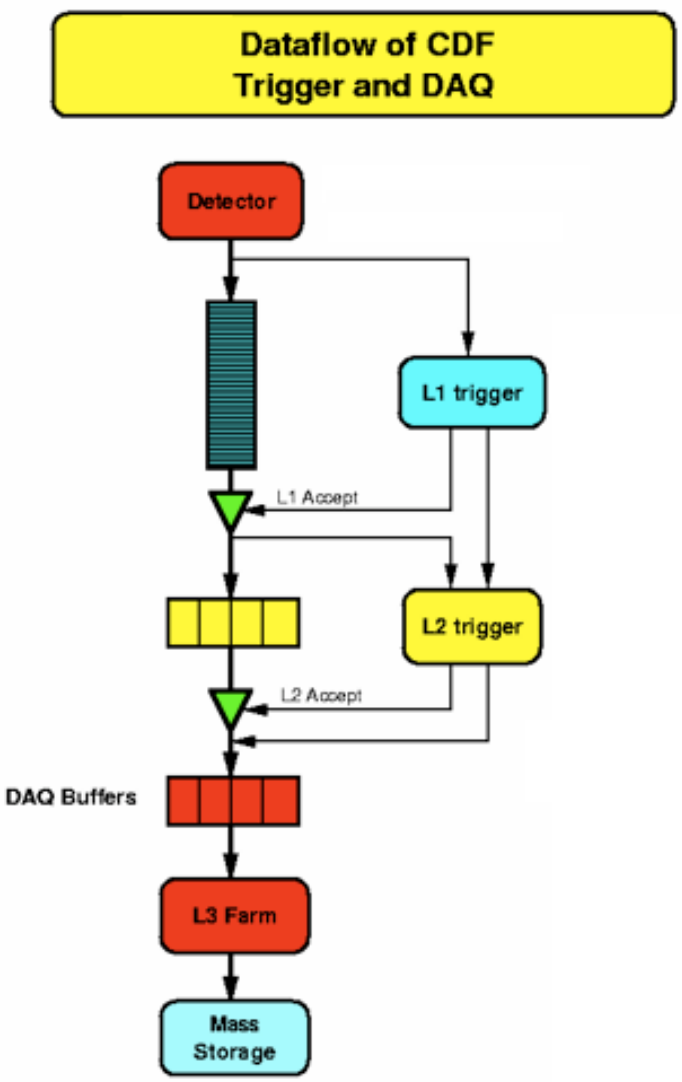


Figure 4.9: Depiction of the flow of CDF II data through the 3-level trigger system to mass storage.

Chapter 5: Object Reconstruction

The combined readout of the CDF detector's subcomponents (roughly 250 kbytes of data per event) is used to form tracks, calorimeter energy clusters, and muon detector signals which in turn form physics objects such as electrons, muons, and jets. The following sections will provide a brief description of the track, electron, muon, $\not\!\!E_T$ and jet reconstruction algorithms used at CDF.

5.1 Track Reconstruction

The momentum, p, of a charged particle traveling through a constant magnetic field is proportional to the radius of curvature of the particle's path :

$$p = 0.3 \times B \times R \tag{5.1}$$

where p is in units of GeV/c, B is the magnitude of the magnetic field in Tesla, and R is the radius of curvature in meters. To measure R, the path of a charged particle through the CDF tracking detectors must be reconstructed. CDF employs several algorithms for track reconstruction depending on which components of the detector a particle travels through; due to the cylindrical geometry of the CDF detector, a charged particle traveling through the detector with $|\eta| \leq 1.1$ (central) will pass through more layers of tracking detector than a forward particle ($|\eta| > 1.1$).

Central tracks are reconstructed using the outside-in (OI) [58] track reconstruction algorithm. An OI track is formed by identifying a "seed" (*i.e.* starting) hit in an outer COT layer. Starting from the seed hit, additional COT hits are added to form a track moving inwards toward the $p\bar{p}$ interaction point. The track reconstructed in the COT is then required to match a track reconstructed separately from silicon detector signals. The

OI algorithm is used to reconstruct tracks for particles that pass through at least half of the COT in the r direction. For such tracks, the COT tracking efficiency (with the OI algorithm) has been measured to be close to one ($98.62 \pm 0.12\%$ for tracks with $p_T \geq 10\,\mathrm{GeV}/c$) [59] in a sample of W candidates formed from calorimeter based $W \to e\nu$ reconstruction. Events in this W sample are required to also contain a silicon track. The efficiency is measured as the fraction of the events in which the W is linked to the silicon track by an OI track. Measurements of the efficiency in simulated $W \to e\nu$ events find an efficiency of $97.2 \pm 0.2\%$ for tracks with $p_T \geq 10\,\mathrm{GeV}/c$ [60]. The slight difference in the efficiencies measured in data and in simulated events is attributed to Bremsstrahlung radiation. In the simulation, the likelihood that an electron which emits a photon has a silicon track that correctly extrapolates to the calorimeter is slightly overestimated. Therefore, the efficiency measurement in the simulation includes Bremsstrahlung events that are less likely to enter the efficiency measurement in data.

In order to reconstruct the path of charged particles at $|\eta| > 1.1$ an inside-out (IO) [61] algorithm is employed. In this algorithm, each silicon track is propagated outward in r towards the exterior of the COT forming a potential path. Next, each layer of the COT is examined for hits close to this path and (if hits are present) a χ^2 fit is performed to compute the likely track parameters. IO tracking is used in the reconstruction of muon candidates at $\eta > 1.0$ and is greater than 95% efficient for high p_T tracks.

A third algorithm is applied to reconstruct the tracks of electron candidates which enter the forward calorimeter systems. The PHOENIX [62] algorithm begins by locating both an energy cluster in the calorimeter and a silicon track. The silicon track is used to identify the particle's point of origin. All detector signals along the probable path between the point of origin and the calorimeter deposit are fit to two likely trajectories; one for each sign of charge ($\pm$). The trajectory formed under the charge hypothesis which best fits the detector signals is taken as the particle's track. The PHOENIX track reconstruction algorithm has an efficiency of approximately 92%, but is limited by a significant (approx. 30% for $|\eta| > 2$)

charge misidentification rate.

In addition to the offline, *i.e.* post-trigger selection, track reconstruction algorithms presented above, the CDF trigger system utilizes the eXtremely Fast Tracker (XFT) [63] system for fast, online track reconstruction. The XFT processor forms tracks from COT hits and performs track extrapolations to calorimeter or muon detector signals; forming electron or muon candidates used in online trigger selection. In addition to forming trigger objects, cuts on XFT track parameters (such as the number of hits, XFT fit χ^2, etc.) are also utilized in analysis level object selection (for example requiring a minimum number of XFT identified COT hits for a muon candidate).

5.2 Electrons

Electrons are formed by matching a calorimeter energy deposit (called a "cluster") to either an OI or PHOENIX track. The CDF calorimeter systems are divided in η and ϕ into 478 CEM, 384 CHA, 288 WHA, 960 PEM, and 864 PHA regions known as "towers". The energy content of each tower is read out by a pair of PMTs, each responsible for a particular range of the tower in ϕ.

Electron energy clusters [64] are formed around "seed" towers with an electromagnetic energy (EM) of at least 2 GeV. Seed towers are organized in descending EM E_T, and beginning with the highest E_T tower, clusters are formed by combining seeds with all neighboring towers with EM energy greater than 100 MeV. A seed tower adjacent to another seed tower of greater energy is removed from the list of seeds. Towers that neighbor more than one seed tower are clustered with the seed with higher energy.

Once a cluster is formed, it is required to meet the following criteria :

- $EM\ E_T \geq 5$ GeV and
- the ratio of Hadronic (Had) energy to EM energy is less than or equal to 0.125 or
- $EM\ E_T \geq 100$ GeV

The towers assigned to clusters which fail these requirements are released and can be used to construct alternate clusters around seeds of lower energy.

Electron candidates are characterized by the following quantities :

- E_T - The transverse energy of the electron candidate is measured from the combined energies of calorimeter deposits in up to two adjacent calorimeter towers.

- p_T - The transverse momentum of the reconstructed track.

- fiduciality - Due to the segmented structure of the CDF calorimeter, there are uninstrumented regions every 15° in ϕ and at η values of 0 and ± 1.1 (as indicated in Fig. 5.1). Objects formed from tracks that point to these regions are assigned a fiduciality value of zero.

- Region - This quantity indicates the location of the calorimeter deposit in $|\eta|$; central clusters are assigned a Region value of 0 while forward objects have a value of 1.

- HadEm - The ratio of the amount of energy deposited by the electron candidate in the hadronic calorimeter to that deposited in the electromagnetic calorimeter. Electrons have small values of HadEm.

- conversion - Electrons produced in photon conversions are identified by requiring the presence of a second oppositely charged track with low separation between track origins and low combined mass [65]. Electron candidates consistent with coming from $\gamma \rightarrow e^+e^-$ conversions are assigned a conversion value of 1.

- Isolation - An electron's isolation is defined as the ratio of the calorimeter cluster's energy (hadronic and EM) to that of all calorimeter energy within a cone of $\Delta R \leq 0.4$ (ΔR is defined in Eq. 4.3) around the cluster center.

- Phoenix - A Phoenix value of 1 is assigned to electron candidates with tracks formed by the PHOENIX algorithm.

- L_{shr} - The consistency of the electron candidate's observed shower pattern to that of a hadronic particle as observed in test beam studies [66].

- track $Z0$ - This is the distance of closest approach between the track and the interaction point.

- Axial and Stereo layers/hits - the number of layers in the COT with hits and the total number of hits in all layers.

- E/P - This is the ratio of the energy measured in the calorimeter to the momentum of the track.

- χ^2 - This quantity indicates the quality of the track fit to the observed detector signals.

- ΔX - Distance between the track extrapolated to the calorimeter and the calorimeter shower in the $r - \phi$ plane.

- ΔZ - Distance between the track extrapolated to the calorimeter and the calorimeter shower in the $r - z$ plane.

  The details of electron selection using the above quantities are presented in Chapter 7.

5.3 Muons

Muons candidates are formed by matching a muon detector signal to either an OI or an IO track. Muon candidates can be characterized by the following quantities :

- p_T - The transverse momentum of the reconstructed track.

- hadEn and emEn - The amount of energy associated with the muon candidate in the hadronic (hadEn) and electromagnetic (emEn) calorimeter systems.

- track $Z0$ - This is the distance of closest approach between the track and the interaction point.

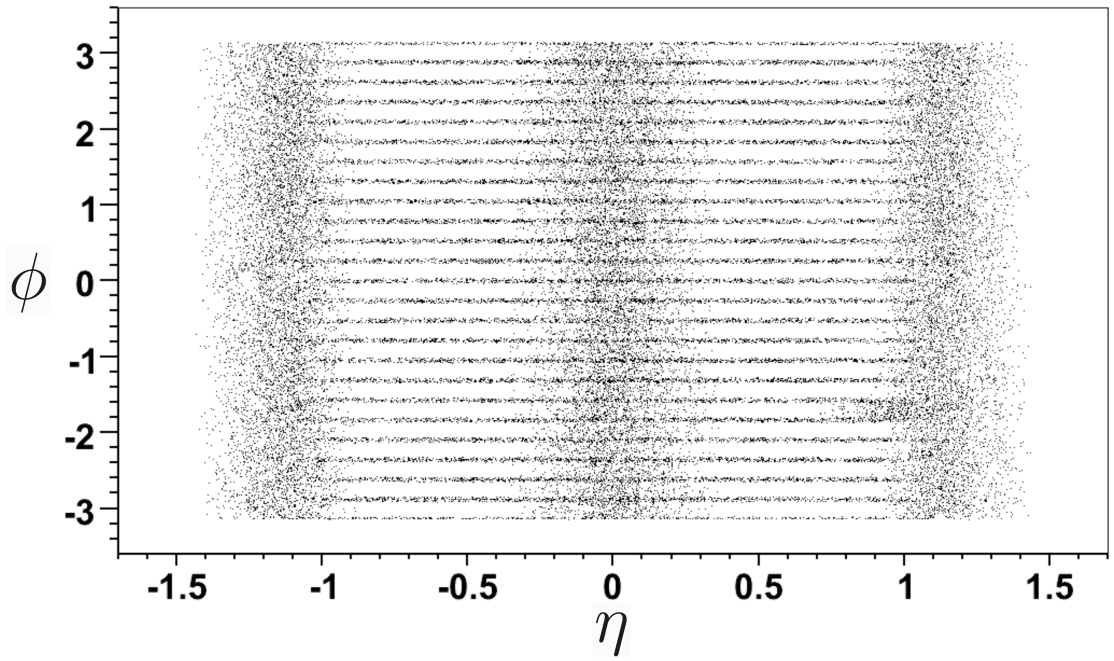


Figure 5.1: ϕ vs. η of tracks extrapolated to non-fiducial regions of the CDF calorimeter in CDF data. The dots outline the location of unistrumented calorimeter space.

- d_0 - This is the track's impact parameter to the beam axis.
- ρ - The radius at which the track exits the COT.
- ΔX_{cmx} - Distance between the extrapolated track and the CMX detector stub in the $r - \phi$ plane.
- ΔX_{cmu} - Distance between the extrapolated track and the CMU detector stub in the $r - \phi$ plane.
- ΔX_{cmp} - Distance between the extrapolated track and the CMP detector stub in the $r - \phi$ plane.
- Isolation Fraction - This is the fraction of the total momentum within a cone of $\Delta R \leq 0.4$ around the muon candidate track due to the presence of additional tracks.

- Axial and Stereo Segments - the number of COT track segments identified by the XFT system.

The details of muon selection using the above quantities are presented in Chapter 7.

5.4 Jets

A quark or gluon created in the CDF II detector is observed as a cone of particles called a "jet". Since quarks and gluons carry color charge, when either of these objects is created QCD interactions pull additional particles out of the vacuum to create colorless particles, a process known as hadronization; meaning that the original colored particle manifests itself as a collection of colorless hadrons. When the original quarks and gluons carry large momentum, the collection of hadrons is boosted into a cone, creating a spray of particles termed a jet. For analysis, jets are treated as the manifestations of quarks and gluons. While several algorithms for the reconstruction and estimation of the energy of a jet are currently employed in high energy physics, the algorithm used in the reconstruction of jet objects in this analysis is known as the JetClu (Jet Cluster) [67] algorithm.

The version of the JetClu algorithm used in this study reconstructs jet objects from calorimeter towers with a cone size of $\Delta R \leq 0.4$. The JetClu algorithm proceeds as follows :

- Each tower with $E_T > 1\,\mathrm{GeV}$ is identified as a seed tower; these are ordered in decreasing E_T.
- Beginning from the highest E_T seed tower, all towers within $\Delta R \leq 0.4$ of the seed tower are combined to form initial calorimeter clusters. Each seed tower is assigned to at most one cluster.
- The centroid and transverse energy of each cluster are computed.
- New clusters are formed from all towers within $\Delta R \leq 0.4$ of the initial cluster centroids (in decreasing E_T). Once again, each seed tower is assigned to at most one cluster.

- The centroids for the new clusters are computed, and the previous step is repeated until the list of towers assigned to each cluster is stable.

- The energy from shared towers (*i.e.* non-seed towers) between each cluster pair is computed; if the energy is more than 75% of the the energy of the lower E_T cluster, the pair of clusters is merged.

- The jets are taken as the final set of clusters, with the jet E_T, η and ϕ computed over the sum of contributing towers.

The initial jet E_T computed from calorimeter tower energies is corrected [68] for the following effects :

- detector effects - The calorimeter energy scale is corrected by setting the measured $Z \rightarrow e^+e^-$ reconstructed mass to LEP averages. The corrections are updated to account for PMT gains over time.

- η-dependent corrections - Due to the segmentation (and varying coverage, materials etc.) of the CDF calorimeter there is an η dependence to the measured jet energy. This dependence is reduced by applying a correction derived by comparing the energy of jets in events with an exclusive 2 jet final state (where the transverse energy of the two jets should be equal). Jet energies measured inside the region $0.2 < |\eta| < 0.6$ (far from uninstrumented regions of the calorimeter) are used to scale the energies of jets at other values of $|\eta|$.

- multiple interactions - There are approximately six $p\bar{p}$ interactions per bunch crossing. The majority of these interactions are "soft" scattering interactions; in that no particles are produced with significant p_T. However, energy from these interactions will increase the measured energy of a jet coming from one particular $p\bar{p}$ interaction. The average calorimeter tower energy from soft interactions is measured in "minimum bias" events which contain two charged particles close to the beampipe (*i.e.* one at $3.2 < \eta < 5.9$

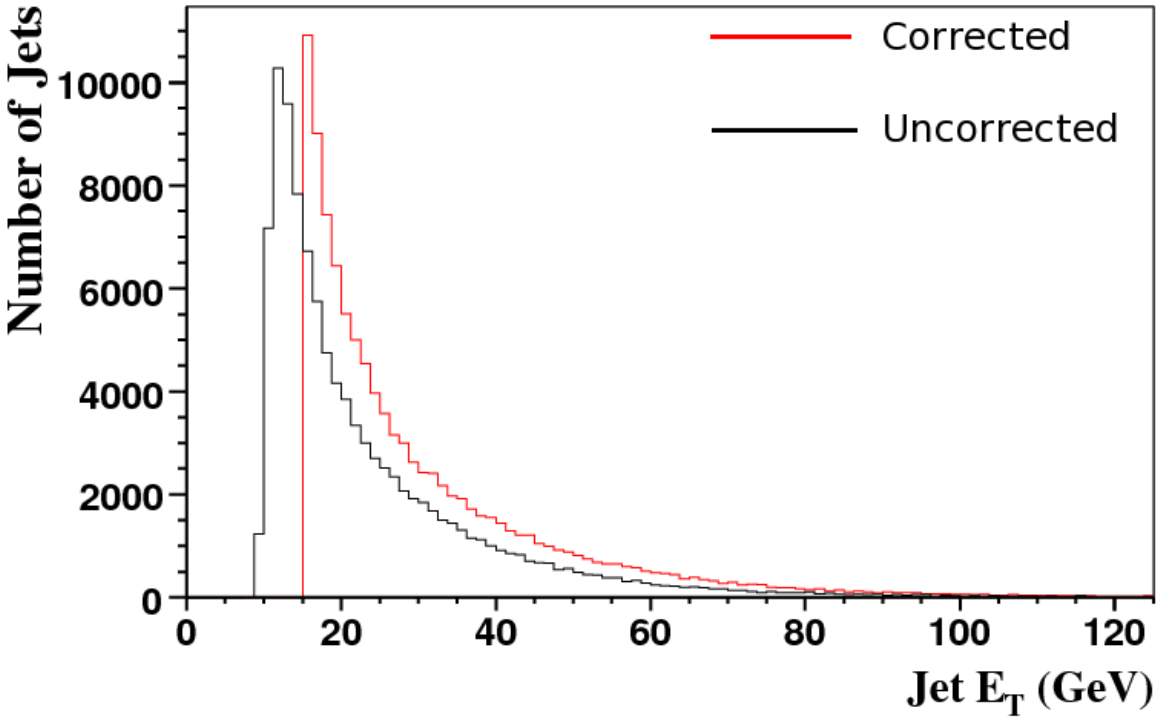


Figure 5.2: Comparison of corrected (red) to uncorrected (black) jet E_T for jets with a corrected jet $E_T > 15\,\text{GeV}$ and $|\eta| < 2.0$ in CDF data.

and one at $-5.9 < \eta < -3.2$). This average tower energy is subtracted from the towers contributing to a jet's energy.

The net effect of these corrections on jet transverse energies is shown in Fig. 5.2.

5.5 Missing Transverse Energy

The missing transverse energy $\vec{\not\!\!E}_T$ is computed as the sum over all calorimeter towers :

$$\vec{\not\!\!E}_T = -\sum_i E_T^i \hat{n}_i \tag{5.2}$$

where the index i runs over all calorimeter towers with $|\eta| < 3.6$, $\hat{n}_i$ and is a unit vector perpendicular to the beam axis and pointing at the ith calorimeter tower. We also define

$\not\!\!E_T = |\vec{\not\!\!E}_T|$.

In events with muons of significant p_T, the above missing E_T equation must be corrected for muon track momentums. As minimum ionizing particles, muons do not deposit significant energy in the calorimeter. With this correction Eq. 5.2 becomes :

$$\vec{\not\!\!E}_T = -\sum_i E_T^i \hat{n}_i - c\sum_j \vec{p_T}^j \tag{5.3}$$

where index j runs over all muons with $\vec{p_T} > 10\,\mathrm{GeV}/c$.

Chapter 6: Data Sample and Model

The following chapter describes the data sample and data model used in this search. We select our sample of candidate ZH events as follows :

- Events with high p_T muons or electrons are selected by the CDF trigger system.
- We identify events containing $Z \rightarrow \ell^+\ell^-$ candidates reconstructed from muon or electron pairs.
- Each event is required to contain two or more jets with significant E_T. Events meeting the above requirements are used to check our model of the data. We call these events the PreTag sample.
- We further require that the events contain a $H \rightarrow b\bar{b}$ candidate, by demanding that at least one jet is consistent with a b quark. We refer to such jets as b-tagged jets. Events meeting the Z, jet, and b-tag requirements form our signal region sample.

The full details of the analysis selection are presented in Chapter 7.

6.1 Data Samples

This search includes CDF Run II data recorded between February 2002 and February 2009 collected with the inclusive high p_T lepton triggers: ELECTRON_CENTRAL_18, CMUP_18 and CMX_18. These triggers are designed to select events with at least one central, energetic track which leaves a signature in either the calorimeter or muon detector systems. Events are also accepted from the Z_NOTRACK trigger, designed to identify events with pairs of electron candidates. The full details of the trigger selection will be presented in Chapter 7.

The initial data sample corresponds to a total integrated luminosity of $4.4\,\mathrm{fb}^{-1}$. Approximately 7% of this sample, collected while a major component of the CDF II

detector (silicon, calorimeter, or muon detector) was inoperable, is rejected leaving a sample of $4.1\,\mathrm{fb}^{-1}$ for further analysis.

A continuous period of detector operation is assigned a sequential number called the run number. The data used in this search spans the range in run numbers from 138425 to 272214. We include only "good runs" for which all major components of the CDF II detector were operational.

Data are also organized by data periods which divide the total sample into segments of time during which the detector was operated under a consistent configuration (trigger settings, average instantaneous luminosity, system calibrations, etc. can be changed over time).

The data periods are defined in Tables 6.1. The total luminosity of the datasets is shown by trigger in Table 6.2.

The analysis selection is applied to all events in the HIGH_PT_ELECTRON data stream passing the ELECTRON_CENTRAL_18 and/or the Z_NOTRACK triggers and meeting good run requirements. Events in the HIGH_PT_MUON data stream are skimmed (by the CDF Top Group [69]) for the presence of one or more muons. The analysis selection is applied to all events passing this initial skim which meet the CMUP_18 and/or CMX_18 trigger and good run requirements.

6.2 Data Model

6.2.1 Simulated Processes

Signal events are modeled with PYTHIA [71] Monte Carlo (MC) samples generated for Higgs masses between 100 and $150\,\mathrm{GeV}/c^2$ in $5\,\mathrm{GeV}/c^2$ steps. These samples restrict the Higgs to decay to $b\bar{b}$, with the Z decaying to $e^+e^-, \mu^+\mu^-$ or $\tau^+\tau^-$. Signal samples with varied amounts of initial and final state radiation (ISR/FSR) are generated assuming a Higgs mass of $120\,\mathrm{GeV}/c^2$. The signal MC datasets used are listed in Table 6.3 along with the standard

Table 6.1: Integrated luminosity and calendar range by data period. Numbers are from [70]. The integrated luminosity is shown for run numbers with operational silicon, muon and calorimeter systems and active ELECTRON_CENTRAL_18 and Z_NOTRACK triggers.

data period	dates collected	integrated luminosity pb^{-1}
0	04 Feb 02 - 22 Aug 04	331.47
1-4	07 Dec 04 - 04 Sep 05	362.94
5-7	05 Sep 05 - 22 Feb 06	258.37
8	9 June 06 - 1 Sept 06	166.29
9	1 Sept 06 - 22 Nov 06	156.76
10	24 Nov 06 - 30 Jan 07	243.19
11	31 Jan 07 - 30 Mar 07	234.99
12	1 Apr 07 - 13 May 07	162.01
13	13 May 07 - 4 Aug 07	280.86
14	28 Oct 07 - 3 Dec 07	32.01
15	5 Dec 07 - 27 Jan 08	161.87
16	27 Jan 08 - 27 Feb 08	101.81
17	28 Feb 08 - 16 Apr 08	183.56
18	18 Apr 08 - 1 Jul 08	304.88
19	1 Jul 08 - 24 Aug 08	206.98
20	24 Aug 08 - 04 Oct 08	226.92
21	12 Oct 08 - 01 Jan 09	435.59
22	2 Jan 09 - 10 Feb 09	265.67

Table 6.2: Luminosity by trigger. The silicon systems are required to be operational. The CMUP_18 luminosity is reduced (in comparison to the ELECTRON_CENTRAL_18/Z_NOTRACK luminosity) due to a rejection of runs with an improper CMUP_18 trigger configuration in period 18. The lower CMX_18 trigger luminosity is the result of the CMX trigger not being fully implemented for run numbers before 15014.

Luminosity	Trigger
4.11617 fb^{-1}	ELECTRON_CENTRAL_18
4.11617 fb^{-1}	Z_NOTRACK
4.1071 fb^{-1}	CMUP_18
4.06287 fb^{-1}	CMX_18

model production cross sections and $H \to b\bar{b}$ branching ratios.

Background processes possess a detector signature similar to the signal; two leptons and two (or more) jets in the final state. The dominant background process is Z+jets, with Z+ light flavor jets (u,d,s) forming the major background component before b-tag requirements are imposed. In events with one or more b-tags, $Z + b\bar{b}$ and $Z + c\bar{c}$ are the main backgrounds with Z+mistagged light jets contributing more as tag requirements are loosened. Z+ jets processes are modeled with ALPGEN [72] for the hard scattering and with PYTHIA for the hadronization and showering. After b-tagging the contribution from Z+mistagged light jets is modeled from the data (Section 6.2.2). Diboson processes (ZZ, WZ, WW) and $t\bar{t}$ are modeled with PYTHIA. The $t\bar{t}$ simulation uses a top quark mass of $175\,\mathrm{GeV}/c^2$, a little above the present world average of $173.1\,\mathrm{GeV}/c^2$ [6]. The background MC samples for this analysis are listed in Table 6.4.

MC events enter the model with a weight given by the following formula :

$$\mathcal{E}vent\ Weight = \int \mathcal{L}\ dt \times \epsilon_{trigger} \times \mathcal{SF}_{recon.} \times \mathcal{SF}_{b-tag} \times \frac{\sigma_{process} \times BR}{N_{evt}} \qquad (6.1)$$

where $\int \mathcal{L}\ dt$ is the integrated luminosity, $\sigma_{process} \times BR$ is the process cross-section multiplied by the appropriate branching ratios, N_{evt} is the total number of events in the MC sample, $\epsilon_{trigger}$ is the ratio of the trigger efficiency in MC to that in data, $\mathcal{SF}_{recon.}$ is the ratio of the lepton reconstruction efficiency in MC to that in data, and $\mathcal{SF}_{b-tag}$ is the ratio of the b-tag efficiency in MC to that in data. Process cross sections are taken from Refs. [73, 74, 75]. The b-tag scale factors, muon reconstruction scale factors, and trigger efficiencies are computed from the CDF Joint Physics Scale Factor Class [76]. The b-tag scale factors are set equal to 1.0 before b-tag requirements are imposed. Analysis specific electron reconstruction scale factors are computed for each electron category.

Table 6.3: Signal MC samples. Numbers are taken from [77] where the cross-sections (σ_{ZH}) are computed using the method outlined in Ref. [78] and the branching ratios are computed with the HDECAY code discussed in Ref. [79]. A 5% uncertainty is assumed on all cross-sections.

$M_H(GeV/c^2)$	σ_{ZH}(pb)	$BR(H \to b\bar{b})$
100	0.16851	0.8121
105	0.14457	0.7957
110	0.12458	0.7702
115	0.10778	0.7322
120	0.09353	0.6789
125	0.08139	0.6097
130	0.07109	0.5271
135	0.06222	0.4362
140	0.05468	0.3436
145	0.04811	0.2556
150	0.04240	0.1757
120 ISR+	0.09353	0.6789
120 ISR-	0.09353	0.6789
120 FSR+	0.09353	0.6789
120 FSR-	0.09353	0.6789

Table 6.4: Backgrounds modeled with MC. The $\sigma \times BR$ for $Z+$ lf (light flavor) jets, $c\bar{c}$ and $b\bar{b}$ are from [73], and vary according to $\ell^+\ell^-$ mass and the jet multiplicity. These samples are scaled by an additional factor of 1.4 to account for the difference between leading and next to leading order predictions. WW, WZ and ZZ cross-sections are from [74]. The $t\bar{t}$ cross-section is from [75].

Process	MC Generator	$\sigma \times BR$
$(Z \to \ell\ell)+$ lf jets	ALPGEN+PYTHIA	3.3 fb to 1514 pb
$(Z \to \ell\ell) + c\bar{c}$ + jets	ALPGEN+PYTHIA	107 to 1280 fb
$(Z \to \ell\ell) + b\bar{b}$ + jets	ALPGEN+PYTHIA	38.5 to 625 fb
WW	PYTHIA	12.4 pb
WZ	PYTHIA	3.7 pb
ZZ	PYTHIA	3.8 pb
$t\bar{t}$	PYTHIA	6.7 pb

6.2.2 Data-Derived Backgrounds

Events where one or more jets are incorrectly classified as a muon (resulting in a "fake" $Z \to \mu\mu$ event) are estimated from data events meeting all selection requirements and containing two like-sign muons. The CDF muon detectors utilized in this search cover $|\eta| < 1.0$, a region where adequate tracking coverage ensures accurate measurement of muon charge. For electrons, we utilizes calorimeter detectors extending to $|\eta| < 3.6$. In forward regions, charge measurement is of limited accuracy due to insufficient tracking coverage. This renders a like-sign electron sample an insufficient model for "fake" $Z \to ee$ events. Instead, the probability that a jet fakes an electron is measured in jet triggered data for each of our electron categories. These fake rates and their application will be discusses in Chapter 7.

In order to estimate the fraction of $Z+$light flavor jets events entering our final b-tagged sample, each data event possessing a Z candidate and 2 or more jets is assigned a mistag weight and associated uncertainty. The mistag weights represent the probability for a

particular jet to be incorrectly identified as a b-jet by a particular b-tagging algorithm. These weights are computed for individual jets in various data and MC samples by the CDF High p_T B-Tag Group [80]. The mistag probability for a given jet is parameterized in seven quantities listed in Table 6.5.

Event kinematics are computed using the Z candidate and the two highest E_t jets, while the event's mistag probability is computed by properly accounting for each possible combination of taggable jet(s) in the event. Taggable jets have at least two tracks associated with a secondary vertex and an $E_t \geq 10\,\text{GeV}$. The mistag background is thus modeled using re-weighted $Z+ \geq 2$ jets data.

Table 6.5: Variables entering the mistag probability calculation

Jet E_T
number of tracks in the jet
jet $\lvert\eta\rvert$
number of vertices in the event
z coordinate of the primary event vertex
run number
sum of the E_T of all jets in the event with $E_T \geq 10\,\text{GeV}$

Chapter 7: Event Selection

Signal $ZH \rightarrow \ell\ell b\bar{b}$ events are distinguished by the presence of two oppositely charged energetic leptons and two $b-$jets with high transverse momentum. The transverse energies of the leading and next-to-leading leptons and jets for signal MC events are shown in Fig. 7.1. The "leading" lepton (jet) is defined as the lepton (jet) with the largest E_T, while the "next-to-leading" lepton (jet) is the lepton (jet) with the second largest E_T in the event. In order to compensate for the low ZH production cross-section ($\sigma_{ZH} \sim 0.1\,\text{pb}$) and the low branching fraction of $Z \rightarrow \ell^+\ell^-$ we devise electron identification cuts with the goal of maximizing Z reconstruction efficiency. Lepton selection is designed to meet the minimum requirements of the CDF high p_T lepton triggers (achieving high signal acceptance), while jet selection is designed to ensure $Z+ \geq 2$ jet kinematics are well understood. $Z+$low ($E_T < 15\,\text{GeV}$) jets are not well modeled by our simulation and are removed by the jet E_T requirements. A combination of b-tagging algorithms is used to achieve a high signal–to–background purity. Details of lepton identification, Z reconstruction, jet selection and b-tagging are discussed in the following sections.

7.1 Trigger Selection

Candidate events pass the requirements of one of four CDF high p_T lepton triggers: ELECTRON_CENTRAL_18, Z_NOTRACK, MUON_CMUP_18, or MUON_CMX_18. To fire ELECTRON_CENTRAL_18, an event must contain at least one electromagnetic deposit of $E_t \geq 18\,\text{GeV}$ in the central region of the detector, with an associated track of $p_T \geq 9\,\text{GeV}/c$. The ratio of energy deposited in the hadronic calorimeter to that in the electromagnetic calorimeters must be less than 12.5%, and the cluster's shower pattern must be consistent with an electron shower. The Z_NOTRACK trigger is satisfied by events containing two or more calorimeter deposits of $E_t \geq 18\,\text{GeV}$ restricted to the region

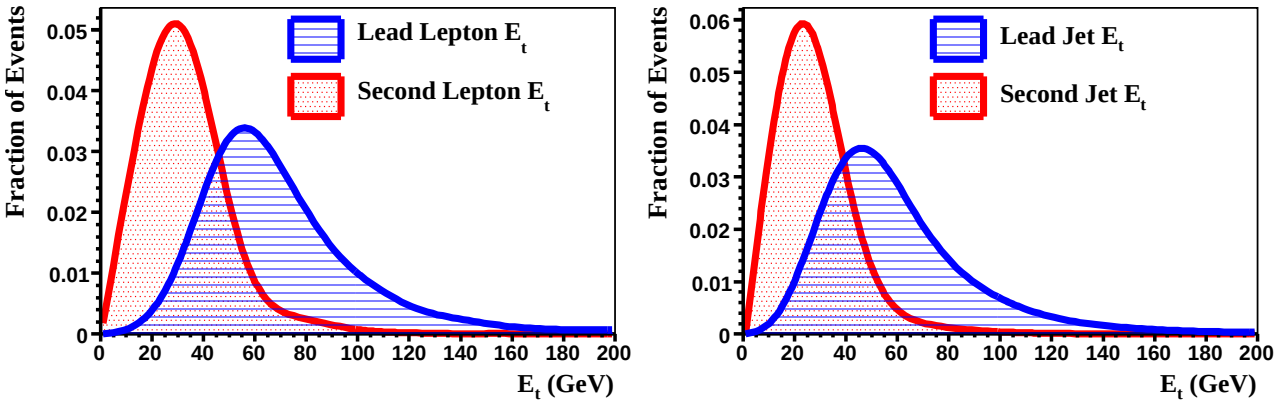


Figure 7.1: Distribution of leading and next-to-leading lepton (left) and jet (right) transverse energies in simulated $ZH \to \ell\ell b\bar{b}$ events. The simulation has a Higgs mass of $120\,\mathrm{GeV}/c^2$.

$|\eta| \leq 3.6$. The MUON_CMUP_18 and MUON_CMX_18 triggers fire on events with "stubs" in the appropriate muon system matched to a track with $p_T \geq 18\,\mathrm{GeV}/c$. The trigger requirements are summarized in Tables 7.1 through 7.4.

7.1.1 Trigger Efficiencies

The probability for a given event to fire a particular trigger can be parameterized in terms of various quantities including run number, calorimeter deposit η, and track p_T. These trigger probabilities are computed by the CDF Joint Physics Group and are available in the Joint Physics Scale Factor Class [76].

Each MC event is assigned a trigger efficiency ($\epsilon_{trigger}$ in Eq. 6.1) based on the event's lepton content. Events containing multiple leptons or capable of firing multiple triggers are assigned weights computed using all combinations of appropriate triggers and leptons. Since the simulated data periods in the MC samples available do not directly match the data periods in data (max run # in MC is 237795 while the analysed data go to run 272214) a weighted average of run dependent efficiencies is applied to each MC event.

Table 7.1: ELECTRON_CENTRAL_18 trigger path

Level 1
One central EM energy cluster with $E_T \geq 8\,\mathrm{GeV}$
Track hits identified by XFT in at least 4 COT layers
XFT track $p_T \geq 8.34\,\mathrm{GeV/c}$
Level 2
Central energy cluster $E_T > 18$ GeV
cluster $\lvert\eta\rvert \leq 1.317$
Track hits identified by XFT in at least 4 COT layers
XFT track $p_T \geq 8.34\,\mathrm{GeV/c}$
Level 3
EM object with $E_T \geq 18$ GeV
$L_{shr} \geq 0.4$
Central track with $p_T \geq 9\,\mathrm{GeV}/c$
EM object with $HadEm \leq 0.125$

Table 7.2: Z_NOTRACK trigger path

Level 1
Two EM energy clusters
a plug or central EM cluster with $E_T \geq 8\,\mathrm{GeV}$
central EM cluster with $HadEm \leq 0.125$
central EM cluster with $HadEm \leq 0.0625$
Level 2
Two seed towers at $\lvert\eta\rvert < 3.6$ and with EM energy $\geq 8\,\mathrm{GeV}$ each
Each seed tower has a neighboring tower with EM energy $\geq 7.5\,\mathrm{GeV}$
Total E_T in two seed towers > 16 GeV
Level 3
Two EM objects
each $E_T > 18$ GeV

Table 7.3: MUON_CMX_18 trigger path

Level 1
CMX stub with associated track $p_T \geq 6\,\mathrm{GeV}/c$
Track hits identified by XFT in at least 4 COT layers
XFT track $p_T \geq 8.34\,\mathrm{GeV}/c$
Level 2
Track hits identified by XFT in at least 4 COT layers
XFT track $p_T \geq 14.77\,\mathrm{GeV}/c$
Level 3
$\lvert\Delta X_{cmx}\rvert \leq 10\,\mathrm{cm}$
Track with $p_T \geq 18\,\mathrm{GeV}/c$

Table 7.4: MUON_CMUP_18 trigger path

Level 1
CMU or CMP stub with associated track $p_T \geq 6\,\mathrm{GeV}/c$
XFT track matched to stub with $p_T \geq 4.09\,\mathrm{GeV}/c$
Level 2
Track hits identified by XFT in at least 4 COT layers
XFT track $p_T \geq 14.77\,\mathrm{GeV}/c$
Level 3
$\lvert\Delta X_{cmp}\rvert \leq 20$ cm
$\lvert\Delta X_{cmu}\rvert \leq 10$ cm
Track with $p_T \geq 18\,\mathrm{GeV}/c$

7.2 Muon Identification

The search defines three muon categories CMUP, CMX and CMIO. The CMUP and CMX categories are designed to meet the requirements of high p_T central muon triggers and match the CDF Top Group's definitions closely. Each $Z \rightarrow \mu\mu$ candidate event must contain at least one CMUP or CMX muon. The CMUP and CMX selection requires a high quality track with $p_T \geq 20\,\mathrm{GeV}/c$. The full CMUP/CMX selection is summarized in Tables 7.5 and 7.6. The third muon category, CMIO, is defined by loose selection requirements. CMIO muon candidates must have track $p_T \geq 10\,\mathrm{GeV}/c$ and no "stub" requirements are imposed. The CMIO selection requirements are summarized in Table 7.7.

7.3 Electron Identification

To increase the efficiency of $Z \rightarrow ee$ reconstruction in events passing ELECTRON_CENTRAL_18 or Z_NOTRACK we define nine electron classes listed in

Table 7.5: CMX muon selection.

CMX Muon Selection
$p_T \geq 20 GeV$ /c
Had Energy $\leq$ 6 GeV
Em Energy $\leq$ 2 GeV
CMX $\rho >$ 140 cm
$\geq$ 3 axial and $\geq$3 stereo segments
Isolation fraction $\leq$ 0.1
$\lvert\Delta X_{cmx}\rvert <$ 6.0 cm
Impact parameter $d_0 <$ 0.02 w/Silicon hits (0.2 w/out)
Track origin (Z_0) is within 60 cm of z=0

Table 7.6: CMUP muon selection.

CMUP Muon Selection
$p_T \geq 20 GeV$ /c
Had Energy $\leq$ 6 GeV
Em Energy $\leq$ 2 GeV
>= 3 axial and >=3 stereo segments
Isolation fraction $<$ 0.1
$\lvert\Delta X_{cmu}\rvert <$ 3.0 cm
$\lvert\Delta X_{cmp}\rvert <$ 5.0 cm
Impact parameter $d_0 <$ 0.02 w/Silicon hits (0.2 w/out)
Z_0 is within 60 cm of z=0

Table 7.7: The loose muon selection criteria for CMIO muons.

CMIO Muon Selection
$p_T \geq 10 GeV$ /c
Had Energy $\leq$ 6 GeV
Em Energy $\leq$ 2 GeV
Isolation fraction $<$ 0.1
Impact parameter $d_0 <$ 0.02 w/Silicon hits (0.2 w/out)
$\geq$ 2 axial and $\geq$2 stereo segments
$\geq$ 1 COT hits
No stub requirements
Z_0 is within 60 cm of z=0

Table 7.8: The electron types we consider in this analysis.

Tc	Tight-central
Lc1	Loose-central 1
Lc2	Loose-central 2
Tp	Tight Plug Phoenix
Lp	Loose Plug
LpPhx	Loose Plug Phoenix
LpNphx25	Loose Plug Non-phoenix, 25 GeV E_T cut
LpNphx33	Loose Plug Non-phoenix, 33 GeV E_T cut
CrkTrk	Crack Track, no Em requirement

Table 7.8. The categories are based on the 'tight' and 'loose' CEM/PHX electrons defined in Ref. [81], with relaxed requirements for non-trigger and plug electrons. The following subsections discuss the electron selection, calculation of reconstruction scale factors, and fake rate estimates in detail.

7.3.1 Central Electron Identification

We define four categories of central electrons ($|\eta| \leq 1.1$) :

- A tight central electron (Tc).
- A loose central electron (Lc1) with intermediate selection requirements.
- A very loose central electron (Lc2) with minimal selection requirements.
- A track based electron category (CrkTrk) reconstructed from high quality tracks that do not point to instrumented regions of the calorimeter. This category is based on the electron CrkTrk used in Ref. [82].

The selection criteria for each central electron type are presented in Tables 7.9 to 7.12.

Table 7.9: Tight central (Tc) electron selection.

Tc (Tight central) Selection
Region = 0
Fiducial = 1 or 2
$E_T \geq 18$
Trk $P_T \geq 9$ GeV
Not a conversion
$\lvert TrackZ_0 \rvert \leq 60$ cm
≥ 2 Ax layers ≥ 10 hits
≥ 2 St layers ≥ 10 hits
HadEm $\leq 0.055 + 0.00045 \times$ En
Isolation – $0.35 \times (N_{vtx} - 1) \leq 3 + 0.02 \times E_T$
$L_{shr} \leq 0.2$
$\frac{E}{P} \leq 2.5 + 0.015 \times E_T$ or TrkPt ≥ 50 GeV
$\lvert \Delta Z \rvert \leq 3$ cm
$-3 \leq \Delta X \times Q \leq 1.5$
$\chi^2_{strip} \leq 25$

Table 7.10: Loose central 1 (Lc1) electron selection.

Lc1 (Loose central 1) Selection
Region = 0
Not a Tc
$E_T \geq 18$
Trk $P_T \geq 9$ GeV
Not a conversion
$\lvert TrackZ_0 \rvert \leq 60$ cm
HadEm $\leq 0.055 + 0.00045 \times$ En
Isolation – $0.35 \times (N_{vtx} - 1) \leq 4 + 0.02 \times E_T$
$L_{shr} \leq 0.2$

Table 7.11: Loose central 2 (Lc2) electron selection.

Lc2 (Loose central 2) Selection
Region = 0
Not a Tc or Lc1
$E_T \geq 10$
Trk $P_T \geq 5$ GeV
$\lvert TrackZ_0 \rvert \leq 60$ cm
HadEm $\leq 0.055 + 0.00045 \times$ En
Isolation – $0.35 \times (N_{vtx} - 1) \leq 4 + 0.02 \times E_T$

Table 7.12: Crack Track (CrkTrk) electron selection.

CrkTrk (Crack Track) Selection
$\lvert Z_0 \rvert \leq 60$ cm
$\lvert D_0 \rvert \leq 0.2$ cm
≥ 3 Ax layers & ≥ 15 hits
≥ 3 St layers & ≥ 15 hits
Not a CMUP or CMX muon
Not Track Fiducial
Not a conversion
$p_T \geq 20$
$\frac{Isolation}{p_T} < 0.1$ or $\frac{EmIsolation}{p_T} < 0.1$

Table 7.13: Tight plug (Tp) electron selection.

Tp (Tight plug) Selection
Region = 1
$E_T \geq 25$
Phoenix = 1
$1.2 \leq \lvert\eta\rvert \leq 2.8$
if En $\leq$ 100 GeV, HadEm ≤ 0.05
if En $>$ 100 GeV, HadEm $\leq 0.05 + 0.026 \times \log(\text{En}/100)$
Isolation $- 0.35 \times (\text{N}_{vtx} - 1) \leq 1.6 + 0.02 \times E_T$
$\chi^2_{3x3} \leq 25$

7.3.2 Plug Electron Identification

In the "plug" regions of the detector ($|\eta| > 1.1$) we define five electron classes, each with an $E_T \geq 18$ GeV requirement, to match the E_T requirement of the Z_NOTRACK trigger. The five categories of plug electron are :

- A tight plug electron (Tp) is required to have $E_t \geq 25\,\text{GeV}$ and consist of a phoenix track pointing to a calorimeter deposit phoenix track match [83].
- A loose plug electron with a phoenix track match (LpPhx) and $E_t \geq 18\,\text{GeV}$ which is not a Tp.
- A loose plug electron without a phoenix track match (LpNphx33) and $E_t \geq 33\,\text{GeV}$.
- A loose plug electron without a phoenix track match (LpNphx25) and $33\,\text{GeV} > E_t \geq 25\,\text{GeV}$.
- A loose plug electron without a phoenix track match (Lp) and $25\,\text{GeV} > E_t \geq 18\,\text{GeV}$.

The selection criteria for each plug electron type are listed in Tables 7.13 through 7.17.

Table 7.14: Loose plug phoenix(LpPhx) electron selection.

LpPhx (Loose plug Phoenix) Selection
Region = 1
Not a Tp
Phoenix = 1
$E_T \geq 18$
if En $\leq$ 100 GeV, HadEm $\leq$ 0.05
if En $>$ 100 GeV, HadEm $\leq$ 0.05 + 0.026 $\times$ log(En/100)
Isolation – 0.35 $\times$ (N_{vtx} – 1) $\leq$ 2.5 + 0.02 $\times E_T$
$\frac{Isolataion}{E_T} < 0.1$

Table 7.15: Loose plug non-phoenix(LpNphx25) electron selection.

LpNphx25 (Loose plug Non-phoenix) Selection
Region = 1
Not a Tp or LpPhx
Phoenix = 0
$E_T \geq 25$
if En $\leq$ 100 GeV, HadEm $\leq$ 0.05
if En $>$ 100 GeV, HadEm $\leq$ 0.05 + 0.026 $\times$ log(En/100)
Isolation – 0.35 $\times$ (N_{vtx} – 1) $\leq$ 2.5 + 0.02 $\times E_T$
$\frac{Isolataion}{E_T} < 0.1$

Table 7.16: Loose plug non-phoenix(LpNphx33) electron selection.

LpNphx33 (Loose plug Non-phoenix) Selection
Region = 1
Not a Tp or LpPhx or LpNphx25
Phoenix = 0
$E_T \geq 33$
if En $\leq$ 100 GeV, HadEm $\leq$ 0.05
if En $>$ 100 GeV, HadEm $\leq$ 0.05 + 0.026 $\times$ log(En/100)
Isolation – 0.35 $\times$ (N_{vtx} – 1) $\leq$ 2.5 + 0.02 $\times E_T$
$\frac{Isolataion}{E_T} < 0.1$

Table 7.17: Loose plug (Lp) electron selection.

Lp (Loose plug) Selection
Region = 1
Not a Tp
$E_T \geq 18$
if En $\leq$ 100 GeV, HadEm $\leq$ 0.05
if En $>$ 100 GeV, HadEm $\leq$ 0.05 + 0.026 $\times$ log(En/100)
Isolation – 0.35 $\times$ (N_{vtx} – 1) $\leq$ 2.5 + 0.02 $\times E_T$

7.3.3 Electron Reconstruction Scale Factors

To ensure that the fraction of the total electron sample derived from each electron category is identical in the data and simulation, we apply a correction in the form of a weight scale factor to each MC electron. For each electron type we define scale factors using the method outlined in Ref. [84]. We define loose probe electrons (Tables 7.18,7.19) in the central and plug regions. Probe Z candidates are formed from Tc+central probe and Tp+plug probe electron candidates. The scale factor for a given central (plug) electron category is calculated using the ratio of the number of Z candidates formed from an electron of that category and a Tc (Tp) electron to the number of probe Z candidates. After subtracting off background contamination from sidebands, the scale factor is taken as the ratio obtained in data divided by the ratio from MC. Scale factors are computed in three Z mass regions, with the difference in the values used to set a systematic uncertainty on the scale factor. The results are listed in Table 7.20. Electrons in MC are weighted (used to compute $\mathcal{SF}_{recon.}$ in Eq. 6.1) by the scale factor computed in the $76 - 106\,\mathrm{GeV}/c^2$ mass range, with a 1% systematic uncertainty assigned to cover the average deviation in scale factors measured in different mass windows. For the CrkTrk category, we use the scale factor found in Ref. [82], 0.951 averaged over 22 run periods.

Table 7.18: Probe (CEM) electron selection.

Probe (CEM) Selection
Region = 0
$E_T \geq 10$
$Trkp_T \geq 10$
HadEm $\leq 0.055 + 0.00045 \times E_n$
Isolation – $0.35 \times (\mathrm{N}_{vtx} - 1) \leq 4.5 + 0.02 \times E_T$

Table 7.19: Probe (PEM) electron selection.

Probe (PEM) Selection
Region = 1
$E_T \geq 18$
if En $\leq$ 100 GeV, HadEm ≤ 0.05
if En $>$ 100 GeV, HadEm $\leq 0.05 + 0.026 \times \log(\mathrm{En}/100)$
Isolation – $0.35 \times (\mathrm{N}_{vtx} - 1) \leq 3.0 + 0.02 \times E_T$

Table 7.20: Electron scale factors in different mass regions.

e Type	66-116 GeV/c^2	76-106 GeV/c^2	86-96 GeV/c^2
Tc	0.987063	0.985325	0.98961
Lc1	1.00452	1.00415	1.00296
Lc2	1.00098	1.00088	1.00033
Tp	0.961057	0.937155	0.972057
Lp	0.998407	0.9969	0.994712
LpPhx	1.01429	1.01449	1.01405
LpNphx25	1.00468	1.01075	1.00326
LpNphx33	1.00617	1.03572	1.01415

7.3.4 Fake Rates

We calculate the probability to misidentify a jet as an electron for each of our electron categories (fake rate) in the jet triggered data samples : jet20, jet50, jet70, and jet100. These jet samples are collected with triggers that require energetic jets of E_T greater than 20, 50, 70 and 100 GeV respectively. In order to suppress the presence of real electrons from W processes in the jet triggered samples, we impose a cut on the transverse missing energy ($\not\!\!E_T < 15\,\mathrm{GeV}$). Similarly we require jet triggered events to have no more than one electron candidate to remove real electrons from Z decays. For the "denominator" in our fake rate calculation we use all cone 0.4 jets with sufficient E_T to meet the electron category's requirement. Denominator jets must also be in the correct region of the detector (central/plug) for a given electron class. The lead (trigger) jet is not included in the denominator to avoid trigger bias. For "numerator" objects we use EM objects which pass our electron selection and are matched to a denominator jet within $\Delta R < 0.4$. The fake rate is calculated as the ratio of numerator objects to denominator objects in each bin of jet E_T. We use variable-sized bins to maintain statistics, and we find that a 50% systematic uncertainty is required to span the fake rates coming from each jet sample. The electron fake rate functions for each electron type are shown in Figs. 7.2, 7.3, and 7.4. For the CrkTrk category like-sign data is used to estimate the fake contribution in the Z to Tc + CrkTrk subsample.

When a jet is misidentified as an electron, it can be erroneously paired with a real electron or a similarly misidentified jet. The result is a fake $Z \rightarrow e^+e^-$ event. To derive a model of the fake $Z \rightarrow e^+e^-$ background, we identify all jets and electrons in our electron triggered event sample. Each jet is assigned a fake probability using the fake rate functions which are parameterized in jet E_T. We identify all combinations of electron+jet and jet+jet which produce a fake Z. An event enters the fake $Z \rightarrow e^+e^-$ model once for every combination identified, with a weight reflecting the probability of misidentifying the jet(s) and forming a fake Z. Each time an event enters the model, its kinematics are recomputed for the

appropriate arrangement of electrons and jets.

When used in the reconstruction of a fake Z, the transverse energy of a jet is adjusted to match the value expected for an electron. The degree of adjustment is derived from the ratio of the E_T of the EM object to that of the matched jet as observed in jet triggered data. We fit the distribution of the ratio, and use the resulting Gaussian to smear the jet energies in our fake Z model. Separate fits are performed for central and plug objects. The ratio of EM to jet E_T is shown in Fig. 7.5 by detector region.

The expected fake fraction for each Z category is shown in Table 7.21. The relative fake fractions indicate the categories with the highest non-Z background rates. We demonstrate that there is no real Z contamination in our fake estimate by showing the dilepton mass distribution for our expected fake background shape in Figs. 7.6, and 7.7.

Table 7.21: For each dilepton category, we list the percentage of Z's in data, as well as the expected fake percentage.

Combination	% of Total Z's Candidates in data	% fakes
TcTc	12.18	0.015
TcLc	5.73	4.06
TcTp	16.22	0.034
TcLp	15.87	3.5
Lc1Lc1	0.35	4.0
TpLc	4.24	8.1
Lc1Lp	2.34	14.75
TpTp	5.1	0.03
TpLp	5.2	0.09
LpLp	1.08	5.95
TcCrkTrk	5.73	4.8
MuonMuon	25.9	0.27
All	100	2

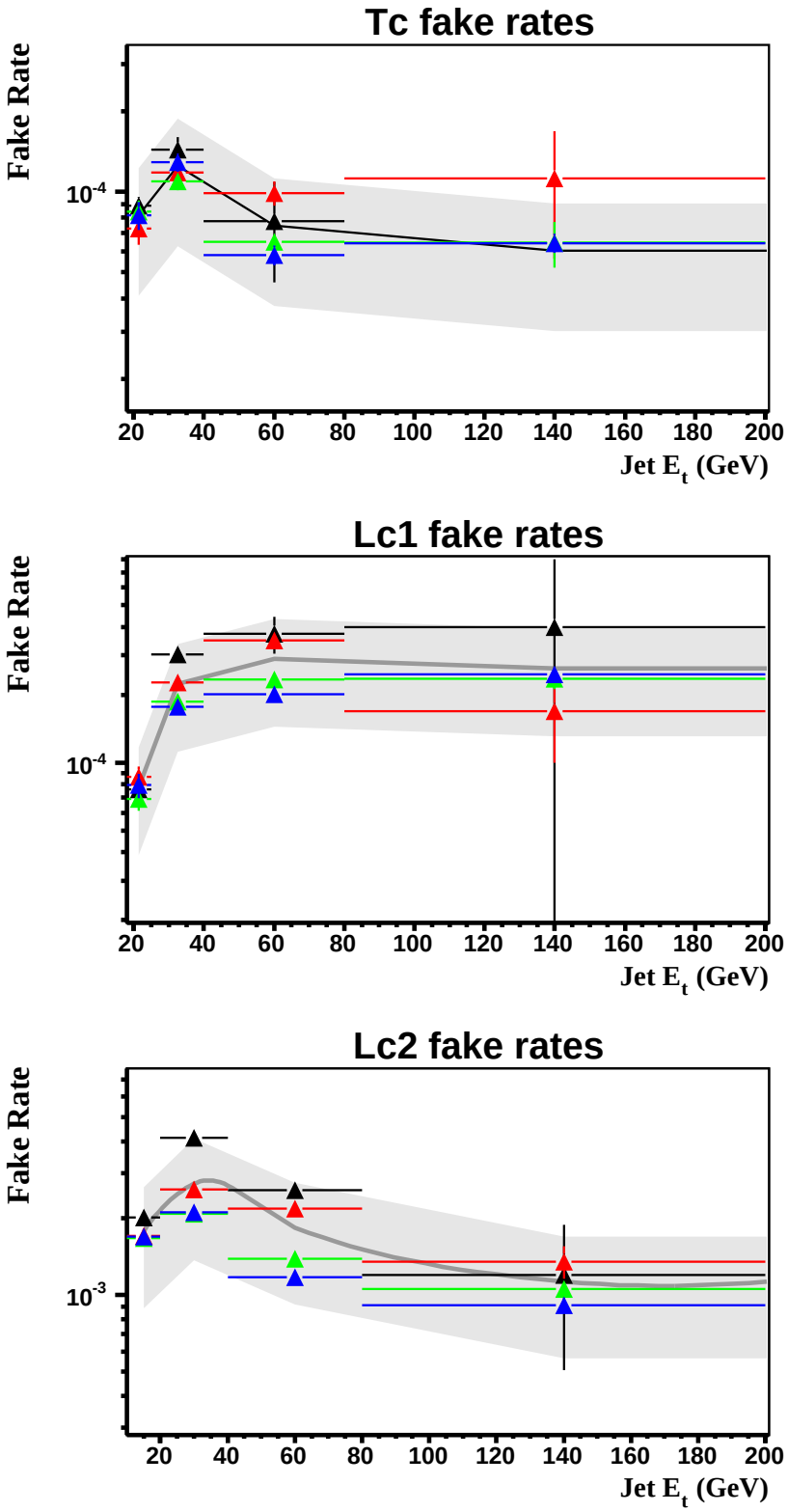


Figure 7.2: Fake rates for Tc, Lc1 and Lc2 type electrons. The mean fake rate from jet20 (black) , jet50 (red), jet70 (green) and jet100 (blue) data is applied to the high p_T electron data. The 50 % error band is shown in grey.

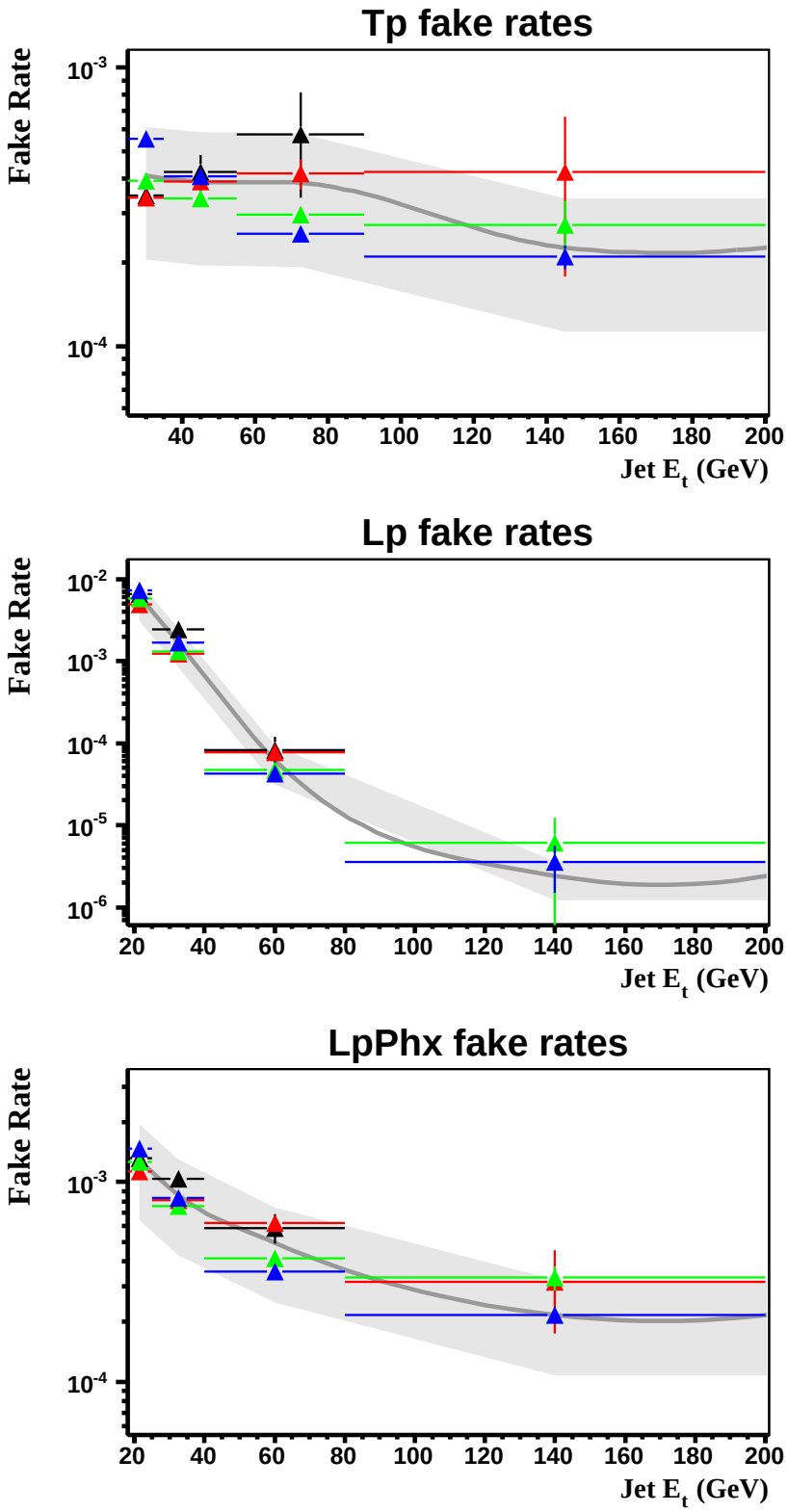


Figure 7.3: Fake rates for Tp, Lp and LpPhx type electrons.

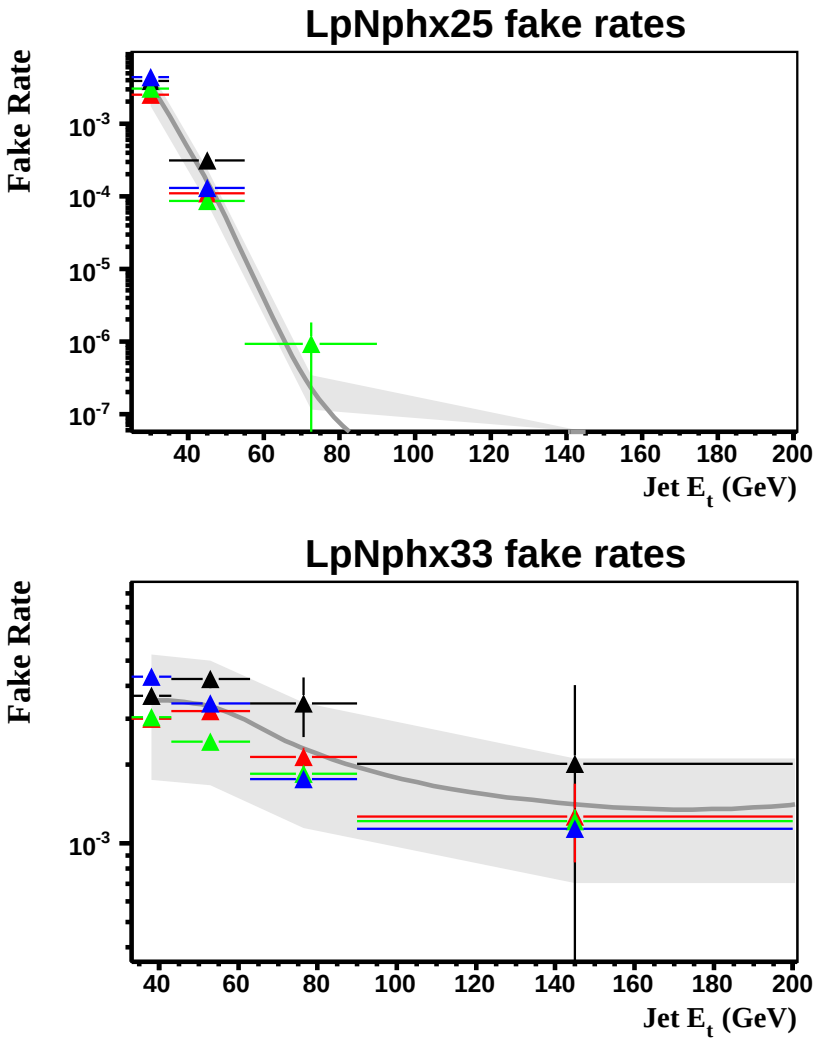


Figure 7.4: Fake rates for LpNphx25 and LpNphx33 type electrons.

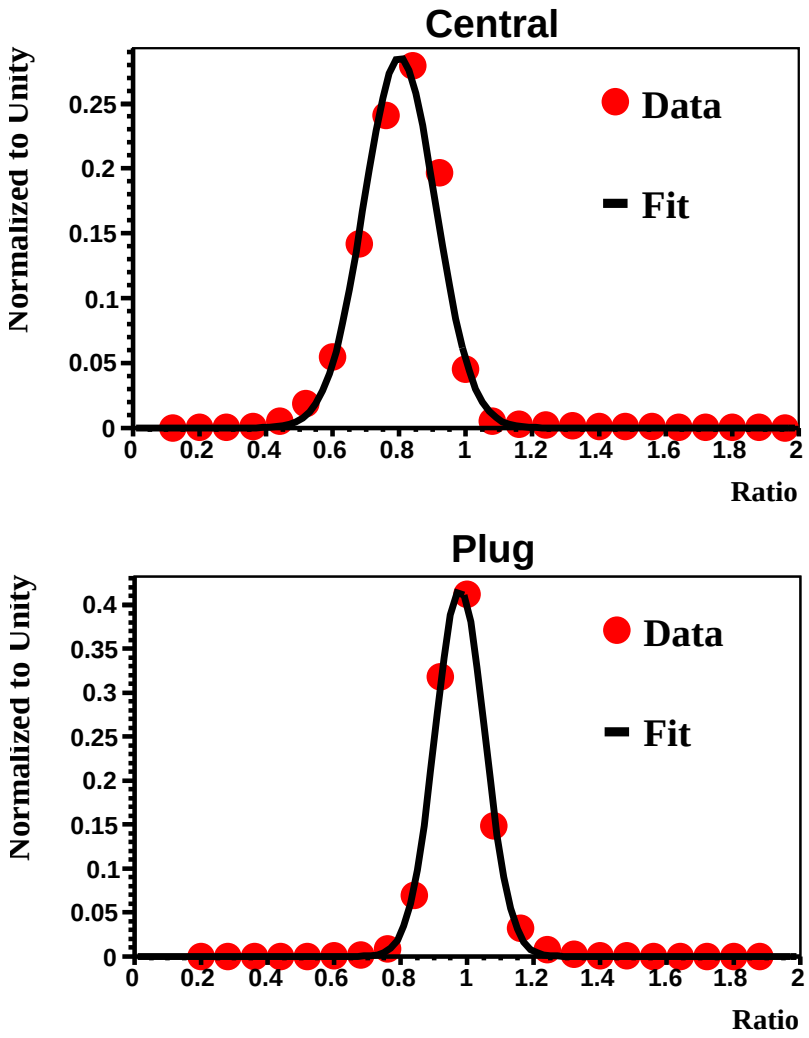


Figure 7.5: Ratio of Em object E_t to jet E_t in jet triggered data.

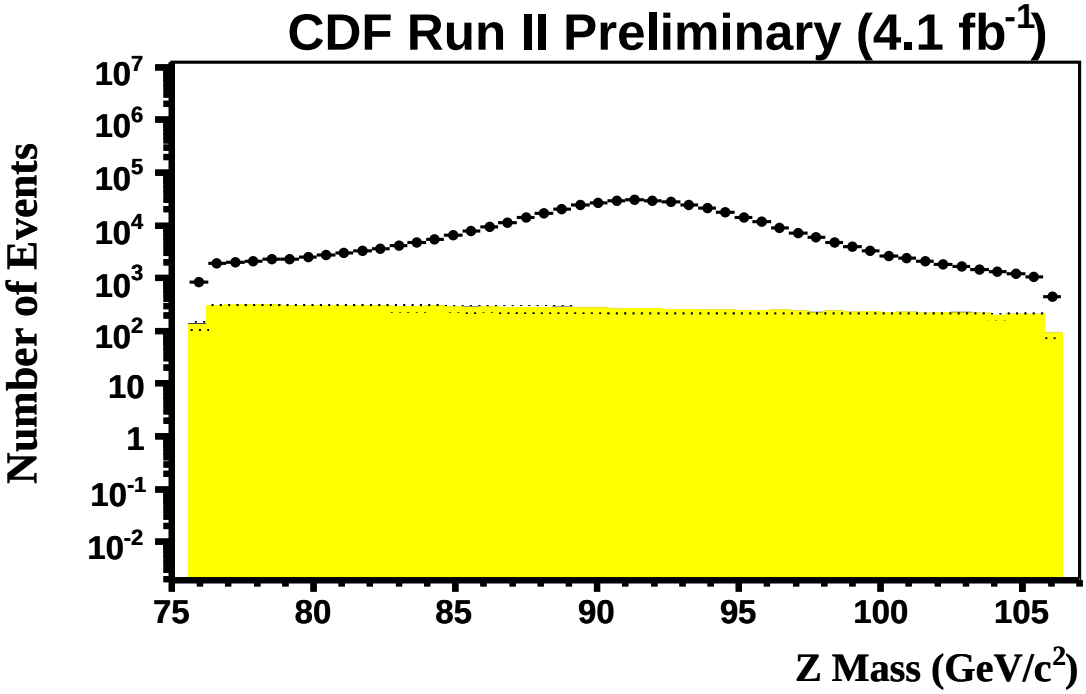


Figure 7.6: Fake (rates measured in jet triggered events) component (yellow) of total data (black) for $Z \rightarrow ee$.

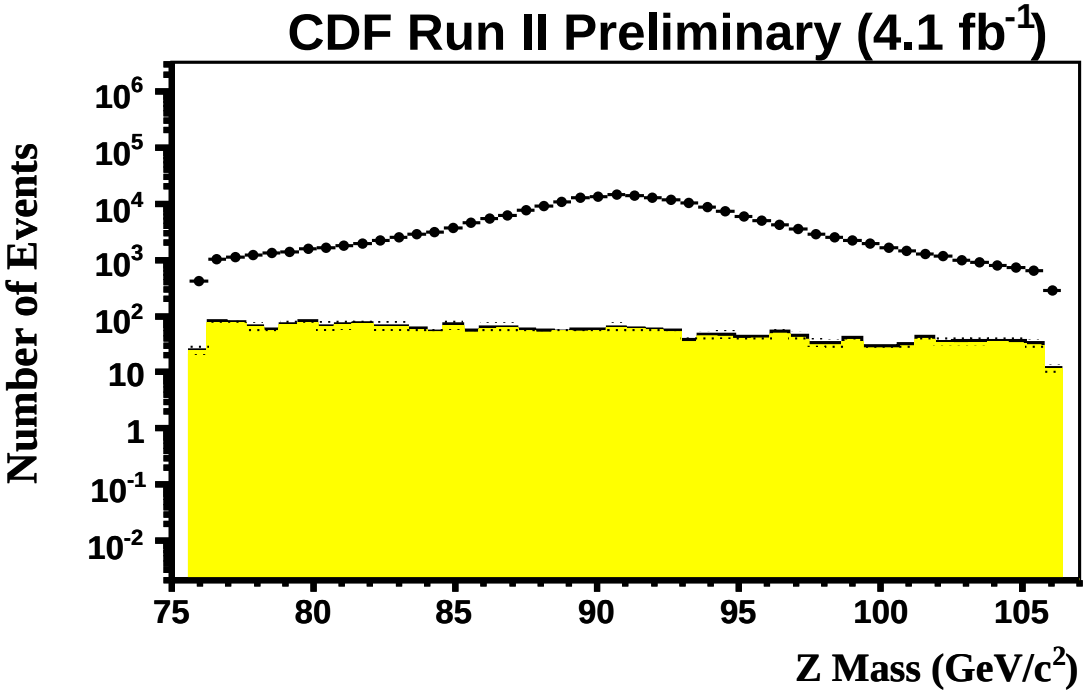


Figure 7.7: Fake (from like-sign events) component (yellow) of total data (black) for $Z \rightarrow \mu\mu$ and $Z \rightarrow Tc + CrkTrk$

7.4 Z Reconstruction and Classification

The pairs of leptons used to form Z candidates in this analysis are limited to the combinations listed in Table 7.22. We require the dilepton mass of the two electron or muon candidates to fall in the mass window $76 \leq M_{ll} \leq 106\,\mathrm{GeV}/c^2$. For events with more than two identified leptons, the Z candidate with largest transverse momentum is used. We impose an opposite charge requirement for all $\mu\mu$ and central-central electron pairs. No charge requirement is imposed when one (or more) of the candidate electrons is in the plug region of the detector. When both electron candidates contain a track we require that distance between their z positions at closest approach to the beamline $|\Delta\text{Track } Z_0| \leq 4\,\mathrm{cm}$. The mass of the reconstructed Z candidate is shown in Fig. 7.8 for $Z \to ee$ and in Fig. 7.9 for $Z \to \mu\mu$. Event totals are tabulated in Tables 7.23 and 7.24.

Table 7.22: Allowed dilepton types considered in this analysis.

Combination	1st Lepton	2nd Lepton	Additional Cuts
TcTc	Tc	Tc	opposite charge
TcLc	Tc	Lc1, Lc2	opposite charge
TcTp	Tc	Tp	none
TcLp	Tc	LpPhx, LpNphx25, LpNphx33, Lp	none
Lc1Lc1	Lc1	Lc1	opposite charge
TpLc	Tp	Lc1, Lc2	none
Lc1Lp	Lc1	LpPhx, LpNphx25, LpNphx33	none
TpTp	Tp	Tp	none
TpLp	Tp	LpPhx, LpNphx33	none
LpLp	LpPhx	LpPhx, LpNphx33	both $E_t >= 30$
TcCrkTrk	Tc	CrkTrk	opposite charge
MuonMuon	CMUP, CMX	CMUP,CMX,CMIO	opposite charge

7.4.1 Additional Requirements on very Loose Z's

We apply additional requirements to improve the quality of Z's formed from very loose electron pairs. The requirements are listed in Table 7.25. The effect of the additional

Table 7.23: Event totals for $Z \to ee$ selection. The systematic uncertainty on the background is about 20%

Events Satisfying $Z \to ee$ Selection Requirements

Source	≥ 2 leptons	$76 \leq M_{ee} \leq 106\,\mathrm{GeV}/c^2$
$t\bar{t}$	198.47	67.07
WW	318.73	99.81
WZ	251.95	214.24
ZZ	262.45	229.73
$Z \to \tau\tau$	2318.79	231.24
Z+jets ($b\bar{b}$)	1846.41	1748.56
Z+jets ($c\bar{c}$)	4093.59	3884.81
Z+jets (lf)	537504	481673
fakes	39842.6	13509.8
ZH ($120\,\mathrm{GeV}/c^2$)	4.16	3.74
Total Background	586636.99	501658.26
Data	567260	475927

Table 7.24: Event totals for $Z \rightarrow \mu\mu$ selection. The systematic uncertainty on the background is about 20%

Events Satisfying $Z \rightarrow \mu\mu$ Selection Requirements

Source	≥ 2 leptons	$76 \leq M_{\mu\mu} \leq 106\,\mathrm{GeV}/c^2$
$t\bar{t}$	87.34	26.82
WW	133.99	40.82
WZ	120.21	107.74
ZZ	133.87	120.36
$Z \rightarrow \tau\tau$	617.63	27.59
Z+jets ($b\bar{b}$)	881.6	857.23
Z+jets ($c\bar{c}$)	1841	1789.43
Z+jets (lf)	207197	186837
fakes	2329	575
ZH ($120\,\mathrm{GeV}/c^2$)	2.4	2.25
Total Background	213341.64	190382.0
Data	199767	174058

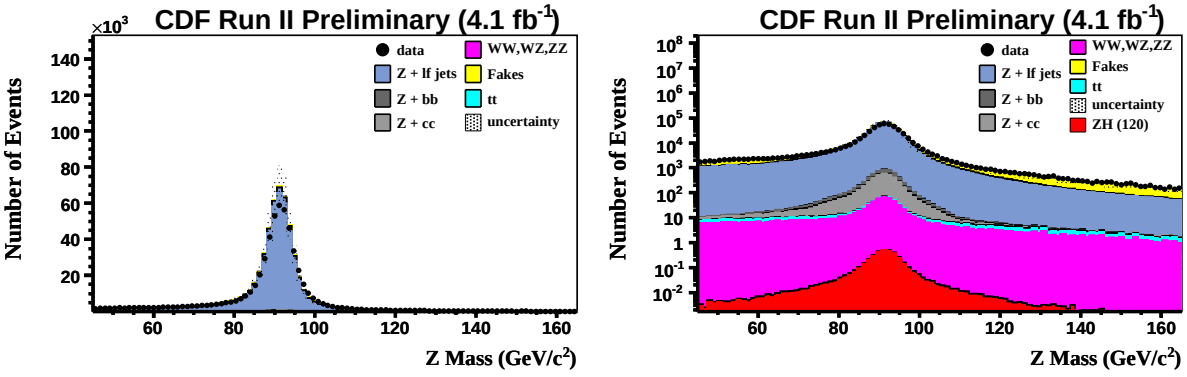


Figure 7.8: Dilepton mass in Z candidates reconstructed from two electron candidates.

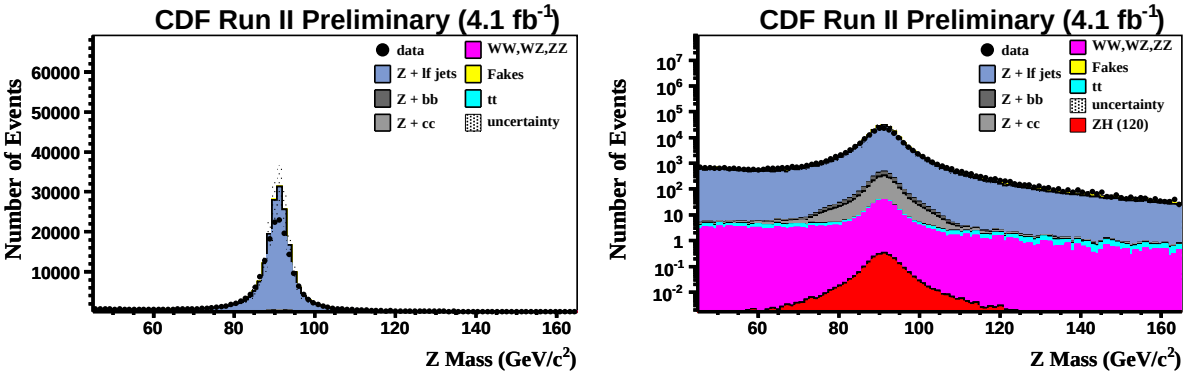


Figure 7.9: Dilepton mass in Z candidates reconstructed from two muon candidates.

requirements is to suppress poorly modeled events and improve the agreement between data and MC as visible in Fig. 7.10.

7.4.2 EM energy corrections

As seen in Fig. 7.11 , the Z mass distribution does not peak at the same value in data and MC. Period specific scale factors are computed from the difference between Z mass fits

Table 7.25: Summary of additional requirements on electrons forming loose Z's.

For a Lc1-Lc1 Z :

one of the Lc1 electrons should meet the following requirements :

number of track axial segments is ≥ 1
number of track axial hits is ≥ 5
number of track stereo segments is ≥ 1
number of track stereo hits is ≥ 5
$|\Delta Z| \leq 3$
$\chi^2 \leq 25$

For a Lc1-Lp Z

the Lc1 electron should pass :

number of track axial segments is ≥ 1
number of track axial hits is ≥ 5
number of track stereo segments is ≥ 1
number of track stereo hits is ≥ 5
$|\Delta Z| \leq 3$
$\chi^2 \leq 25$

or the Lp electron should pass :

$2.4 \geq |\eta| \geq 1.2$
$(En \leq 100 \ \& \ Hadem \leq 0.05)$ *or* $(En > 100 \ \& \ Hadem \leq 0.05 + 0.026 \log En/100.0)$
$(Isolation - 0.35 * (N_{vtx} - 1)) \leq (1.6 + 0.02 * E)$

For a Tp-Lp Z

the Lp electron should pass :

$2.4 \geq |\eta| \geq 1.2$
$(En \leq 100 \ \& \ Hadem \leq 0.05)$ *or* $(En > 100 \ \& \ Hadem \leq 0.05 + 0.026 \log En/100.0)$
$(Isolation - 0.35 * (N_{vtx} - 1)) \leq (1.6 + 0.02 * E)$

For a Lp-Lp Z

one Lp electron should pass :

$2.4 \geq |\eta| \geq 1.2$
$(En \leq 100 \ \& \ Hadem \leq 0.05)$ *or* $(En > 100 \ \& \ Hadem \leq 0.05 + 0.026 \log En/100.0)$
$(Isolation - 0.35 * (N_{vtx} - 1)) \leq (1.6 + 0.02 * E)$

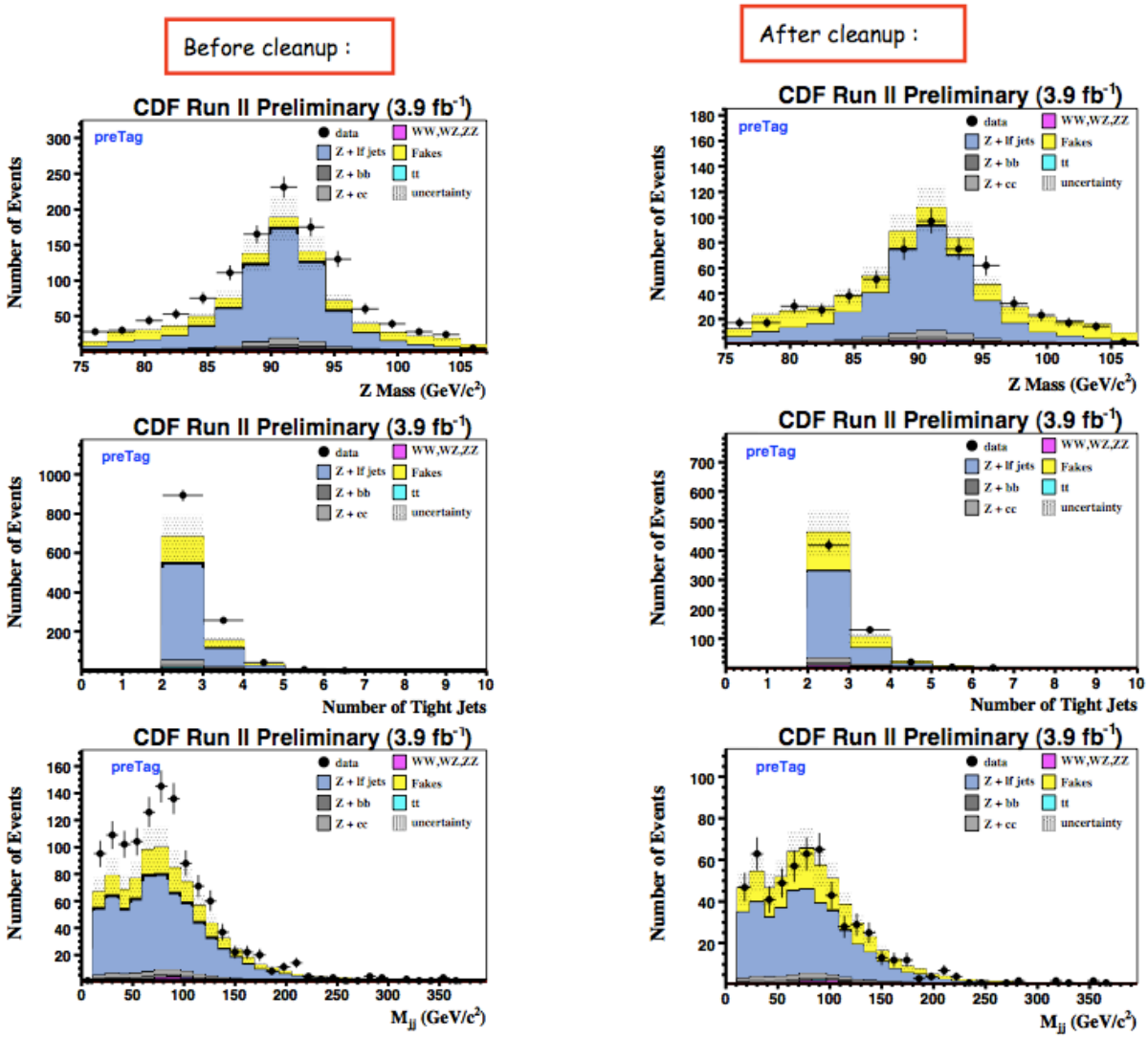


Figure 7.10: Distributions of the Z mass, dijet Mass and Number of tight jets for Lc1Lc1, LpLp, TpLp, and Lc1Lp Z's before and after the additional cuts listed in Table 7.25 are applied.

for data and MC following the method outlined in [84]. The Z mass distribution in the $76 \leq M_{ee} \leq 106\,\mathrm{GeV}/c^2$ mass range is fit to a Gaussian in events with a Z candidate and less than two jets. The ratio of the mean of the Z mass fit in MC to that in data is the correction scale factor. The central (plug) scale factors are computed from events in which both electrons are central (plug). The scale factors presented in Table 7.26 are applied to the electrons in data. Application of the scale factors produces a small change in our acceptance. To ensure that we account for the effect of incorrect lepton energy measurement on our MC normalization we include a 1.5% systematic uncertainty.

Table 7.26: Scale factors for central and plug electron energies.

Data Period	Central Scale Factor	Plug Scale Factor
0	1.004	0.9989
1-4	1.007	1.006
5-7	1.009	1.004
8	1.008	1.013
9	1.009	0.9953
10	1.009	1.005
11	1.012	1.006
12	1.013	1.001
13	1.013	0.9978
14	1.009	1.001
15	1.012	1.010
16	1.013	1.006
17	1.013	0.9946
18	1.007	0.9992
19	1.004	1.003
20	1.006	1.003
21	1.007	1.008
22	1.005	1.004

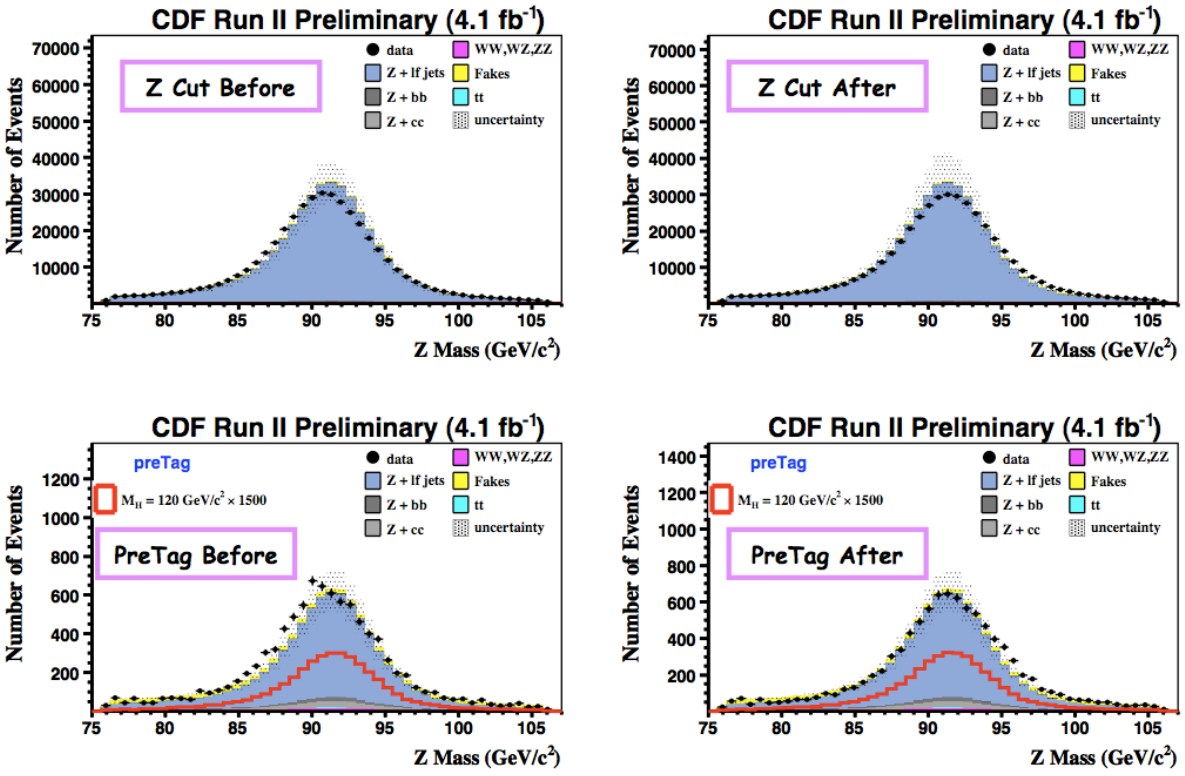


Figure 7.11: Z mass in electron triggered data before and after corrections are applied.

7.5 Jet Selection

After requiring a Z candidate, we impose the requirement that events have two or more Cone 0.4 jets with $E_T \geq 15\,\text{GeV}$and $|\eta| \leq 2.0$. Jets meeting this requirement are referred to as "tight jets". We further require that 1 or more of these jets have $E_T \geq 25\,\text{GeV}$.

Events passing the $Z+ \geq 2$ tight jets selection, with at least one of $E_T \geq$25 GeV, form the "PreTag" sample. These events are further divided into two categories (high and low S/B) to maintain high sensitivity depending on which Z selection the events satisfy. Events passing only the muon triggers or containing a Tc electron are placed in the 'high' class, while events containing a CrkTrk or passing only from the Z_NOTRACK trigger enter the 'low' S/B category. Table 7.27 summarizes the Z categories and high/low grouping. We find

good agreement between our background model and observed data in our PreTag samples in both the predicted number of events (Tables 7.28 and 7.29) and in the shapes of various distributions (Figs. 7.12 through 7.17).

Table 7.27: Z type and $\frac{S}{\sqrt{B}}$ category.

High	Low
TcTc	Lc1Lc1
TcLc	TpLc
TcTp	Lc1Lp
TcLp	TpTp
μ μ	TpLp
	LpLp
	TcCrkTrk

7.6 b-tagging

In order to significantly diminish the $Z + jets$ background we impose b-tag requirements on our PreTag sample; that is we require at least one jet in the event to be identified as the product of a b quark's hadronization.

Table 7.28: Preselection event totals (high S/B).

Source	PreTag high S/B
$t\bar{t}$	53.01 ± 11.26
WW	5.22 ± 0.71
WZ	117.89 ± 15.95
ZZ	118.14 ± 15.98
$Z \rightarrow \tau\tau$	2.98 ± 1.21
Z+jets ($b\bar{b}$)	370.93 ± 150.71
Z+jets ($c\bar{c}$)	682.59 ± 277.34
Z+jets (lf)	9977.08 ± 1995.42
fakes	541.02 ± 270.51
ZH ($120\,\mathrm{GeV}/c^2$)	4.25 ± 0.32
Total Background	11868.9 ± 2038.4
Data	11806

Table 7.29: Preselection event totals (low S/B).

Source	PreTag low S/B
$t\bar{t}$	27.12 ± 5.76
WW	4.3 ± 0.58
WZ	27.04 ± 3.66
ZZ	23.28 ± 3.15
$Z \rightarrow \tau\tau$	4.33 ± 1.76
Z+jets ($b\bar{b}$)	74.51 ± 30.28
Z+jets ($c\bar{c}$)	142.25 ± 57.79
Z+jets (lf)	2206.9 ± 441.38
fakes	504.44 ± 252.22
ZH ($120\,\mathrm{GeV}/c^2$)	0.67 ± 0.05
Total Background	3014.17 ± 512.6
Data	3061

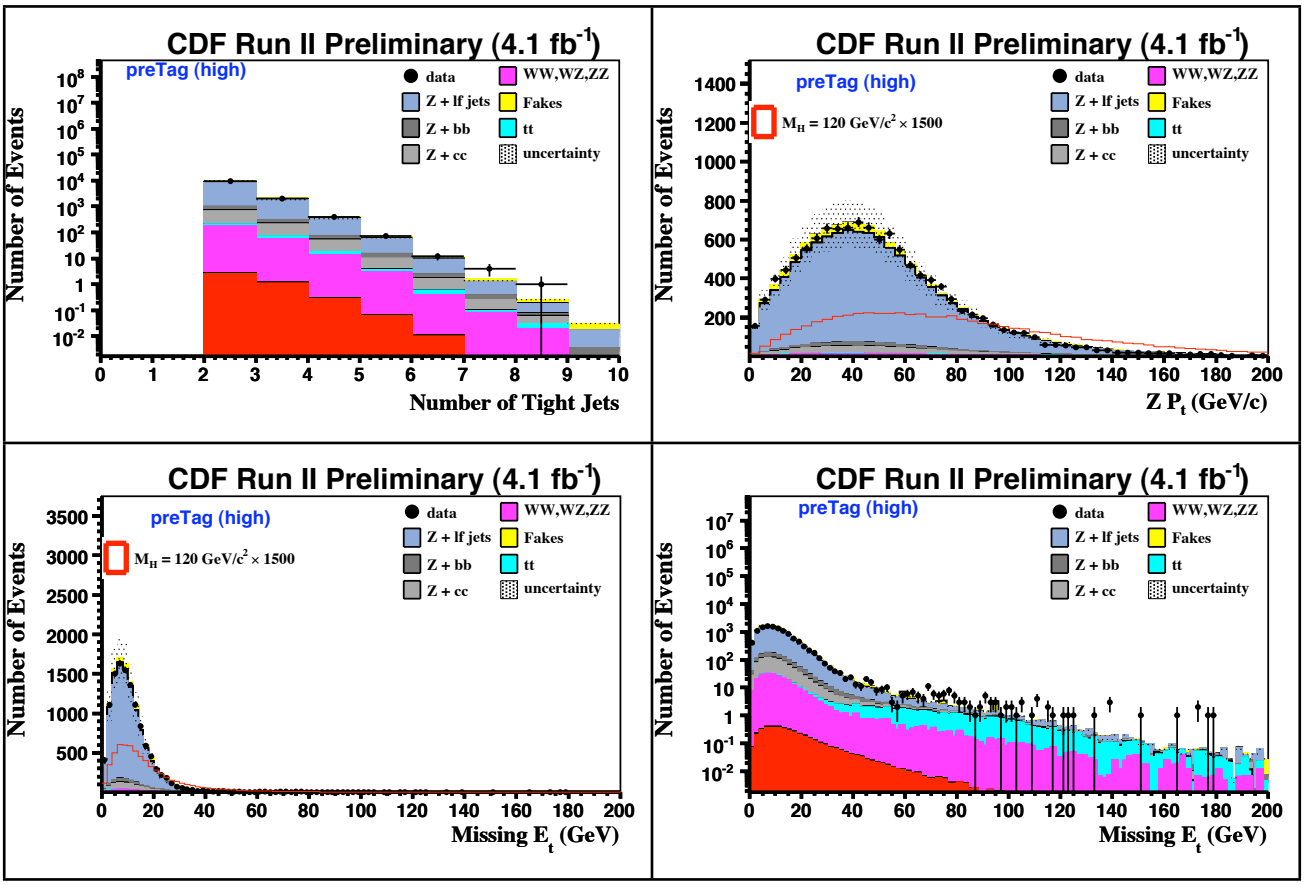


Figure 7.12: Distributions of the number of tight jets, Z p_T and $\not\!\!E_T$ in preTag events in the high S/B Z category.

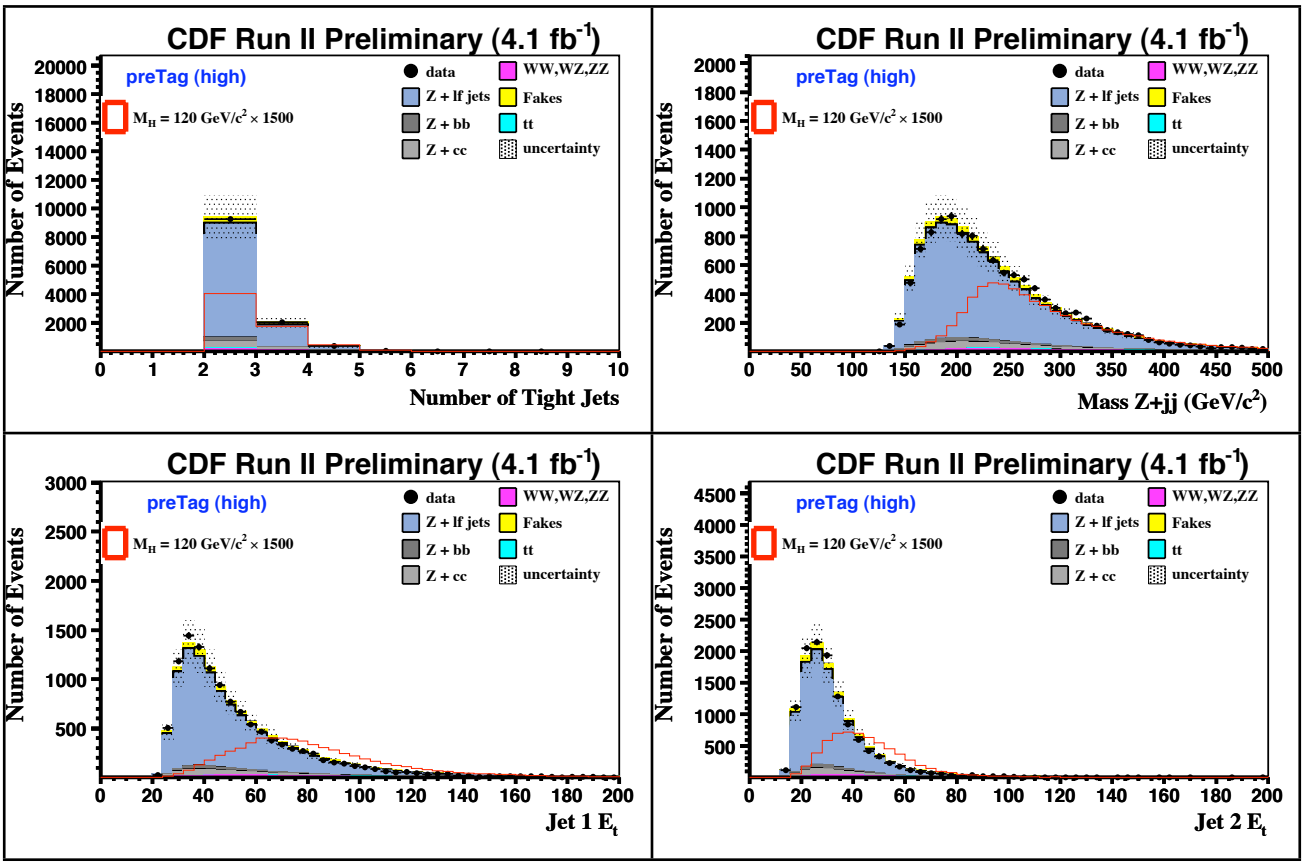


Figure 7.13: Distributions of the number of tight jets, $Z+2$ jet Mass and jet E_T's in preTag events in the high S/B Z category.

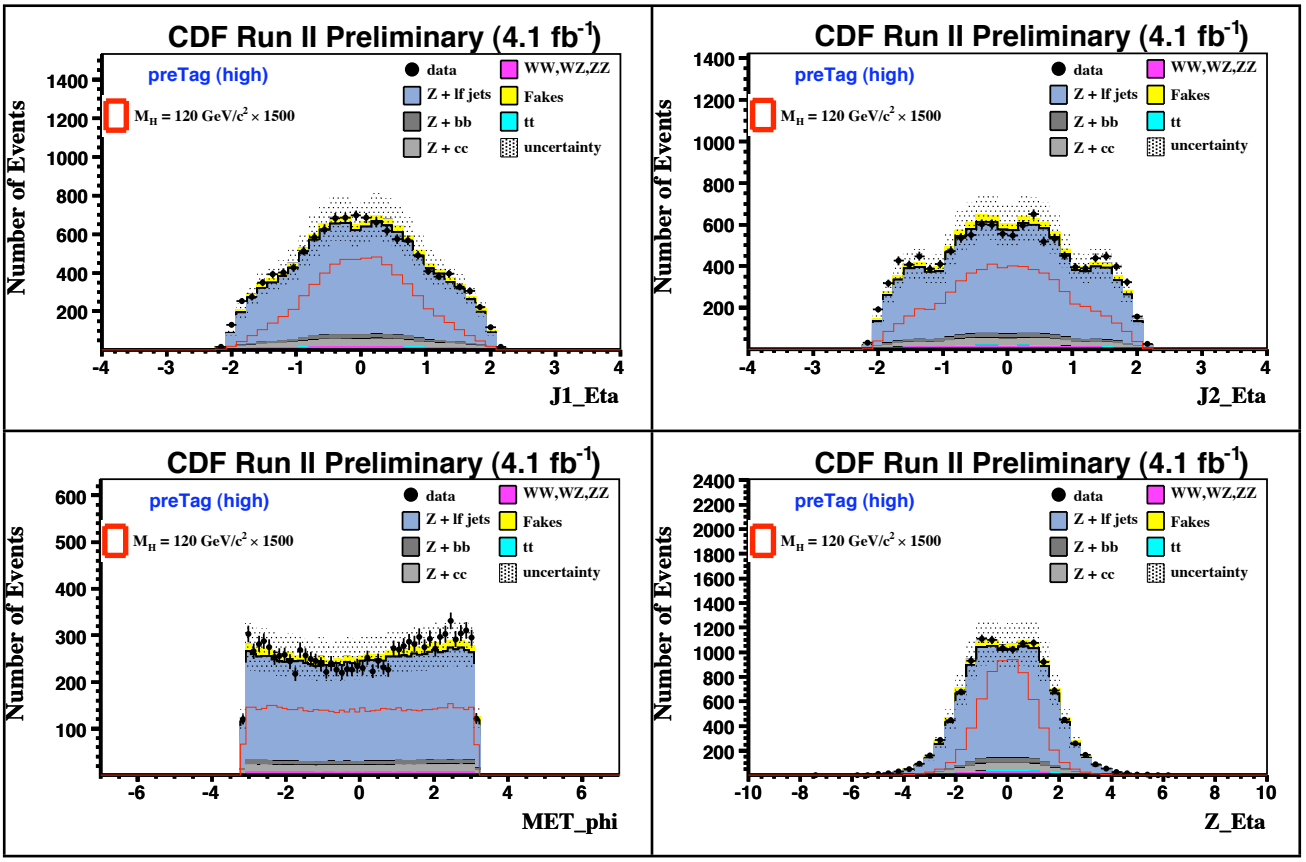


Figure 7.14: Jet 1, Jet 2, and Z ηs and $\not\!\!E_T$ ϕ in preTag events in the high S/B Z category.

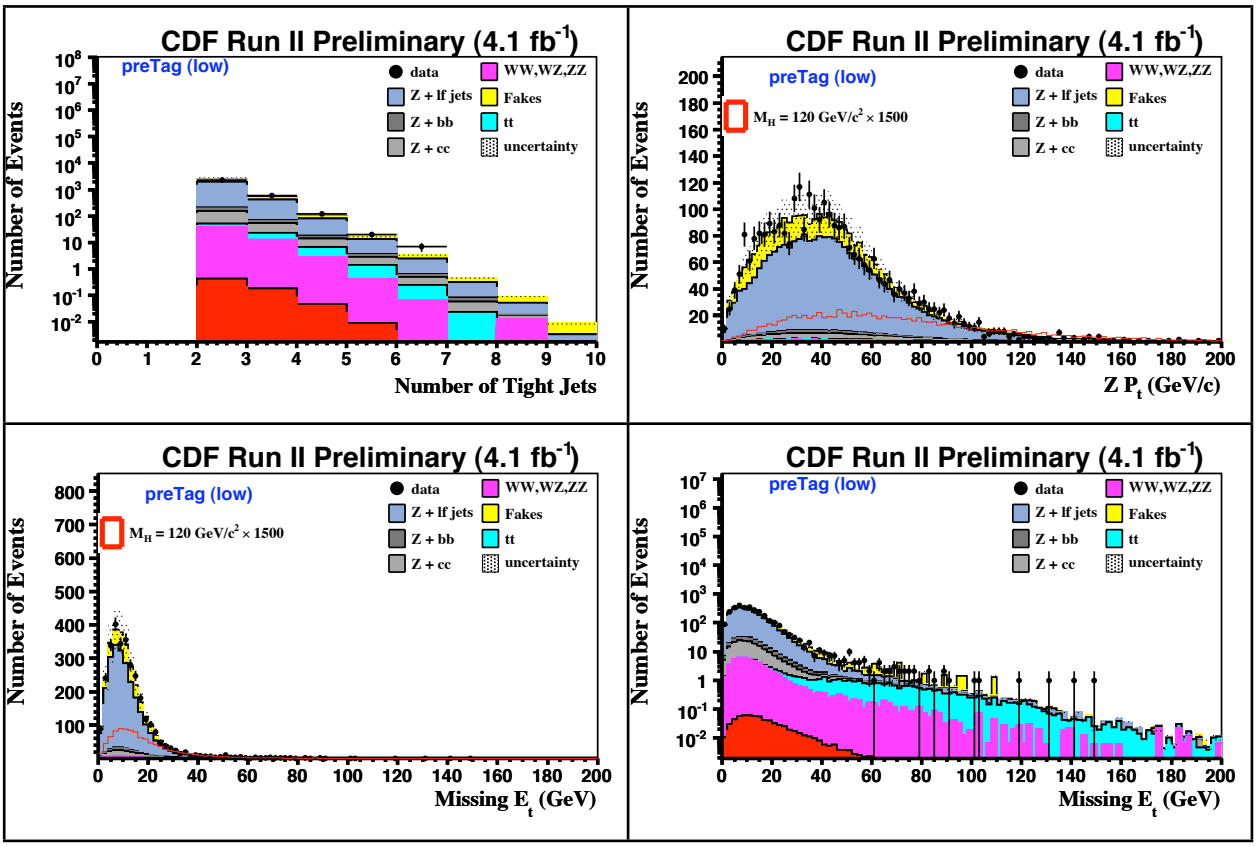


Figure 7.15: Distributions of the number of tight jets, Z p_T and $\not\!\!E_T$ in preTag event in the low S/B Z category.

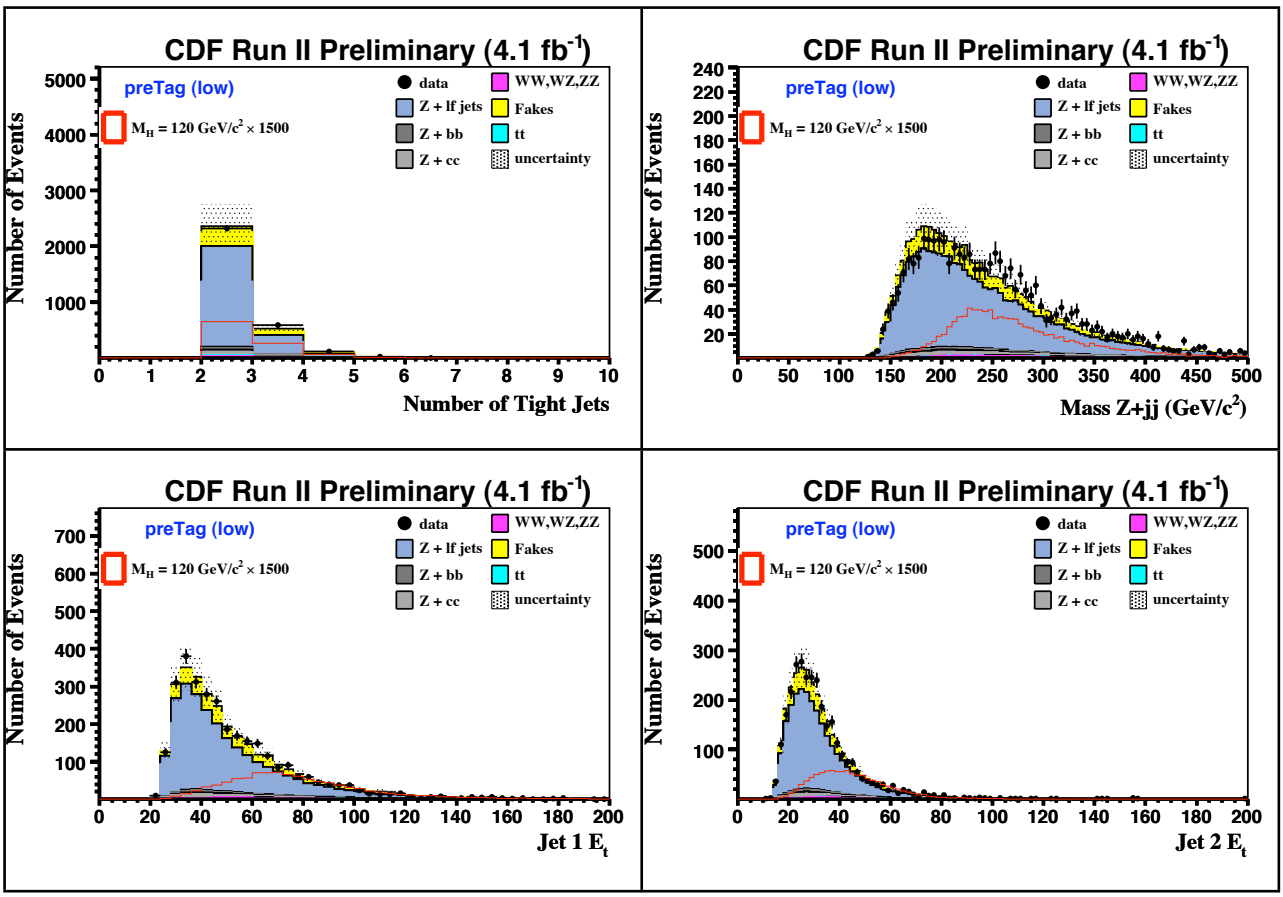


Figure 7.16: Distributions of the number of tight jets, $Z+2$ jet Mass and jet E_T's in preTag events in the low S/B Z category.

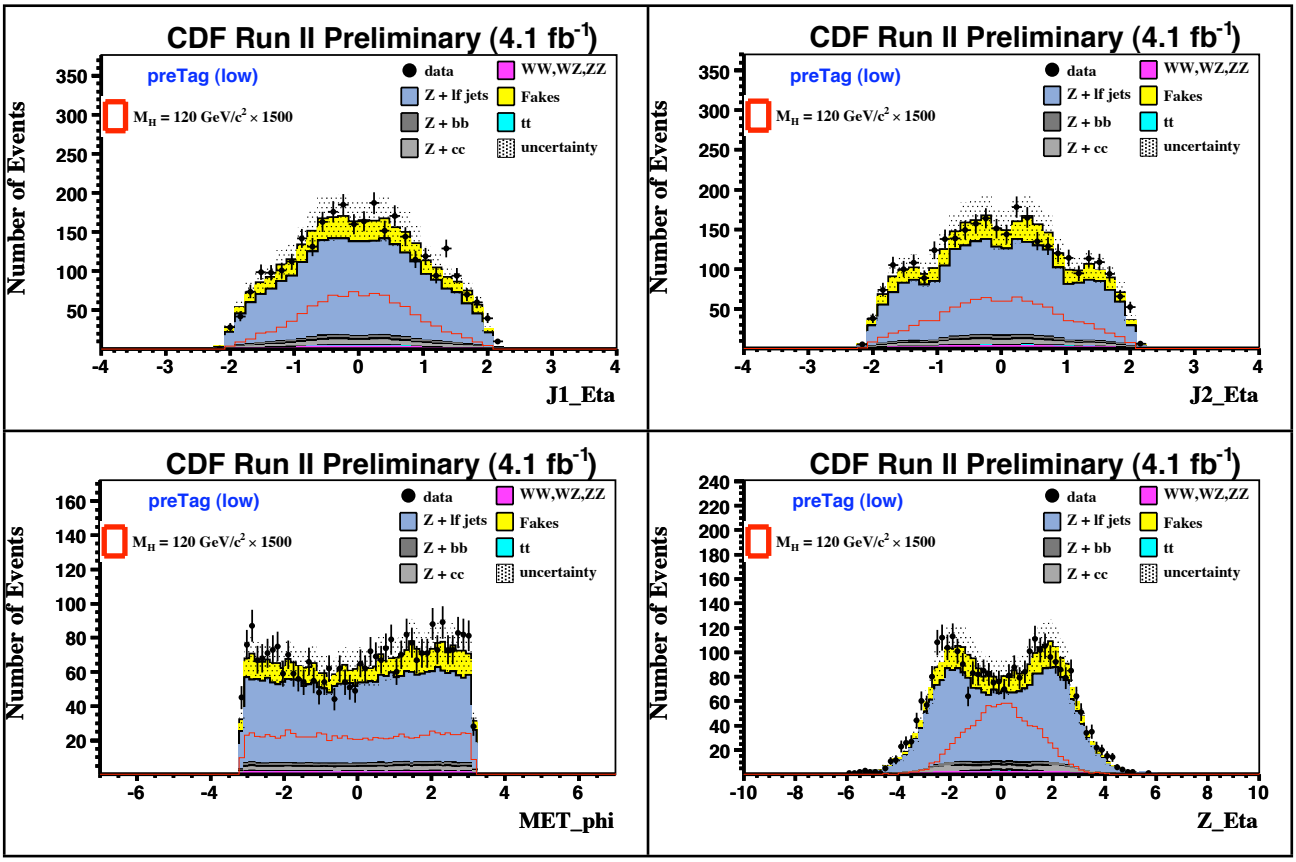


Figure 7.17: Jet 1, Jet 2, and Z ηs and $\not\!\!E_T$ ϕ in preTag events in the low S/B Z category.

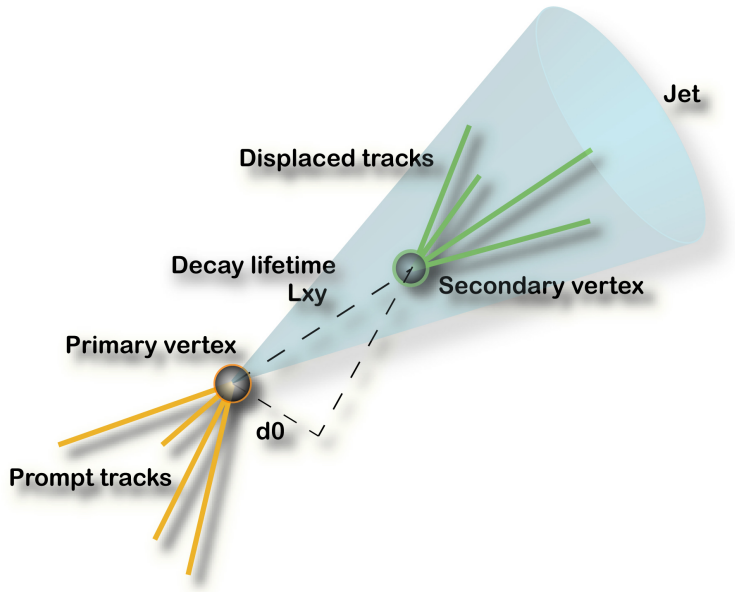


Figure 7.18: Depiction of a b jet with secondary vertex. d0 is the impact parameter of a displaced track. For simplicity, prompt tracks are not shown. Image from [85].

When a b quark is produced in the CDF II detector, it will hadronize producing a jet containing B hadrons. These hadrons possess a lifetime long enough to allow them to travel a short distance (about 1 cm) before decaying. Therefore, a b quark manifests as a jet which contains several tracks pointing to a displaced (secondary) vertex within the jet cone. This topology is depicted in Fig. 7.18.

We examine each tight jet for signs of a picosecond lifetime hadron: a displaced vertex, or tracks with large impact parameters. These characteristics are typical of jets resulting from b-quark hadronization. We use two algorithms to identify (tag) b jets: one based on evidence for a displaced vertex and one based on track impact parameters:

- The secondary vertex (SecVtx) [86] tagging algorithm tags b jets using displaced vertex information. If two or more tracks associated with the jet are found to originate from a secondary vertex within the jet, the algorithm tags the jet as a b jet. We use both the "Tight" and "Loose" SecVtx operating points which differ in track and vertex quality

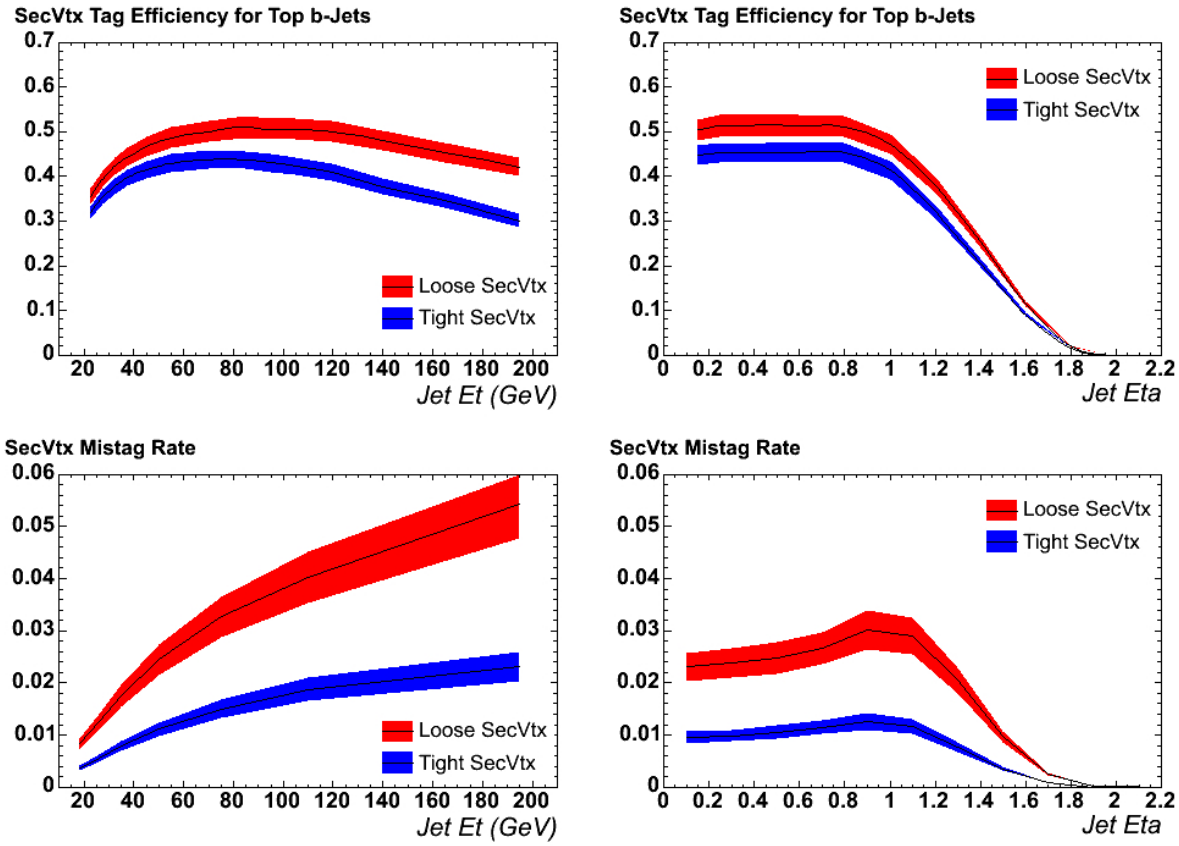


Figure 7.19: Tagging efficiency for b jets in $t\bar{t}$ MC as a function of jet E_T (top left) and η (top right). Mistag rates for jets in jet data as a function of jet E_T (bottom left) and η (bottom right). Figures from [87].

requirements. The efficiency for tagging b jets in $t\bar{t}$ MC events and the likelihood of tagging non-b jets in jet triggered data is presented in Fig. 7.19.

- The jet probability (JP) [88] tagging algorithm uses the signed impact parameters (Fig 7.20) of tracks associated with a jet to compute the likelihood that the tracks in the jet originate from the primary vertex. Light flavor jets are uniformly distributed in JP output between 0 and 1, while the algorithm is more likely to return small values for heavy flavor jets. In order for a jet to be considered a b jet in this search, we require the JP algorithm to return a value less than or equal to 0.05.

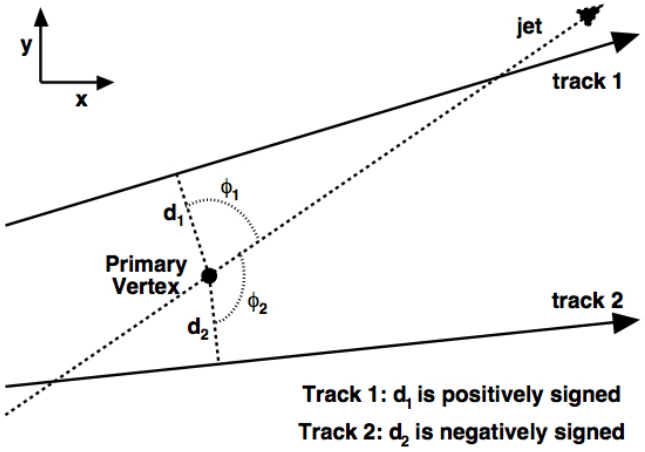


Figure 7.20: The sign of the impact parameter of a track. The impact parameter is assigned a negative (positive) sign if the angle ϕ is greater (less) than $\pi/2$. Figure from [88].

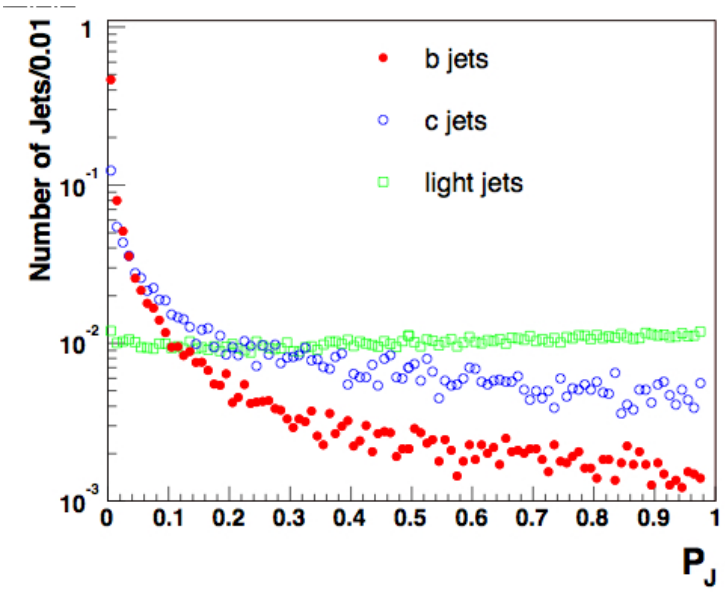


Figure 7.21: Distribution of values returned by the JP algorithm for simulated b (red), c (blue) and light flavor jets (green). Figure from [88].

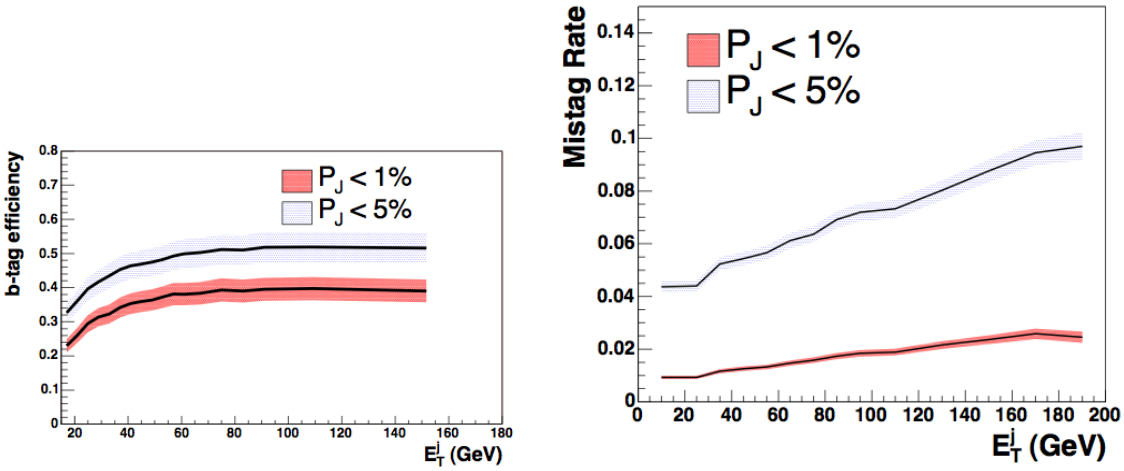


Figure 7.22: (left) Efficiency of the JP tagging algorithm on b jets in $t\bar{t}$ MC as a function of jet E_T. The efficiency is shown for $P_J < 0.01$ (red) and $P_J < 0.05$ (blue) selection. (right) The likelihood of mistagging light flavor jets in jet data as a function of jet E_T. The mistag rate is shown for $P_J < 0.01$ (red) and $P_J < 0.05$ (blue) selection. We impose the $P_J < 0.05$ b tag cutoff (as opposed to 0.01) due to the significantly increased b tag efficiency. Figures from [88].

Table 7.30: Summary of jet/b-tag selection.

Jet Selection for PreTag Region
2 or more Cone 0.4 jets $E_T > 15$ GeV, $\|\eta\| < 2$ and 1 of these jets with $E_T > 25$ GeV $\rightarrow$ (PreTag)
b-tagging Selection for Signal Region
2 or more tight SECVTX tags $\rightarrow$ (TT) If not found 1 Loose SECVTX tag and 1 JetProbability Tag (5%) $\rightarrow$ (L+JP) If not found 1 tight SECVTX tag $\rightarrow$ (T)

While there is some overlap in the information used by the two algorithms, we find that about 10% of all tight jets in our signal sample are exclusively tagged by the JP algorithm.

In order to identify the pair of jets in our PreTag event sample most likely to have been produced in an $H \rightarrow b\bar{b}$ decay, we consider the SecVtx and JP algorithm outputs for each possible pair of tight jets in the event in which at least one of the jets has $E_T \geq 25\,\text{GeV}$. If both jets in the pair meet the Tight SecVtx b tag requirements we classify the pair as having a double tight tagged pair (TT). If the pair fail the TT requirement, and one jet is Loose SecVtx tagged and the other has a JP value less than or equal to 0.05, we classify the pair as having a loose plus jet probability tagged pair (L+JP). If the pair fails to meet the TT or L+JP selection, it will receive a single tight classification (T) if one of the jets is Tight SecVtx tagged.

In events with multiple jet pairs classified as T, L+JP, or TT we preferentially choose the TT pair as the $H \rightarrow b\bar{b}$ candidate over L+JP or T pairs. Similarly L+JP pairs are chosen over T pairs. This selection preference follows naturally from the $Z + b\bar{b}$ to Z+mistagged jet ratios of the three tagging classifications: 14.9, 1.9, and 0.6 for TT, L+JP and T respectively. When two jet pairs have the same classification (for example TT and TT) the pair with the highest combined E_T is selected as the $H \rightarrow b\bar{b}$ candidate. Our jet and b-tag requirements are summarized in Table 7.30.

7.7 Final Analysis Channels

With two Z categories (high S/B , low S/B) and three b-tag regions (TT, L+JP, and T) we form six final event samples, as listed in Table 7.31.

Table 7.31: Analysis channels.

Channel	Lepton ID	b-Tag Requirements
Double Tag High	High S/B	Two Tight SecVtx Tags
Loose + 5% JP High	High S/B	One Loose SecVtx Tag & One 5% JetProbability Tag
Single Tag High	High S/B	One Tight SecVtx Tags
Double Tag Low	Low S/B	Two Tight SecVtx Tags
Loose + 5% JP Low	Low S/B	One Loose SecVtx Tag & One 5% JetProbability Tag
Single Tag Low	Low S/B	One Tight SecVtx Tags

Chapter 8: Multvariate Techniques

Multivariate techniques combine multiple quantities to form a combined signal/background discriminant and have become a standard tool of high energy physics. This search relies on artificial neural networks (NN) and matrix element probabilities (ME) to maximize the separation of ZH signal from background events.

8.1 Artificial Neural Networks

NNs have been utilized in previous searches for the SM Higgs in [89, 90, 91, 92] and others. Figure 8.1 depicts a simple feed-forward NN with input layer, hidden layer, and output layer. In a feed-forward NN information flows forward through the network, from the input layer to the output layer. Each "node" in the input layer represents a measured or calculated quantity such as the mass or momentum of a particle. Let the set $\{X_1, X_2, .., X_i, ..\}$ denote a collection of such quantities.

Acting as NN inputs, the values X_i are scaled by a collection of weight factors W_{ij} before being passed to the nodes in the hidden layer (denoted by the set $\{Y_1, Y_2, .., Y_j, ..\}$), such that the "$j^{th}$" node in the hidden layer accepts a weighted sum of the inputs :

$$Y_j^{IN} = \sum_i W_{ij} \times X_i \tag{8.1}$$

Before the values, $\{Y_j^{IN}\}$, are passed to the output layer, an activation function (g) is applied. The activation function is typically of the form :

$$g_j(x) = (1 + e^{-2[x+\phi_j^h]})^{-1} \tag{8.2}$$

chosen to approximate the activation behavior of neurons in biological systems. The parameters ϕ_j^h are known as the "threshold" values of hidden nodes Y_j.

Each node in the hidden layer passes a value Y_j^{OUT} :

$$Y_j^{OUT} = g_j(Y_j^{IN}) \tag{8.3}$$

to the output layer.

Each node in the output layer (denoted by the set $\{Z_1, Z_2, .., Z_k, ..\}$) accepts a weighted sum :

$$Z_k^{IN} = \sum_j M_{jk} \times Y_j^{OUT} \tag{8.4}$$

where the set $\{M_{jk}\}$ is the collection of weight factors applied between the hidden and output layers. The final output values of the NN are formed by applying a second collection of activation functions $\{g'\}$:

$$g'_k(x) = (1 + e^{-2[x+\phi_k^O]})^{-1} \tag{8.5}$$

where the parameters ϕ_k^O are the threshold values of output nodes Z_j.

For the simple NN architecture depicted in Fig. 8.1, the NN output value is given by :

$$Z_1^{OUT} = g'_1\Big(M_{11} \times g_1(W_{11}X_1 + W_{21}X_2 + W_{31}X_3) + M_{21} \times g_2(W_{12}X_1 + W_{22}X_2 + W_{32}X_3)\Big) \tag{8.6}$$

The NN achieves discriminating power by fitting the combined set of thresholds and weights, $\{\omega\} = \{W_{ij}, M_{jk}, \phi_j^h, \phi_k^O\}$, to produce the desired output values for a given set of training events. In this search we utilize the back-propogation (BP) algorithm [93] to fix $\{\omega\}$.

In BP a NN is formed with random initial values $\{\omega_0\}$. This NN is applied to a sample of training patterns (in our search MC events), and the performance of the NN is rated

according to the average error :

$$E = \frac{1}{2N_p} \times \sum_{p=1}^{N_p} \sum_{k} \left(O_k^p - T_k^p \right)^2 \tag{8.7}$$

where k runs over each node in the output layer, N_p is the total number of training patterns, T_k^p is the target (or desired) response from output node k on pattern p, and O_k^p is the observed response from output node k on pattern p. In BP, $\{\omega\}$ is updated after the NN is exposed to a collection of training patterns according to :

$$\{\omega_1\} = \{\omega_0\} - \frac{\partial E_0}{\partial \omega} \tag{8.8}$$

where stabalization and momentum terms [94] are omitted for simplicity.

BP is repeated until the NN achieves the desired performance. At regular intervals called epochs (in our case, defined as exposure to all training events) the NN is exposed to an independent "test" sample of events. To avoid over-fitting $\{\omega\}$ to the training patterns, the average error is required to have similar values for both test and training samples.

8.2 Matrix Elements

Matrix element calculations have previously been applied in top mass measurements [95, 96] and Higgs boson searches [97, 36]. Here we present a brief overview of the method. A detailed description of matrix element calculations in the context of a CDF measurement is presented in [98].

Given a set of observables $\vec{y}$ (in our case the 4-momenta of the $\ell^+, \ell^-, b, \bar{b}$ and the x and y components of the $\not\!\!\vec{E}_T$), we form the likelihood that the observed quantities are the result

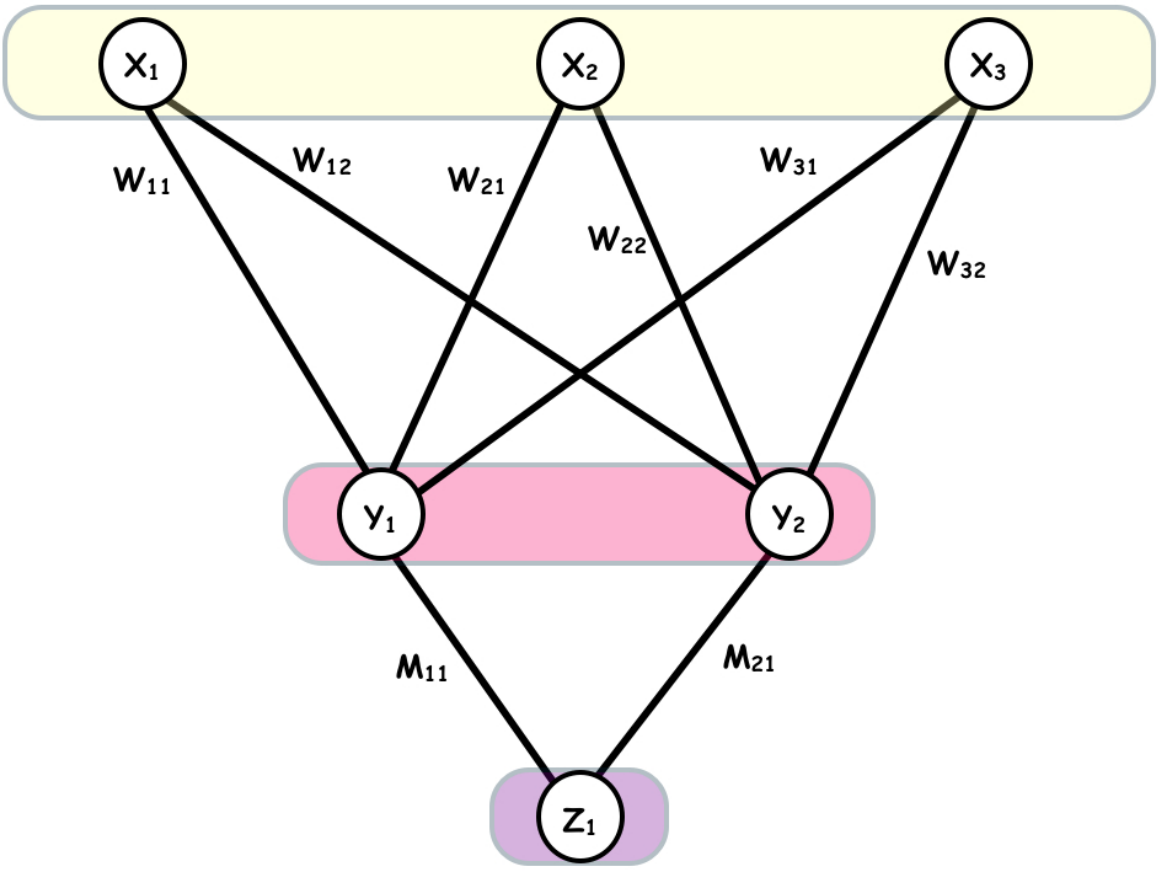


Figure 8.1: Structure of a NN with 3 nodes in the input layer (yellow band), 2 nodes in a single hidden layer (pink band) and 1 node in the output layer (purple band). Connections between nodes are represented by solid black lines with weights W_{ij} and M_{jk} indicated. The values X_1, X_2, X_3 represent experimentally observed quantities. Equations 8.1 and 8.3 show expressions for the values Y_j, while the final NN value Z_1 is given in Eq. 8.6.

of some SM process, P^{SM} (such as ZH, $t\bar{t}$, etc.) as :

$$\mathcal{L}(\vec{y}|P^{SM}) = \int |M(P^{SM},\vec{x})|^2 \mathcal{F}(\vec{y}|\vec{x}) P(\vec{x}) d\vec{x} \tag{8.9}$$

where M is the matrix element for the process P^{SM}, $\vec{x}$ is the parton level momenta of the final state particles, $\mathcal{F}$ is the "transfer function" which relates the probability of observing $\vec{y}$ given the parton quantities $\vec{x}$, and P is the prior distribution of $\vec{x}$.

In this search, M is computed at leading-order for $ZH \to \ell^+\ell^- b\bar{b}$, $t\bar{t}$, and Z+jets processes using MCFM [99]. While leptons are assumed to be perfectly measured, $\mathcal{F}$ are constructed for jet energies by comparing the energy of generator level quarks in MC to the resulting jet energy after the full detector simulation is applied [100]. The prior, P is determined by the product of the parton density functions of the incoming proton and antiproton.

The following sections detail the application of NN's and ME's to this search.

8.3 NN Jet Energy Corrections

The dijet mass (M_{jj}) is one of the most useful quantities to discriminate ZH from $Z + jets$ and is even more powerful with improved jet energy resolution. In general, incorrect measurement of jet energies can result in overestimation of $\not\!\!E_T$. To improve the dijet mass resolution we correct jet energies by a factor which depends on the $\not\!\!E_T$ direction and magnitude, and projections onto the jet directions as described in Ref. [101].

We train a NN to correct jet energies back to generator (parton) values. The NN is trained on a range of Higgs masses and $Z + jets$ samples to ensure proper performance on a variety of event types. The NN is given the measured (lead and second) jet transverse energies along with information about the $\not\!\!E_T$ content. The full list of inputs is presented in Table 8.1. The NN returns correction factors for the lead and second jet. The improvement of the dijet mass resolution can be seen in Figs. 8.2 and 8.3.

To retain the strong ZH versus $t\bar{t}$ discriminating power of the $\not\!\!E_T$, we do not adjust the

Table 8.1: Variables used to correct jet energies to parton level.

L5 Jet 1 E_T
L5 Jet 2 E_T
Jet 1 η
Jet 2 η
$\Delta\phi(jet1, jet2)$
$\Delta\phi(\not\!\!E_T, jet1)$
$\Delta\phi(\not\!\!E_T, jet2)$
Jet 1 Projection onto $\not\!\!E_T$
Jet 2 Projection onto $\not\!\!E_T$
$\not\!\!E_T$ magnitude
number of tight jets
the projection of the Z boson onto the lead E_T jet
the projection of the Z boson onto the second E_T jet

$\not\!\!E_T$ for the change in jet energies after NN correction. Figure 8.4 shows $\not\!\!E_T$ with and without the Neural Network corrections. All other event quantities are re-calculated using the NN corrected jet energies.

8.4 Karlsruhe Flavor Separator

In previous iterations [35, 36] of this analysis the single tag categories have suffered from low S/B due to the presence of large (about 40%) backgrounds from incorrectly tagged light flavor jets. In order to increase the ability of our final analysis discriminants (2D-NN discussed below) to separate this 'mistag' background from signal, we include the output of the Karlsruhe Neural Network (KNN) $b - tagger$ [102]. The KNN is applied to jets with a tight SevVtx tag. This NN is trained to separate b jets from c quark and light flavor jets by returning high values (near 1) for b jets and low values (near -1) for incorrectly tagged jets. The KNN output for the single tag channels (T) is shown in Fig. 8.5. The KNN output for the lead E_T jet in (TT) events is included for comparison.

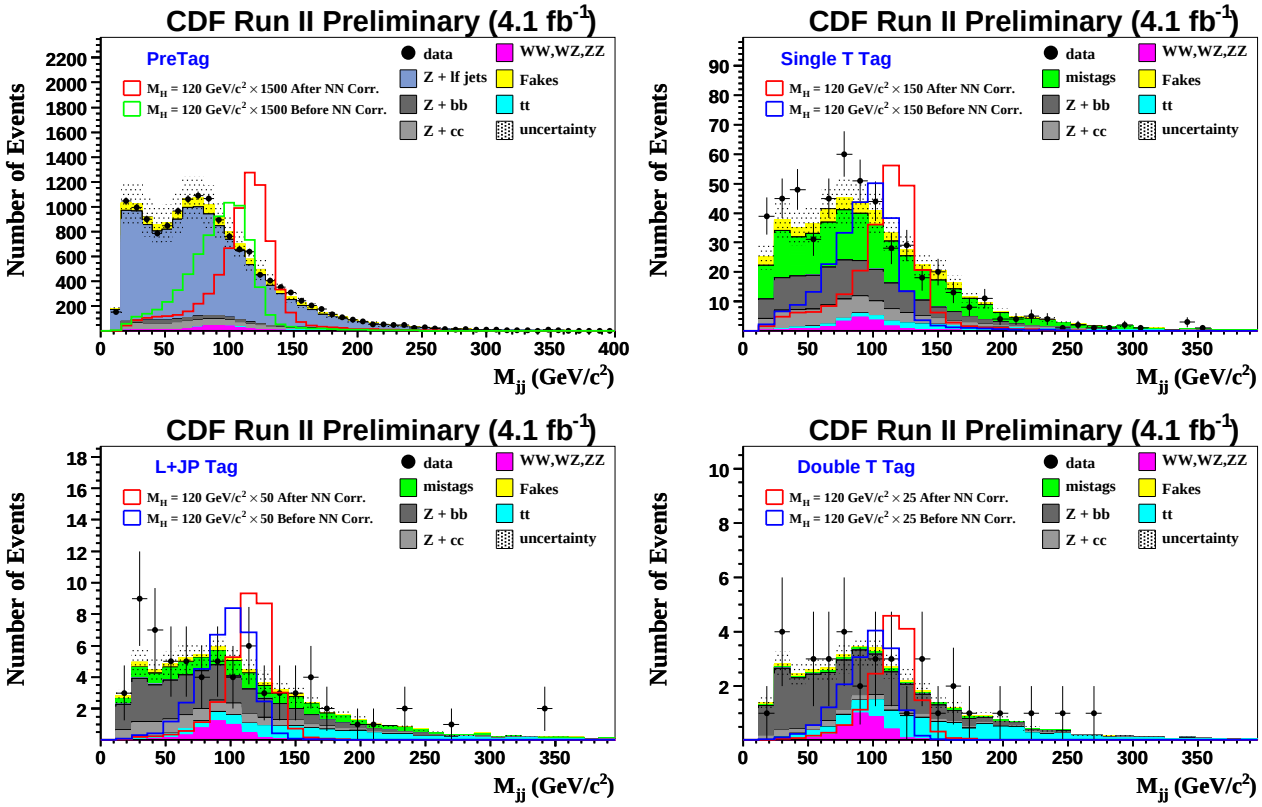


Figure 8.2: Effect of jet energy NN corrections on signal. The blue or green line shows the signal dijet mass before corrections. The red line shows dijet mass after corrections. The dijet mass is shown for the data and background model after corrections are applied.

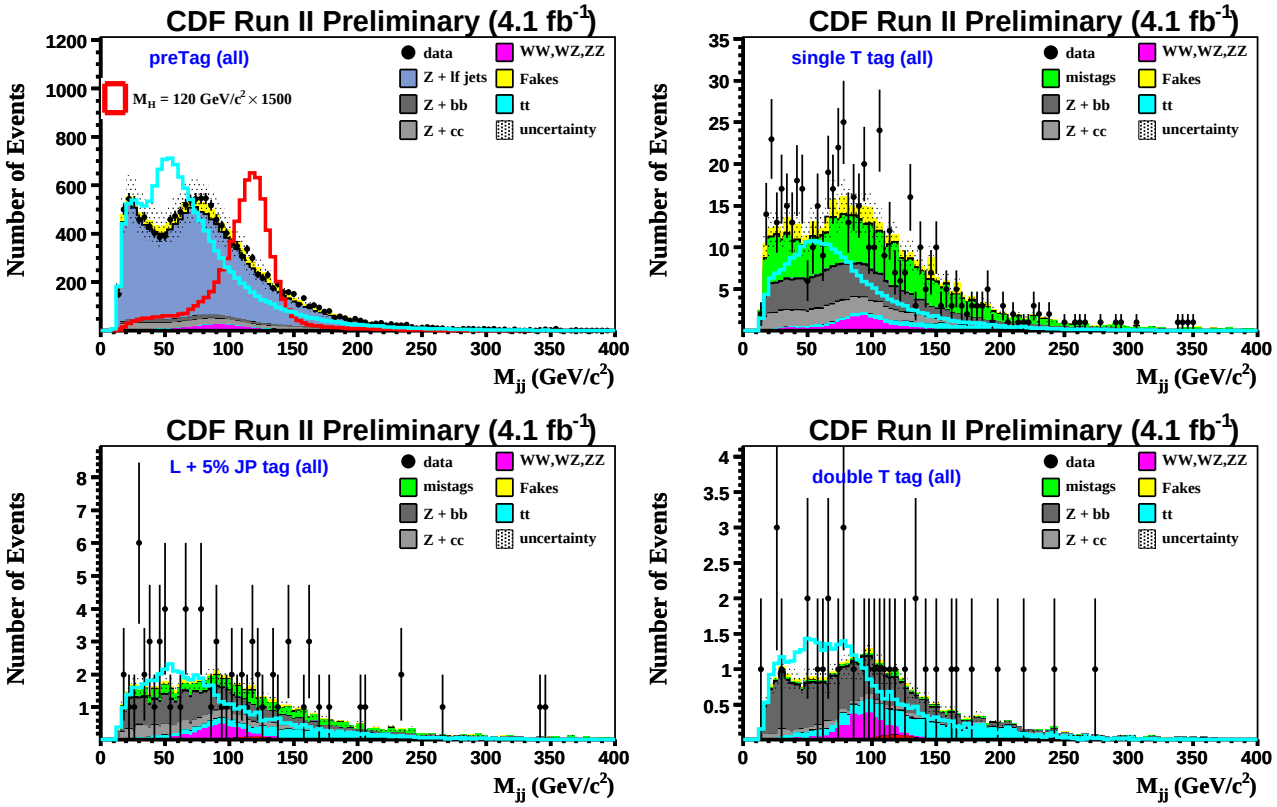


Figure 8.3: Effect of jet energy NN corrections on backgrounds. The teal line shows the background dijet mass before corrections. The teal line does not include the mistag background in the tag level plots.

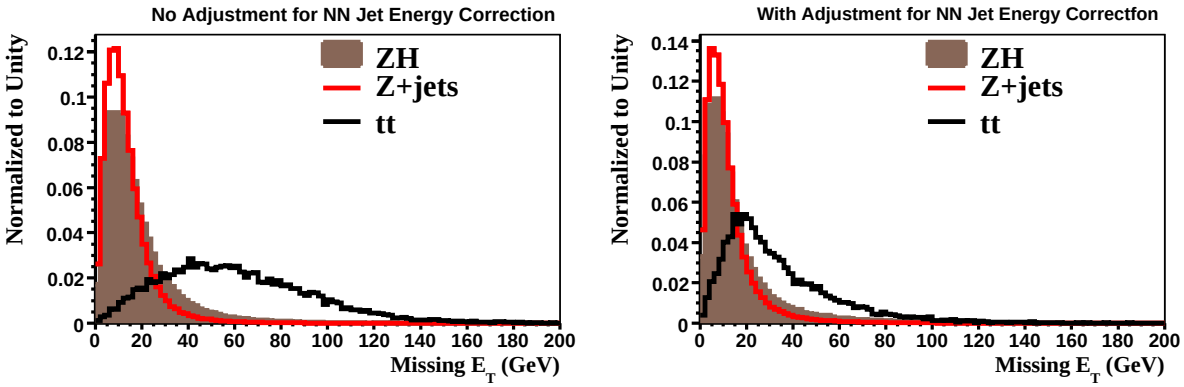


Figure 8.4: The magnitude of missing $\vec{E_T}$ without (left) and with (right) Neural Network corrections. Missing E_T is shown for ZH(brown), $t\bar{t}$ (black) and Z+jets (red) MC.

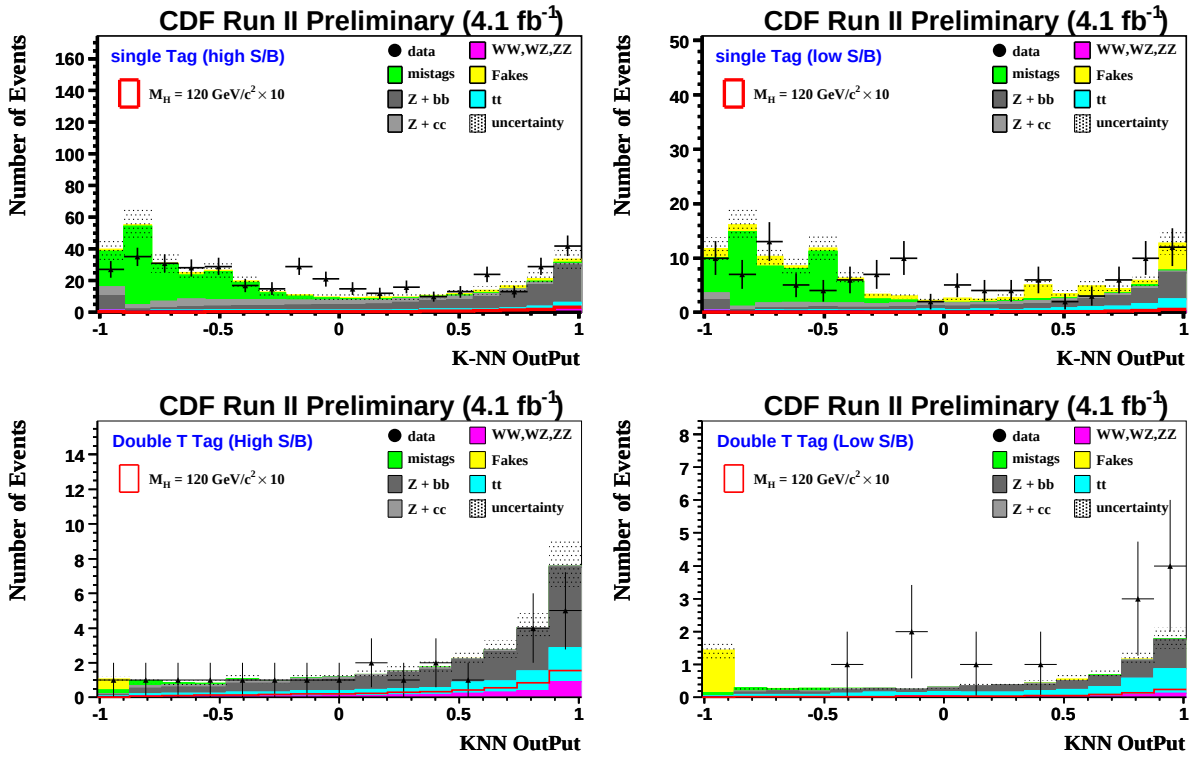


Figure 8.5: Distribution of KNN flavor separator output for single b-tag events (top row) and for the lead E_T jet in double b-tag events (bottom row).

8.5 Matrix Element Probabilities

As discussed in Sec. 8.2, we form matrix element probabilities (ME) as potential inputs to our final 2D-NNs. Matrix element discriminants were first employed in the search for $ZH \rightarrow \ell\ell b\bar{b}$ in [103]. We use the same calculation to determine the ME's for this analysis. We compute matrix element probabilities for the processes Z+jets, $t\bar{t}$ and ZH signal, and refer to the resulting likelihoods as P_{jj}, P_{tt} and P_{zh} respectively. We display the logarithm of the ME probabilities by b tag and S/B category in Figs. 8.6 through 8.11. In general, $\log(P_{zh})$ tends closer to zero for ZH events than it does for background events. Similarly, $t\bar{t}$ and Z+jets events tend to have larger values in $\log(P_{tt})$ and $\log(P_{jj})$ repectively, than events from other processes.

8.6 Two Dimensional NNs for S/B Discrimination

The two largest background classes at tag level are Z+jet events (Zbb, fakes, Zcc, mistags etc.) and $t\bar{t}$. We train two-dimensional NN's to simultaneously separate signal events from $Z + jets$ and $t\bar{t}$. The NN's are designed to return two values (NNx,NNy) for a given sample. For signal the NN targets the values (NNx=1,NNy=0), for $Z + jets$ the NN targets (NNx=0,NNy=0) and for $t\bar{t}$, the target is (NNx=1,NNy=1). We optimize three NNs (one for each tag category T,L+JP, and TT) with each NN trained on the same sample of signal ($M_H = 120\,\mathrm{GeV}/c^2$) and $t\bar{t}$ while the $Z + jets$ training samples are constructed to reflect the amount of light flavor, bb and cc events in each tag category.

We utilize a sequential input algorithm which automatically selects the most powerful discriminants as NN inputs. This algorithm begins by forming single input NNs (considering each of 41 available inputs) and finds the single input which produces the best performing (lowest testing error) NN. Once the best single input is found, the algorithm loops through the remaining pool of inputs to find the best two input NN. The algorithm continues in this way until the addition of inputs no longer improves the testing error. The results of

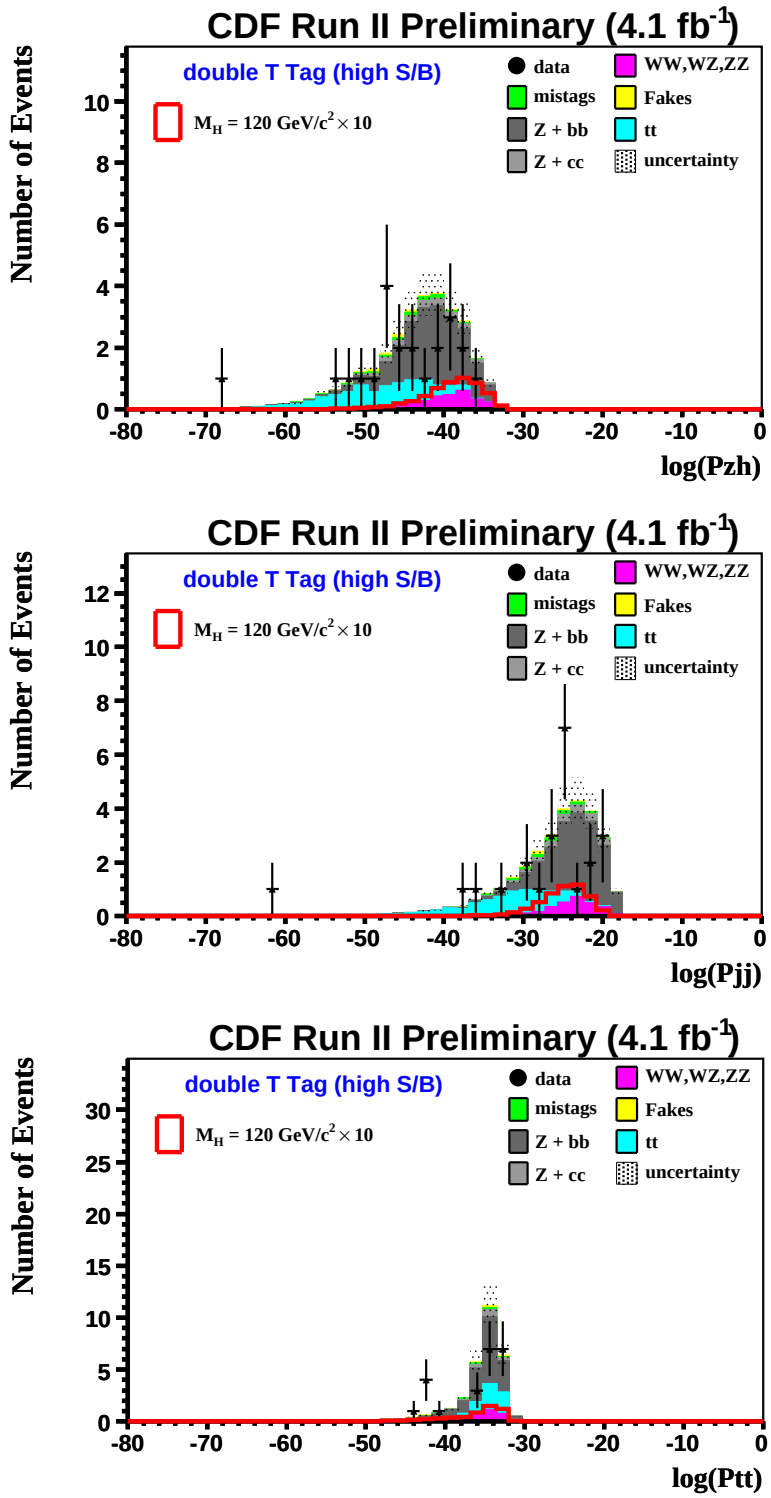


Figure 8.6: Distribution of matrix element probabilities for events in the TT tag high S/B sample.

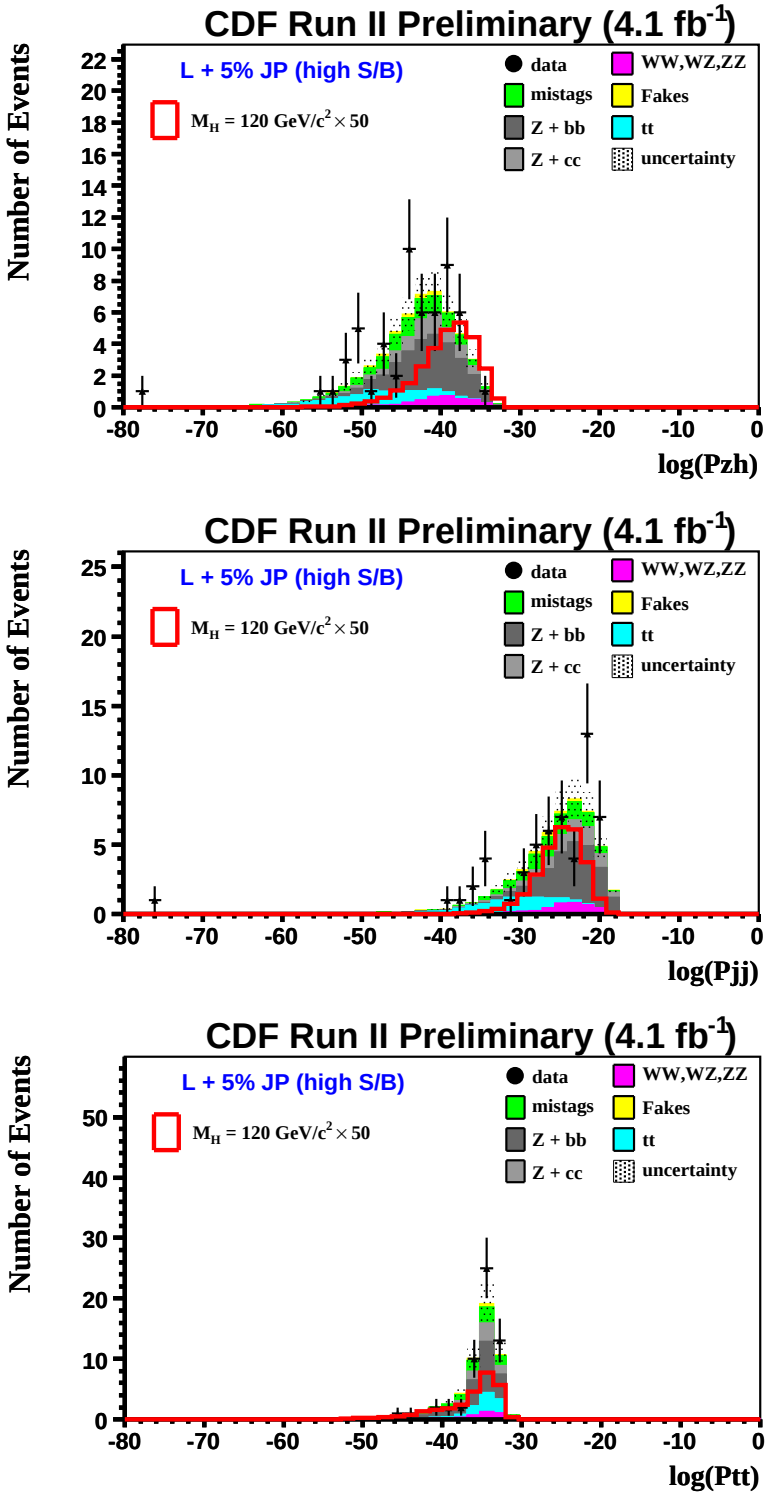


Figure 8.7: Distribution of matrix element probabilities for events in the L+JP tag high S/B sample.

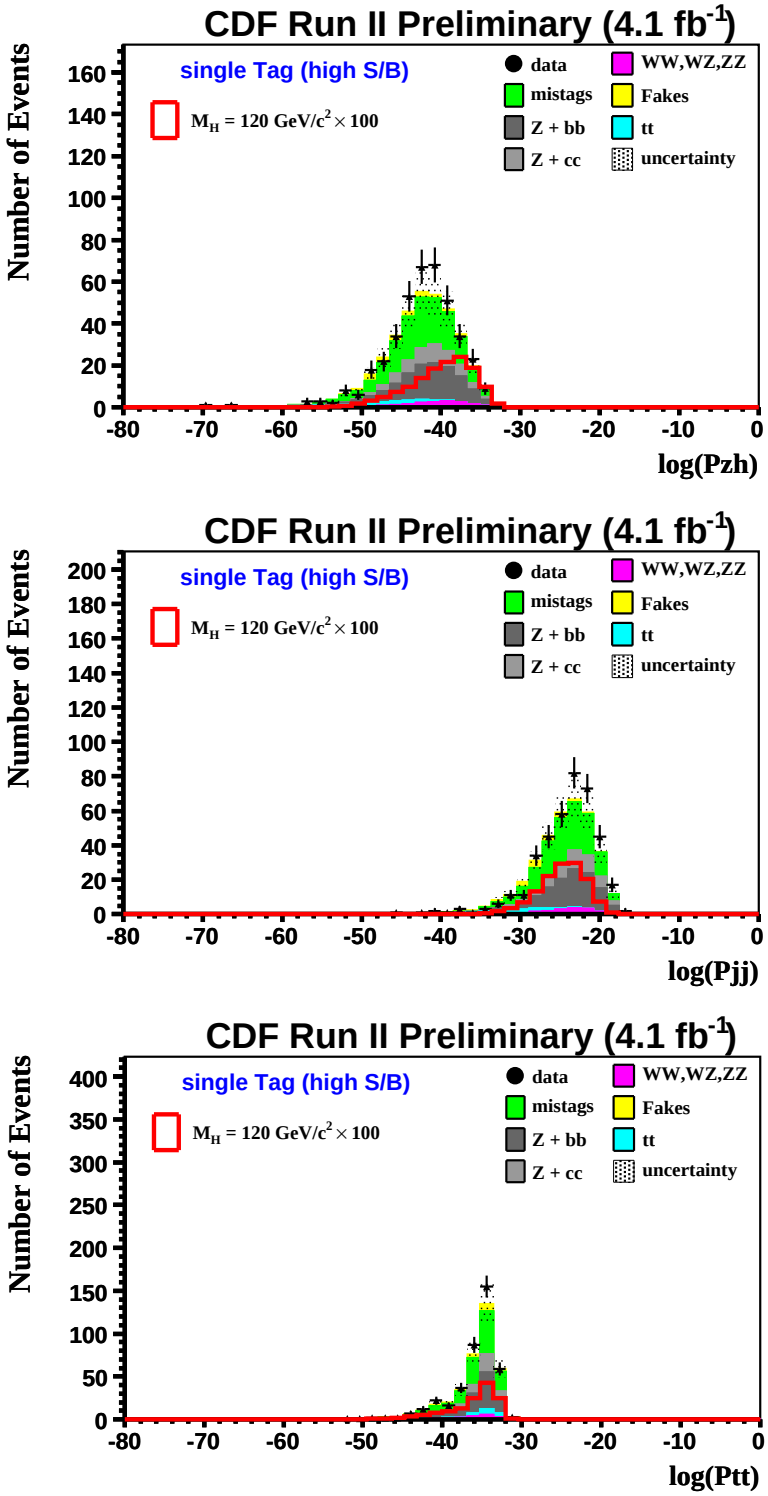


Figure 8.8: Distribution of matrix element probabilities for events in the T tag high S/B sample.

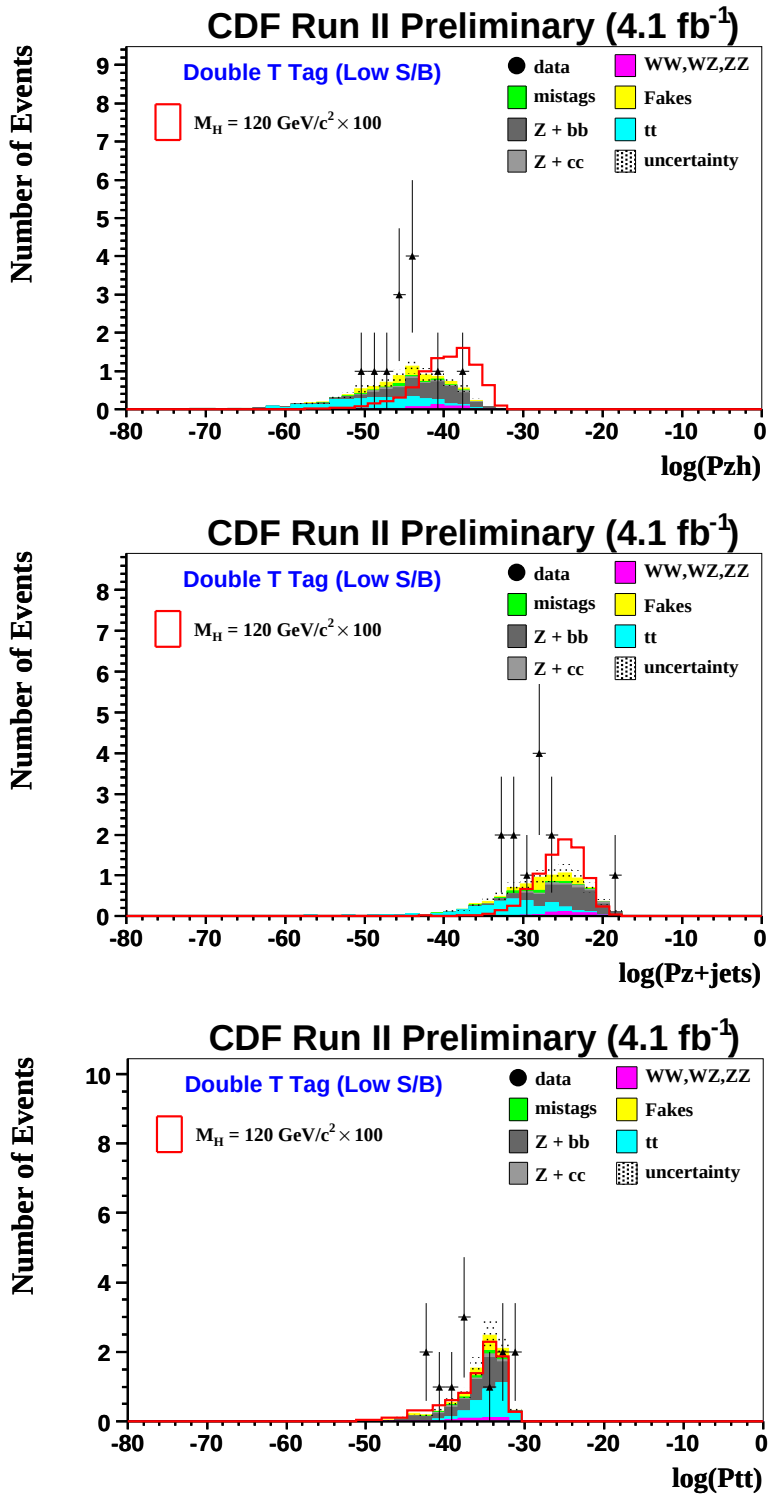


Figure 8.9: Distribution of matrix element probabilities for events in the TT tag low S/B sample.

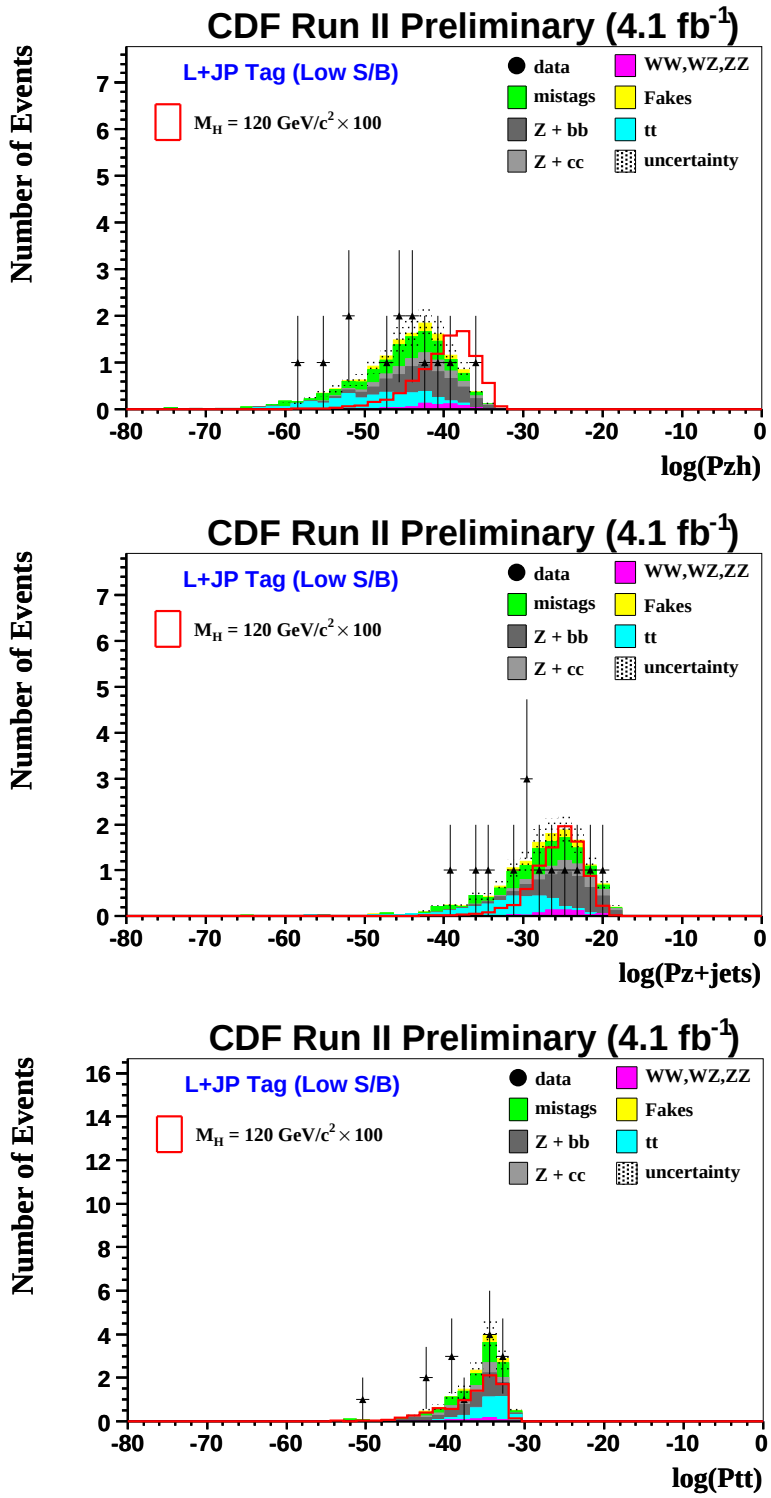


Figure 8.10: Distribution of matrix element probabilities for events in the L+JP tag low S/B sample.

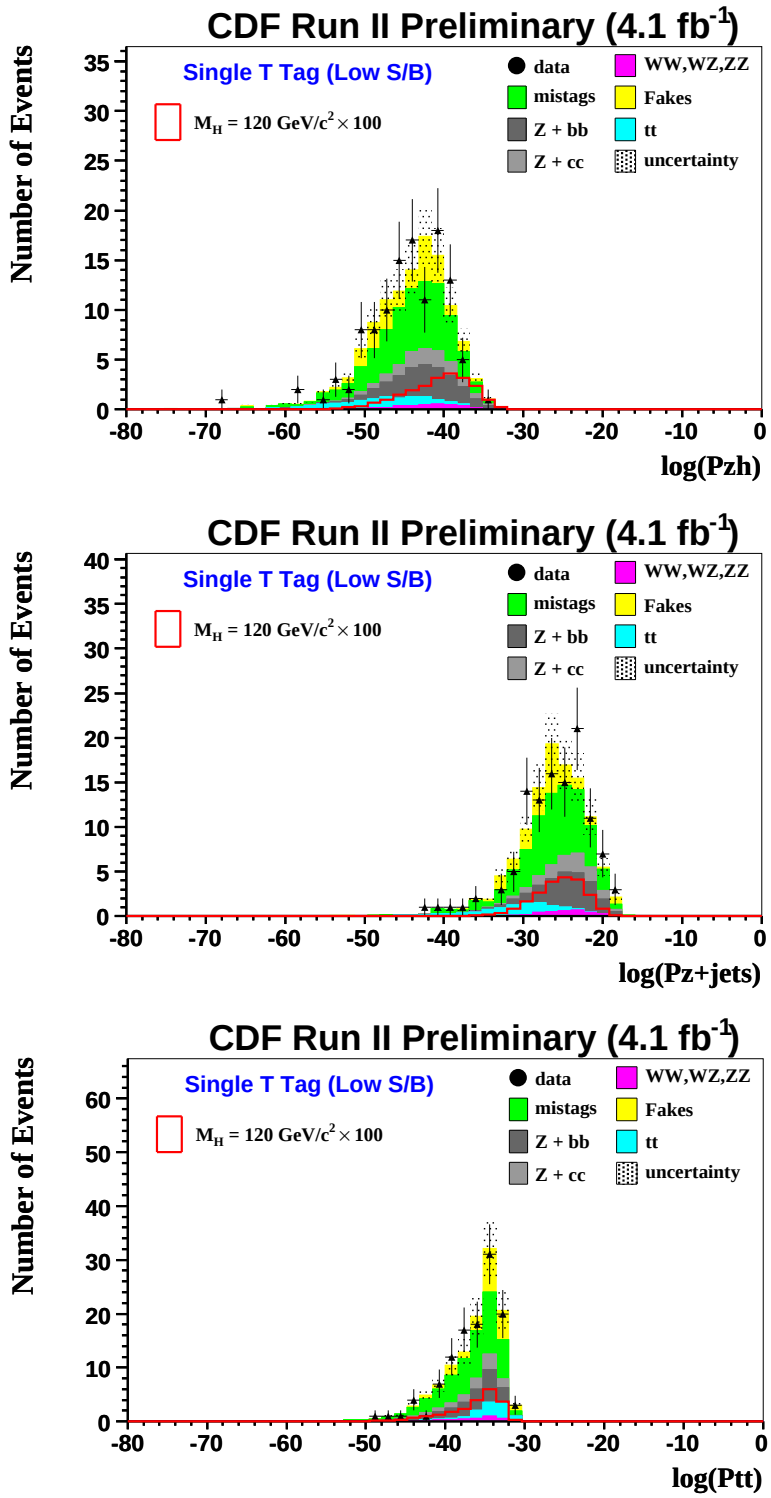


Figure 8.11: Distribution of matrix element probabilities for events in the T tag low S/B sample.

the variable selection algorithm are shown in Fig. 8.12. Once the algorithm has found the optimal inputs for each b-tag category the final NN's are trained.

We ensure that kinematic variables used as NN inputs are well described by our model at the PreTag selection level. While trained on simulated MC events, we find that the NN displays similar performance on data. We check the agreement of our data and model in selected PreTag NN inputs and NN outputs which can be seen in Figs. 8.13 through 8.18. We do not show the output of the KNN for PreTag events; it is only applicable to SevVtx tight tagged jets. Similarly, the output of the NN optimized for the single T tag events is not shown for PreTag events; it requires KNN output as an input.

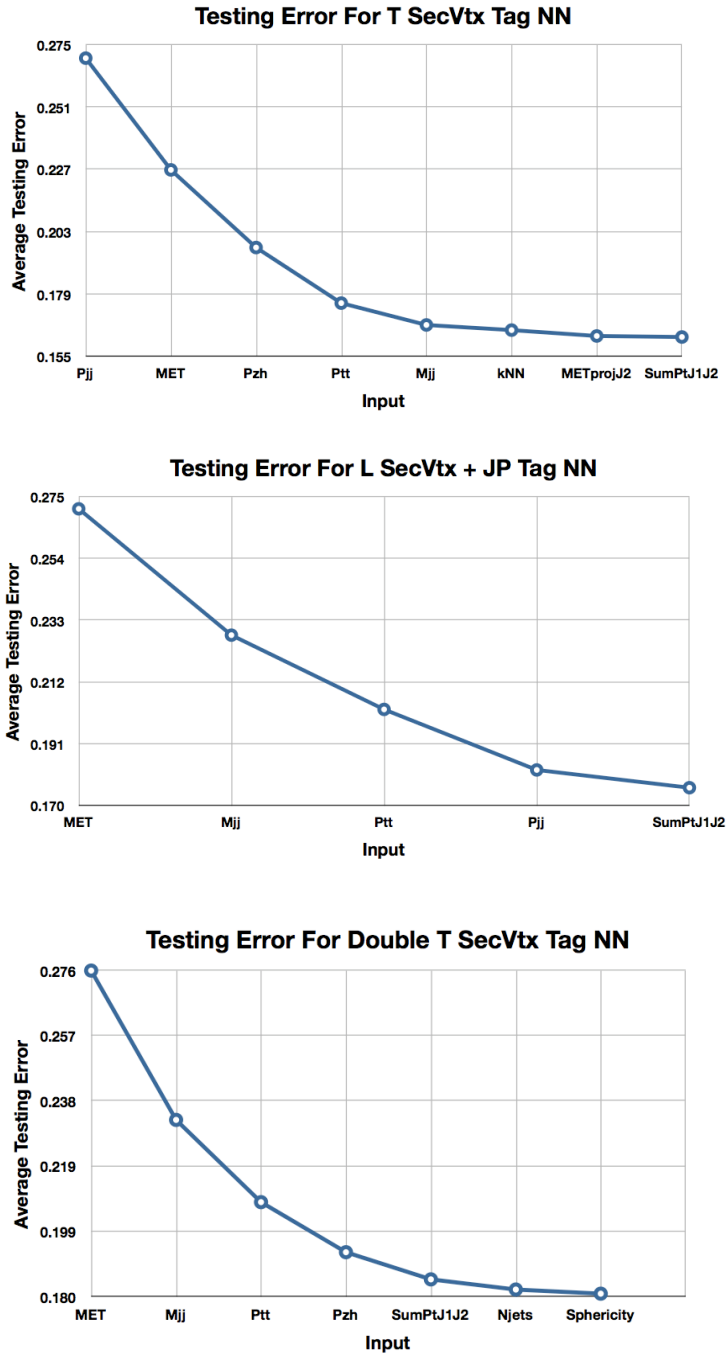


Figure 8.12: The selected NN variables, in order of selection from left to right, and the average testing error produced by their addition.

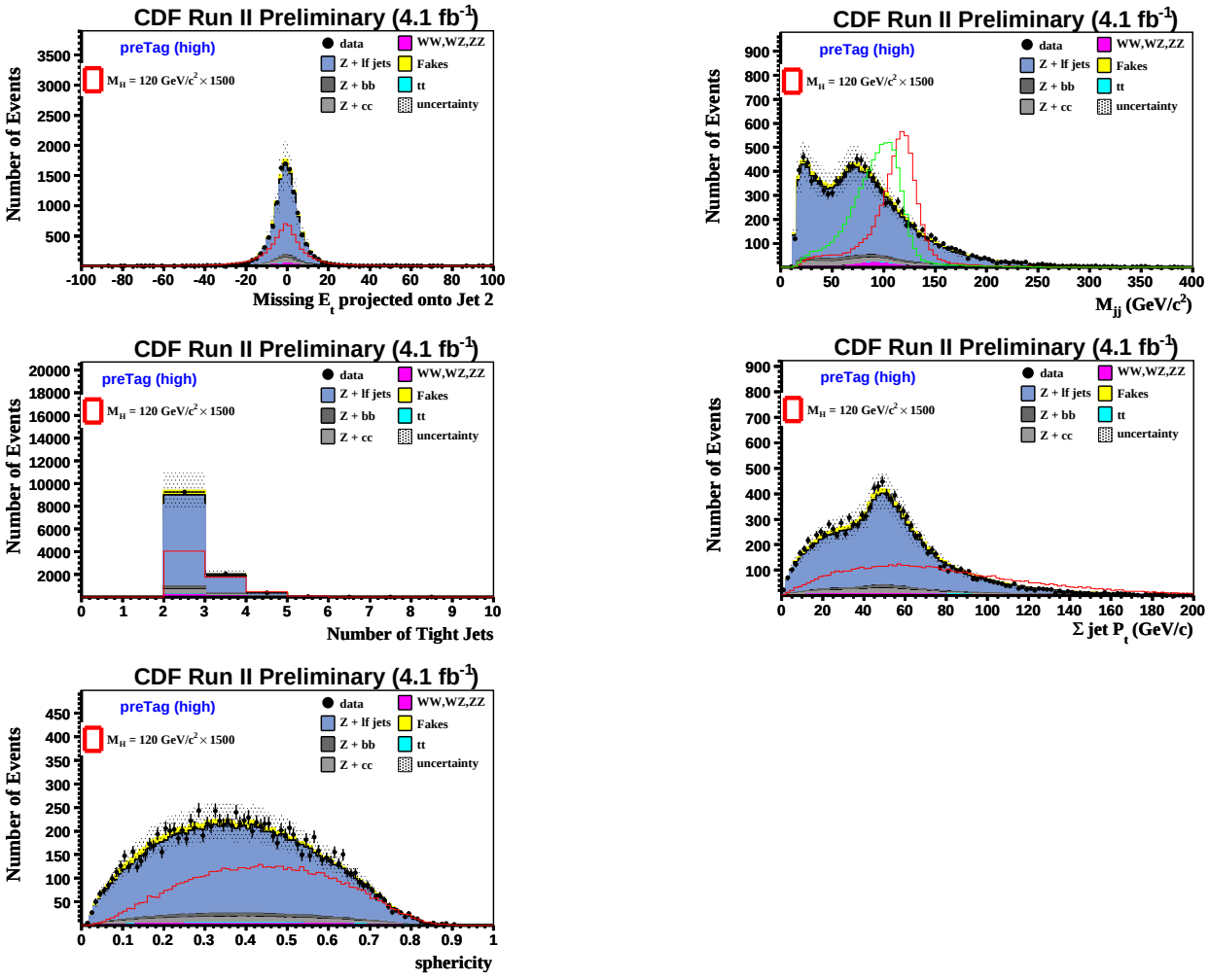


Figure 8.13: Pre-Tag high S/B NN inputs

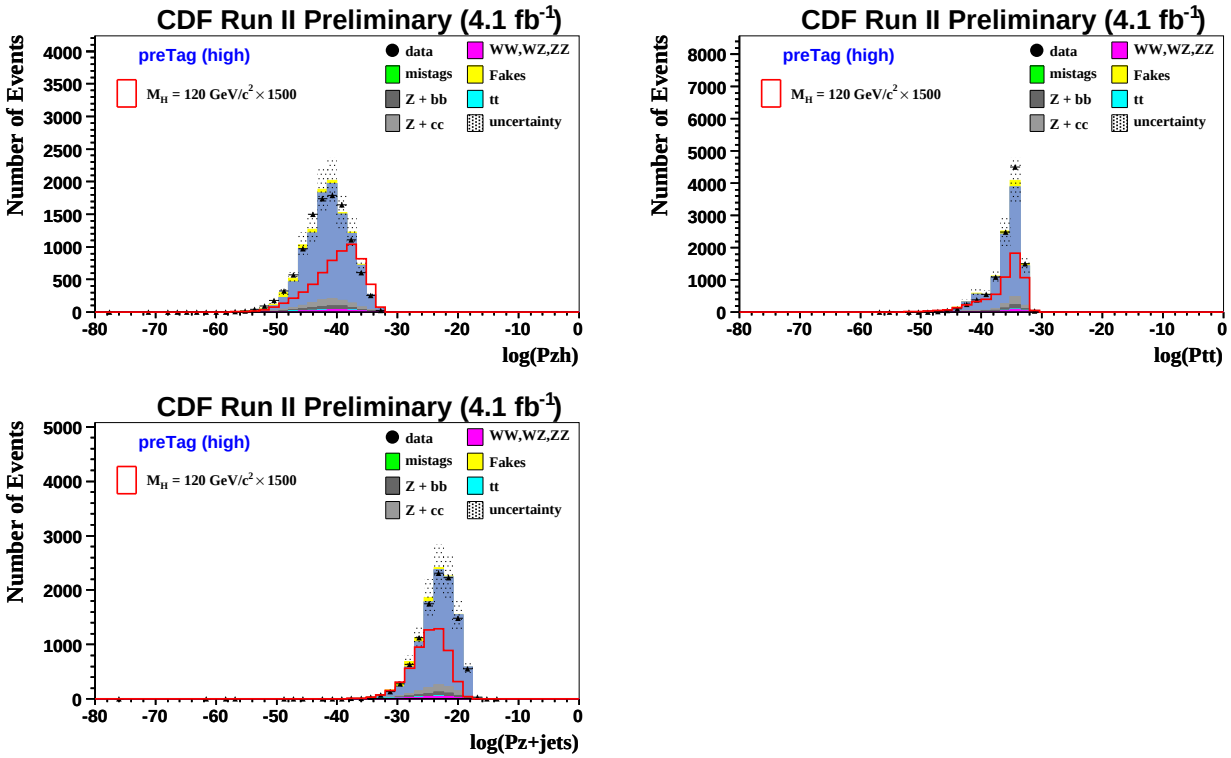


Figure 8.14: Pre-Tag high S/B NN inputs. Due to computing time required for ME calculations, a random 10% of each MC background is shown (scaled by 10).

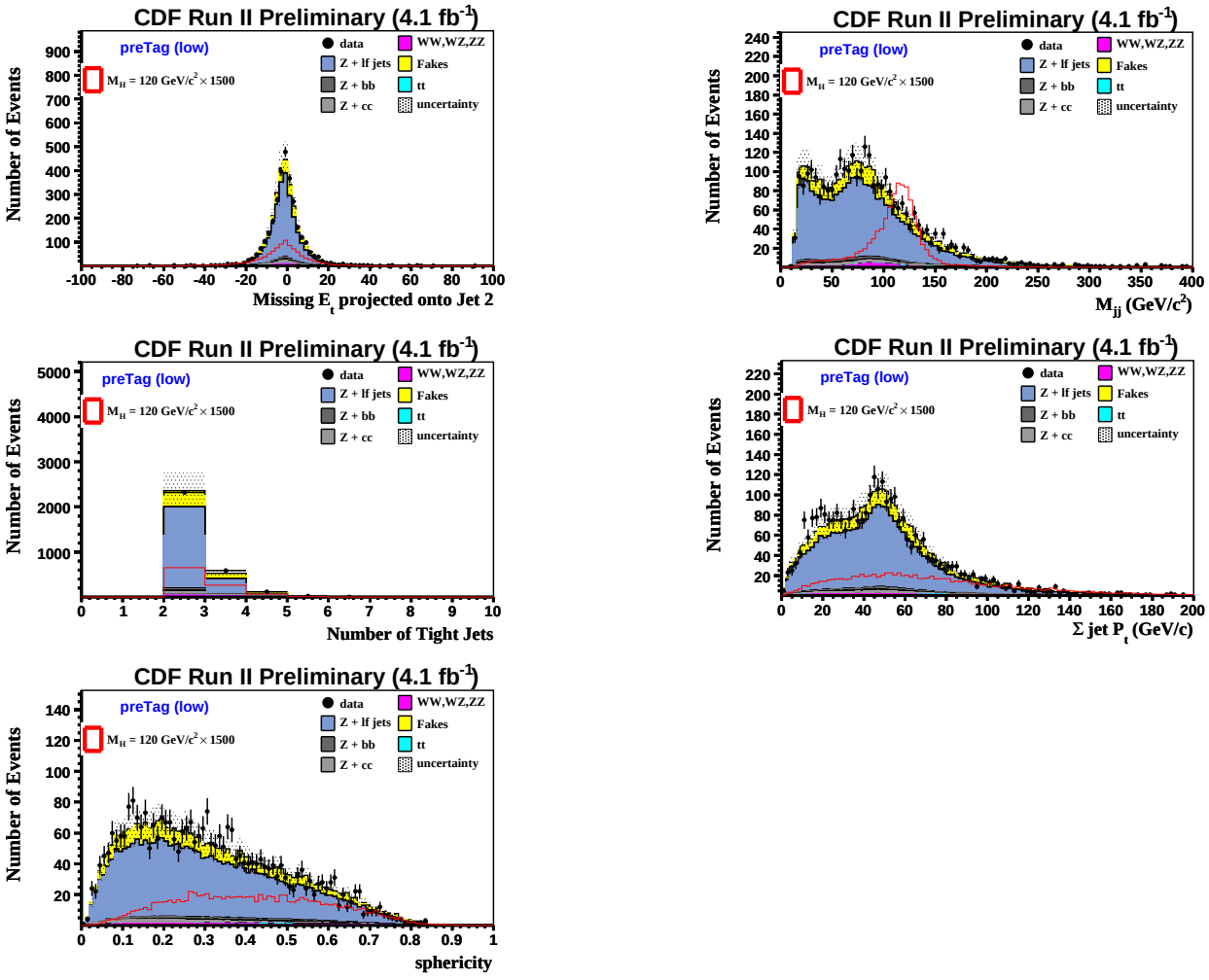


Figure 8.15: Pre-Tag low S/B NN inputs

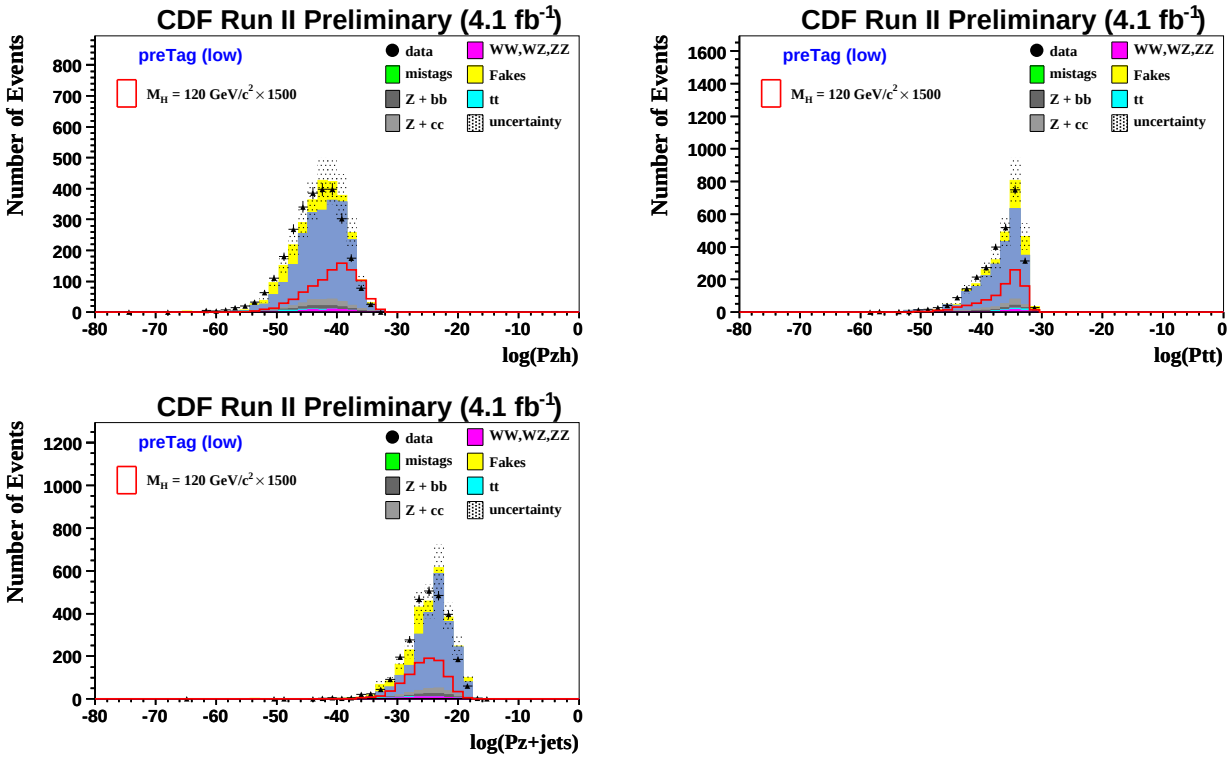


Figure 8.16: Pre-Tag low S/B NN inputs. Due to computing time required for ME calculations, a random 10% of each MC background is shown (scaled by 10).

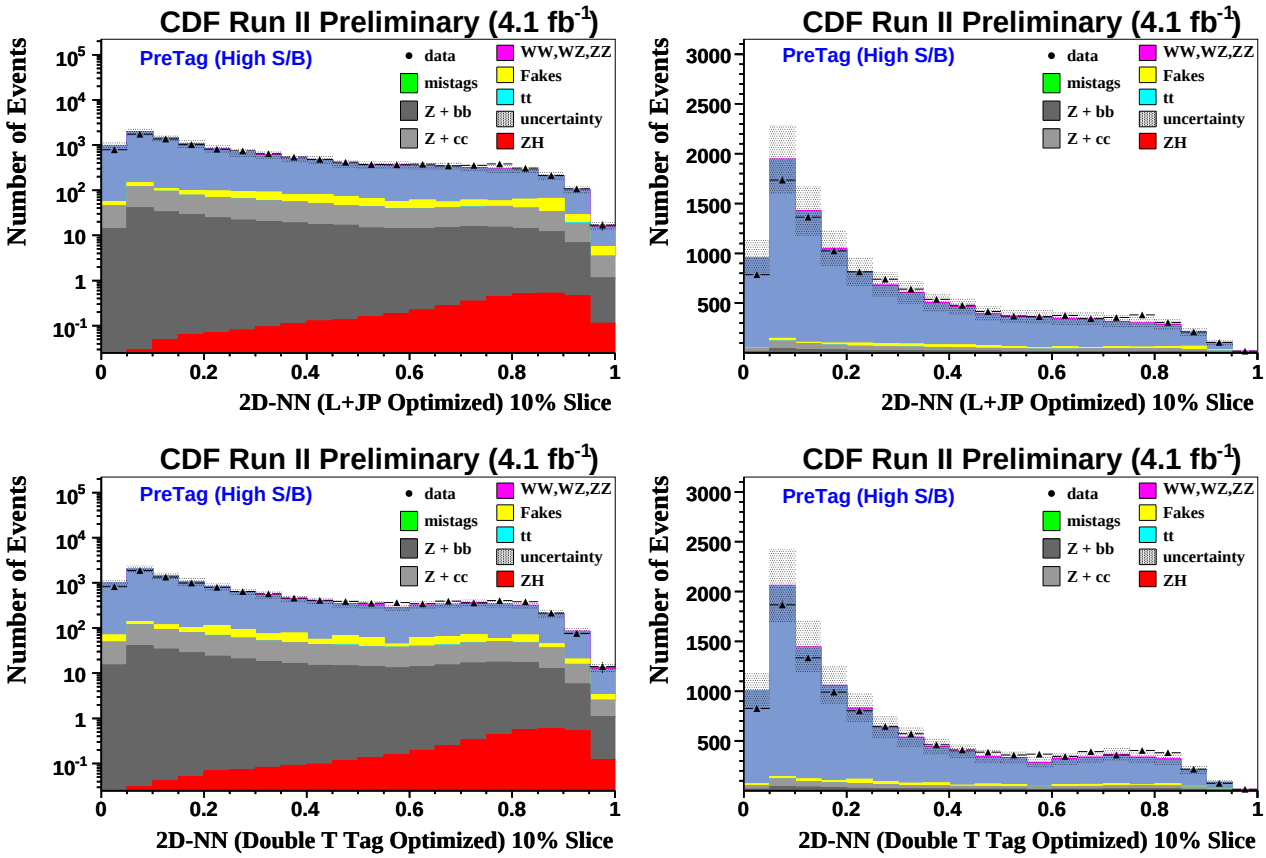


Figure 8.17: Two dimensional NN output projections for high S/B pretag data. A cut on NNy<0.1 is made to highlight the signal region. Due to computing time required for ME calculations, a random 10% of each MC background is shown (scaled by 10).

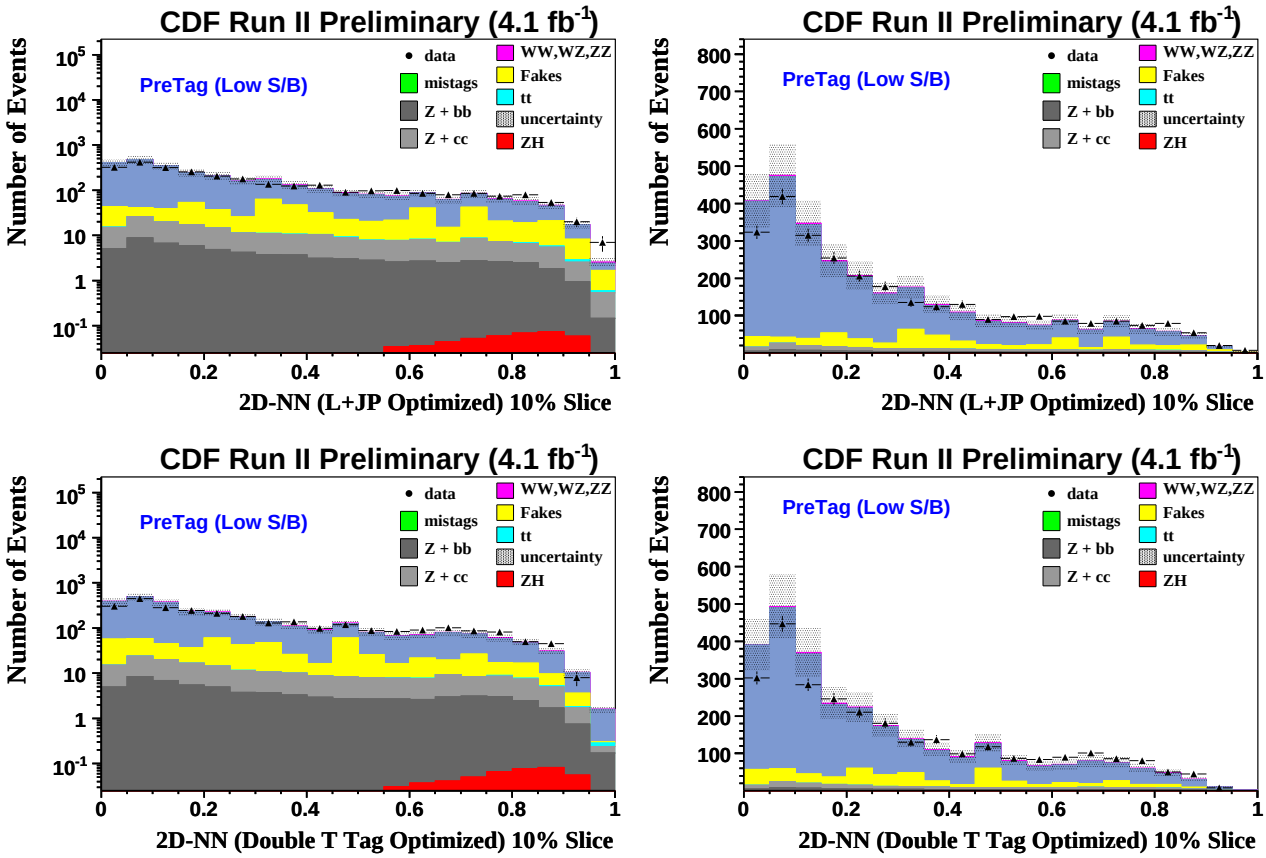


Figure 8.18: Two dimensional NN output projections for low S/B pretag data. A cut on NNy<0.1 is made to highlight the signal region. Due to computing time required for ME calculations, a random 10% of each MC background is shown (scaled by 10).

Chapter 9: Systematic Uncertainties

To identify a potential ZH signal, we compare the two dimensional NN output observed in data to the model of the output expected for the SM background. However, the NN output of our model is affected by several sources of systematic uncertainty. The effects of the uncertainties on NN output can be characterized by their affect on the NN output:

- "Rate" uncertainties change the normalization of NN outputs without affecting the shape.
- "Shape" uncertainties change the shape of the of NN output and can also affect the normalization.

To assess the agreement between the observed data and the model of SM background, or to quantify the significance of any potential excess, the NN outputs for data, signal, and each background are input to the MCLIMIT [104] program. MCLIMIT contains classes for calculation of signal significance and upper limits on the signal content in the observed data. Details of the calculation performed with MCLIMIT for this search are presented in Chapter 10.

In MCLIMIT calculations, the sources of uncertainty on the NN output, called nuisance parameters, are integrated over (utilizing a MC numerical integration method) assuming that they are Gaussian distributions. For each rate uncertainty, the central value of the SM prediction for the NN normalization is the mean of the associated Gaussian, while the standard deviation is set by a 1σ variation under the rate uncertainty. Shape uncertainties are included by providing MCLIMIT with three versions of each NN shape (template) varied under the uncertainty :

- A 'Default' shape with no systematic shift applied.
- An 'UP' shape : the given uncertainty is shifted 1σ upwards from the default shape, and all event kinematics are computed under this shift. Event selection is applied after

Table 9.1: Summary of systematic uncertainties in terms of fractional acceptance change on samples.

Systematic Uncertainty		Samples Affected
Tevatron Luminosity	0.05	All MC
CDF Luminosity	0.04	All MC
Z+h.f cross-section	0.40	$Z + b\bar{b}$, $Z + c\bar{c}$
$t\bar{t}$ cross-section	0.20	$t\bar{t}$
Diboson cross-sections	0.115	ZZ, ZW, WW
Mistag uncertainty	Histogram Shape & Acceptance	Mistags
Trigger uncertainty	0.01	All MC
Lepton Reconstruction	0.01	All MC
b-tag scale factor	0.04	All single tag MC (T)
	0.08	All double SecVtx tag MC (TT)
	0.11	All Loose + JP tag MC (L+JP)
Fakes	0.50	Fake ee, $\mu\mu$
JES	Histogram Shape & Acceptance	All MC
ISR & FSR	Histogram Shape & Acceptance	Signal MC
ZH cross-section	0.05	Signal MC
lepton energy/momentum	0.015	All MC

the kinematics are shifted (allowing normalizations to change). Next, the modified kinematic distributions are fed into our 2D-NNs producing output shapes that carry the effect of the 1σ upward shift.

- A 'Down' shape : the given uncertainty is shifted 1σ down from the default shape.

The mean of the Gaussian assumed for a given shape nuisance parameter is derived from the normalization of the default shape, while the standard deviation is derived from the normalizations of the up and down shapes.

The following sections discuss the sources of rate and shape uncertainties. Table 9.1 summarizes the systematic uncertainties applied in our limit calculations.

9.1 Rate Uncertainties

Following Joint Physics Group [105] recommendations, we assume a 5% uncertainty on the integrated luminosity (the "Tevatron Luminosity" systematic) to cover the uncertainty on the total inelastic $p\bar{p}$ cross-section. An additional 4% systematic is applied to cover the uncertainty in the CDF luminosity measurement (the "CDF Luminosity" systematic).

We apply uncertainties of 4% (T), 8% (TT) and 11% (L+JP) to our b-tagged MC samples to account for the systematic errors associated with the calculation of b-tag efficiencies and scale factors. A 1% rate systematic is applied to all MC samples to cover the uncertainty in measured trigger efficiencies. Similarly, an uncertainty of 1% is assigned for discrepancies in lepton reconstruction efficiency between data and MC. An additional uncertainty of 1.5% is applied to MC normalization to account for the effect of lepton energy or momentum measurement on selection efficiency.

In order to cover the spread of fake rates measured from different jet triggered data samples we assign a 50% uncertainty on our total fake estimate. We apply a 40% uncertainty to $Z + bb$ and $Z + cc$ samples to cover the theoretical uncertainty on the Z+heavy flavor jets cross-section. Similarly we apply an 11.5% cross-section uncertainty to all diboson samples. For $t\bar{t}$ samples we include a conservative 20% uncertainty to account for both the theoretical uncertainty on the process cross-section and the difference between our simulated top mass ($175\,\mathrm{GeV}/c^2$) and current experimental measurements. For signal MC, we apply a 5% normalization systematic to cover theoretical cross-section uncertainty. The signal cross-section uncertainty is applied only when presenting limits in relation to SM cross-section $\times$ branching ratio. Limits presented in picobarns are calculated without the 5% uncertainty.

9.2 Uncertainties Affecting Normalization and NN Output Shapes

Uncertainties which affect the value of measured quantities or which affect the weight applied to a given event can result in uncertainties on the shape of NN output distributions. We include uncertainties on the jet energy scale (JES), the amount of initial and final state radiation (ISR/FSR) and the mistag event weighting as shape uncertainties in our limit calculation.

The JES-varied NN outputs are produced by shifting the Joint Physics energy correction factor applied to the lead and second E_T jets by $\pm 1\sigma$. This shift can cause a given event to migrate into or out of our final event sample, for example a low E_T jet can pass the jet E_T cut under a $+1\sigma$ correction shift, causing this systematic to also affect normalizations. The ISR/FSR uncertainties are included by generating specific MC samples with increased or decreased amounts of ISR/FSR. Following the Tevatron Higgs convention, this uncertainty is only applied to signal; although we did find the effect of applying this (as a bin by bin rate error) to all background MCs was negligible. The mistag shape uncertainties are included by shifting the assigned mistag weights by $\pm 1\sigma$ as specified by the mistag matrices. Figures 9.1 through 9.6 show one dimensional projections of 2D NN output (templates) and the effect of shape uncertainties. The low statistics observed in the WW templates do not significantly affect the MCLIMIT calculation due to the small contribution of WW to the total background model. Projections of the observed data are displayed for comparison.

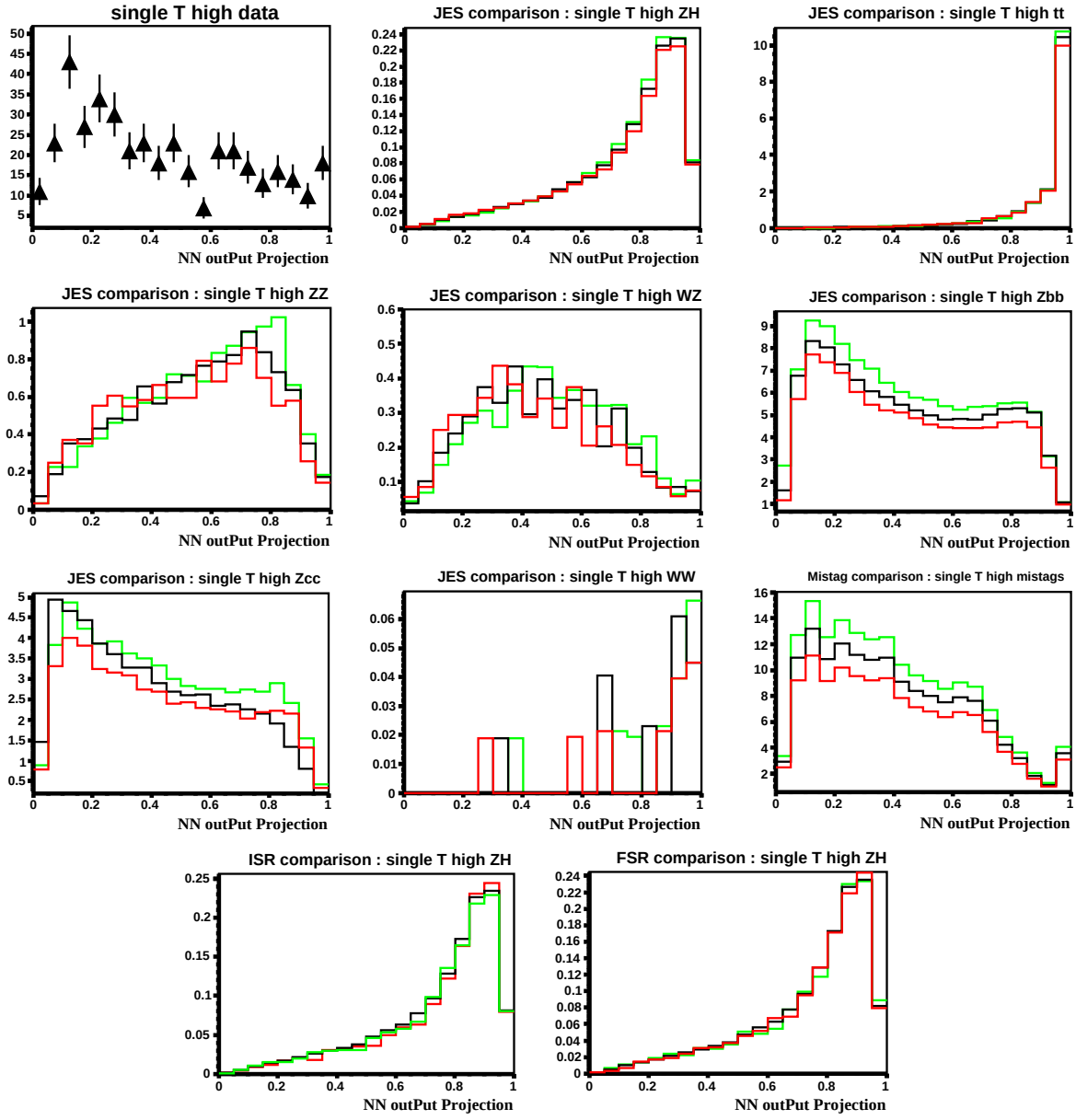


Figure 9.1: Projections of templates affected by shape uncertainties used in the calculation of limits for the single tag high S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts.

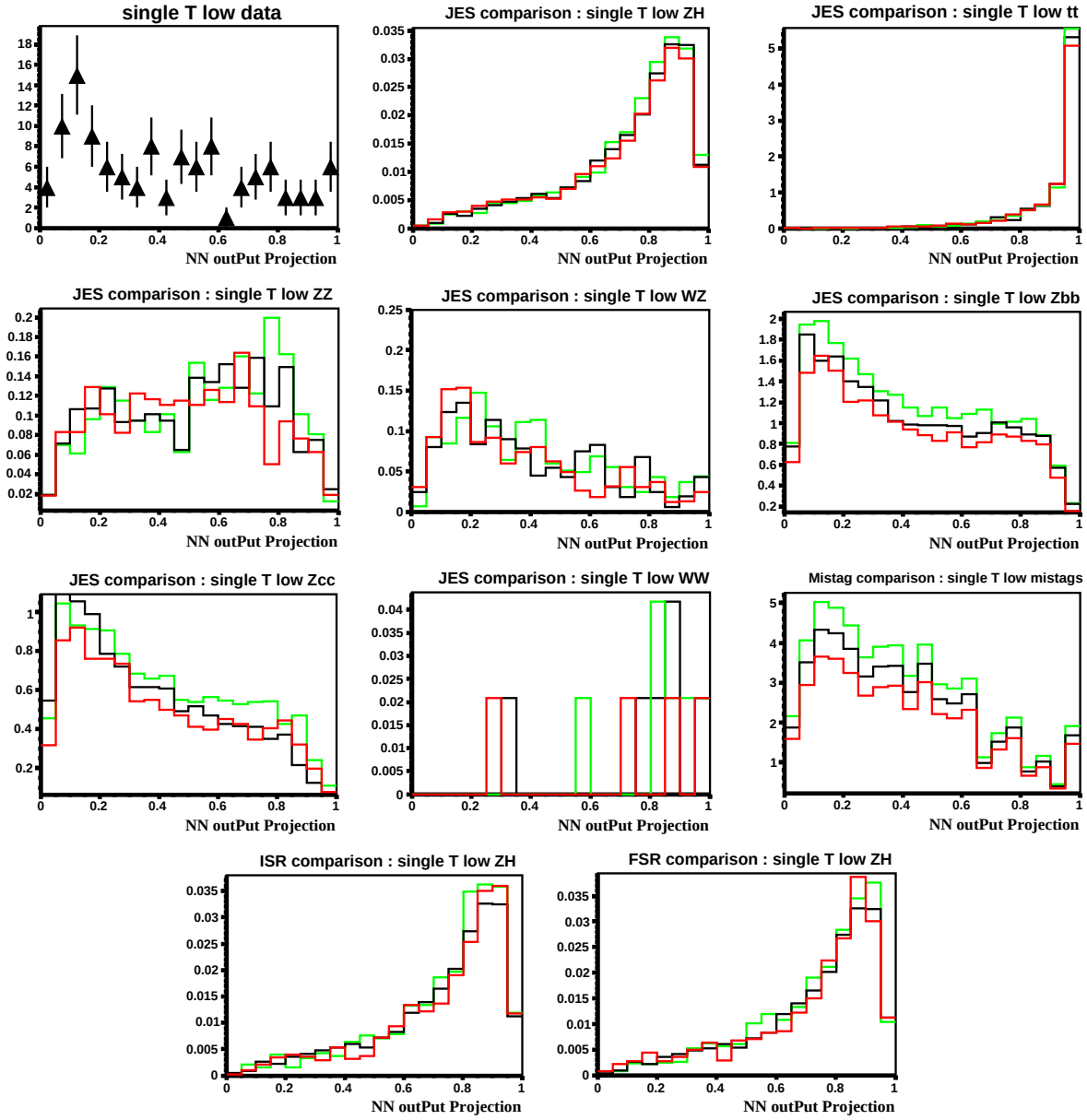


Figure 9.2: Projections of templates affected by shape uncertainties used in the calculation of limits for the single tag low S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts.

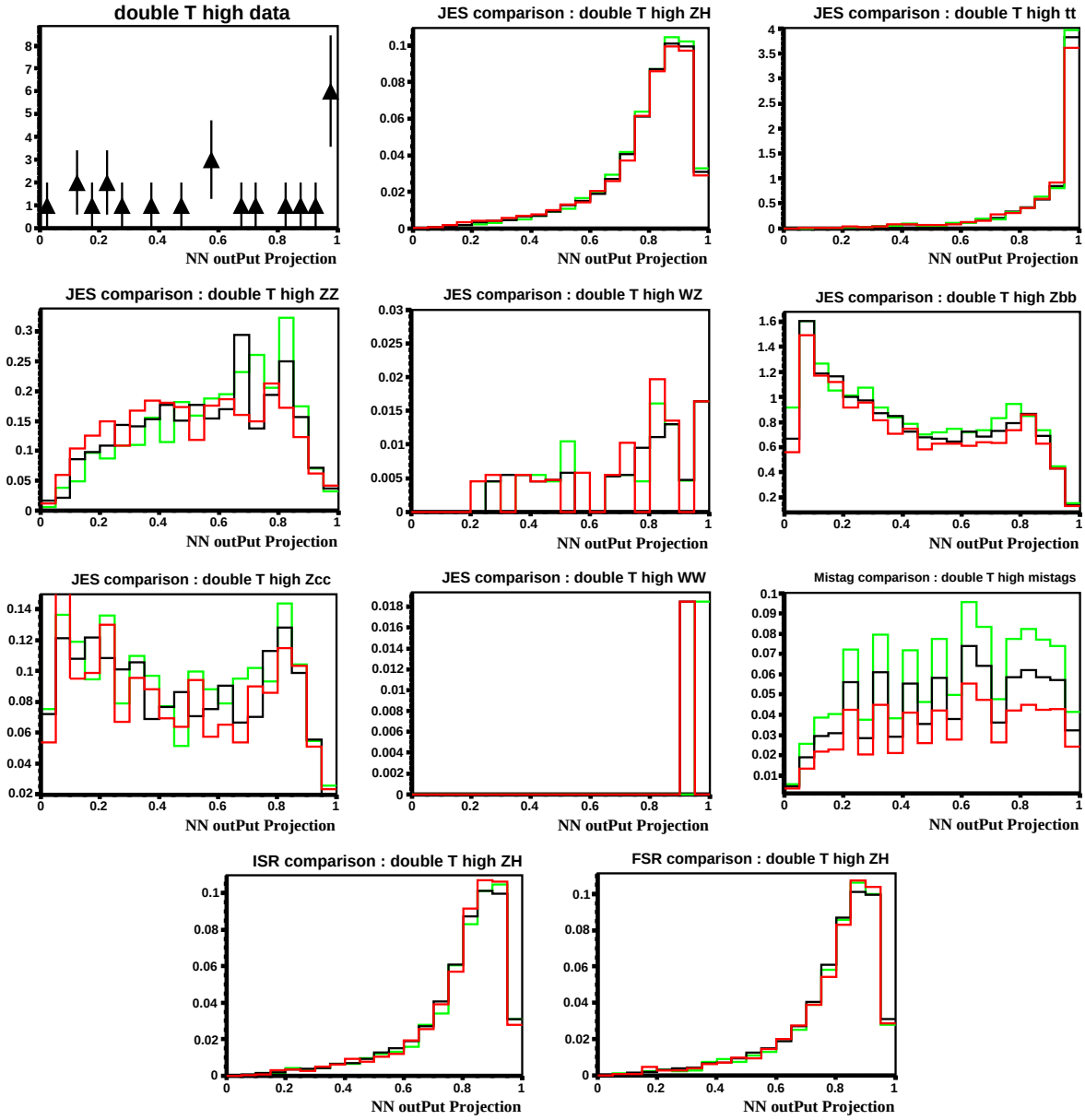


Figure 9.3: Projections of templates affected by shape uncertainties used in the calculation of limits for the double tag high S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts.

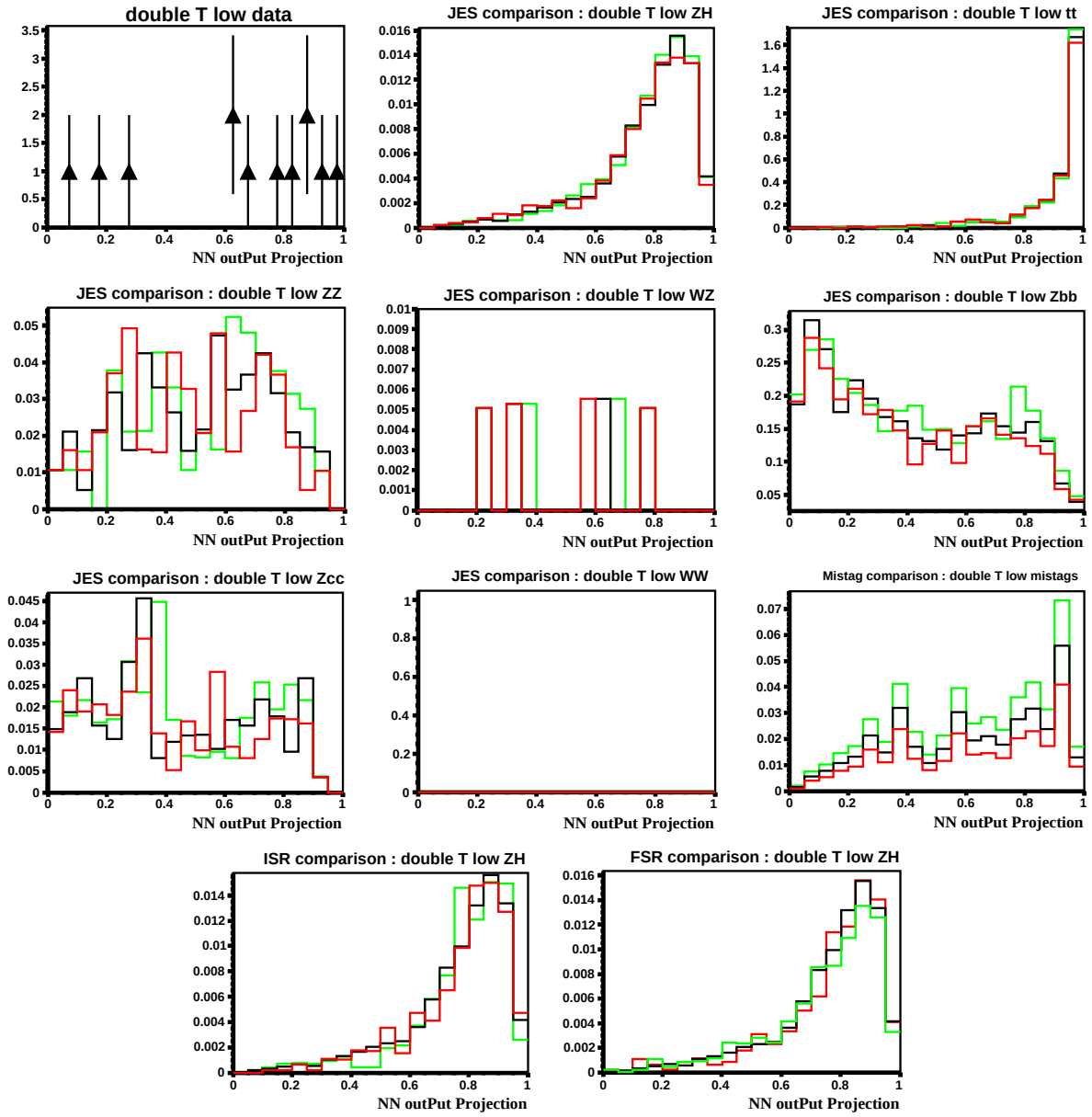


Figure 9.4: Projections of templates affected by shape uncertainties used in the calculation of limits for the double tag low S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts.

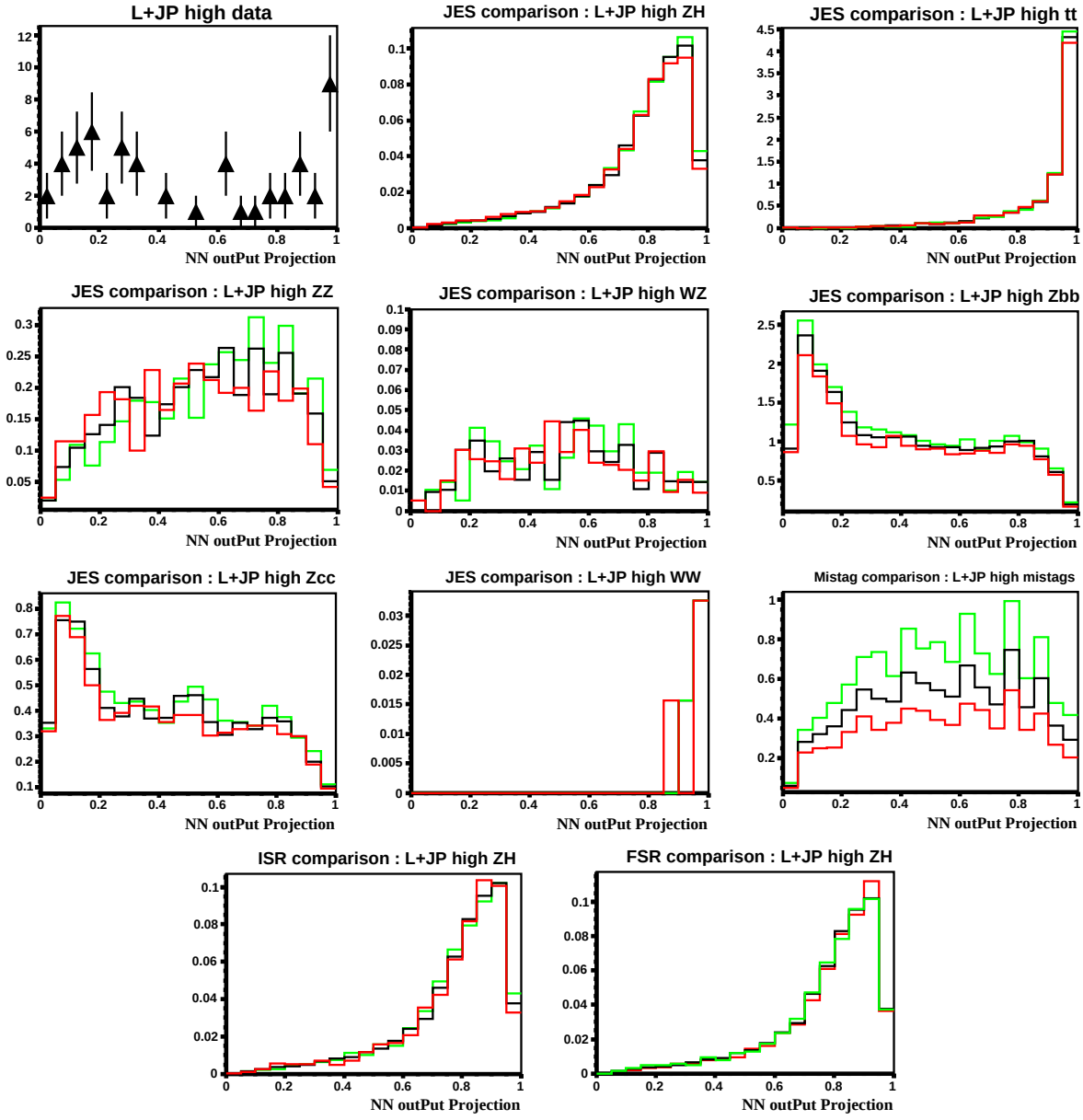


Figure 9.5: Projections of templates affected by shape uncertainties used in the calculation of limits for the L+JP tag high S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts.

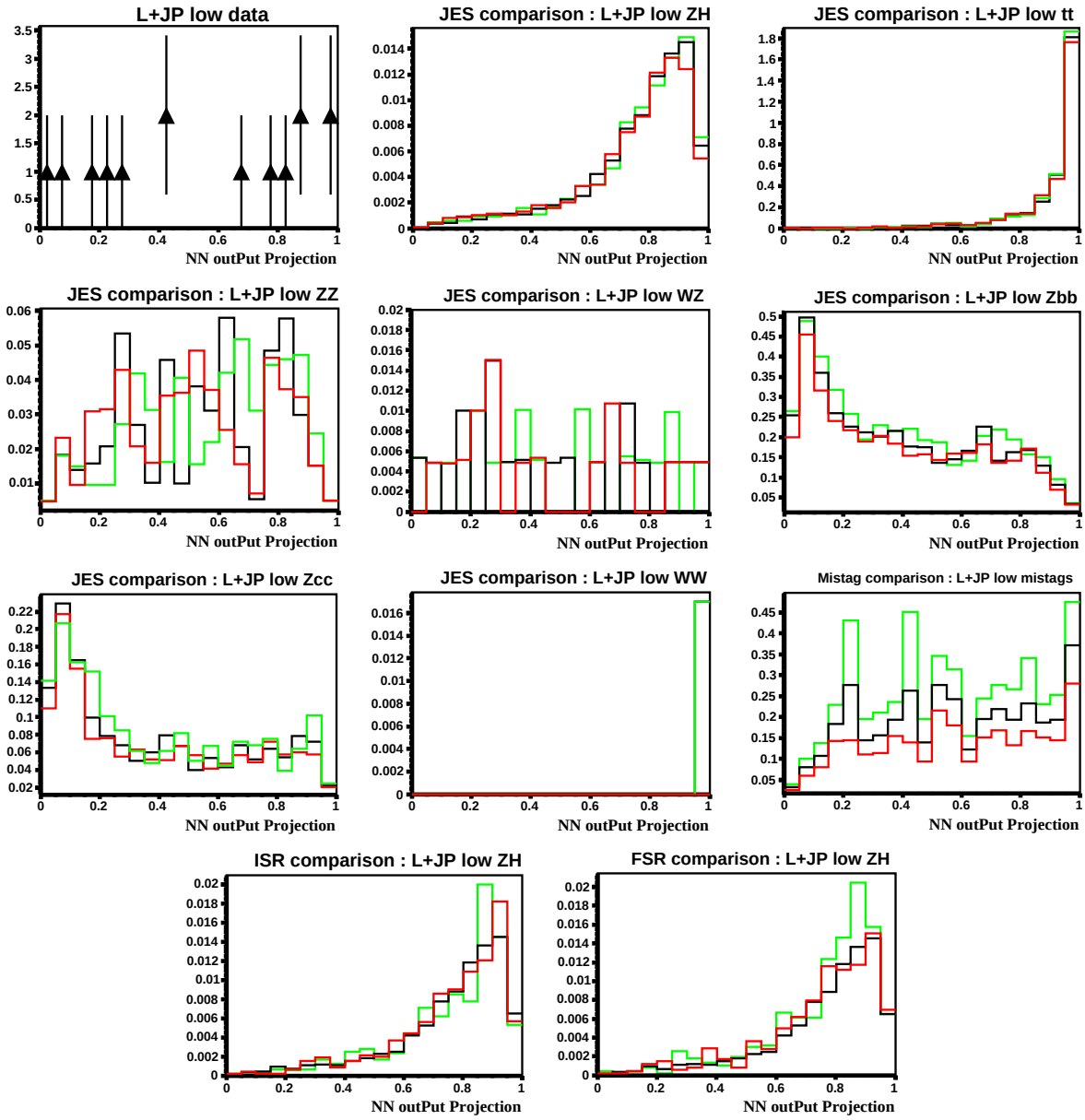


Figure 9.6: Projections of templates affected by shape uncertainties used in the calculation of limits for the L+JP tag low S/B category. Green lines show $+1\sigma$ shifted templates, while red lines show -1σ shifts.

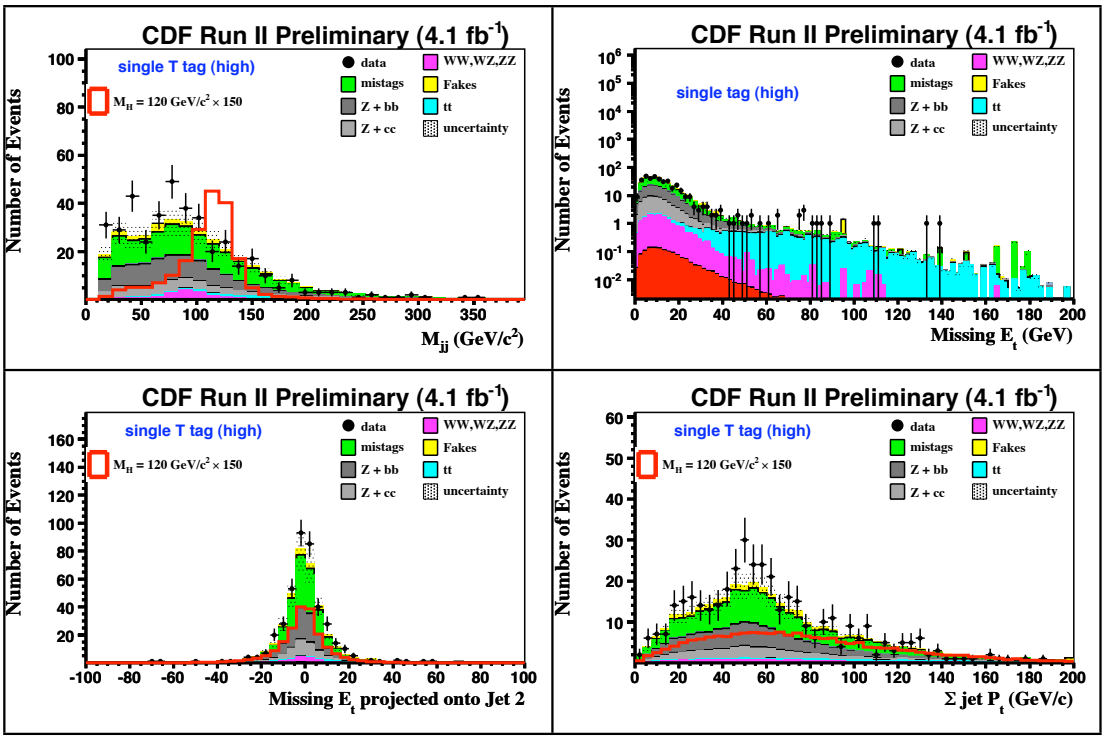
Chapter 10: Results

10.1 Results

Applying the b-tag selection to the data and background samples yields the final event samples. Tables 10.1 and 10.2 show the expected and observed number of events in each our final analysis channels. Figures 10.1 through 10.6 show the NN input distributions. Figures 10.7 through 10.18 show the full NN output, while Figs. 10.19 and 10.20 shows NN output projections.

Table 10.1: Comparison of observed and predicted event totals for tag level high S/B selection.

	(High S/B Categories)		
Source	Double T Tag	L+JP Tag	Single T Tag
$t\bar{t}$	7.0 ± 1.5	8.1 ± 1.9	17.3 ± 3.6
WW	0.02 ± 0.003	0.1 ± 0.01	0.2 ± 0.03
WZ	0.1 ± 0.01	0.5 ± 0.1	4.8 ± 0.7
ZZ	2.7 ± 0.4	3.4 ± 0.6	11.1 ± 1.5
Z+jets ($b\bar{b}$)	16.1 ± 6.8	21.5 ± 9.2	105.4 ± 44.3
Z+jets ($c\bar{c}$)	1.8 ± 0.7	8.0 ± 3.3	53.7 ± 22.6
Z+Mistags	0.9 ± 0.3	9.4 ± 3.2	151.6 ± 22.7
fakes	0.7 ± 0.3	1.8 ± 0.9	22.0 ± 11.0
ZH ($120\,\mathrm{GeV}/c^2$)	0.5 ± 0.1	0.6 ± 0.1	1.4 ± 0.1
Total Background	29.3 ± 7.0	52.8 ± 10.5	366.1 ± 55.9
Data	23	56	406



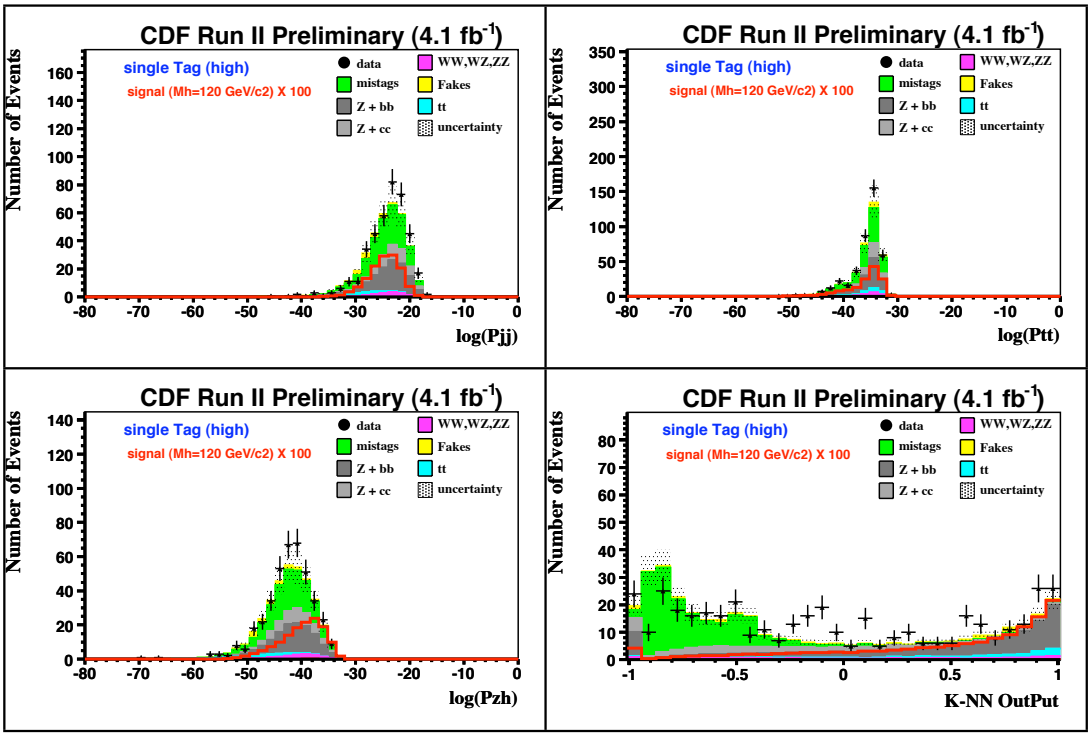
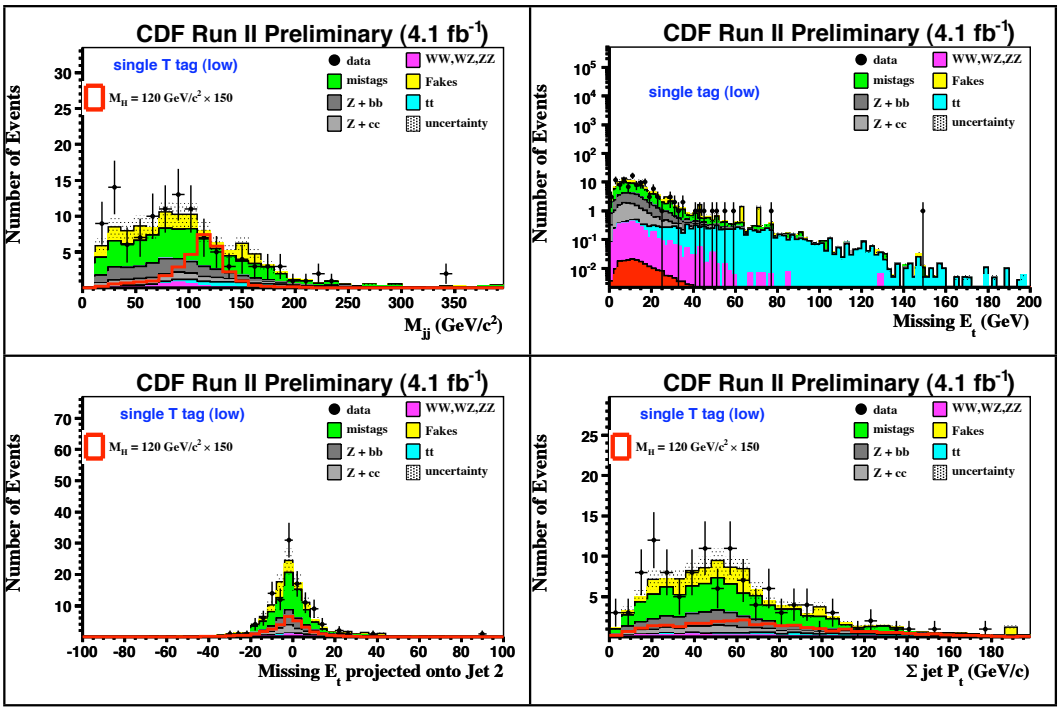


Figure 10.1: NN input distributions for the single tag high S/B channel.



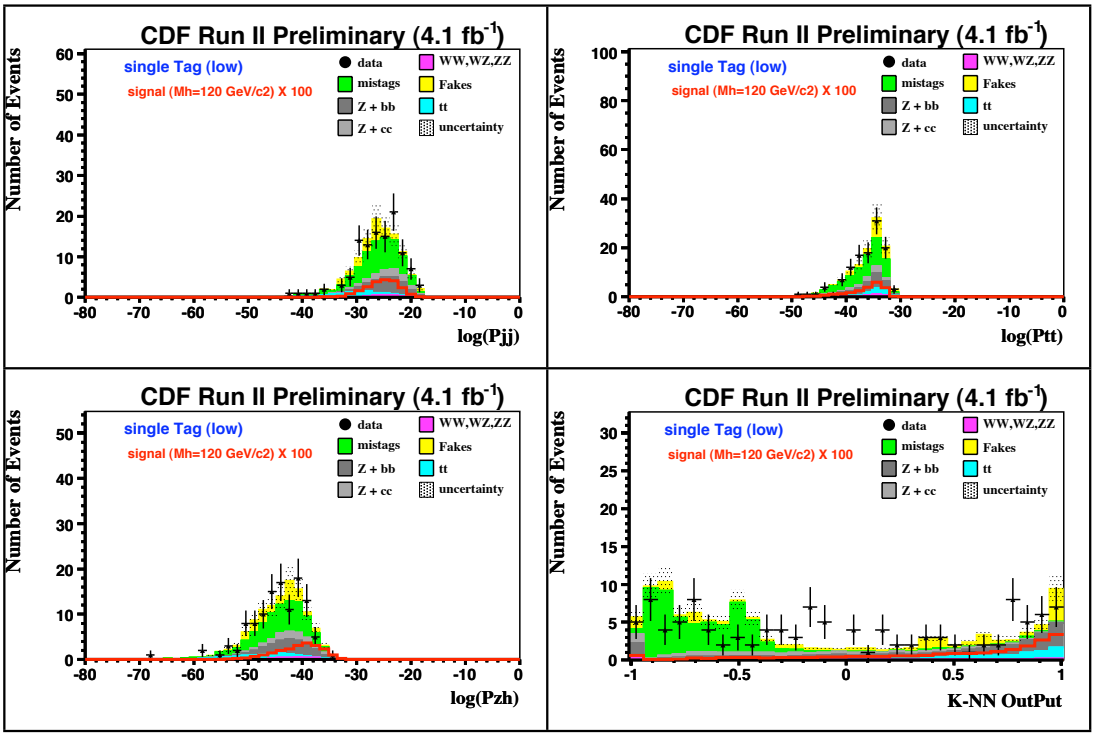
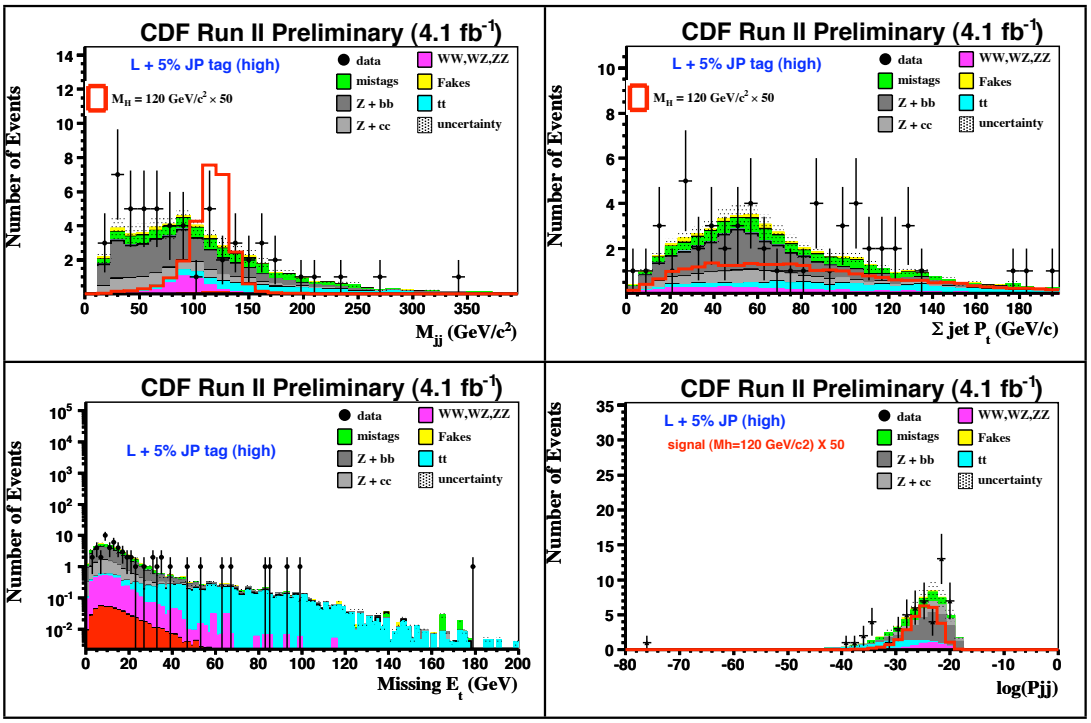


Figure 10.2: NN input distributions for the single tag low S/B channel.



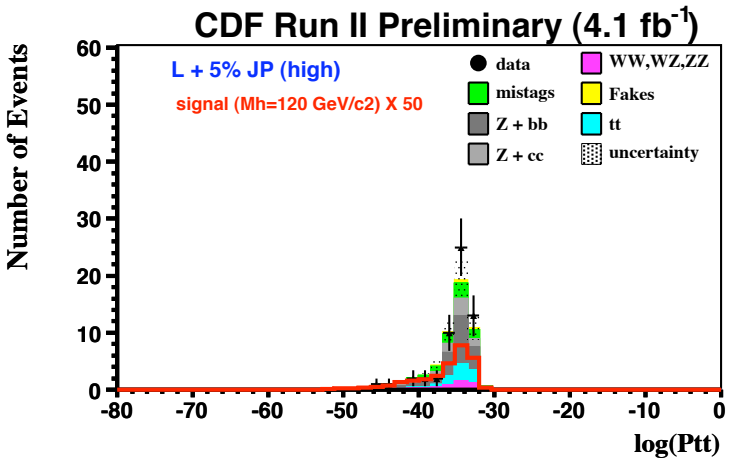
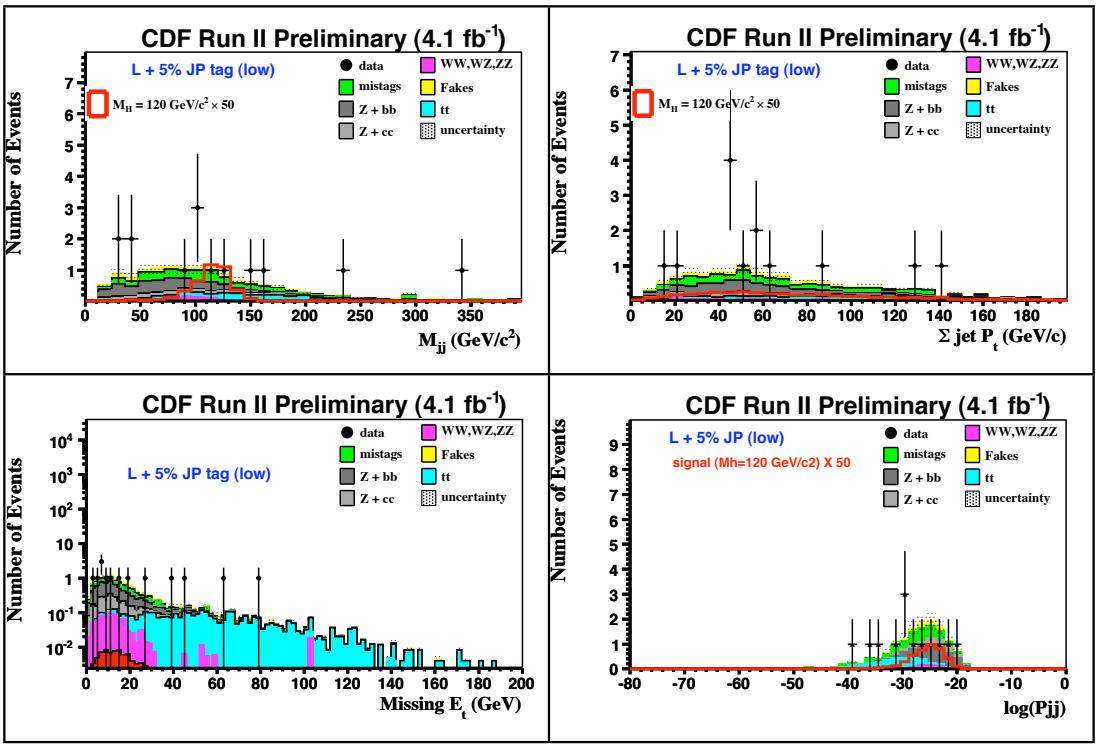


Figure 10.3: NN input distributions for the L+JP tag high S/B channel.



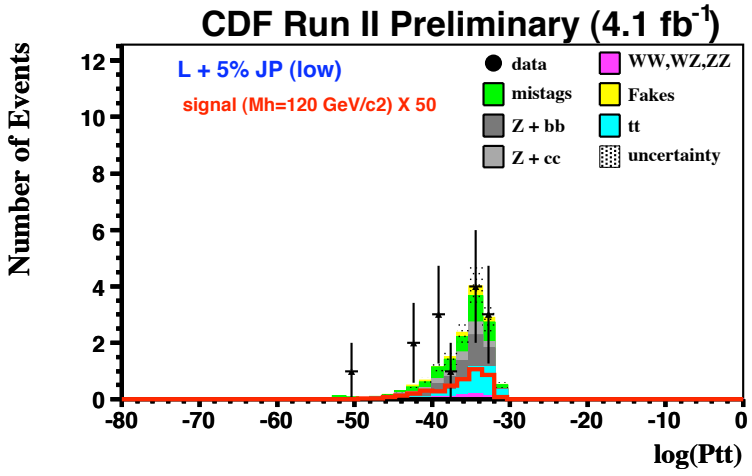


Figure 10.4: NN input distributions for the L+JP tag low S/B channel.

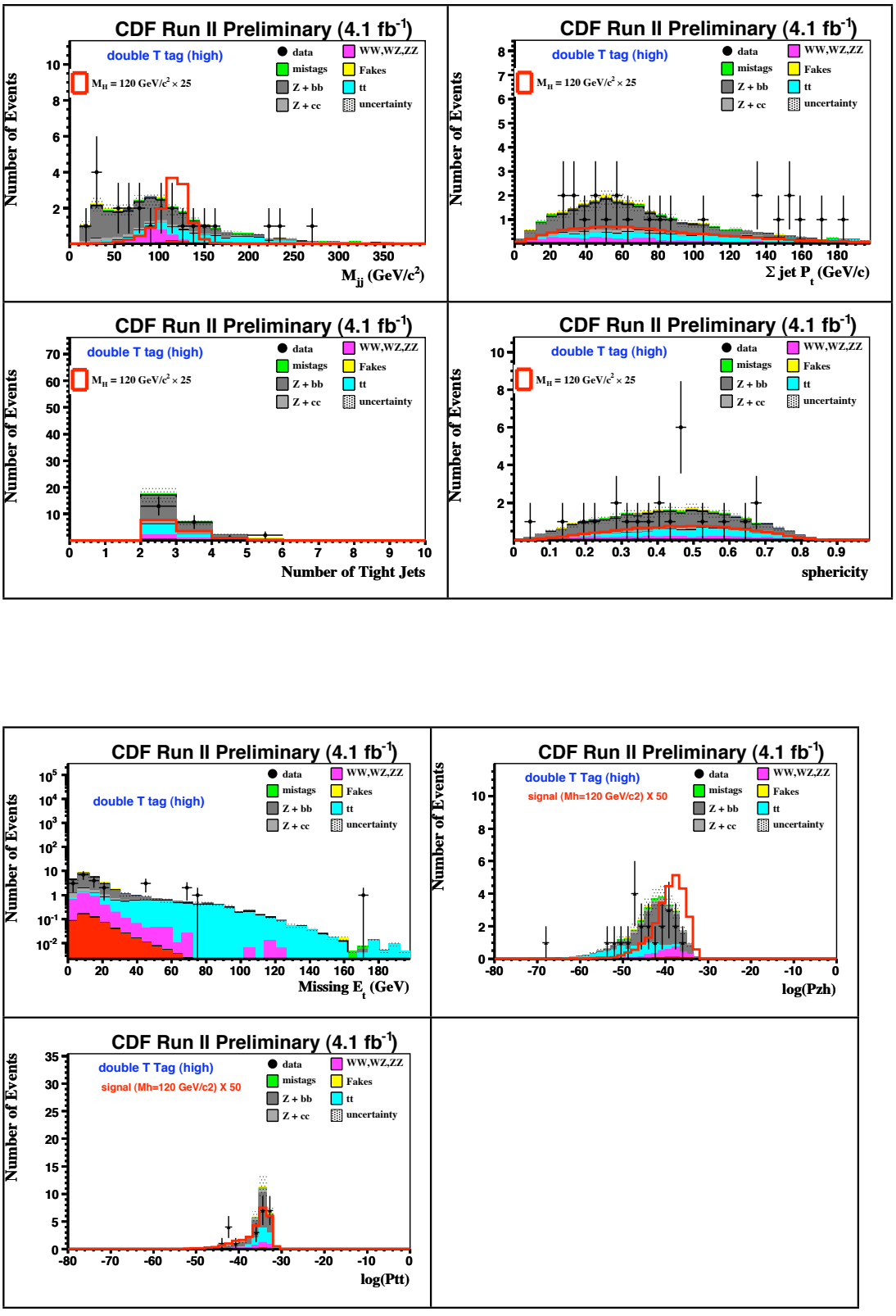
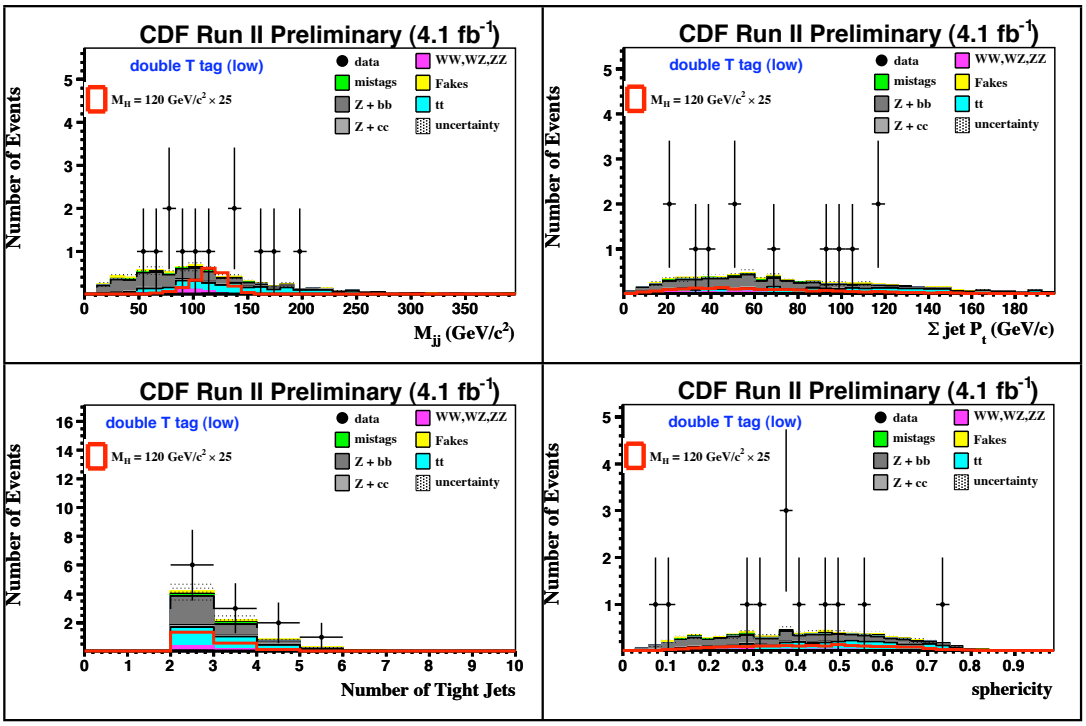


Figure 10.5: NN input distributions for the double tag high S/B channel.



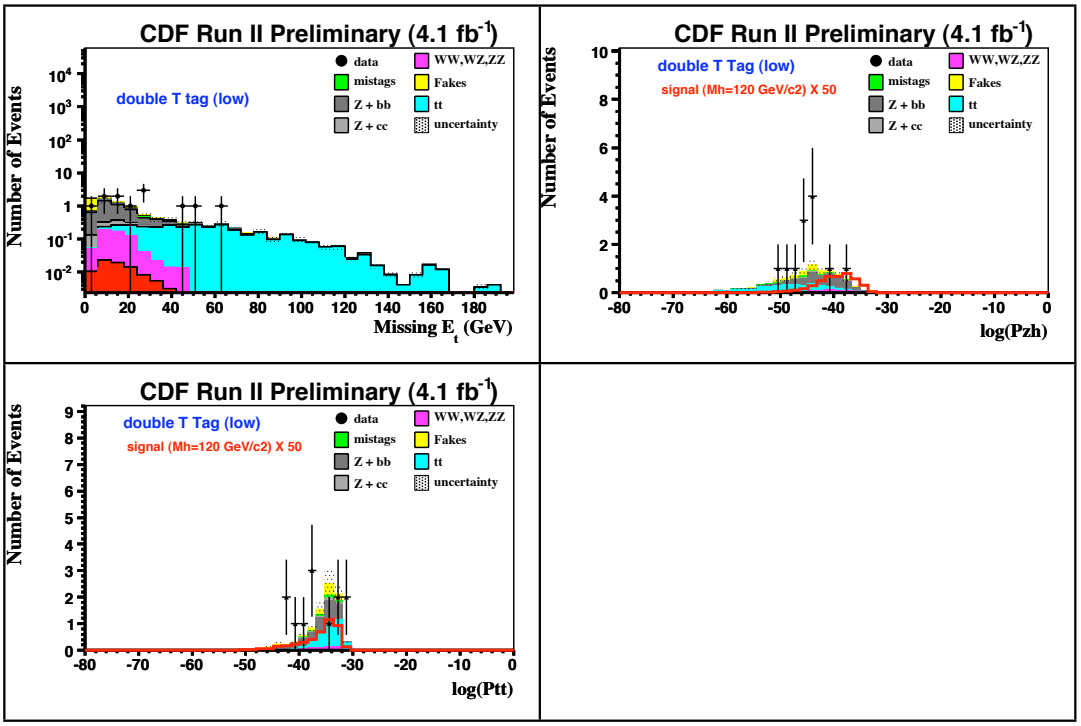


Figure 10.6: NN input distributions for the double tag low S/B channel.

Table 10.2: Comparison of observed and predicted event totals for tag level low S/B selection. Blank entries denote negligible contributions.

	(Low S/B Categories)		
Source	Double T Tag	L+JP Tag	Single T Tag
$t\bar{t}$	2.9 ± 0.6	3.2 ± 0.8	8.9 ± 1.9
WW		0.02 ± 0.003	0.1 ± 0.02
WZ		0.1 ± 0.02	1.2 ± 0.2
ZZ	0.5 ± 0.1	0.5 ± 0.1	2.0 ± 0.3
Z+jets ($b\bar{b}$)	3.2 ± 1.4	4.0 ± 1.7	21.1 ± 8.9
Z+jets ($c\bar{c}$)	0.3 ± 0.1	1.6 ± 0.7	11.0 ± 4.6
Z+Mistags	0.4 ± 0.1	3.8 ± 1.3	50.0 ± 7.5
fakes	1.4 ± 0.7	1.1 ± 0.5	22.5 ± 11.3
ZH ($120\,\mathrm{GeV}/c^2$)	0.1 ± 0.01	0.1 ± 0.02	0.2 ± 0.03
Total Background	8.7 ± 1.7	14.3 ± 2.4	116.8 ± 17.0
Data	12	14	116

10.2 Observed and Expected Limits

The observed event yields in data agree with the events predicted by our background model and we do not find a significant excess consistent with a Higgs signal. We use the MCLIMIT [104] machinery to quantify the maximum allowed ZH signal in the data when compared to the expected SM background in each bin of the 2D NN output. The MCLIMIT software package allows for the computation of Bayesian limits [8, 106] across multiple channels, background sources, and uncertainties.

To compute the 95% CL upper limit on the signal content of our data, we form the binned likelihood as :

$$\mathcal{L} = \prod_{i=1}^{N_{chan}} \times \prod_{j=1}^{N_{bins}} \Big[\frac{(s'_{ij} + \sum_{k=1}^{N_{bkg}} b'_{ijk})^{n_{ij}} \times e^{-(s'_{ij} + \sum_{k=1}^{N_{bkg}} b'_{ijk})}}{(n_{ij})!} \Big] \tag{10.1}$$

where N_{chan} is the number of analysis channels (6 in our case), N_{bins} is the total number of bins in the full two-dimensional NN output ($21 \times 21 = 441$ total including histogram overflow bins) of each channel, n_{ij} is the number of data events observed in bin j of channel i, N_{bkg} is equal to the total number of background processes contributing to the data model, b'_{ijk} represents the number of background events in bin j of channel i from background source k, and s'_{ij} stands for the number of signal events in bin j of channel i.

We include uncertainties on the background and signal estimates as Gaussian distributions, with the lower bound of the integration set to zero. This eliminates negative fluctuations in sample normalization. Assuming a flat (constant) prior on the number of signal events in the data, we form the Bayesian prior as :

$$\pi = \pi(\vec{s}) \times \prod_{k=1}^{N_{bkg}} \prod_{\ell=1}^{N_{unc}} \frac{1}{\sqrt{2\pi\sigma^B_{ijk\ell}}} e^{-\frac{(b_{ijk} - b'_{ijk})^2}{2(\sigma^B_{ijk\ell})^2}} \times \prod_{\ell=1}^{N_{unc}} \frac{1}{\sqrt{2\pi\sigma^S_{ij\ell}}} e^{-\frac{(s_{ij} - s'_{ij})^2}{2(\sigma^S_{ij\ell})^2}} \tag{10.2}$$

where $\pi(\vec{s})$ is the flat prior on the total signal content $\vec{s}$, N_{unc} is the number of uncertainties considered, $\sigma^B_{ijk\ell}$ represents the one-sigma uncertainty from systematic ℓ on

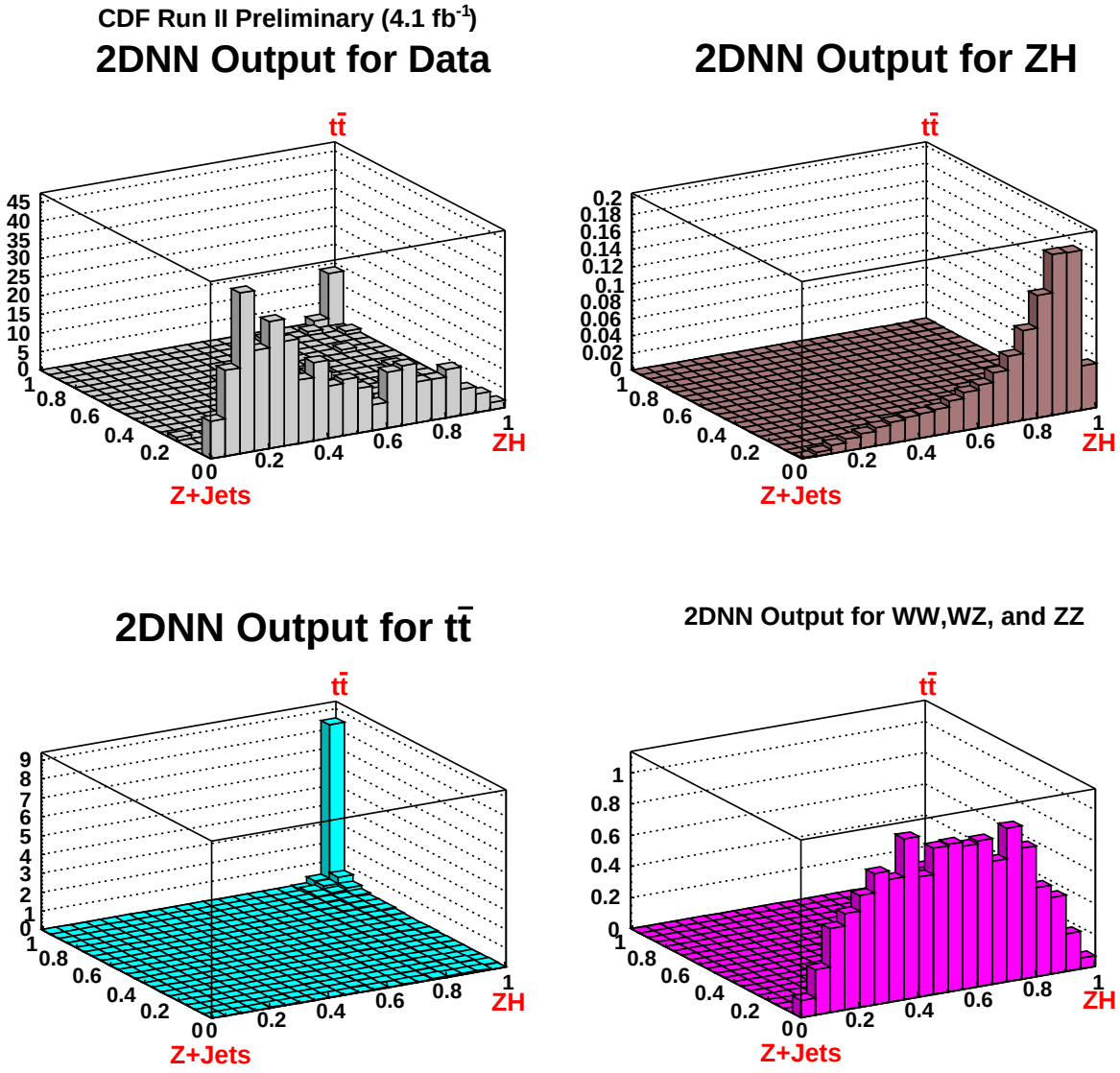


Figure 10.7: NN output for high S/B single tag (T) channel data, ZH ($M_H = 120\text{GeV}/c^2$), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

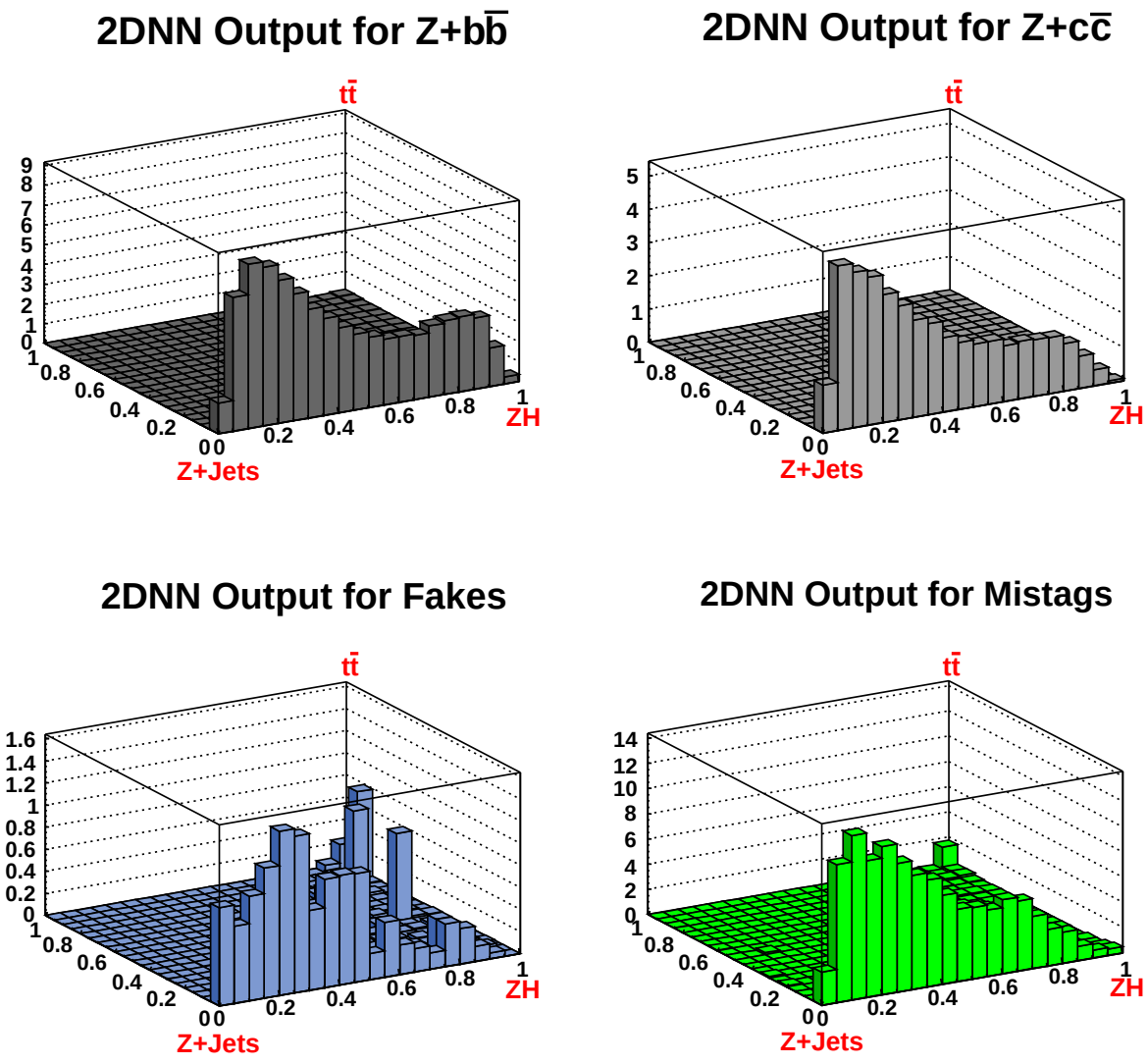


Figure 10.8: NN output for high S/B single tag (T) channel $Z+b\bar{b}$, $Z+b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

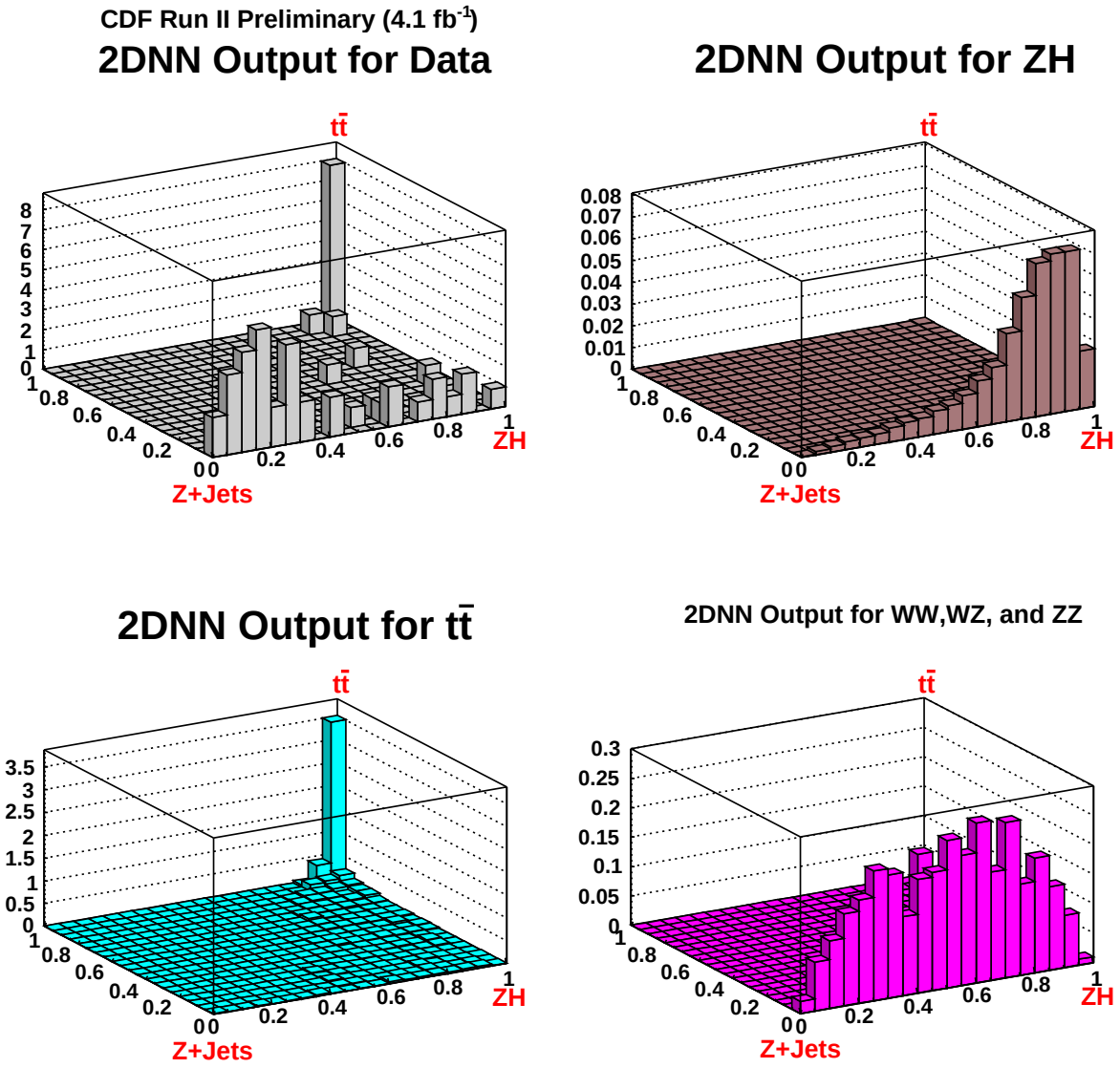


Figure 10.9: NN output for high S/B single tag (L+JP) channel data, ZH (M_H = 120GeV/c^2), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

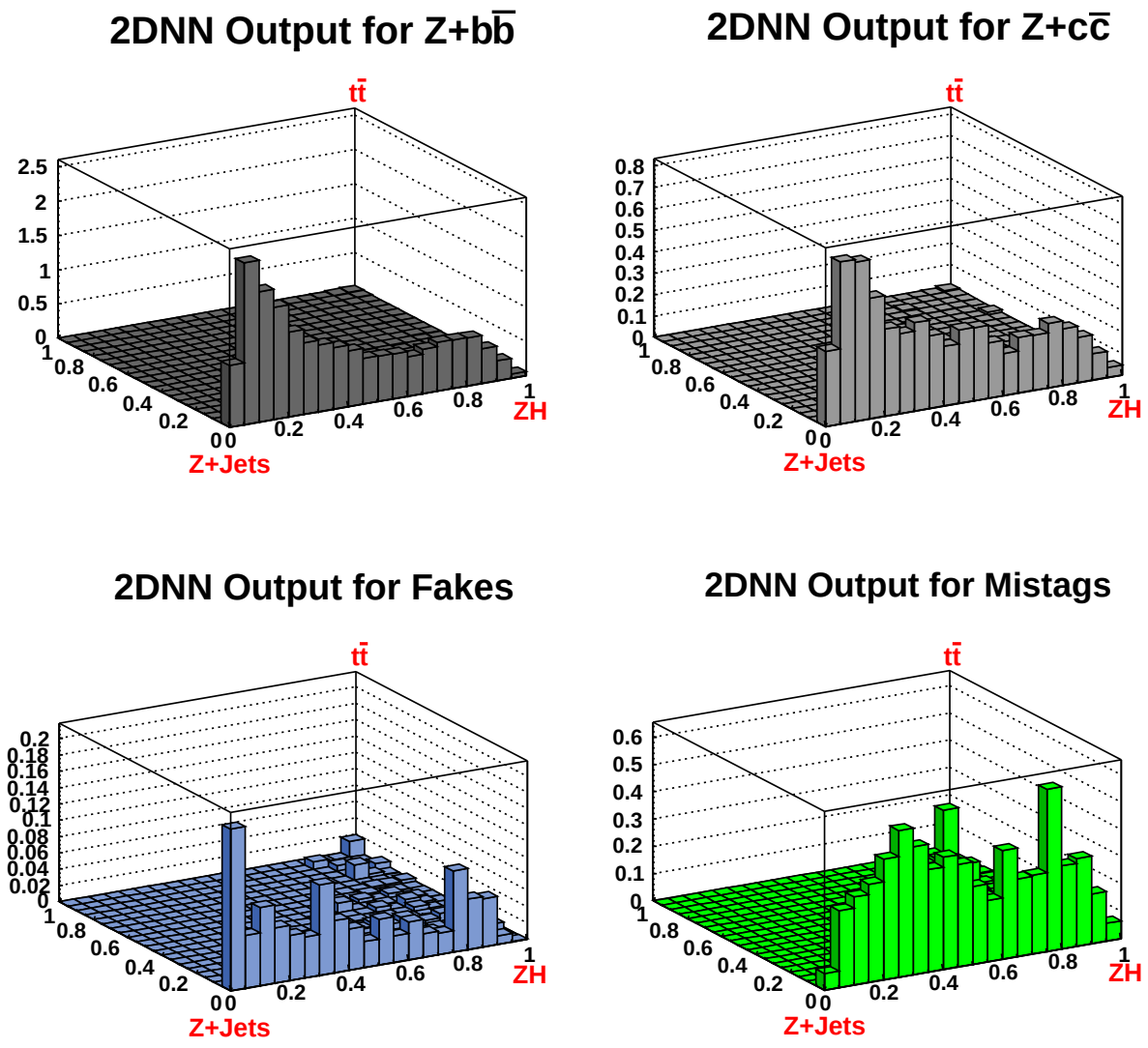


Figure 10.10: NN output for high S/B single tag (L+JP) channel $Z + b\bar{b}$, $Z + b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

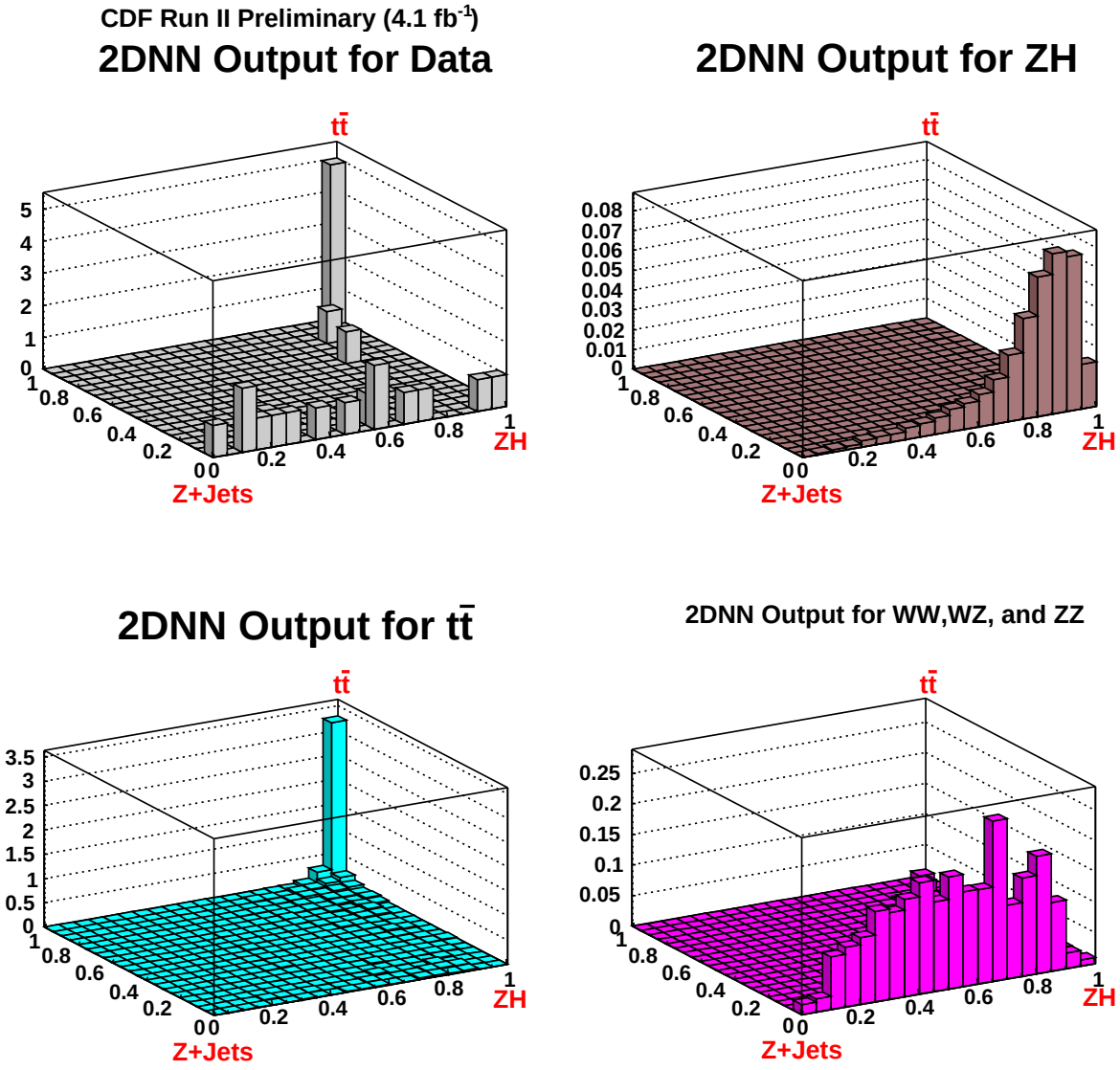


Figure 10.11: NN output for high S/B single tag (TT) channel data, ZH ($M_H = 120\text{GeV}/c^2$), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

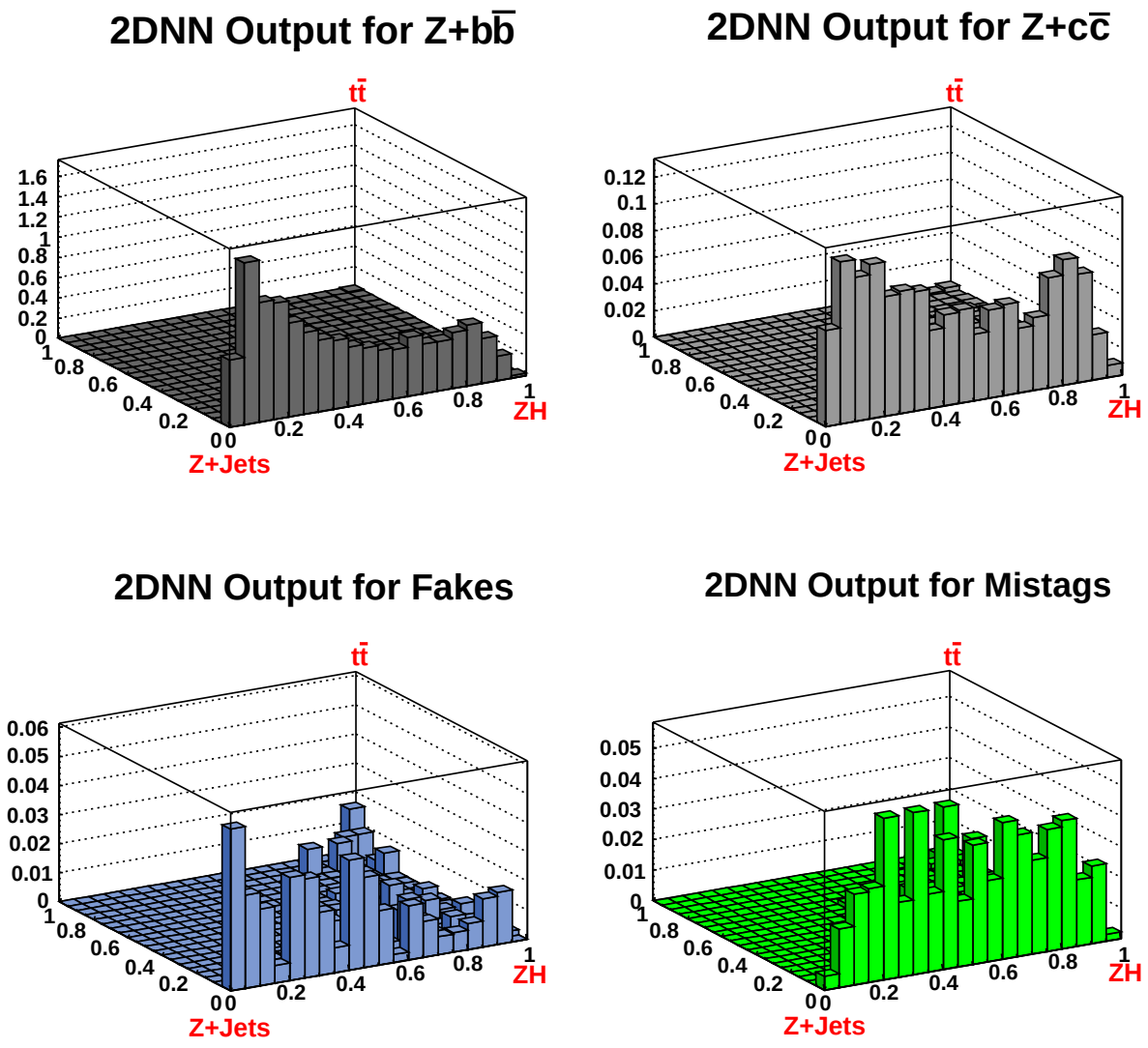


Figure 10.12: NN output for high S/B single tag (TT) channel $Z + b\bar{b}$, $Z + b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

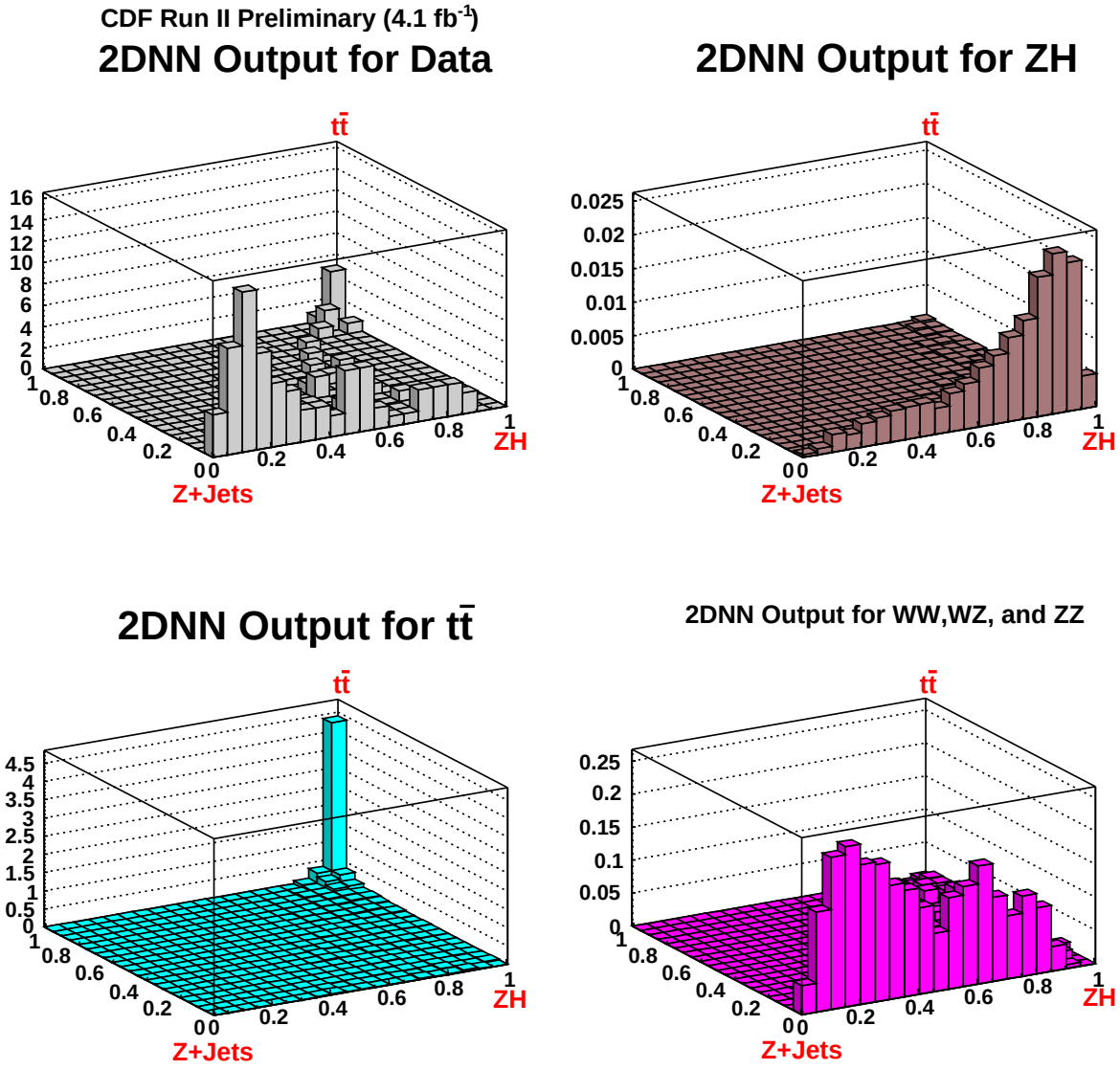


Figure 10.13: NN output for low S/B single tag (T) channel data, ZH ($M_H = 120\text{GeV}/c^2$), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

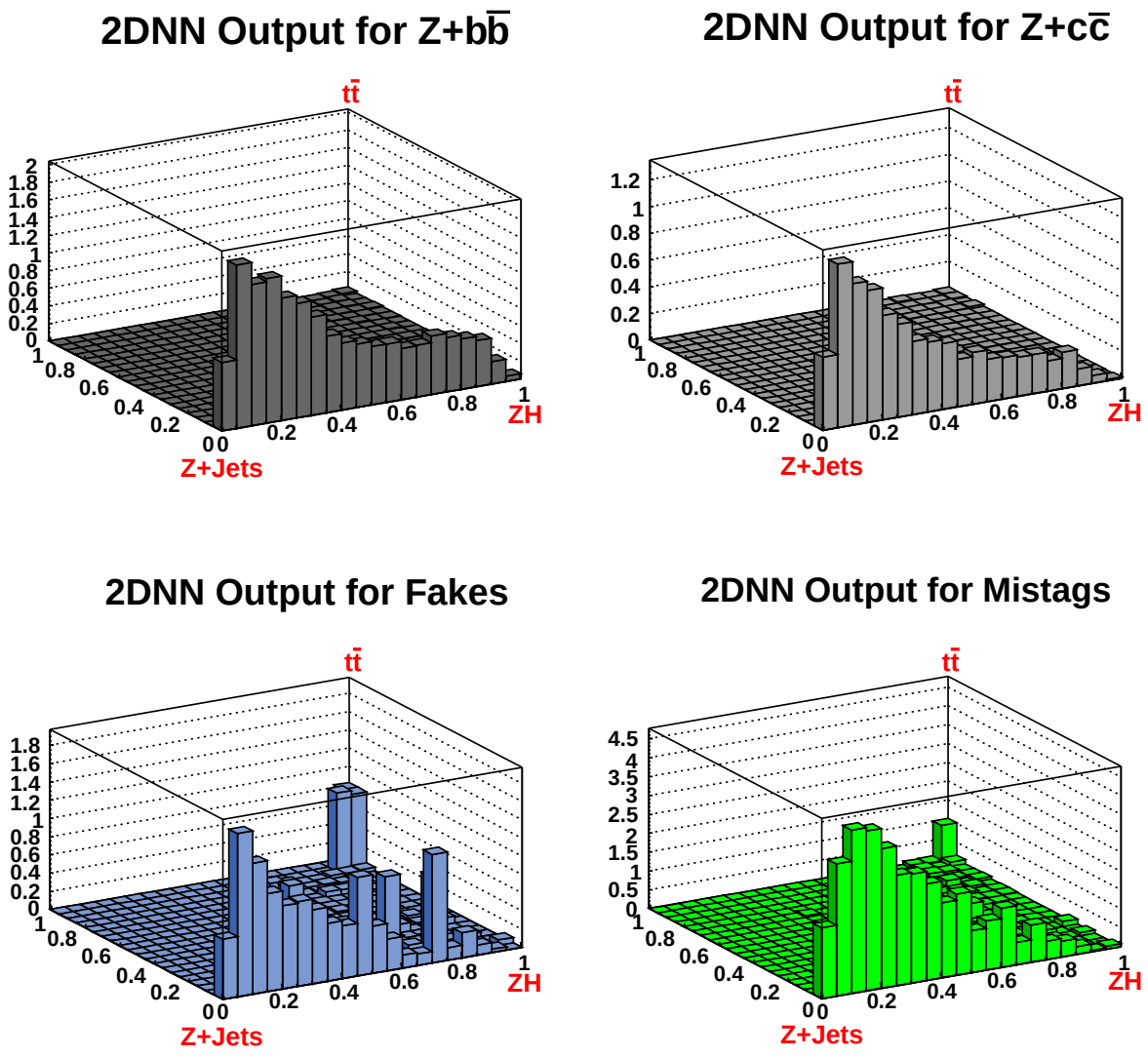


Figure 10.14: NN output for low S/B single tag (T) channel $Z+b\bar{b}$, $Z+b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

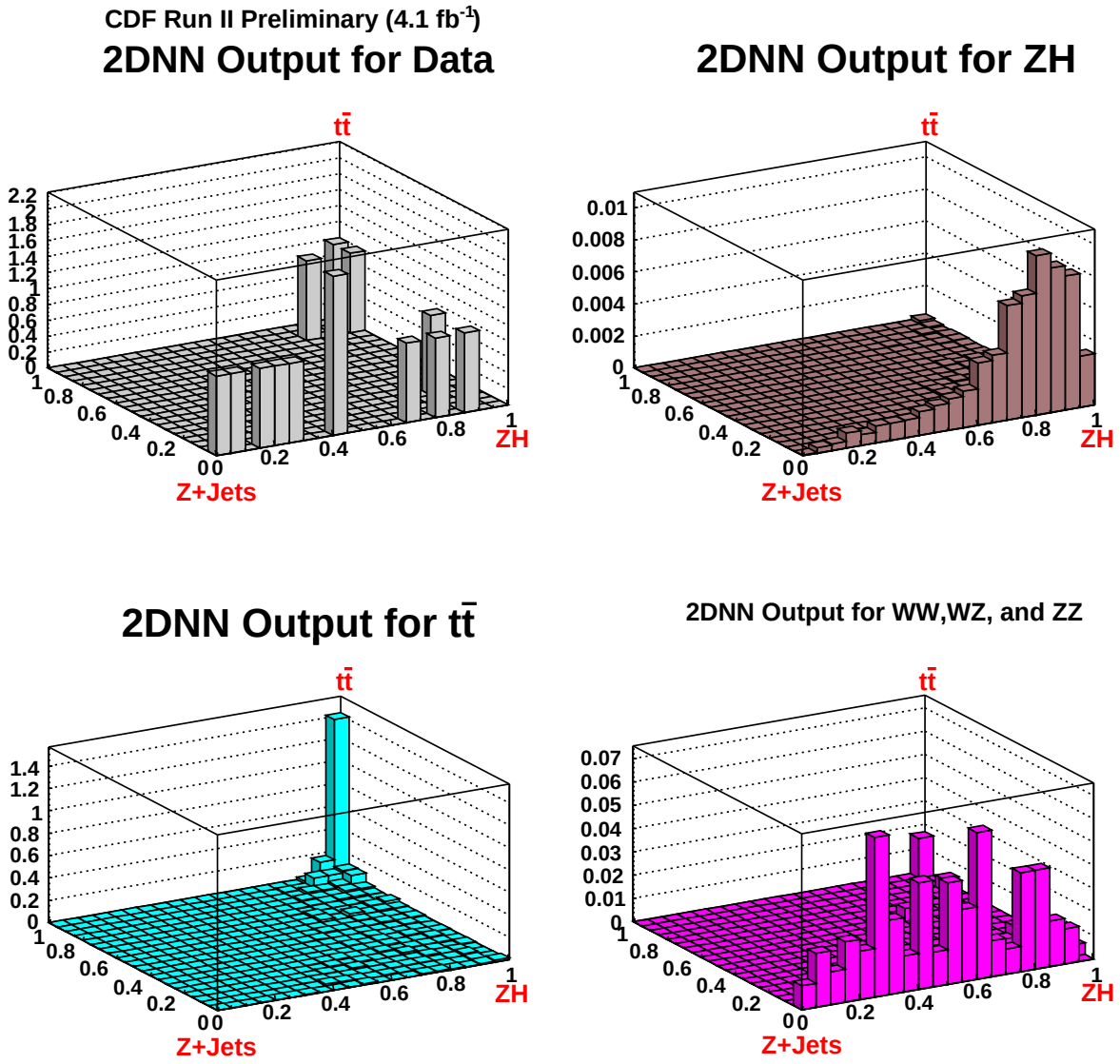


Figure 10.15: NN output for low S/B single tag (L+JP) channel data, ZH (M_H = 120GeV/c^2), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

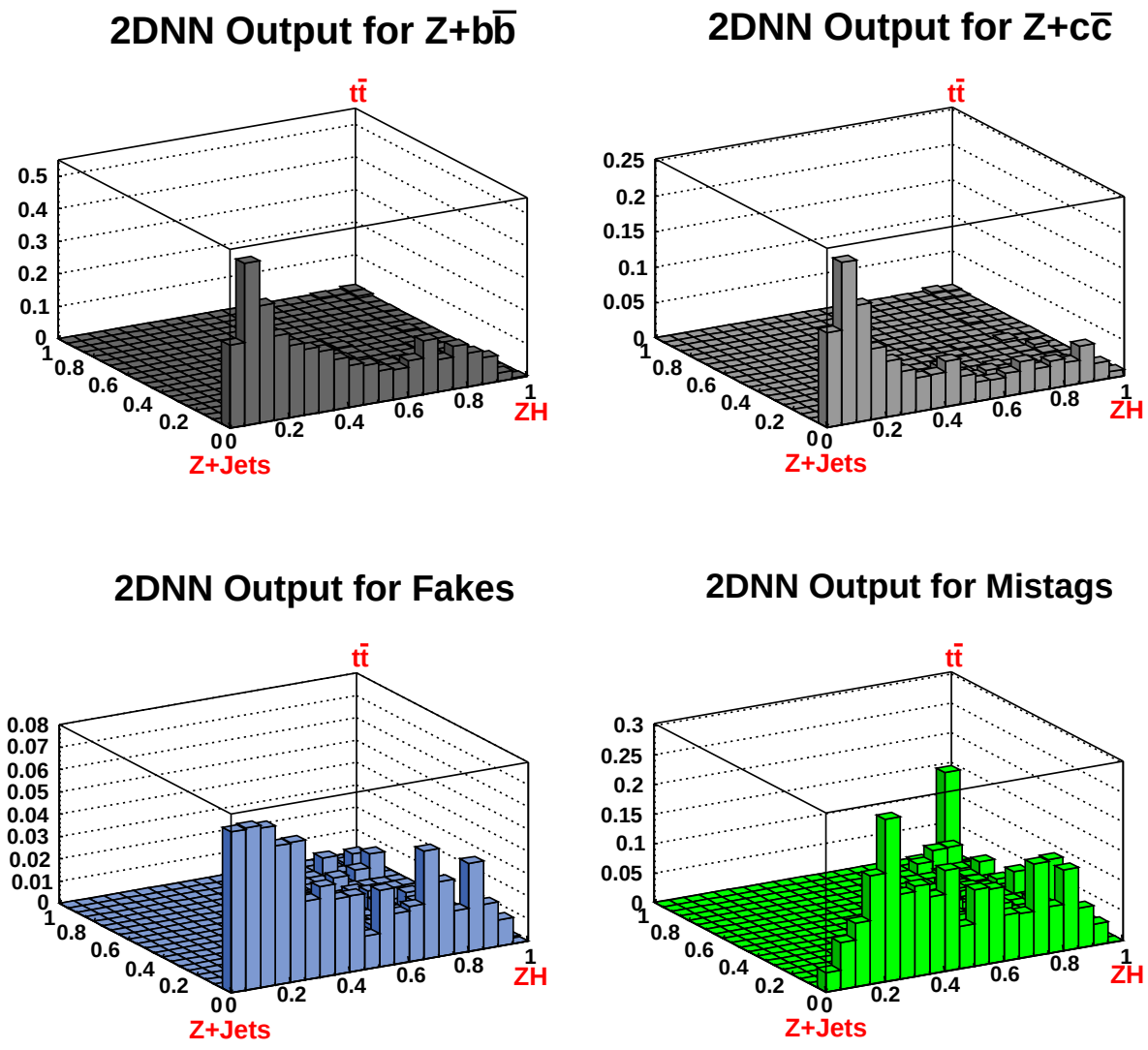


Figure 10.16: NN output for low S/B single tag (L+JP) channel $Z + b\bar{b}$, $Z + b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

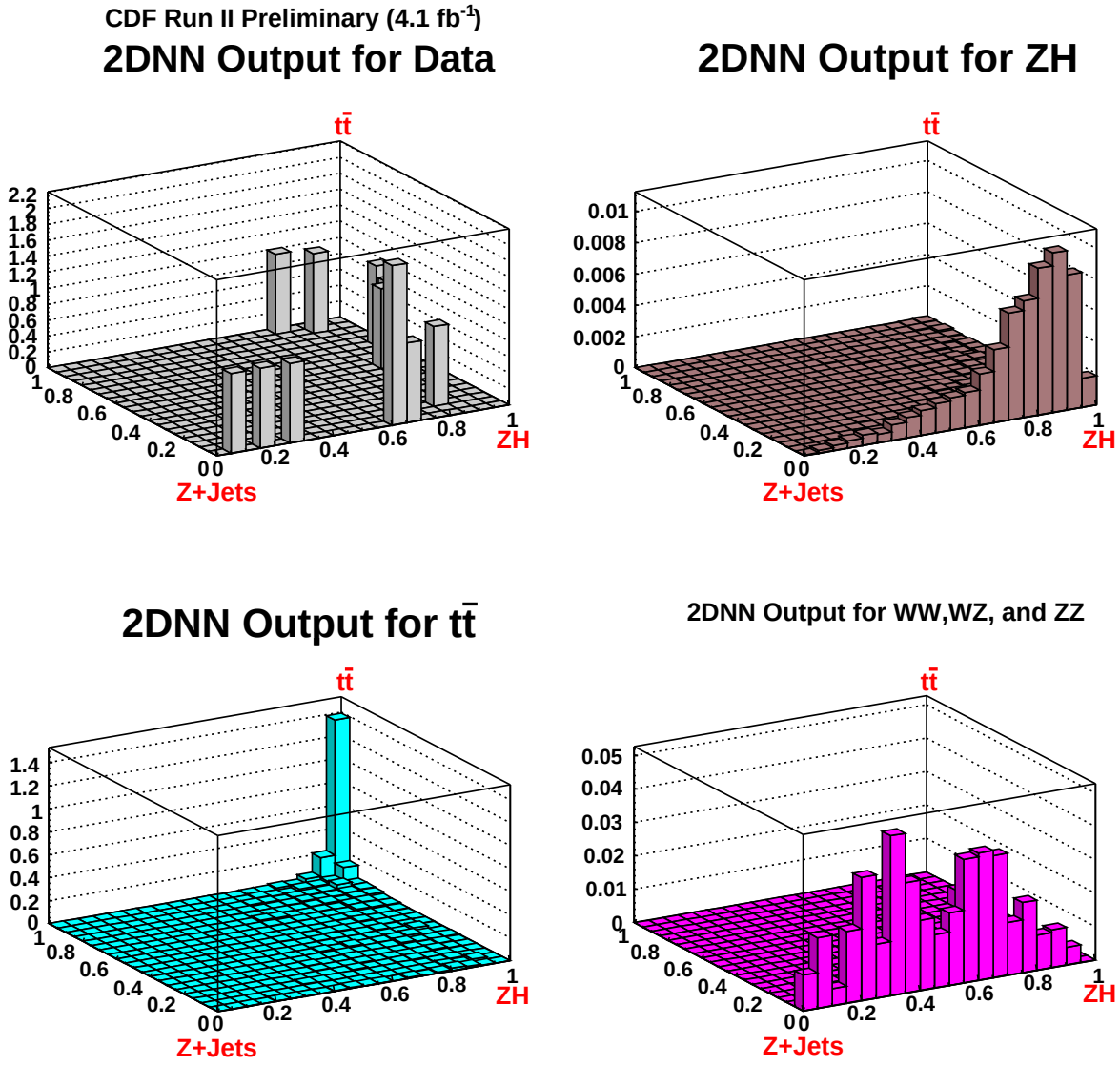


Figure 10.17: NN output for low S/B single tag (TT) channel data, ZH ($M_H = 120\text{GeV}/c^2$), $t\bar{t}$ and Diboson events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

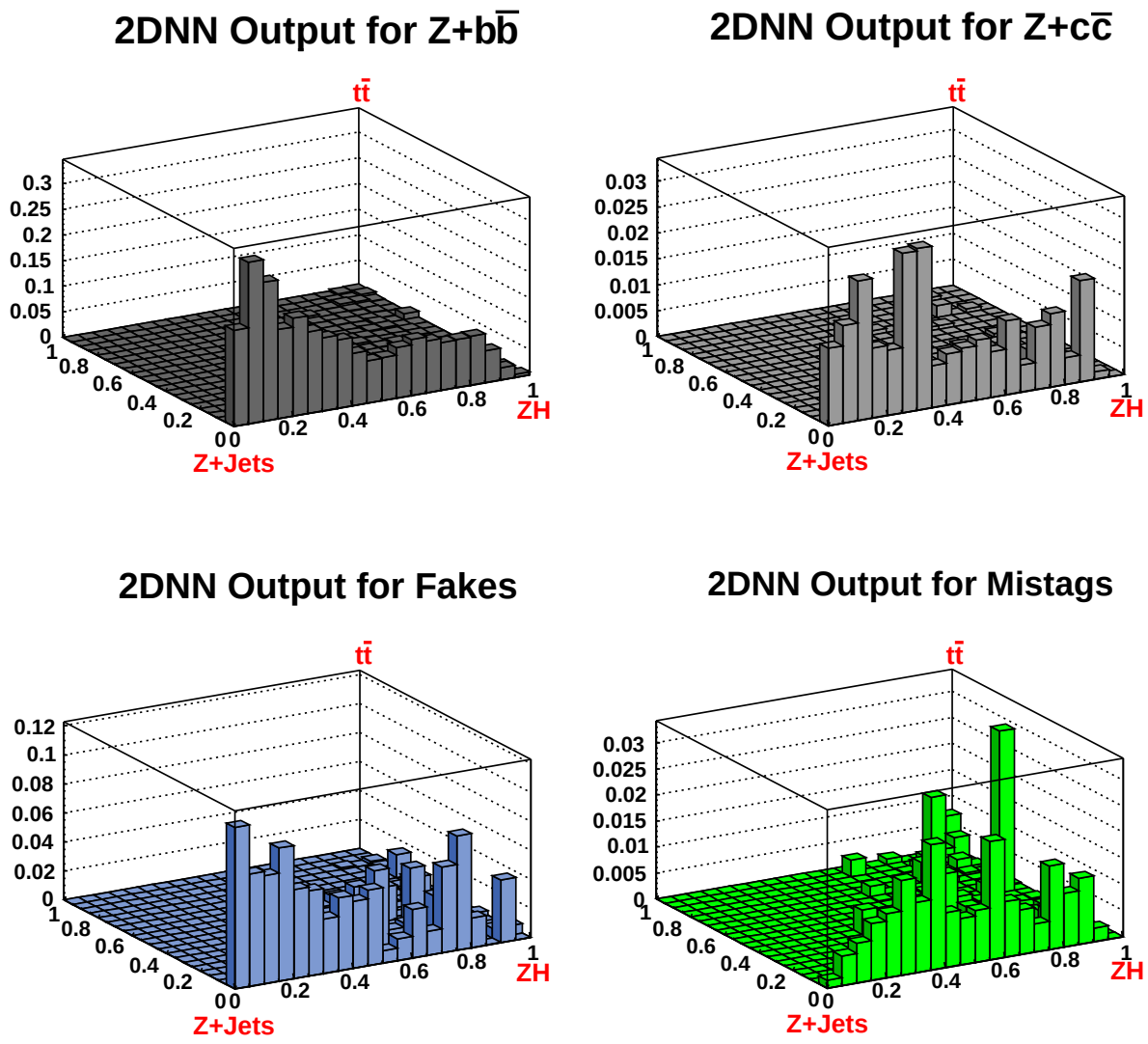


Figure 10.18: NN output for low S/B single tag (TT) channel $Z + b\bar{b}$, $Z + b\bar{c}$, Fake, and mistag events. The Z+jets, ZH, and $t\bar{t}$ corners are indicated in red text.

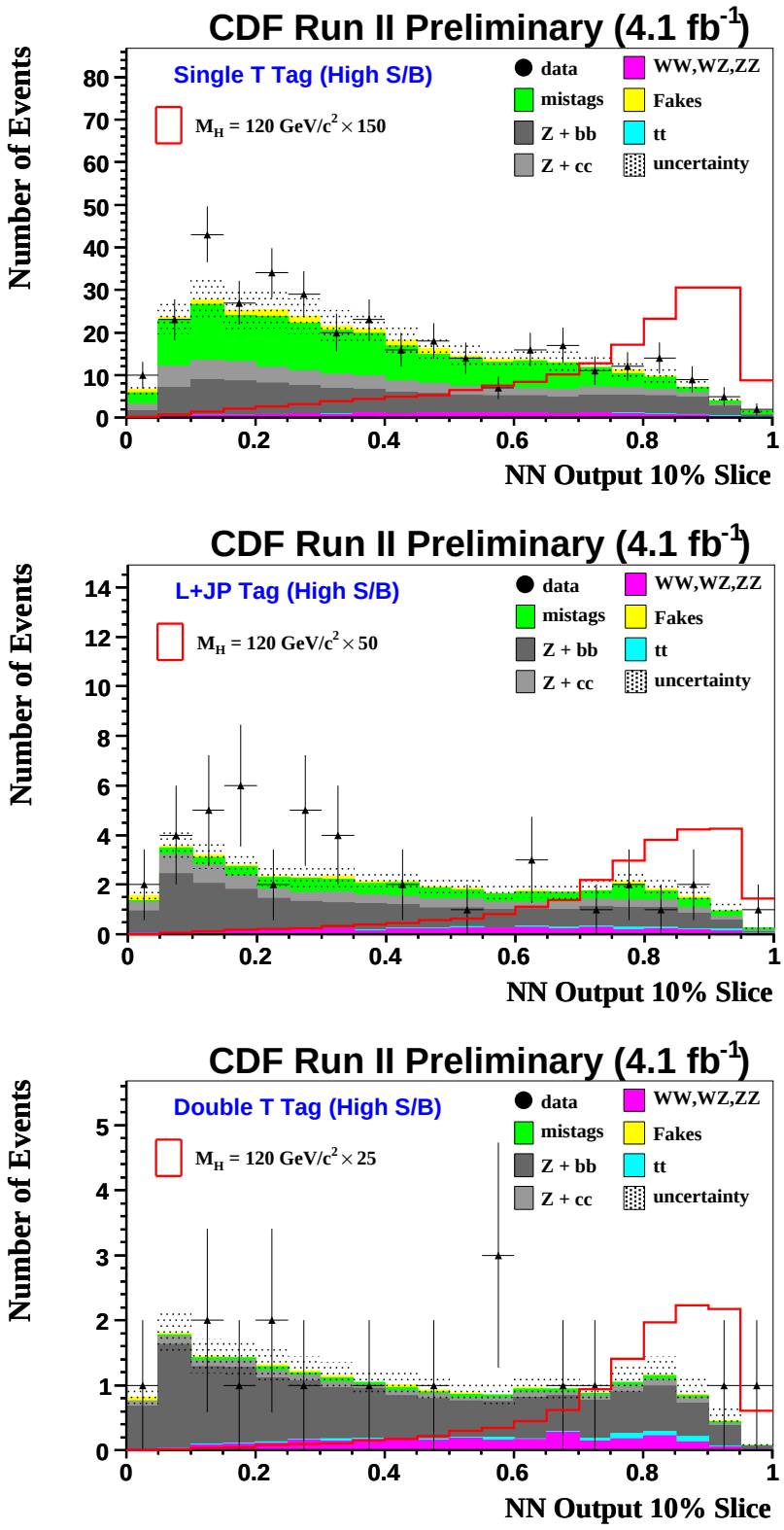


Figure 10.19: NN output projections for the final high (S/B) analysis channels onto the Z+jets – ZH (x) axis of the 2D output. A cut of NNy<0.1 is made to emphasize the signal region.

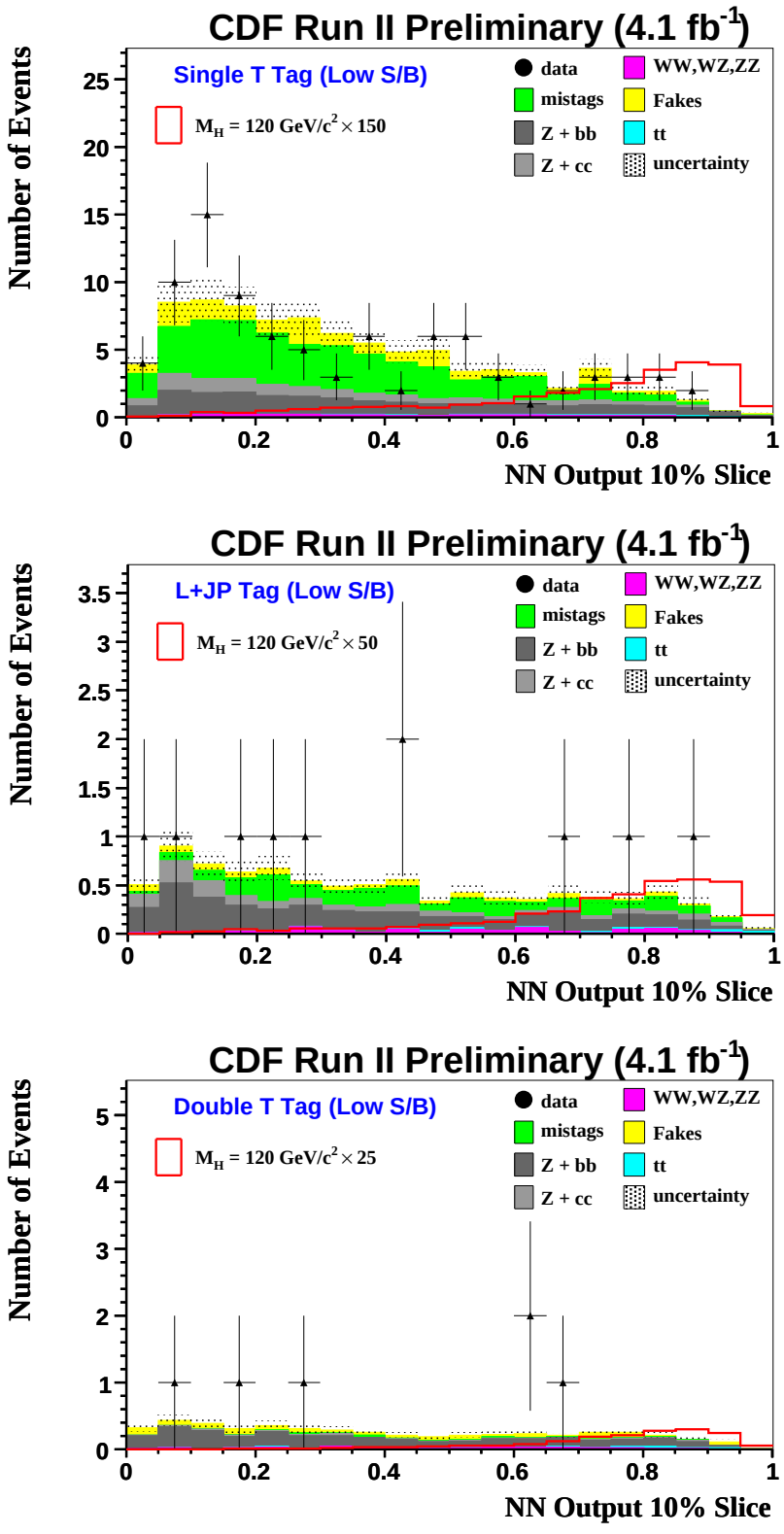


Figure 10.20: NN output projections for the final low (S/B) analysis channels onto the Z+jets – ZH (x) axis of the 2D output. A cut of NNy<0.1 is made to emphasize the signal region.

background k in bin j of channel i, b_{ijk} is the SM prediction for the number of events from background k in bin j of channel i, $\sigma^S_{ij\ell}$ represents the one-sigma uncertainty from systematic ℓ on the signal in bin j of channel i, and s_{ij} is an additional parameter representing the number of signal events in bin j of channel i.

The 95% confidence level upper limit on the number of total signal events $\vec{s}_{up}$ is obtained by integrating over the likelihood convoluted with the Bayesian prior :

$$0.95 = \frac{\int_0^{\vec{s}_{up}} \int_0^{\infty} \int_0^{\infty} [\mathcal{L} \times \pi] \; d\vec{b'} \; d\vec{s'} \; d\vec{s}}{\int_0^{\infty} \int_0^{\infty} \int_0^{\infty} [\mathcal{L} \times \pi] \; d\vec{b'} \; d\vec{s'} \; d\vec{s}} \tag{10.3}$$

The value of $\vec{s}_{up}$ obtained from Eq. 10.3 is divided by the total signal content (at each Higgs mass considered) to express the upper limits as multiples of the predicted SM value for $\sigma_{ZH} \times BR(H \rightarrow b\bar{b})$. Upper limits expressed in picobarns are obtained by dividing $\vec{s}_{up}$ by the total integrated luminosity.

As mentioned in Chapter 9, a Gaussian is included in Eq. 10.2 to account for uncertainty on the SM cross-sections when the upper limits are presented as multiples of $\sigma_{ZH} \times BR(H \rightarrow b\bar{b})$; this term is removed when calculating limits in picobarns.

To compare the observed limits to the distribution of upper limits predicted by our background model and systematics alone, we run 1000 pseudo-experiments at each Higgs mass. For each pseudo-experiment, the background templates are fluctuated within their uncertainties to generate a pseudo-data set (replacing the n_{ij} in equation 10.1). The distribution of limits obtained for $M_H = 115\,\mathrm{GeV}/c^2$ is shown in Fig. 10.21. The median of the pseudo-experiment limit distribution is the expected 95% CL upper limit.

Expected and observed limits for individual channels assuming a Higgs mass of $115\,\mathrm{GeV}/c^2$ are shown in Table 10.3. Expected and observed limits for all six channels combined for $100 \leq M_H \leq 150\,\mathrm{GeV}/c^2$ are shown in Table 10.4 and Fig. 10.22 as multiples of $\sigma_{ZH} \times BR(H \rightarrow b\bar{b})$. Table 10.5 and Fig. 10.23 present the limits in pb. Variation in the pb limits is due to the dependence of the signal systematic uncertainties on Higgs mass (the $\sigma^S_{ij\ell}$ in Eq. 10.2 depend on the Higgs mass).

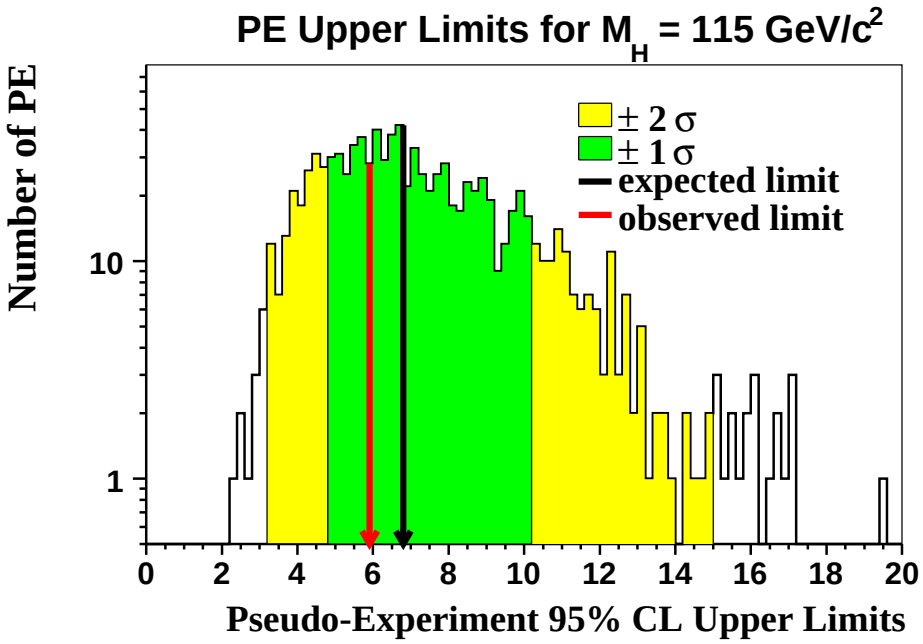


Figure 10.21: Distribution of 95% CL upper limits from 1000 pseudo-experiements at $M_H = 115\,\mathrm{GeV}/c^2$ with 1σ (green) and 2σ (yellow) bands indicated. The expected limit (black arrow) is the median of the pseudo-experiement limit distribution. The observed limit (red arrow) is shown for comparison.

Table 10.3: Expected and observed limits divided by the SM value for $\sigma_{ZH} \times BR(H \to b\bar{b})$ for individual channels at $M_H = 115\,\mathrm{GeV}/c^2$. The total number of observed data events, predicted total background events, and the $S/\sqrt{B}$ are also shown.

Channel	Observed Events	Total Bkg.	$S/\sqrt{B}$	Expected$^{+1\sigma}_{-1\sigma}$	Observed
TT High S/B	23	29.3 ± 7.0	0.13	$12.1\ ^{+\ 5.7}_{-\ 5.5}$	11.3
L+JP High S/B	56	52.8 ± 10.5	0.09	$15.98\ ^{+\ 6.95}_{-\ 7.6}$	10.6
T High S/B	406	366.1 ± 55.9	0.09	$15.5\ ^{+\ 7.1}_{-\ 7.98}$	16.9
TT Low S/B	12	8.7 ± 1.7	0.04	$49.2\ ^{+\ 18.95}_{-\ 19.9}$	58.2
L+JP Low S/B	14	14.3 ± 2.4	0.03	$50.6\ ^{+\ 21.6}_{-\ 21.99}$	71.1
T Low S/B	116	116.8 ± 17.0	0.02	$41.6\ ^{+\ 19.99}_{-\ 20.04}$	38.5
Combined				$6.8\ ^{+\ 3.22}_{-\ 2.04}$	5.91

Table 10.4: Expected and observed 95% CL upper limits on $\sigma_{ZH} \times BR(H \to b\bar{b})$ divided by the SM values.

$M_H(\,\mathrm{GeV}/c^2)$	Expected$^{+1\sigma}_{-1\sigma}$	Observed
100	$6.7\ ^{+\ 2.91}_{-\ 2.18}$	4.53
105	$6.38\ ^{+\ 2.67}_{-\ 1.94}$	4.6
110	$6.34\ ^{+\ 3.17}_{-\ 1.9}$	5.25
115	$6.8\ ^{+\ 3.22}_{-\ 2.04}$	5.91
120	$8.49\ ^{+\ 3.58}_{-\ 2.57}$	7.89
125	$10.21\ ^{+\ 3.99}_{-\ 3.18}$	8.14
130	$12.79\ ^{+\ 6.27}_{-\ 3.9}$	10.3
135	$18.74\ ^{+\ 8.34}_{-\ 5.79}$	14.41
140	$28.49\ ^{+\ 12.22}_{-\ 8.67}$	19.27
145	$45.34\ ^{+\ 18.76}_{-\ 13.31}$	24.22
150	$73.72\ ^{+\ 37.3}_{-\ 23.07}$	42.93

Table 10.5: Expected and observed 95% CL upper limits on $\sigma_{ZH} \times BR(H \to b\bar{b})$. The SM values are included for comparison.

M_H(GeV/c^2)	Expected$^{+1\sigma}_{-1\sigma}$ (pb)	Observed (pb)	SM Value (pb)
100	$0.87\ ^{+\ 0.4}_{-\ 0.27}$	0.63	0.1368
105	$0.72\ ^{+\ 0.32}_{-\ 0.23}$	0.58	0.1150
110	$0.62\ ^{+\ 0.26}_{-\ 0.2}$	0.52	0.0960
115	$0.53\ ^{+\ 0.26}_{-\ 0.17}$	0.44	0.0789
120	$0.53\ ^{+\ 0.22}_{-\ 0.16}$	0.49	0.0635
125	$0.48\ ^{+\ 0.21}_{-\ 0.14}$	0.44	0.0496
130	$0.49\ ^{+\ 0.21}_{-\ 0.15}$	0.41	0.0375
135	$0.5\ ^{+\ 0.21}_{-\ 0.16}$	0.39	0.0271
140	$0.51\ ^{+\ 0.25}_{-\ 0.16}$	0.34	0.0188
145	$0.54\ ^{+\ 0.23}_{-\ 0.16}$	0.33	0.0123
150	$0.57\ ^{+\ 0.25}_{-\ 0.17}$	0.35	0.0074

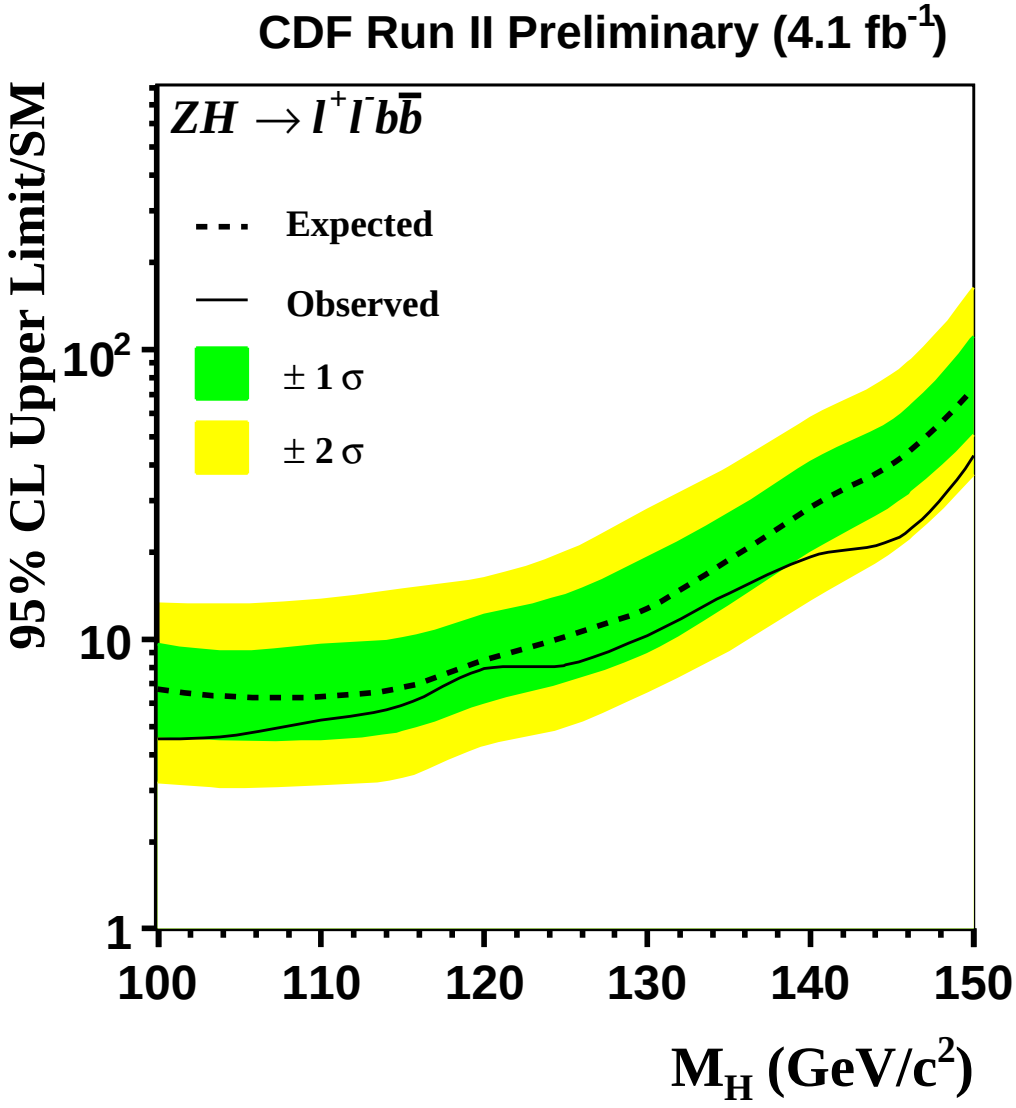


Figure 10.22: Expected and observed limits with $\pm$ 1 and 2 σ bands.

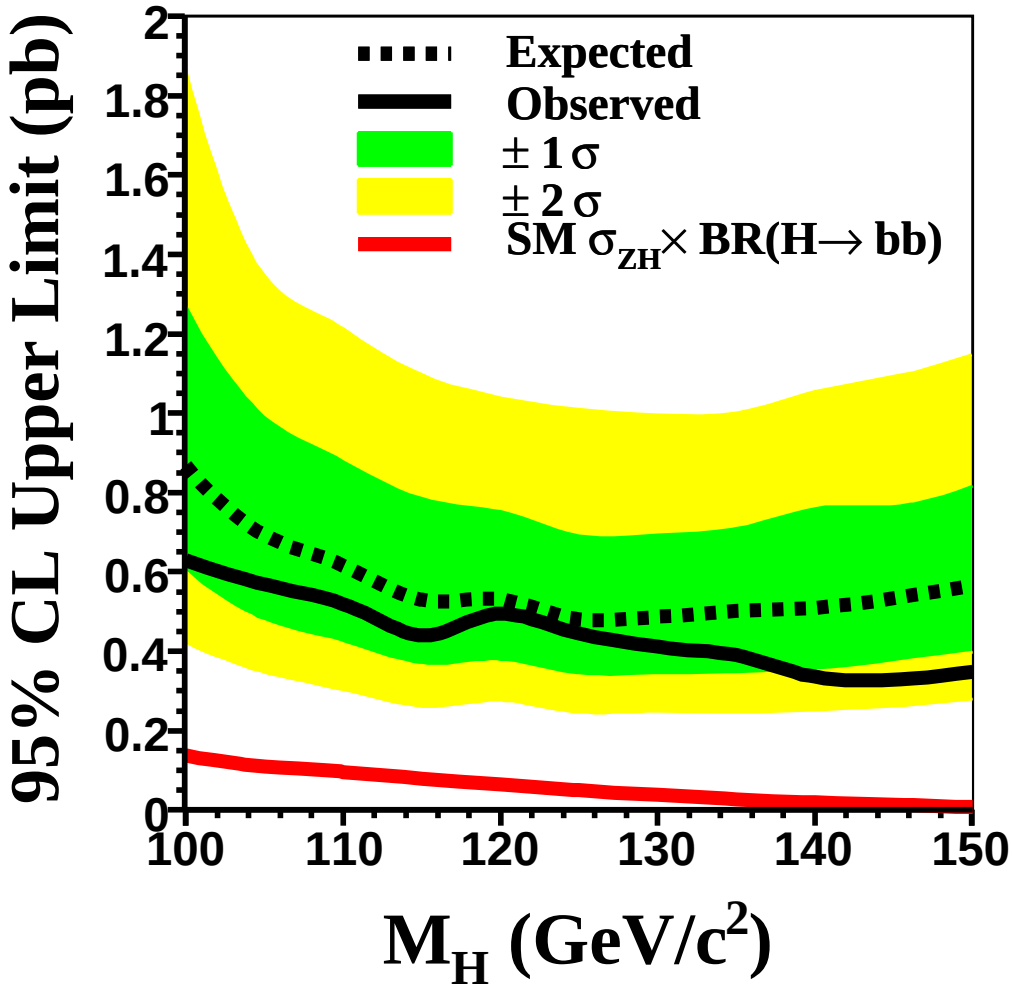


Figure 10.23: Expected and observed limits with $\pm$ 1 and 2 σ bands. The SM values of $\sigma_{ZH} \times BR(H \to b\bar{b})$ are indicated by the red line.

Chapter 11: Conclusions

11.1 Sensitivity Gain

The SM Higgs search presented in this dissertation achieves significant improvement in sensitivity over previous CDF II searches in the mode $ZH \to \ell^+\ell^- b\bar{b}$ [35, 36], producing the world's strongest upper limits on the $\sigma_{ZH}^{SM} \times BR(H \to b\bar{b})$. Table 11.1 shows the degree of improvement achieved by this search compared to previous CDF II efforts.

Table 11.1: Comparison of observed 95% CL upper limits on $\sigma_{ZH}^{SM} \times BR(H \to b\bar{b})$ for CDF II searches @ $M_H = 115\,\mathrm{GeV}/c^2$. The absolute improvement is calculated by dividing previous limits by the current limit (in $4.1\,\mathrm{fb}^{-1}$). The projected limits are calculated under the assumption that sensitivity improves with increased integrated luminosity ($\mathcal{L}$) as $1/\sqrt{\mathcal{L}}$. The relative improvement is the ratio of the projected limits to the current.

Analysis	Limit (× SM)	Absolute Imp.	Projected Limit (× SM)	Relative Imp.
$1\,\mathrm{fb}^{-1}$ [35]	16	2.7	7.9	1.34
$2.7\,\mathrm{fb}^{-1}$ [36]	8.2	1.4	6.7	1.14
$4.1\,\mathrm{fb}^{-1}$	5.9	-	5.9	-

The improvement in sensitivity is due to the combined effect of several factors :

- increased total integrated luminosity
- expanded trigger selection (this search is the first to use the `Z_NOTRACK` trigger)
- the addition of Z candidates formed from forward electron pairs
- compensation for gaps in calorimeter coverage with the CrkTrk electron category
- incorporation of the Jet Probability b-tagging algorithm for 3 b-tag classes
- division of the final event sample into separate categories by $S/\sqrt{B}$

- improved rejection of incorrectly b-tagged jets through the Karlsruhe NN [102]
- updated NN jet energy corrections
- the combination of matrix element probabilities with classification NNs
- the use of discriminant NNs optimized for specific b-tag classes

11.2 Combination With Other CDF II Higgs Searches

Due to the low signal expectation and large background rates, no single Tevatron SM Higgs mode is expected to achieve discovery or exclusion alone at low mass. To maximize the overall sensitivity of the CDF II experiment to a Higgs signal, a combined search across multiple Higgs production and decay modes is performed.

All NN output histograms (signal, background, and systematic variations) were passed to the CDF Higgs Working Group (CDFHWG) [107] for inclusion in the combined CDF Higgs result [108]. The results presented in this dissertation were cross-checked by the CDFHWG with a similar Bayesian computation and an alternate CL_s method. Due to the large ($21 \times 21 \times 6$) number of bins from the ZH 2D-NN outputs, all CDF ZH histograms were reduced to a 10×10 binning by the CDFHWG. This resulted in a roughly 5% reduction in sensitivity. Despite the rebinning, the $ZH \to \ell^+\ell^- b\bar{b}$ search contributed strongly to the CDF upper limits on SM Higgs production at $M_H = 115\,\mathrm{GeV}/c^2$ of $3.12 \times SM$ observed and $2.38 \times SM$ expected. CDF upper limits for the mass range $200 \geq M_H \geq 100\,\mathrm{GeV}/c^2$ are shown in Figure 11.1.

11.3 Tevatron Combination

To further improve the Tevatron's sensitivity to a SM Higgs signal, a combination of CDF and D0 SM Higgs results was performed by the Tevatron New Phenomena and Higgs Working Group (TEVNPHWG) [109]. The 2D-NN histograms produced for this dissertation

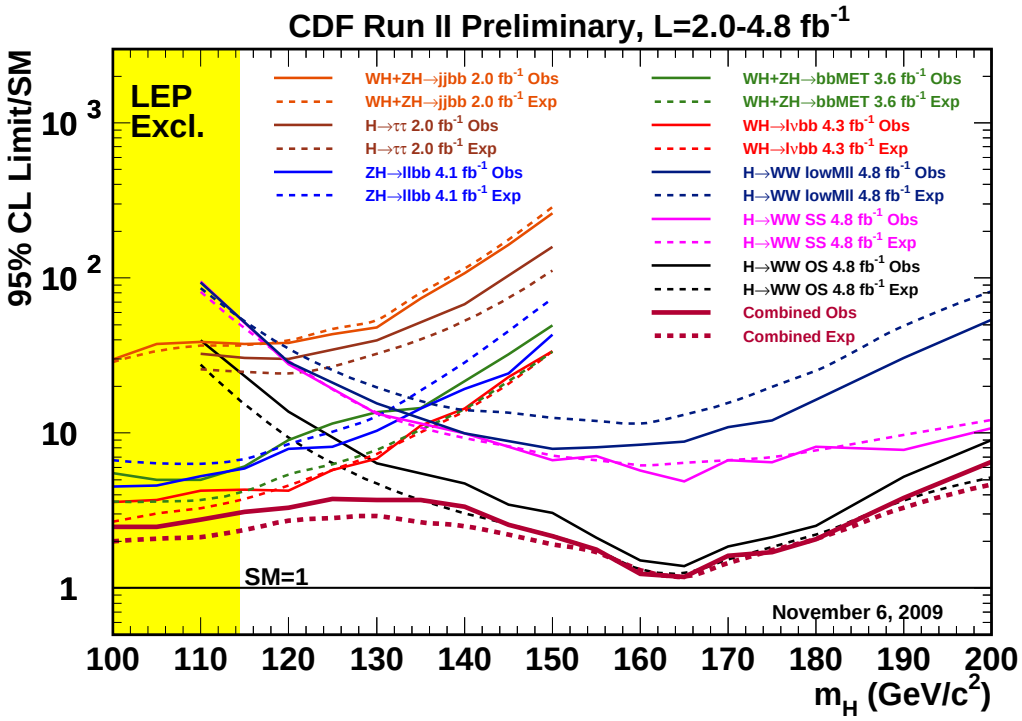


Figure 11.1: CDF Higgs Working Group upper limits on SM Higgs production as a function of Higgs mass. The combined upper limits are shown as solid (observed) and dashed (expected) lines in dark red. The limits combine the eight Higgs modes listed in the legend. The $ZH \to \ell^+\ell^- b\bar{b}$ limits are indicated by the solid (observed) and dashed (expected) blue lines.

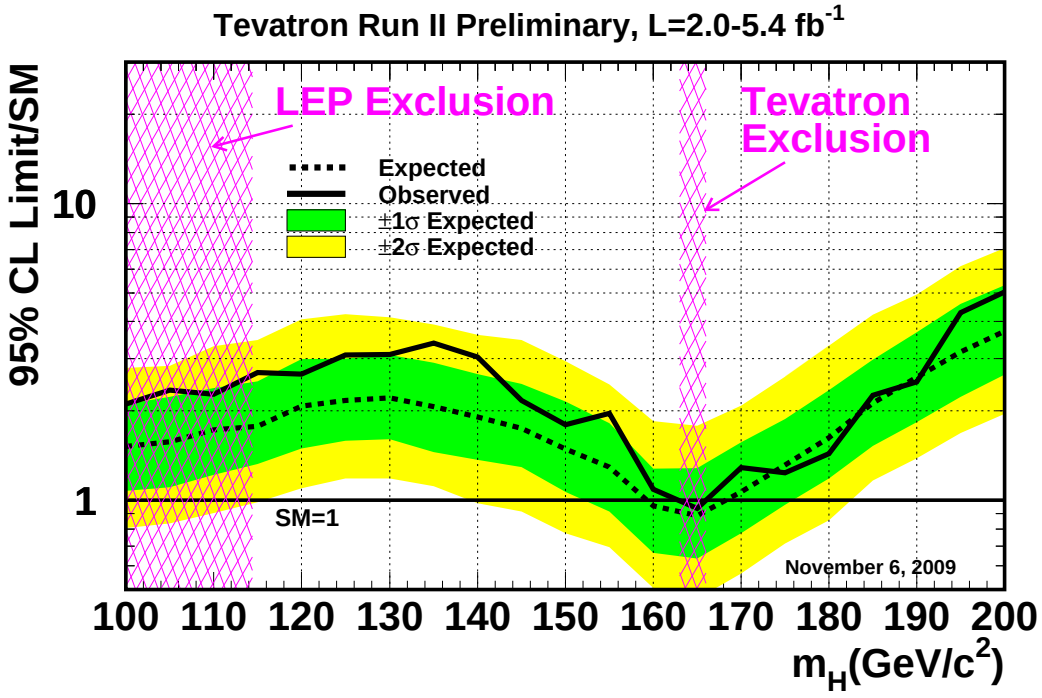


Figure 11.2: Tevatron New Phenomena and Higgs Working Group upper limits on SM Higgs production as a function of Higgs mass. The combined upper limits are shown as solid (observed) and dashed (expected) lines, with $\pm 1\sigma$ (green) and $\pm 2\sigma$ (yellow) bands indicated. The pink shaded regions show the LEP and Tevatron 95% CL exclusions.

were passed to the TEVNPHWG (after rebinning), and the upper limits were cross-checked by members of the $D0$ experiment. The CDF $ZH \to \ell^+\ell^- b\bar{b}$ search contributed to the latest Tevatron limits on SM Higgs production at $M_H = 115\,\mathrm{GeV}/c^2$ of $2.70 \times SM$ observed and $1.78 \times SM$ expected.

The combination of CDF and D0 searches excludes the mass range $166 \geq M_H \geq 163\,\mathrm{GeV}/c^2$ at the 95% CL. The reduced exclusion range (when compared to the March 2009 3.2), results from an upward fluctuation in the latest fb^{-1} of data. The combined Tevatron upper limits for the mass range $200 \geq M_H \geq 100\,\mathrm{GeV}/c^2$ are shown in Figure 11.2.

11.4 Future Prospects

The $ZH \rightarrow \ell^+\ell^- b\bar{b}$ mode will continue to be an important component of the Tevatron SM Higgs search. To date, CDF has recored $2\,\mathrm{fb}^{-1}$ of collision data beyond the $4.1\,\mathrm{fb}^{-1}$ used in this dissertation. The addition of this data is expected to yield a 22% improvement in sensitivity to a ZH signal.

CDF may continue to record Tevatron collisions through the Fall of 2011, bringing the total Run II data sample to about $10\,\mathrm{fb}^{-1}$. Applying the current analysis to a data sample of this size would produce upper limits roughly 36% less than the current values.

Beyond the gain from increased integrated luminosity, future iterations of the CDF $ZH \rightarrow \ell^+\ell^- b\bar{b}$ will feature several improvements :

- expanded use of muon triggers
- likelihood or NN based lepton identification
- NN b-tagging algorithms
- addition of a single loose SecVtx b-tag channel

which combined may net a $5 - 10\%$ sensitivity increase.

In the coming years, LHC searches for the SM Higgs boson may yield a discovery or further increase the experimentally excluded range of M_H. However, due to the increased background from $b\bar{b}$ QCD jet pairs, detection or exclusion of a signal consistent with $ZH \rightarrow \ell^+\ell^- b\bar{b}$ will be a challenge; meaning that Tevatron limits on $\sigma_{ZH}^{SM} \times BR(H \rightarrow b\bar{b})$ may stand for some time.

REFERENCES

[1] G. Arnison *et al.*, "Experimental observation of isolated large transverse energy electrons with associated missing energy at $s^{1/2}$ = 540-GeV," *Phys. Lett.*, vol. B122, pp. 103–116, 1983.

[2] G. Arnison *et al.*, "Experimental observation of lepton pairs of invariant mass around 95-GeV/c^2 at the CERN SPS collider," *Phys. Lett.*, vol. B126, pp. 398–410, 1983.

[3] P. Bagnaia *et al.*, "Evidence for Z0 → e+ e- at the CERN anti-p p collider," *Phys. Lett.*, vol. B129, pp. 130–140, 1983.

[4] S. Abachi *et al.*, "Observation of the top quark," *Phys. Rev. Lett.*, vol. 74, pp. 2632–2637, 1995, hep-ex/9503003.

[5] F. Abe *et al.*, "Observation of top quark production in $\bar{p}p$ collisions," *Phys. Rev. Lett.*, vol. 74, pp. 2626–2631, 1995, hep-ex/9503002.

[6] Tevatron Electroweak Working Group, "Combination of CDF and D0 Results on the Mass of the Top Quark," 2009, hep-ex/0903.2503.

[7] The LEP Electroweak Working Group. `http://lepewwg.web.cern.ch/LEPEWWG/`.

[8] C. Amsler *et al.*, "Review of particle physics," *Phys. Lett.*, vol. B667, p. 1, 2008.

[9] R. Barate *et al.*, "Search for the standard model Higgs boson at LEP," *Phys. Lett.*, vol. B565, pp. 61–75, 2003, hep-ex/0306033.

[10] D. E. Acosta *et al.*, "Measurement of the J/ψ meson and $b-$hadron production cross sections in $p\bar{p}$ collisions at $\sqrt{s}$ = 1960 GeV," *Phys. Rev.*, vol. D71, p. 032001, 2005, hep-ex/0412071.

[11] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967.

[12] O. W. Greenberg, "Spin and unitary spin independence in a paraquark model of baryons and mesons," *Phys. Rev. Lett.*, vol. 13, pp. 598–602, 1964.

[13] R. P. Feynman and M. Gell-Mann, "Theory of the fermi interaction," *Phys. Rev.*, vol. 109, pp. 193–198, 1958.

[14] F. J. Hasert *et al.*, "Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment," *Phys. Lett.*, vol. B46, pp. 138–140, 1973.

[15] Y. Fukuda *et al.*, "Evidence for oscillation of atmospheric neutrinos," *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, 1998, hep-ex/9807003.

[16] B. McElrath, "On The Origin of Neutrino Mass and Mixing in the Standard Model," 2009, hep-ph/0909.3090.

[17] R. N. Mohapatra and A. Y. Smirnov, "Neutrino Mass and New Physics," *Ann. Rev. Nucl. Part. Sci.*, vol. 56, pp. 569–628, 2006, hep-ph/0603118.

[18] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964.

[19] A. Salam and J. C. Ward, "On a gauge theory of elementary interactions," *Nuovo Cim.*, vol. 19, pp. 165–170, 1961.

[20] D. J. Griffiths, *Introduction to Elementary Particles*. USA:Wiley, New York, 1987.

[21] A. Pich, "The Standard model of electroweak interactions," 2007, hep-ph/0705.4264.

[22] K. Puolamaki, "On higgs processes in electron-positron-collider," Feb 1996, HU-SEFT-R-1995-17.

[23] F. Mazzucato, "Search for a standard model higgs at lhc using vector boson fusion," Prepared for 10th International Conference on Supersymmetry and Unification of Fundamental Interactions (SUSY02), Hamburg, Germany, 17-23 Jun 2002.

[24] S. F. Novaes, "Standard model: An introduction," 1999, hep-ph/0001283.

[25] G. Altarelli and G. Isidori, "Lower limit on the higgs mass in the standard model: An update," *Phys. Lett.*, vol. B337, pp. 141–144, 1994.

[26] J. F. Gunion, H. E. Haber, G. L. Kane, and S. Dawson, *The Higgs Hunter's Guide*. Addison-Wesley, New York, 1990.

[27] R. Barate *et al.*, "Search for the standard model higgs boson at lep," *Phys. Lett.*, vol. B565, pp. 61–75, 2003, hep-ex/0306033.

[28] P. Igo-Kemenes, "Searches for higgs bosons," *Journal of Physics G*, vol. 33, 2006.

[29] A. Heister *et al.*, "Final results of the searches for neutral Higgs bosons in e+ e- collisions at $s^{(1/2)}$ up to 209-GeV," *Phys. Lett.*, vol. B526, pp. 191–205, 2002, hep-ex/0201014.

[30] J. Abdallah *et al.*, "Final results from DELPHI on the searches for SM and MSSM neutral Higgs bosons," *Eur. Phys. J.*, vol. C32, pp. 145–183, 2004, hep-ex/0303013.

[31] C. Mariotti, "Search of the standard model higgs at lep," *Nuclear Physics B - Proceedings Supplements*, vol. 117, no. Supplement 1, pp. 202 – 205, 2003.

[32] G. Abbiendi *et al.*, "Search for the standard model Higgs boson with the OPAL detector at LEP," *Eur. Phys. J.*, vol. C26, pp. 479–503, 2003, hep-ex/0209078.

[33] A. L. Read, "Modified frequentist analysis of search results (the cl(s) method)," 2000, CERN-OPEN-2000-205.

[34] Tevatron New Phenomena, Higgs working group, "Combined CDF and DZero Upper Limits on Standard Model Higgs-Boson Production with up to 4.2 fb-1 of Data," 2009, hep-ex/0903.4001.

[35] T. Aaltonen *et al.*, "Search for the Higgs boson produced with $Z \to \ell^+\ell^-$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev. Lett.*, vol. 101, p. 251803, 2008, hep-ex/0807.4493.

[36] T. Aaltonen *et al.*, "A Search for the Higgs Boson Produced in Association with $Z \to \ell^+\ell^-$ Using the Matrix Element Method at CDF II," *Phys. Rev.*, vol. D80, p. 071101, 2009, hep-ex/0908.3534.

[37] D0 Collaboration, "Search for $ZH \to e^+e^-b\bar{b}$ and $ZH \to mu^+mu^-b\bar{b}$ Production in $4.2\,\mathrm{fb}^{-1}$ of data with the D0 Detector in $p\bar{p}$ Collistions at $\sqrt{s} = 1.96\,\mathrm{TeV}$," D0 Conference Note 5876 (2009).

[38] D. Crawford.
`http://www-bdnew.fnal.gov/operations/rookie_books/Tevatron_v2.3.pdf`.

[39] B. Evanger *et al.*
`http://www-bdnew.fnal.gov/operations/rookie_books/Concepts_v3.6.pdf`.

[40] E. Harms *et al.*
`http://www-bdnew.fnal.gov/operations/rookie_books/Pbar_V1_1.pdf`.

[41] P. T. Lukens, "The cdf iib detector: Technical design report," 2003, FERMILAB-TM-2198.

[42] J. Yoh.
`http://www-cdf.fnal.gov/events/detpic/CDFIsometric.jpg`.

[43] J. Yoh.
`http://www-cdf.fnal.gov/events/cdfintro.html`.

[44] D. Acosta *et al.*, "The performance of the CDF luminosity monitor," *Nucl. Instrum. Meth.*, vol. A494, pp. 57–62, 2002.

[45] D. Acosta *et al.*, "The CDF Cherenkov luminosity monitor," *Nucl. Instrum. Meth.*, vol. A461, pp. 540–544, 2001.

[46] J. Elias *et al.*, "Luminosity monitor based on Cherenkov counters for $p\bar{p}$ colliders," *Nucl. Instrum. Meth.*, vol. A441, pp. 366–373, 2000.

[47] C. S. Hill, "Operational experience and performance of the CDFII silicon detector," *Nucl. Instrum. Meth.*, vol. A530, pp. 1–6, 2004.

[48] A. Sill, "CDF Run II silicon tracking projects," *Nucl. Instrum. Meth.*, vol. A447, pp. 1–8, 2000.

[49] A. A. Affolder *et al.*, "Intermediate silicon layers detector for the CDF experiment," *Nucl. Instrum. Meth.*, vol. A453, pp. 84–88, 2000.

[50] A. Boveia, "Status and performance of the CDF Run II silicon detector," *PoS*, vol. HEP2005, p. 377, 2006.

[51] A. A. Affolder *et al.*, "CDF central outer tracker," *Nucl. Instrum. Meth.*, vol. A526, pp. 249–299, 2004.

[52] CDF Collaboration. `http://www-cdf.fnal.gov/internal/detectors/cdf_det.ps`.

[53] L. Balka *et al.*, "The CDF Central Electromagnetic Calorimeter," *Nucl. Instrum. Meth.*, vol. A267, p. 272, 1988.

[54] M. G. Albrow *et al.*, "The CDF plug upgrade electromagnetic calorimeter: Test beam results," *Nucl. Instrum. Meth.*, vol. A480, pp. 524–546, 2002.

[55] S. Bertolucci *et al.*, "The CDF Central and Endwall Hadron Calorimeter," *Nucl. Instrum. Meth.*, vol. A267, p. 301, 1988.

[56] A. Bhatti *et al.*, "Determination of the jet energy scale at the collider detector at Fermilab," *Nucl. Instrum. Meth.*, vol. A566, pp. 375–412, 2006, hep-ex/0510047.

[57] G. Ascoli *et al.*, "CDF Central Muon Detector," *Nucl. Instrum. Meth.*, vol. A268, p. 33, 1988.

[58] CDF Collaboration, "Histogram Tracking in the COT," CDF Internal Note 5562 (2001) [restricted document].

[59] CDF Collaboration, "Determination of the Run II COT Tracking Efficiency Using the W-NO-Track Sample," CDF Internal Note 8010 (2006) [restricted document].

[60] CDF Collaboration, "Determination of the Run II COT Tracking Efficiency Using the W-NO-Track Sample," CDF Internal Note 6866 (2004) [restricted document].

[61] C. P. Hays *et al.*, "Inside-out tracking at CDF," *Nucl. Instrum. Meth.*, vol. A538, pp. 249–254, 2005.

[62] CDF Collaboration, "Silicon Tracking for Plug Electrons," CDF Internal Note 5970 (2002) [restricted document].

[63] E. J. Thomson *et al.*, "Online track processor for the CDF upgrade," *IEEE Trans. Nucl. Sci.*, vol. 49, pp. 1063–1070, 2002.

[64] CDF Collaboration, "Electron Identification for Run II: Algorithms," CDF Internal Note 5456 (2003) [restricted document].

[65] CDF Collaboration, "Photon Conversion Removal Efficiency," CDF Internal Note 8073 (2006) [restricted document].

[66] CDF Collaboration, "Lateral Shower Size for Hadrons," CDF Internal Note 2081 (1993) [restricted document].

[67] G. C. Blazey and B. L. Flaugher, "Inclusive jet and dijet production at the Tevatron," *Ann. Rev. Nucl. Part. Sci.*, vol. 49, pp. 633–685, 1999, hep-ex/9903058.

[68] CDF Collaboration, "Generic Jet Energy Corrections for Run II data used for the Winter Conferences," CDF Internal Note 6280 (2003) [restricted document].

[69] CDF Top Group.
`http://www-cdf.fnal.gov/physics/new/top/top.html`.

[70] V. Sorin.
`http://www-cdf.fnal.gov/internal/physics/top/RunIITopProp/gen6Sum06/lumi_v28.html`.

[71] T. Sjöstrand *et al.*, "High-Energy-Physics Event Generation with PYTHIA 6.1," *Computer Phys. Commun.*, vol. 135, 2001.

[72] M. Mangano *et al.*, "ALPGEN, a generator for hard multiparton processes in hadronic collisions," *Journal of High Energy Physics*, vol. Issue 07, pp. 001, 2003.

[73] CDF Top Group.
`http://www-cdf.fnal.gov/internal/physics/top/RunIIMC/topmc6/index.shtml`.

[74] J. M. Campbell and R. K. Ellis, "An update on vector boson pair production at hadron colliders," *Phys. Rev.*, vol. D60, p. 113006, 1999, hep-ph/9905386.

[75] M. Cacciari *et al.*, "Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC," *JHEP*, vol. 09, p. 127, 2008, hep-ph/0804.2800.

[76] CDF Joint Physics Group.
`http://www-cdf.fnal.gov/internal/physics/joint_physics/instructions/JPScaleFactor`.

[77] J. Conway.
`http://www.physics.ucdavis.edu/~conway/research/higgs/smhiggs-tev.html`.

[78] T. Han and S. Willenbrock, "QCD correction to the p p → W H and Z H total cross-sections," *Phys. Lett.*, vol. B273, pp. 167–172, 1991.

[79] A. Djouadi, J. Kalinowski, and M. Spira, "HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension," *Comput. Phys. Commun.*, vol. 108, pp. 56–74, 1998, hep-ph/9704448.

[80] CDF B-Tag Group.
`http://www-cdf.fnal.gov/internal/physics/top/RunIIBtag/bTag.html`.

[81] CDF Collaboration, "Electron ID Efficiencies and Scale Factors for Periods 9 to 12," CDF Internal Note 9148 (2008) [restricted document].

[82] CDF Collaboration, "Lepton ID for multilepton Diboson Analyses," CDF Internal Note 8538 (2007) [restricted document].

[83] CDF Collaboration, "Forward Electron Tracking with the PhoenixMods Package," CDF Internal Note 6278 (2003) [restricted document].

[84] CDF Collaboration, "Electron Identification in offline release 6.1.2," CDF Internal Note 7950 (2006) [restricted document].

[85] CDF Top Group. `http://www-cdf.fnal.gov/internal/physics/top/feynman.shtml`.

[86] R. D. Harrington. PhD thesis, Northeastern U., 2007, FERMILAB-THESIS-2007-25.

[87] E. Palencia and S. Gristein. `http://www-cdf.fnal.gov/physics/new/top/2004/btag/`.

[88] A. Abulencia *et al.*, "Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$-TeV using Lepton + Jets Events with Jet Probability b^- tagging," *Phys. Rev.*, vol. D74, p. 072006, 2006, hep-ex/0607035.

[89] V. M. Abazov *et al.*, "Search for the standard model Higgs boson in the ZH → vvbb channel in 5.2 fb-1 of p-pbar collisions at sqrt(s)=1.96 TeV," 2009, hep-ex/0912.5285.

[90] T. Aaltonen *et al.*, "A Search for the Higgs Boson Using Neural Networks in Events with Missing Energy and b-quark Jets in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV," 2009, hep-ex/0911.3935.

[91] J. Pursley, "Search for Standard Model Higgs Boson in $H \to WW$ Channel at CDF," 2009, hep-ex/0910.1540.

[92] D. D. Price, "Search for Standard Model Higgs in WH→lvbb at the Tevatron," 2009, hep-ex/0910.3420.

[93] G. H. D.E. Rumelhart and R. Williams, "Learing Internal Representations by Error Propagation," *Parallel Distributed Processing: Explorations in Microstructure of Cognition*, vol. Vol 1, pp. 318–362, 1986.

[94] C. Peterson, T. Rognvaldsson, and L. Lonnblad, "JETNET 3.0: A Versatile artificial neural network package," *Comput. Phys. Commun.*, vol. 81, pp. 185–220, 1994.

[95] A. Abulencia *et al.*, "Top quark mass measurement from dilepton events at CDF II," *Phys. Rev. Lett.*, vol. 96, p. 152002, 2006, hep-ex/0512070.

[96] V. M. Abazov *et al.*, "A precision measurement of the mass of the top quark," *Nature*, vol. 429, pp. 638–642, 2004, hep-ex/0406031.

[97] T. Aaltonen *et al.*, "Search for a Higgs Boson Decaying to Two W Bosons at CDF," *Phys. Rev. Lett.*, vol. 102, p. 021802, 2009, hep-ex/0809.3930.

[98] P. J. Lujan. PhD thesis, U.C. Berkeley and LBL, Berkeley, 2009, FERMILAB-THESIS-2009-26.

[99] Campbell and Ellis.
`http://mcfm.fnal.gov/mcfm.pdf`.

[100] CDF Collaboration, "Measurement of the top quark mass in the dilepton channel 2.0/fb with the Matrix-Element Method," CDF Internal Note 9098 (2008) [restricted document].

[101] CDF Collaboration, "Multi-jet Energy corrections using a Neural Net," CDF Internal Note 8124 (2006) [restricted document].

[102] S. Richter. PhD thesis, Karlsruhe U., EKP, 2007, FERMILAB-THESIS-2007-35.

[103] CDF Collaboration, "A Search for ZH $\rightarrow$ llbb in 2.7 fb-1 using the Matrix-Element Method," CDF Internal Note 9718 (2009) [restricted document].

[104] T. Junk *et al.*
`https://plone4.fnal.gov:4430/P0/phystat/packages/0711001`.

[105] CDF Joint Physics Group.
http://www-cdf.fnal.gov/internal/physics/joint_physics/index.html.

[106] CDF Collaboration, "Upper Limits on Poisson Processes Incorporating Uncertainties in Acceptacne and Background," CDF Public Note 4476 (1998).

[107] The CDF Higgs Working Group.
http://www-cdf.fnal.gov/physics/new/hdg/Welcome.html.

[108] CDF Collaboration, "Combined Upper Limit on SM Higgs Boson Production," CDF Public Note 9999 (2009).

[109] Tevatron New Phenomena and Higgs Working Group.
http://tevnphwg.fnal.gov/.

ABSTRACT

A SEARCH FOR THE STANDARD MODEL HIGGS BOSON IN THE PROCESS $ZH \to \ell^+\ell^- b\bar{b}$ IN 4.1 fb^{-1} OF CDF II DATA

by

SHALHOUT Z. SHALHOUT

August 2010

Advisor: Robert Harr

Major: Physics

Degree: Doctor of Philosophy

This dissertation presents a search for the standard model (SM) Higgs boson in the associated production process $ZH \to \ell^+\ell^- b\bar{b}$ using 4.1 fb^{-1} of Tevatron $p\bar{p}$ collision data collected with the CDF II detector. To increase the sensitivity to a ZH signal over previous CDF searches [35, 36], we implement new electron and expanded b-jet identification algorithms. We utilize neural network classifiers enhanced with matrix element probabilities, a b-jet identifying neural network [102], and multivariate jet energy corrections to maximize the separation of signal from SM backgrounds.

We employ three neural network classifiers separately optimized for each of our three b-tag categories. We find good agreement between the observed data and the predicted SM backgrounds. The neural network output for data is compared to the output for the expected SM background to set 95% confidence level upper limits on the ZH production cross section times the branching ratio for $H \to b\bar{b}$. We consider Higgs boson masses between 100 and 150 GeV/c^2 in 5 GeV steps. For a Higgs boson mass of 115 GeV/c^2 we observe (expect) a 95% confidence level upper limit of 5.9 (6.8) times $\sigma_{ZH}^{SM} \times BR(H \to b\bar{b})$.

AUTOBIOGRAPHICAL STATEMENT

Shalhout Z. Shalhout

Wayne State University

Department of Physics and Astronomy

Detroit, MI 48201

Phone : (313) 269-0119

E-mail: ag5568@wayne.edu

Education

Ph.D. Physics, Wayne State University, *expected* 2010.

Thesis Topic: Search for $ZH \rightarrow \ell^+\ell^- b\bar{b}$ in CDF II Data

Thesis Advisor: Robert Harr

Thesis Committee: Robert Harr, Weisong Shi, Paul Karchin, and Alexey A. Petrov

M.S. Physics, Wayne State University, 2007.

B.S. Physics, Wayne State University, 2004. *Summa Cum Laude*

B.S. Mathematics, Wayne State University, 2004. *Summa Cum Laude*