

# Search for Excess Soft Leptons

S. Wilbur, D. Krop, C. Grosso-Pilcher, H. Frisch  
*University of Chicago*

## Abstract

We present a search for events with multiple low-energy leptons in associated production with a  $W$  boson. We find that . . .

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Signal Kinematics</b>	<b>4</b>
<b>3</b>	<b>Analysis Strategy</b>	<b>4</b>
<b>4</b>	<b>Datasets</b>	<b>5</b>
4.1	Data . . . . .	5
4.2	Monte Carlo . . . . .	6
<b>5</b>	<b>Non W Background Estimation Technique</b>	<b>11</b>
5.1	Antielectron Definition . . . . .	11
5.2	Fits to the Missing Transverse Energy Distribution . . . . .	12
<b>6</b>	<b>Event Selection</b>	<b>12</b>
6.1	Hard Lepton Identification . . . . .	12
6.2	W Selection . . . . .	12
6.2.1	$W \rightarrow e\nu$ events: Data and MC Predictions . . . . .	14
6.2.2	QCD Backgrounds in the $W \rightarrow e\nu$ Event Sample . . . . .	15
6.2.3	Summary of the Comparison of Observed vs Expected in $W \rightarrow e\nu$ Event Sample . . . . .	15
6.2.4	$W \rightarrow \mu\nu$ events . . . . .	16
6.3	Z Selection . . . . .	18
6.3.1	$Z \rightarrow e^+e^-$ . . . . .	18
6.3.2	$Z \rightarrow \mu^+\mu^-$ . . . . .	20
6.4	Summary of Observed and Predicted Numbers of W's and Z's . . . . .	22
6.5	Checking the Ratio of W to Z production 'R', as a Function of Time . . . . .	23
6.5.1	Electron-Triggered Modes . . . . .	23
6.5.2	Muon-Triggered Modes . . . . .	24
6.6	Analysis Sample Selection . . . . .	24
<b>7</b>	<b>Soft Lepton Identification</b>	<b>24</b>
7.1	Soft Electrons . . . . .	24
7.1.1	Training Sample Selection . . . . .	24
7.1.2	Discriminating Variables . . . . .	25
7.1.3	Efficiency and Fake Rate Matrices . . . . .	27
7.1.4	Validation . . . . .	28
7.2	Soft Muons . . . . .	28
<b>8</b>	<b>Track/<math>\pi^0</math> Clustering</b>	<b>28</b>
8.1	Clustering Parameters . . . . .	30
8.2	Single and Di-Tau Likelihoods Ratios . . . . .	30

<b>9 Results</b>	<b>34</b>
9.1 Object Counts . . . . .	34

# 1 Introduction

Various models have been proposed to avoid the large fine-tuning difficulties of the MSSM [ref. needed]. One attractive model that solves the fine-tuning problem is the next to minimal supersymmetric model (NMSSM) [ref. needed]. The NMSSM contains an additional scalar superfield . . . This model is also consistent with the deviation near  $m(H) = 100$  GeV observed in the combined LEP standard model Higgs search as seen in fig. 1.

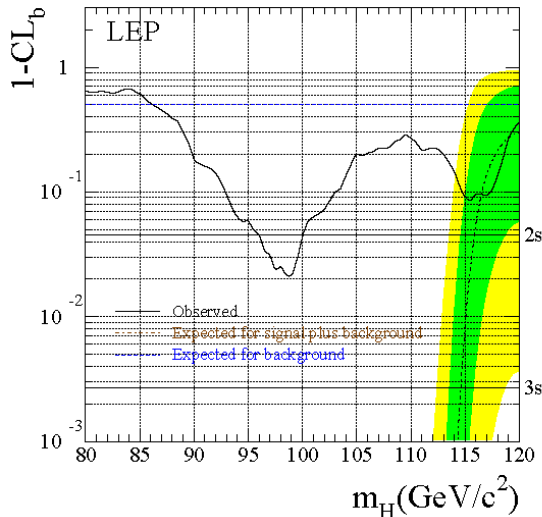


Figure 1: The LEP combined  $1 - CL_b$  plot for standard model Higgs production in  $b\bar{b}$  final states [ref. needed]. Note that the largest deviation from the background hypothesis occurs for  $97 < m(H) < 100$  GeV. This deviation is inconsistent with standard model Higgs production.

We search for the final state predicted to dominate in this model,  $H \rightarrow a^0 a^0 \rightarrow 4\tau$ , as seen in fig. 2.

## 2 Signal Kinematics

A Monte Carlo sample of 100000  $WH$  events was generated using Pythia with  $m(H) = 90$  GeV and  $m(a^0) = 9$  GeV. The  $W$  boson was forced to decay leptonically. (Discuss the colinearity and of the tau's and how this affects tau ID. Discuss the softness of the tau lepton products and how this leads to soft lepton ID).

## 3 Analysis Strategy

We trigger on the lepton coming from the  $W$  so that we start from a well-understood sample.  $W$  and  $Z$  boson events are reconstructed and full background estimations are

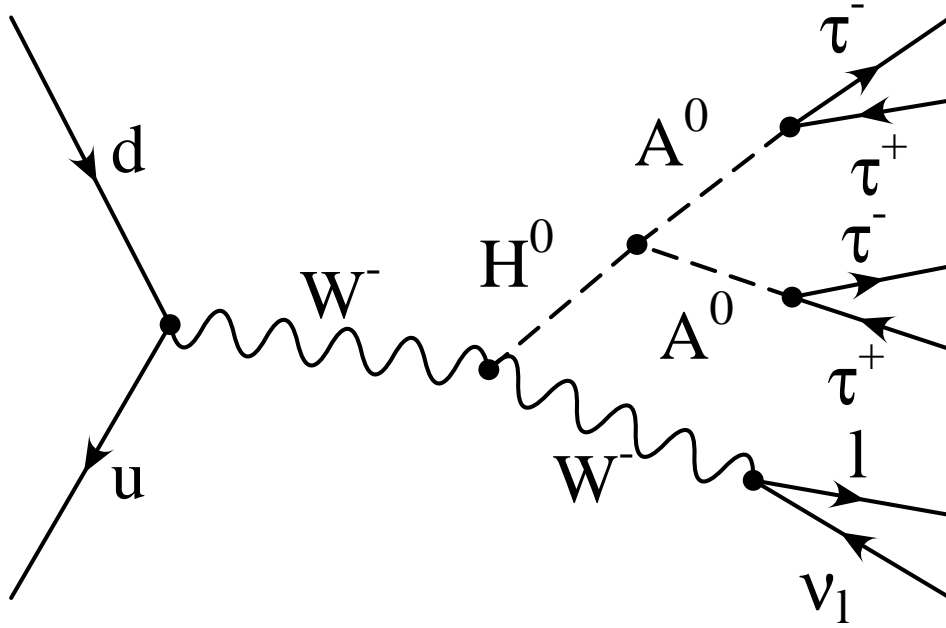


Figure 2: A schematic of the NMSSM Higgs decay searched for in this analysis.

developed for these generic base samples. We validate the high- $p_T$  lepton identification and background estimates using these inclusive  $W$  and  $Z$  events. We then develop specialized identification tools for the collinear and soft  $\tau$  decay topology that we are interested in. These tools are:

- soft electron identification,
- soft muon identification,
- and track- $\pi^0$  clustering.

Because the  $\tau$  production is collinear, the track- $\pi^0$  clustering is tuned to identify hadronic decays of both single and di-tau objects. This tuning is accomplished via the development of two likelihoods: one to discriminate jets versus single taus and another to discriminate jets versus di-taus.

Next we examine control samples obtained by reversing the track- $\pi^0$  likelihood selection to obtain samples that are signal-free. Finally, we look in the signal region for unusual combinations of our three objects that have good signal-to-background ratios. For example, we expect a signal-to-background ratio of XXX for the subsample containing a soft electron, same-sign soft muon, and cluster likely to be from a di-tau decay. By further separating subsamples in which the soft lepton overlaps or does not overlap with a cluster, we end up with many types of subsamples that we can classify by their expected signal-to-background ratio. We then take the most signal-like of these bins and search for an excess.

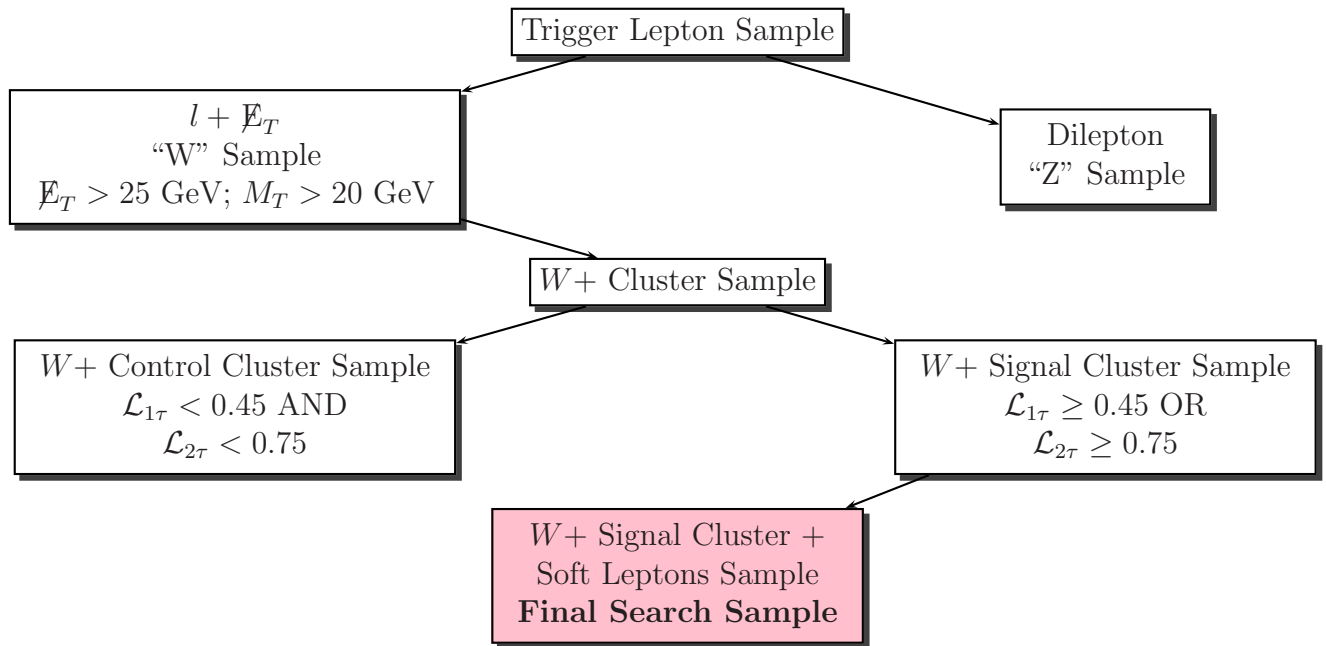


Figure 3: Schematic representation of the various samples used in this analysis. The “Z” sample is used to validate the lepton identification and energy scale. The “W+ Cluster Sample” is used to normalize the background prediction. The “W+ Control Cluster Sample” is used to validate the track- $\pi^0$  cluster and soft lepton likelihood modeling. We search for an excess over the SM background prediction in the “Final Search Sample”.

## 4 Datasets

### 4.1 Data

The analysis uses events selected with the triggers “ELECTRON\_CENTRAL\_18”, “MUON\_CMUP18” and “MUON\_CMX18”. The datasets, runs, run periods, and luminosities are listed in Table 1. The numbers of electron- and muon-triggered events are 60555685 and 19335273, respectively.

### 4.2 Monte Carlo

The main backgrounds are  $W + \text{jets}$ , Drell-Yan,  $Z \rightarrow \tau\tau$ ,  $t\bar{t}$ , single top, and diboson production. We have modeled these using Alpgen or Pythia. The individual W and Drell-Yan datasets for Alpgen are shown in Tables 2 and 3, and the Pythia datasets are shown in Table 4.

Dataset	Runs	Period	Luminosity	
Dataset	Period	Runs	Luminosity	Dates
0h	1	190697 – 195408	363pb <sup>-1</sup>	7 Dec 04 - 18 Mar 05
	2	195409 – 198379		19 Mar 05 - 20 May 05
	3	198380 – 201349		21 May 05 - 19 Jul 05
	4	201350 – 203799		20 Jul 05 - 4 Sep 05
0i	5	203819 – 206989	587pb <sup>-1</sup>	5 Sep 05 - 9 Nov 05
	6	206990 – 210011		10 Nov 05 - 14 Jan 06
	7	210012 – 212133		14 Jan 06 - 22 Feb 06
	8	217990 – 222426		9 Jun 06 - 1 Sep 06
	9	222529 – 228596		1 Sep 06 - 22 Nov 06
0j	10	228664 – 233111	945pb <sup>-1</sup>	24 Nov 06 - 30 Jan 07
	11	233133 – 237795		30 Jan 07 - 31 Mar 07
	12	237845 – 241664		1 Apr 07 - 13 May 07
	13	241665 – 246231		13 May 07 - 4 Aug 07
0k	14	252836 – 254686	484pb <sup>-1</sup>	28 Oct 07 - 3 Dec 07
	15	254800 – 256824		5 Dec 07 - 27 Jan 08
	16	256840 – 258787		27 Jan 08 - 27 Feb 08
	17	258880 – 261005		28 Feb 08 - 16 Apr 08
0m	18	261119 – 264071	1216pb <sup>-1</sup>	18 Apr 08 - 1 Jul 08
	19	264101 – 266513		1 Jul 08 - 24 Aug 08
	20	266528 – 267718		24 Aug 08 - 4 Oct 08
	21	268155 – 271047		12 Oct 08 - 1 Jan 09
	22	271072 – 272214		2 Jan 09 - 10 Feb 09

Table 1: Datasets used in the analysis. Both muon and electron datasets are used, i.e. *0h* refers to bhmubh and bhelbh. The luminosities of the electron and muon datasets are the same because we use unrescaled triggers in this analysis.

Process	Cross Section	Dataset (low-lumi)	Events Generated (low-lumi)	dataset (high-lumi)
$W(e\nu) + 0p$	1.8 nb	pt0sw0	4929337	ut0s00
$W(e\nu) + 1p$	225 pb	pt0sw1	4909767	ut0s01
$W(e\nu) + 2p$	35.3 pb	pt0s2w	918835	ut0s02
$W(e\nu) + 3p$	5.59 pb	pt0s3w	783415	ut0s03
$W(e\nu) + \geq 4p$	1.03 pb	pt0s4w	453531	ut0s04
$W(\mu\nu) + 0p$	1.8 nb	pt0sw5	5010637	ut0s05
$W(\mu\nu) + 1p$	225 pb	pt0sw6	4997783	ut0s06
$W(\mu\nu) + 2p$	35.3 pb	pt0s7w	877801	ut0s07
$W(\mu\nu) + 3p$	5.59 pb	pt0s8w	817043	ut0s08
$W(\mu\nu) + \geq 4p$	1.03 pb	pt0s9w	906265	ut0s09
$W(\tau\nu) + 0p$	1.80 nb	ut0sw0	4868422	ut0s10
$W(\tau\nu) + 1p$	225 pb	ut0sw1	4981403	ut0s11
$W(\tau\nu) + 2p$	35.4 pb	ut0s2w	917094	ut0s12
$W(\tau\nu) + 3p$	5.6 pb	ut0s3w	1008221	ut0s13
$W(\tau\nu) + \geq 4p$	1.03 pb	ut0s4w	986494	ut0s14

Table 2: Alpgen  $W$  + partons dataset names, processes, cross sections, and number of events generated. Note that the cross-sections listed do not include a  $K$ -factor.

Process	Mass Range (GeV)	Cross Section	Dataset (low-lumi)	Events Generated (low-lumi)
$Z(ee) + 0p$	[75 : 105]	158 pb	zt0sp0	2639520
$Z(ee) + 1p$	[75 : 105]	21.6 pb	zt0sp1	2624793
$Z(ee) + 2p$	[75 : 105]	3.46 pb	zt0szb	4595453
$Z(ee) + 3p$	[75 : 105]	0.55 pb	zt0s3p	524261
$Z(ee) + \geq 4p$	[75 : 105]	99.2 fb	zt0s4p	525065
$Z(\mu\mu) + 0p$	[75 : 105]	158 pb	zt0sp5	2659832
$Z(\mu\mu) + 1p$	[75 : 105]	21.6 pb	zt0sp6	2652428
$Z(\mu\mu) + 2p$	[75 : 105]	3.46 pb	zt0szt	4660506
$Z(\mu\mu) + 3p$	[75 : 105]	0.55 pb	zt0s8p	536159
$Z(\mu\mu) + \geq 4p$	[75 : 105]	99.2 fb	zt0s9p	530242
$Z(\tau\tau) + 0p$	[75 : 105]	158 pb	zt0st3	5860164
$Z(\tau\tau) + 1p$	[75 : 105]	21.5 pb	zt0st4	5722772
$Z(\tau\tau) + \geq 2p$	[75 : 105]	4.14 pb	zt0st2	2263107
$DY(ee) + 0p$	[8 : 20]	1514 pb	zt0sl0	531063
$DY(ee) + 1p$	[8 : 20]	19.7 pb	zt0sl1	530980
$DY(ee) + 2p$	[8 : 20]	6.98 pb	zt0sl2	519852
$DY(ee) + 0p$	[20 : 75]	160 pb	xt0s0p	536159
$DY(ee) + 1p$	[20 : 75]	8.39 pb	xt0s1p	530958
$DY(ee) + 2p$	[20 : 75]	1.61 pb	xt0s2p	536159
$DY(ee) + 3p$	[20 : 75]	233 fb	xt0s3p	525670
$DY(ee) + \geq 4p$	[20 : 75]	39.8 fb	xt0s4p	515638
$DY(ee) + 0p$	[105 : 600]	4.07 pb	yt0s0p	519104
$DY(ee) + 1p$	[105 : 600]	705 fb	yt0s1p	524895
$DY(ee) + 2p$	[105 : 600]	117 fb	yt0s2p	513214
$DY(ee) + 3p$	[105 : 600]	18.5 fb	yt0s3p	504749
$DY(\mu\mu) + 0p$	[8 : 20]	1514 pb	zt0sm0	530855
$DY(\mu\mu) + 1p$	[8 : 20]	19.7 pb	zt0sm1	525713
$DY(\mu\mu) + 2p$	[8 : 20]	6.98 pb	zt0sm2	530561
$DY(\mu\mu) + 0p$	[20 : 75]	160 pb	xt0s5p	519237
$DY(\mu\mu) + 1p$	[20 : 75]	8.4 pb	xt0s6p	530696
$DY(\mu\mu) + 2p$	[20 : 75]	1.6 pb	xt0s7p	520703
$DY(\mu\mu) + 3p$	[20 : 75]	233 fb	xt0s8p	509424
$DY(\mu\mu) + \geq 4p$	[20 : 75]	39.8 fb	xt0s9p	523932
$DY(\mu\mu) + 0p$	[105 : 600]	4.07 pb	yt0s5p	530941
$DY(\mu\mu) + 1p$	[105 : 600]	706 fb	yt0s6p	529581
$DY(\mu\mu) + 2p$	[105 : 600]	117 fb	yt0s7p	531006
$DY(\mu\mu) + 3p$	[105 : 600]	18.5 fb	yt0s8p	510246

Table 3: Alpgen Drell-Yan + partons dataset names, processes, cross sections, and number of events generated. The mass range column refers to the generated mass range of the dilepton pair. Note that the cross-sections listed do not include a  $K$ -factor.

Dataset	Process	Cross Section	Events Generated
we0sge	$W(e\nu) + \text{jets}$	2687 pb	11664464
we0s8m	$W(\mu\nu) + \text{jets}$	2687 pb	6923822
we0sat	$W(\tau\nu) + \text{jets}$	2687 pb	6931813
ze1sad	Drell-Yan	493 pb	6931813
ze1s9m	Drell-Yan	493 pb	6937820
ze0sat	$Z \rightarrow \tau\tau$	333 pb	6916050
tt0s75	$t\bar{t}$	7.0 pb	4730477
st0s00	Single Top (t-channel)	0.88 pb	917908
it0sww	Diboson (WW)	12.4 pb	2291309
it0swz	Diboson (WZ)	3.65 pb	2328823
it0szz	Diboson (ZZ)	3.8 pb	2319470

Table 4: Pythia dataset names, processes, cross sections, and number of events generated. The cross-sections listed include a  $K$ -factor of 40%.

## 5 Non W Background Estimation Technique

The  $W$  boson is identified by the presence of a high energy lepton and missing transverse energy. Events containing jets may emulate this signature; a dijet event, for example, may have large  $\cancel{E}_T$  arising from the energy mismeasurement of one jet while the other jet in the event can mimic an electron by leaving a track in the COT associated with an electromagnetic energy deposit. We estimate this “non-W” background by using a data-derived model for these kind of events [3]. This is accomplished by defining an object that is similar to electrons, but has a much larger rate of contamination from jets, labelled “antielectrons”. These objects are then used to create a template which is used to model the non-W background in a fit of the  $\cancel{E}_T$  distribution.

### 5.1 Antielectron Definition

Following the technique described in Ref.[3], we divide the the standard CDF high- $p_T$  electron identification selections [4] into two categories, as shown in Table 5.1. The cuts designed mainly to discriminate electrons from misidentified jets we label as “Identification Cuts” while the selections used enrich the sample with real electrons are labelled “Kinematic Cuts”.

<b>Kinematic Cuts</b>
$\eta_{\text{Detector}} \leq 1.1$ (central) Track must be fiducial to CES $E_T \geq 20$ GeV $p_T \geq 10$ GeV $z_0^{\text{track}} \leq 60$ cm $E/p \leq 2$ unless $p_T \geq 50$ GeV $\text{Iso}(R = 0.4)/E_T \leq 0.1$ Conversion Removal
<b>Identification Cuts</b>
$E(HAD)/E(EM) \leq 0.055 + 0.00045 \times E$ $\chi_{CES}^2 \leq 10$ $L_{shr} \leq 0.2$ $-3.0 \leq Q \times \Delta X_{CES} \leq 1.5$ cm $ \Delta Z_{CES}  < 3$ cm

Table 5: Central tight electron identification cuts [4] divided into two categories: those that enrich the sample with electrons, referred to as “kinematic”, and those that discriminate electrons from misidentified jets, referred to as “identification”.

We define the antielectron sample by requiring that prospective objects must pass all kinematic cuts while simulataneously failing at least two of the identification cuts. This sample has similar kinematics to the high- $p_T$  electron sample but has many fewer real electrons present in it.

## 5.2 Fits to the Missing Transverse Energy Distribution

We obtain the number of events that arise from non- $W$  sources by fitting the  $\cancel{E}_T$  distribution of the data using two templates: an electroweak template obtained from  $W$ + jets,  $Z$ + jets, and diboson Monte Carlo and a QCD template obtained from the anti-electron sample. To obtain the QCD template, we take the anti-electron sample and subtract the expected  $W$  contamination using electroweak Monte Carlo. Figure 4 shows a fit to the electron data using these two templates for events with  $p_T(e) > 20$  GeV,  $m_T > 20$  GeV, and  $\Delta\phi(\cancel{E}_T, e) > 0.5$ . After we obtain the number of non- $W$  events in this sample, we extrapolate this number to the  $W$  signal region with  $\cancel{E}_T > 25$  GeV.

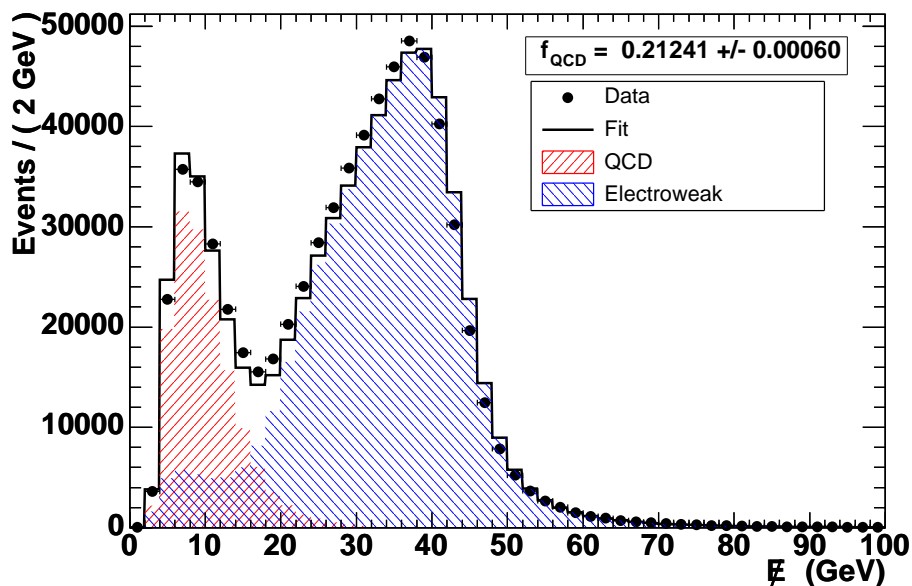


Figure 4: A fit to the  $\cancel{E}_T$  distribution of electron events with  $p_T(e) > 20$  GeV,  $m_T > 20$  GeV, and  $\Delta\phi(\cancel{E}_T, e) > 0.5$ . The “Electroweak” template is obtained from Monte Carlo and the “QCD” template is obtained from the anti-electron data sample. The fit returns a fraction of the events that are non- $W$ ,  $f_{QCD}$ .

## 6 Event Selection

The data encompass CDF Periods 1 to 22, covering the calendar period Dec. 7, 2004 to Feb. 10, 2009. The total luminosity for this dataset is  $3.9 \text{ fb}^{-1}$ .

We use the  $W$  and  $Z$  events in our sample as controls because the cross-sections and kinematics are precisely predicted, and to make sure that we have applied all the correct identification efficiencies and scale factors. In particular, the ratio  $R$  of  $W$  to  $Z$  production is predicted to NNLO with a precision of a few percent [1], providing a very precise test of trigger and lepton efficiencies in the Monte Carlo predictions.

Tight Central Electron Cuts
$\eta_{\text{Detector}} \leq 1.1$
Track must be fiducial to CES
$E_T \geq 20 \text{ GeV}$
$E_{HAD}/E_{EM} \leq 0.055 + 0.00045 \times E$
$\text{Iso}(R = 0.4)/E_T \leq 0.1$
$p_T \geq 10 \text{ GeV}$
$\geq 3$ COT axial segments with $\geq 5$ hits
$\geq 2$ COT stereo segments with $\geq 5$ hits
$z_0^{\text{track}} \leq 60 \text{ cm}$
$E/p \leq 2$ unless $p_T \geq 50 \text{ GeV}$
$\chi_{CES}^2 \leq 10$
$-3.0 \text{ cm} \leq Q \times \Delta X_{CES} \leq 1.5 \text{ cm}$
$ \Delta Z_{CES}  < 3 \text{ cm}$
$L_{shr} \leq 0.2$
Conversion Removal

Table 6: Cuts to identify tight central electrons

Loose Central Electron Cuts
$\eta_{\text{Detector}} \leq 1.1$
Track must be fiducial to CES
$E_T \geq 12 \text{ GeV}$
$E_{HAD}/E_{EM} \leq 0.055 + 0.00045 \times E$
$\text{Iso}(R = 0.4)/E_T \leq 0.1$
$p_T \geq 6 \text{ GeV}$
$\geq 3$ COT axial segments with $\geq 5$ hits
$\geq 2$ COT stereo segments with $\geq 5$ hits
$z_0^{\text{track}} \leq 60 \text{ cm}$
$E/p \leq 2$ unless $p_T \geq 50 \text{ GeV}$
$\chi_{CES}^2 \leq 10$
$-3.0 \text{ cm} \leq Q \times \Delta X_{CES} \leq 1.5 \text{ cm}$
$ \Delta Z_{CES}  < 3 \text{ cm}$
$L_{shr} \leq 0.2$

Table 7: Cuts to identify loose central electrons

## 6.1 Hard Lepton Identification

We use the standard CDF cuts to identify hard ( $> 20 \text{ GeV}$ ) electrons and muons, with the additional requirement that hard muons have silicon hits. We are interested in tight central electrons and CMUP and CMX muons. We are only interested in looser categories (loose central electrons and CMU, CMP, and CMIO muons) in events which already have a tight lepton.

Tight Muon Cuts
Track must be fiducial to CMU, CMP, or CMX
$E_{EM} \leq 2 \text{ GeV}$
$E_{HAD} \leq 6 \text{ GeV}$
$E_{EM} + E_{HAD} \geq 0.1 \text{ GeV}$ for CMIO muons
$\text{Iso}(R = 0.4)/E_T \leq 0.1$
$p_T \geq 10 \text{ GeV}$
$\geq 3$ COT axial segments with $\geq 5$ hits
$\geq 2$ COT stereo segments with $\geq 5$ hits
$\geq 1$ Si hit
$z_0^{\text{track}} \leq 60 \text{ cm}$
$d_0 < 0.02 \text{ cm}$
$\Delta X_{CMU} \leq 3 \text{ cm}$
$\Delta X_{CMP} \leq 5 \text{ cm}$
$\Delta X_{CMX} \leq 6 \text{ cm}$

Table 8: Cuts to identify tight muons

## 6.2 W Selection

The  $W$  selection requires a trigger lepton (see Section 4.1) with  $E_T > 20 \text{ GeV}$ ,  $\cancel{E}_T > 25 \text{ GeV}$ , and  $m_T(l, \cancel{E}_T) > 20 \text{ GeV}$ . The difference in  $\phi$  between the highest-energy lepton and the  $\phi$  of the  $\cancel{E}_T$  is required to be greater than 0.5.

In addition, electrons are required to satisfy the conversion filter, and muons the decay-in-flight algorithm, as described below.

### 6.2.1 $W \rightarrow e\nu$ events: Data and MC Predictions

In addition to passing the kinematic and geometric cuts above, the highest- $E_T$  electron is required to pass a conversion filter. If there is an opposite-sign track in the event that is an electron candidate and has  $\Delta \cot \theta < 0.04$  and  $|sep| < 0.2$ , then the electron is called a conversion. However, if its matching conversion leg also has another conversion partner, then the original candidate is called a trident, and therefore passes the conversion filter. Tight electrons that pass these cuts are defined as ‘good’ electrons.

Table 9 gives the event counts for the  $W \rightarrow e\nu$  cuts applied sequentially, and also the number of events which pass all other cuts except that cut, for both Monte Carlo and data.

### 6.2.2 QCD Backgrounds in the $W \rightarrow e\nu$ Event Sample

We apply the technique described in Section 5 to the 60555685 electron-triggered events, of which 105164 form the anti-electron sample by passing the kinematic cuts and failing at least two identification cuts. We predict 1774 events of the final  $W$  selection (the 636119 events in the last line of Table 9) are due to QCD backgrounds.

Cut	$N_{seq}$ (MC)	$N_{N-1}$ (MC)	$N_{seq}$ (Data)	$N_{N-1}$ (Data)
Number of Triggered Events	NA		60555685	
Tight Electron	2721644	90	3498809	74
Electron is not Conversion	2689656	17479	2503939	25623
$\delta\phi(e, \cancel{E}_T) > 0.5$	2603351	461	2385138	189
$\cancel{E}_T > 25$	2104805	382038	1833721	335678
$m_T > 20$	2103973	832	1833188	533

Table 9: The numbers of  $W$  events with a tight electron which pass each cut in sequence and which pass every cut except a particular one for both data and Monte Carlo (see Tables 2, 3 and 4.) The left-hand set of numbers is the prediction from summing the SM Electroweak Alpgen  $W$  and  $Z$  MC datasets and the Pythia top and diboson datasets as described in Section 4.1 Note that the QCD background estimate is not yet included.

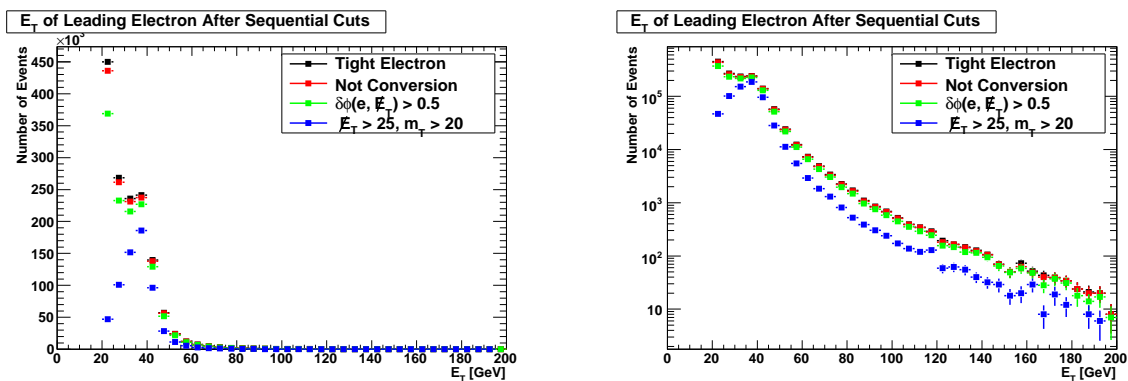


Figure 5: The spectrum in  $E_T$  for the electron data sample after the sequential application of the cuts in Table 9.

### 6.2.3 Summary of the Comparison of Observed vs Expected in $W \rightarrow e\nu$ Event Sample

After adding the QCD background to the predictions, we expect 667432  $W \rightarrow e\nu$  events versus the observed 636119 events. Table 6.2.4 summarizes the predicted numbers of events.

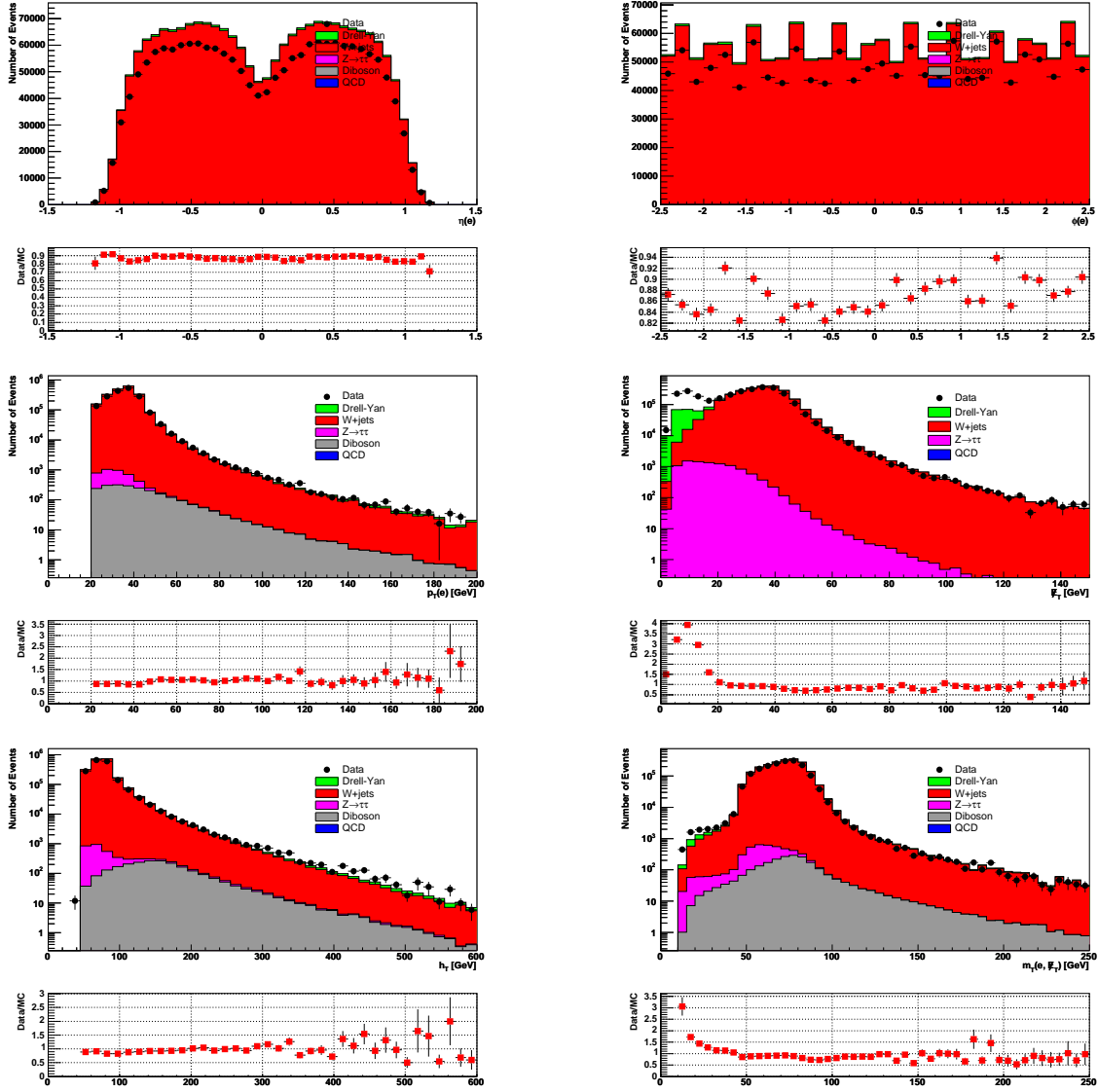


Figure 6:  $W$  validation plots, electron trigger:  $\eta$ ,  $\phi$  and  $E_T$  of the highest- $E_T$  good electron,  $\cancel{E}_T$  and  $h_T$  in the event,  $m_T$  of electron and  $\cancel{E}_T$ . There are cuts requiring  $\cancel{E}_T > 25$  GeV (except on the  $\cancel{E}_T$  plot) and  $m_T > 20$  GeV (except on the  $m_T$  plot)

Source	Number of Events
$W$ +jets	2114160
Drell-Yan	18413
$Z \rightarrow \tau\tau$	2558
$t\bar{t}$	1400
Single Top	284
Diboson	2634
QCD	5114
Expected total	2144560
Observed	1833180

Table 10: Summary of the predictions for  $W^\pm \rightarrow e^\pm\nu$ .

### 6.2.4 $W \rightarrow \mu\nu$ events

Table 11 gives the event counts for the  $W \rightarrow \mu\nu$  cuts applied sequentially, and also the number of events which pass all other cuts except that cut, for both Monte Carlo and data. An event passes the the decay-in-flight (DIF) removal algorithm if at least one muon that passes the previous cuts is not tagged as a DIF (and is therefore a ‘good’ muon). The DIF algorithm requires the  $\chi^2/\text{d.f.}$  of the track fit to be less than 3 and requires that the impact parameter of the track be less than 0.02 cm if the track has silicon hits, or less than 0.1 cm if the track doesn’t have silicon hits.

Cut	$N_{seq}$ (MC)	$N_{N-1}$ (MC)	$N_{seq}$ (Data)	$N_{N-1}$ (Data)
Number of Triggered Events	NA		19335273	
Tight CMUP Muon	1387902	101	1518028	115
Muon is not DIF	1268505	66208	1294527	114137
$\delta\phi(\mu, \cancel{E}_T) > 0.5$	1252327	285	1259405	318
$\cancel{E}_T > 25$	1035584	187398	924937	259444
$m_T > 20$	1035265	319	924564	373
Number of Triggered Events	NA		19335273	
Tight CMX Muon	901051	82	1168023	60
Muon is not DIF	858988	18975	917130	133795
$\delta\phi(\mu, \cancel{E}_T) > 0.5$	848236	132	888182	151
$\cancel{E}_T > 25$	708016	120388	640591	187424
$m_T > 20$	707819	197	640331	260

Table 11: The numbers of W events with a tight muon which pass each cut in sequence and which pass every cut except a particular one for both data and Monte Carlo (see Tables 2, 3 and 4.) The left-hand set of numbers is the prediction from summing the SM Electroweak Alpgen W and Z MC datasets and the Pythia top and diboson datasets as described in Section 4.1.

Source	Number of Events
$W$ +jets	1057740
Drell-Yan	105051
$Z \rightarrow \tau\tau$	1408
$t\bar{t}$	775
Single Top	153
Diboson	1419
QCD	593
Expected total	1167140
Observed	1023620

Table 12: Summary of the predictions for  $W^\pm \rightarrow \mu^\pm\nu$  with a CMUP trigger

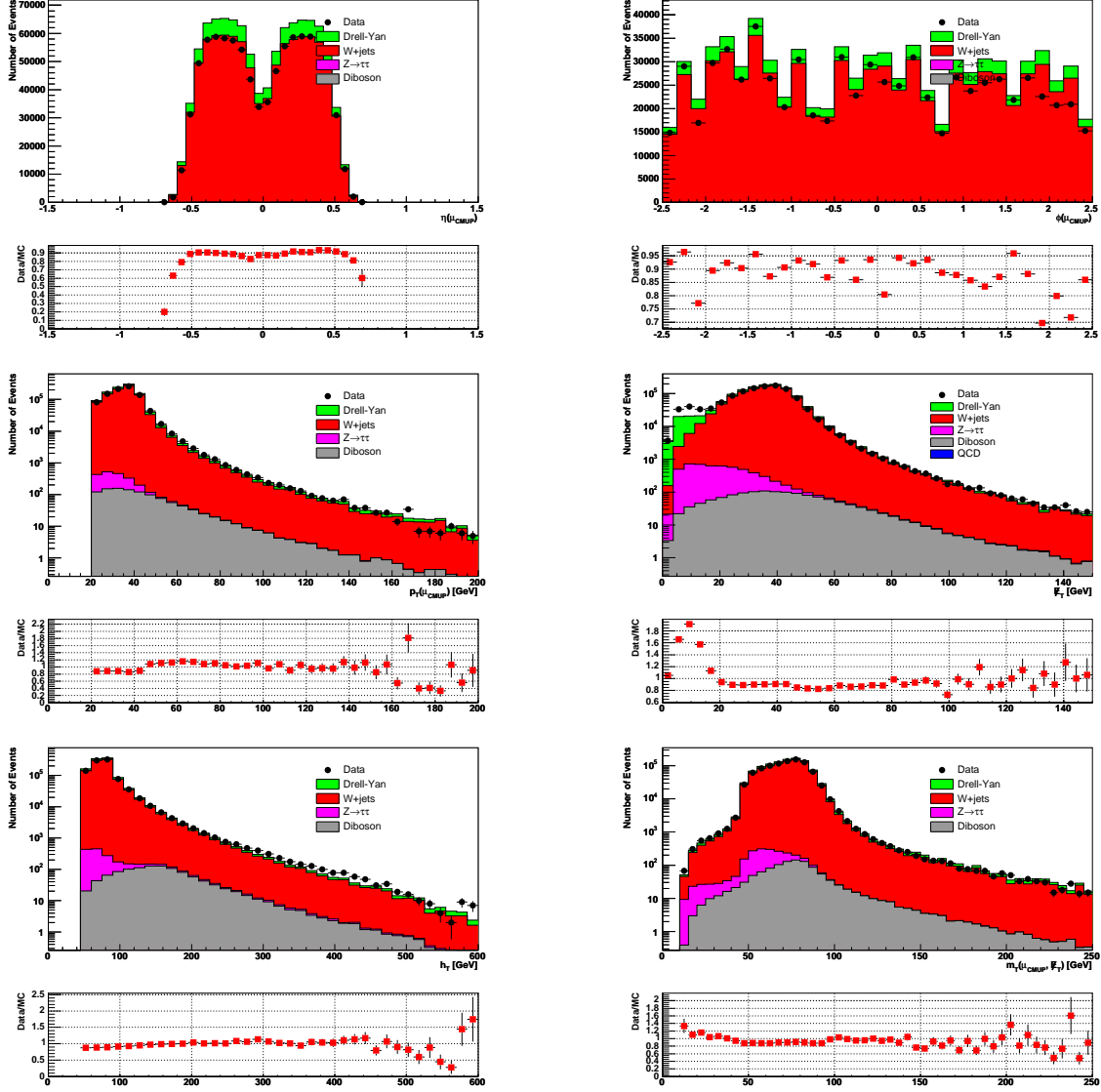


Figure 7:  $W$  validation plots, CMUP trigger:  $\eta$ ,  $\phi$  and  $p_T$  of the highest- $E_T$  good muon,  $\cancel{E}_T$  and  $h_T$  in the event,  $m_T$  of the good highest- $E_T$  muon and  $\cancel{E}_T$ . There are cuts requiring  $\cancel{E}_T > 25$  GeV (except on the  $\cancel{E}_T$  plot) and  $m_T > 20$  GeV (except on the  $m_T$  plot)

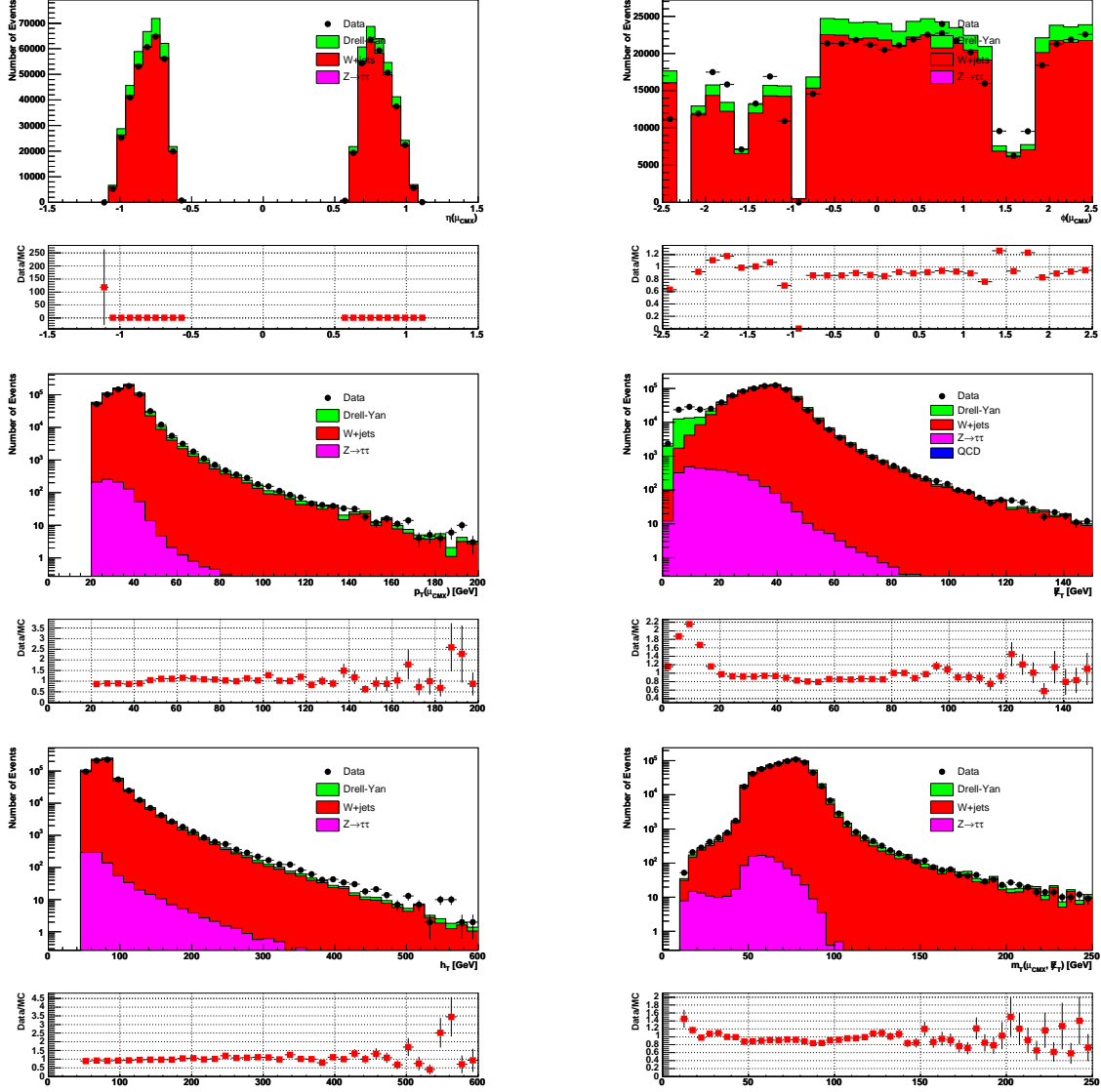


Figure 8: W validation plots, CMX trigger:  $\eta$ ,  $\phi$  and  $p_T$  of the highest- $E_T$  good muon,  $\cancel{E}_T$  and  $h_T$  in the event,  $m_T$  of the good highest- $E_T$  muon and  $\cancel{E}_T$ . There are cuts requiring  $\cancel{E}_T > 25$  GeV (except on the  $\cancel{E}_T$  plot) and  $m_T > 20$  GeV (except on the  $m_T$  plot)

Source	Number of Events
$W$ +jets	666291
Drell-Yan	63845
$Z \rightarrow \tau\tau$	874
$t\bar{t}$	354
Single Top	72
Diboson	775
QCD	447
Expected total	732663
Observed	677994

Table 13: Summary of the predictions for  $W^\pm \rightarrow \mu^\pm\nu$  with a CMX trigger

### 6.3 Z Selection

The  $Z$  selection requires two opposite-sign leptons with  $p_T > 20$  GeV, of which one is tight (Tables 6 and 8), and the other can be loose (Table 7).

#### 6.3.1 $Z \rightarrow e^+e^-$

Table 14 gives the event counts for the cuts applied sequentially, and also the number of events which pass all other cuts except that cut, for both Monte Carlo and data.

The conversion filter is applied to both electrons.

Cut	$N_{seq}$ (MC)	$N_{N-1}$ (MC)	$N_{seq}$ (Data)	$N_{N-1}$ (Data)
Number of Triggered Events	NA		60555685	
Tight Electron	1862007	0	2271874	0
Electron is not Conversion	1861847	0	1308755	0
Second Electron	94486	0	83964	0
Second Electron is not Conversion	91677	0	82809	0
Opposite Sign Pair	91308	369	82531	278

Table 14: Number of events which pass each cut in sequence and which pass every cut except a particular one. Note that the cut of no additional leptons (e or  $\mu$ ) is applied here for the Z control sample, but is not applied below in the search. Also the QCD background estimate is not yet included.

We make plots in two dielectron categories: After selecting a tight electron, we look for either a tight or loose second electron. Figures 9 and 13 show plots of the first (tight) and second (tight or loose) electrons respectively.

Source	Number of Events
$W$ +jets	39
Drell-Yan	76815
$Z \rightarrow \tau\tau$	446
$t\bar{t}$	50
Single Top	0
Diboson	171
QCD	0
Expected total	77523
Observed	68949

Table 15: Summary of the predictions for  $Z^0 \rightarrow e^+e^-$  with both electrons tight

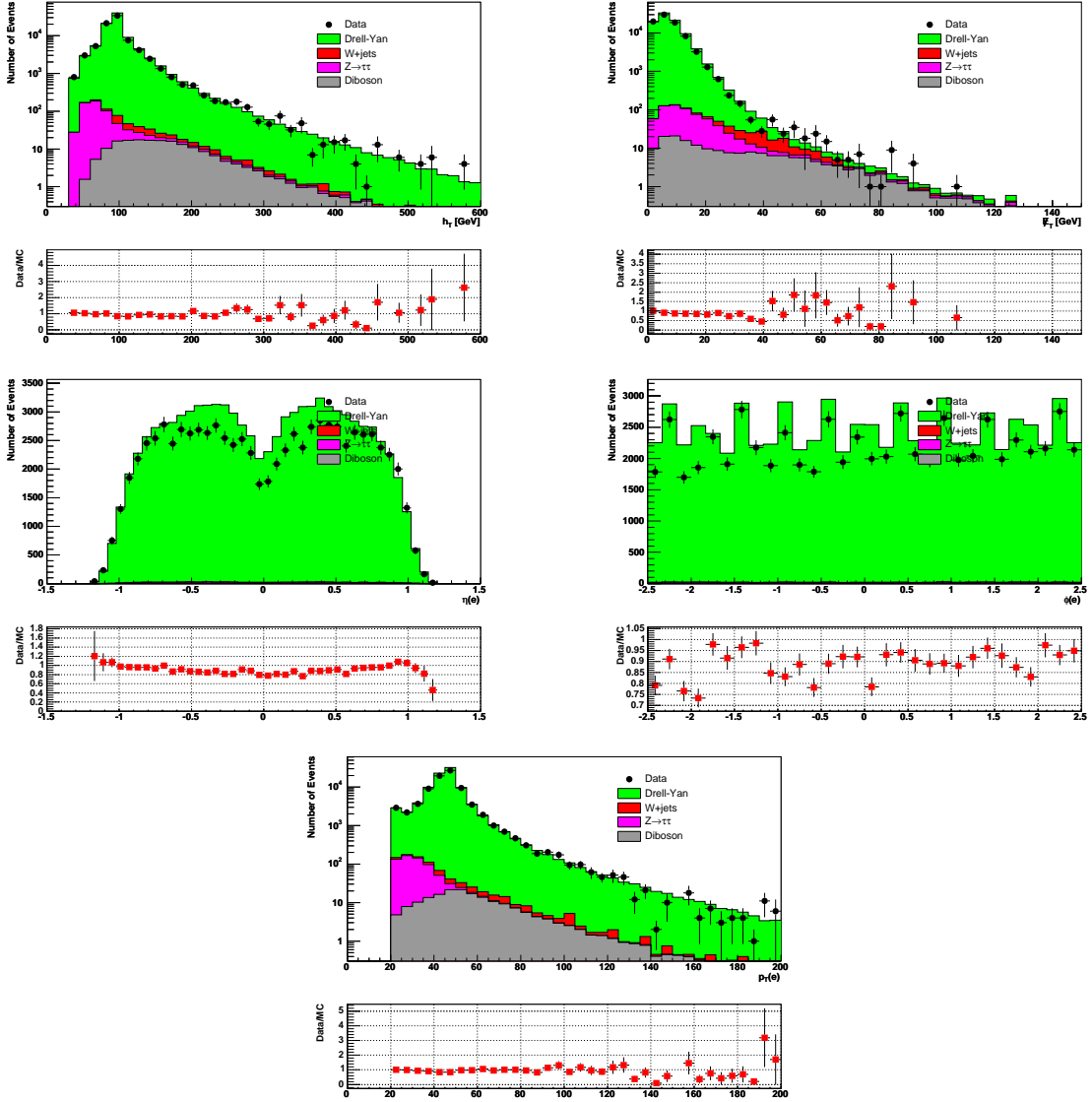


Figure 9: Plots of e-triggered Z events: the  $h_T$  in the event, the  $\cancel{E}_T$  in the event, and the  $\eta$ ,  $\phi$ , and  $p_T$  of the leading electron

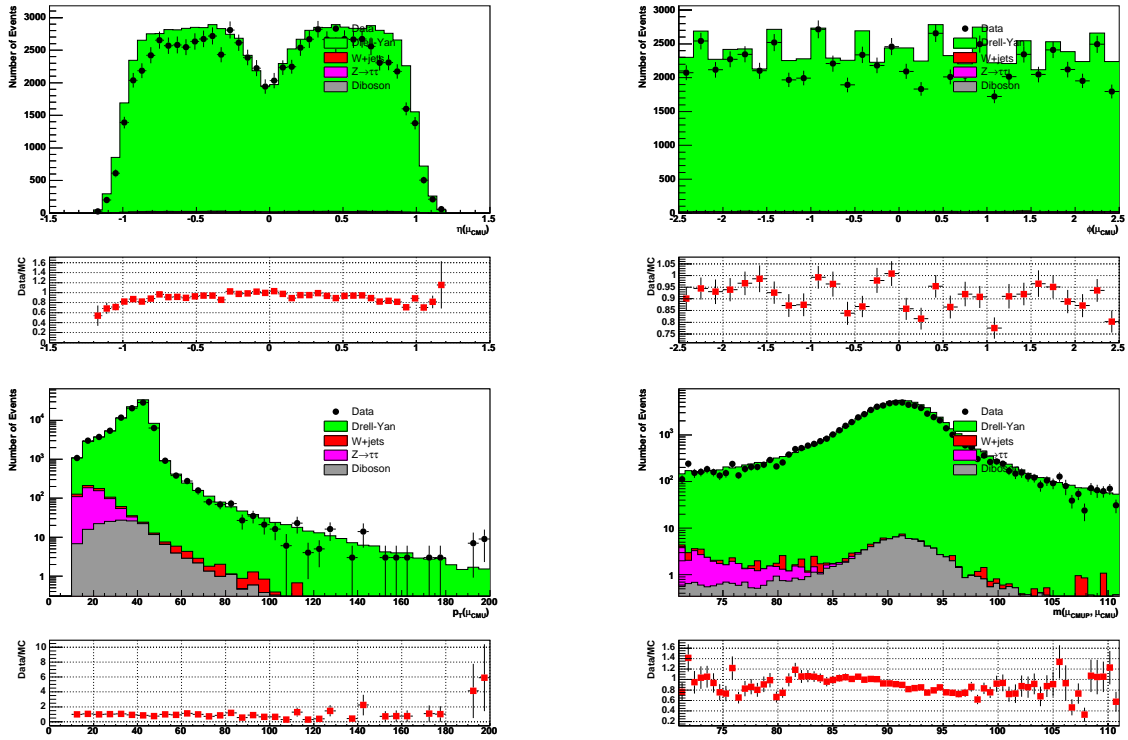


Figure 10: Plots of the second electron in Z events with two electrons:  $\eta$ ,  $\phi$  and  $p_T$  of the second electron and the dilepton mass peak.

Source	Number of Events
$W$ +jets	77
Drell-Yan	12678
$Z \rightarrow \tau\tau$	127
$t\bar{t}$	10
Single Top	0
Diboson	28
QCD	0
Expected total	12922
Observed	13556

Table 16: Summary of the predictions for  $Z^0 \rightarrow e^+e^-$  with both electrons tight

### 6.3.2 $Z \rightarrow \mu^+ \mu^-$

Table 17 gives the event counts for the cuts applied sequentially, and also the number of events which pass all other cuts except that cut, for both Monte Carlo and data. The DIF filter is applied to both muons.

Cut	$N_{seq}$ (MC)	$N_{N-1}$ (MC)	$N_{seq}$ (Data)	$N_{N-1}$ (Data)
Number of Triggered Events	NA		19335273	
Tight CMUP Muon	1195608	1	1507017	2
Second Muon	47887	0	43426	0
Opposite Sign Pair	47862	25	43262	164
Number of Triggered Events	NA		19335273	
Tight CMX Muon	726035	0	1156736	1
Second Muon	29128	0	28579	0
Opposite Sign Pair	29114	13	28469	110

Table 17: Number of events which pass each cut in sequence and which pass every cut except a particular one. Note that the cut of no additional leptons (e or  $\mu$ ) is applied here for the Z control sample, but is not applied below in the search. Also the QCD background estimate is not yet included.

After selecting a good CMUP or CMX muon, we look for a second muon in any of the CMU, CMP, CMUP, CMX, or CMIO categories.

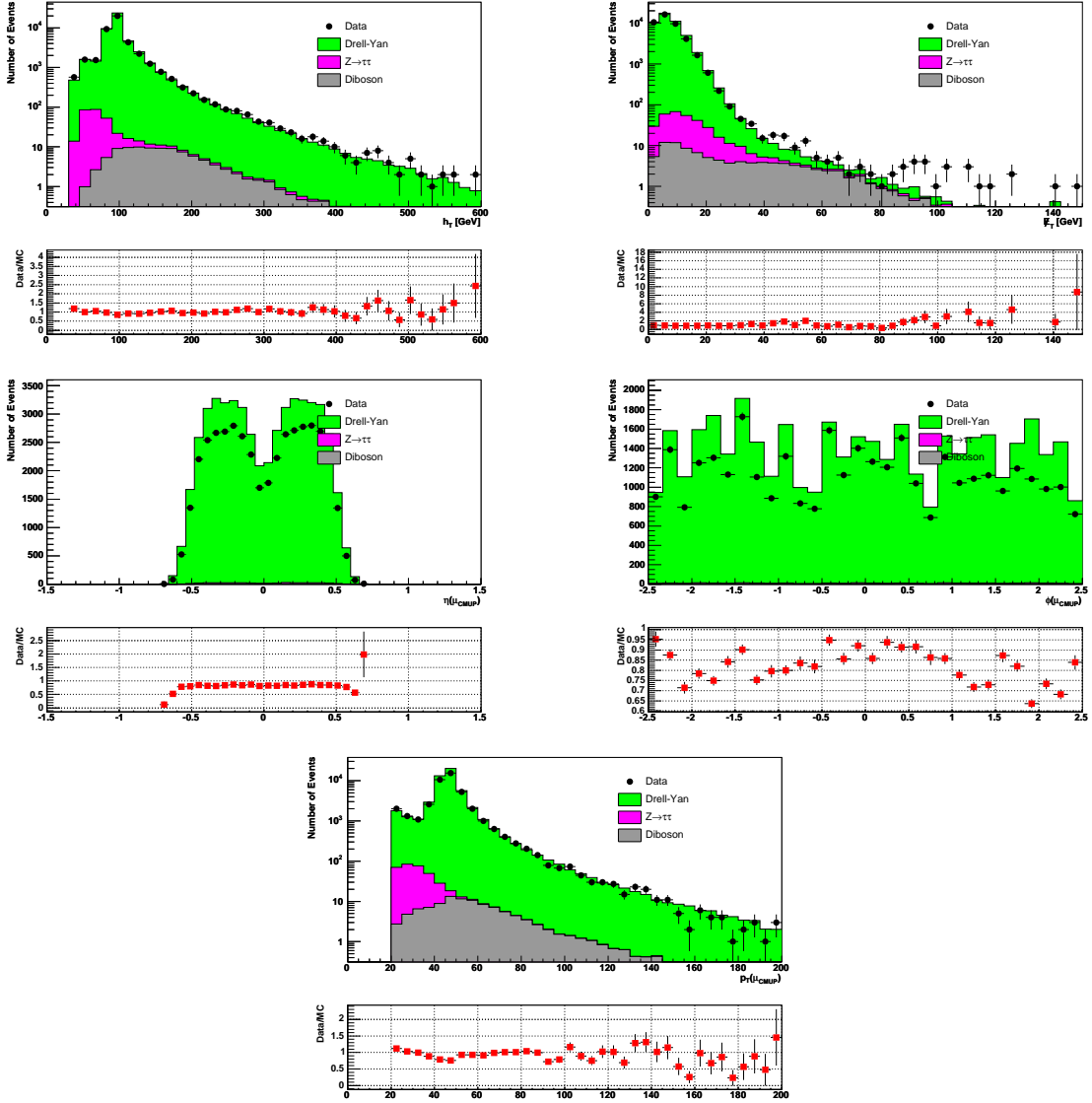


Figure 11: Plots of CMUP  $\mu$ -triggered Z events: the  $h_T$  in the event, the  $\cancel{E}_T$  in the event, and the  $\eta$ ,  $\phi$ , and  $p_T$  of the leading muon

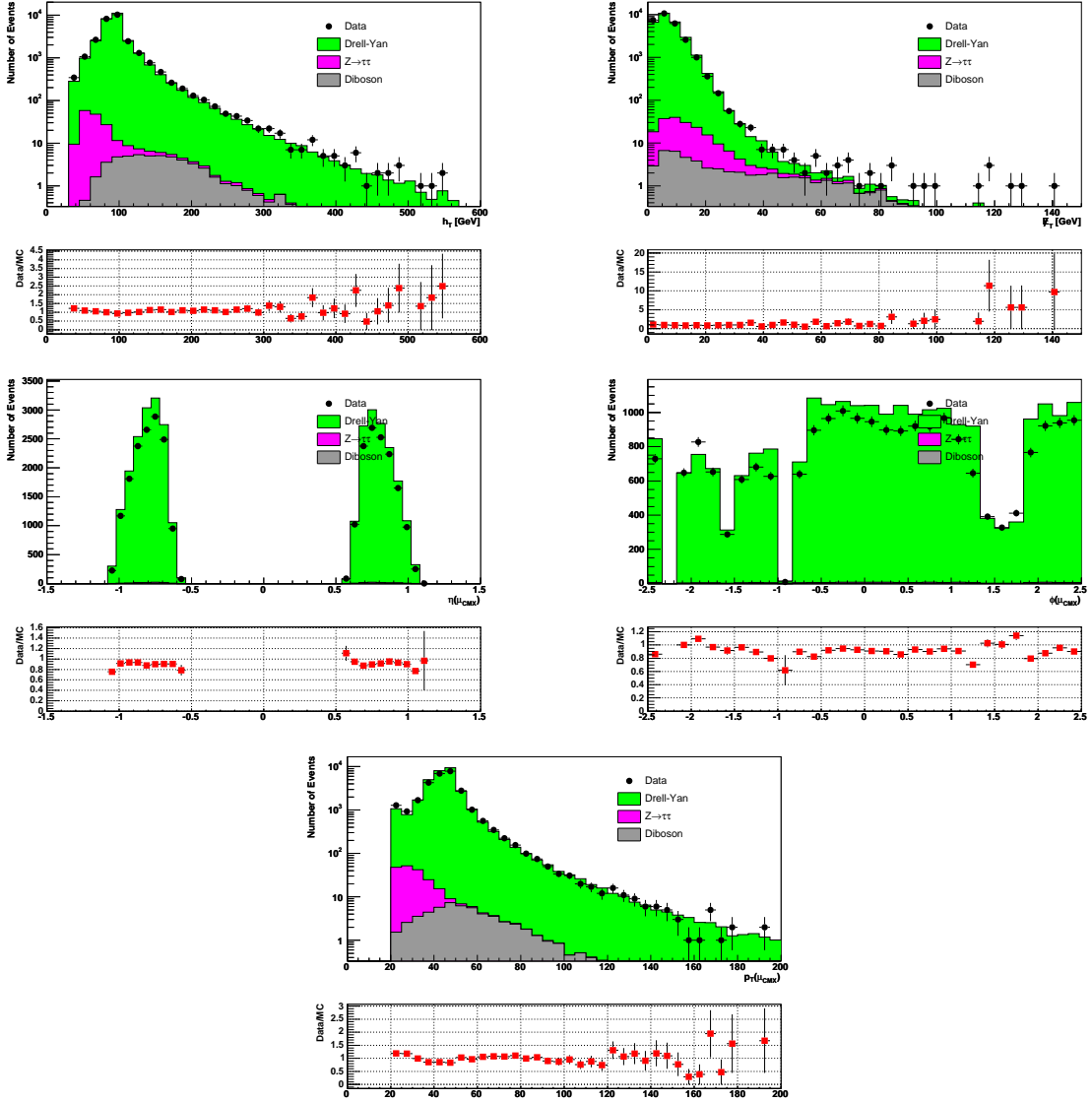


Figure 12: Plots of CMX  $\mu$ -triggered Z events: the  $h_T$  in the event, the  $E_T$  in the event, and the  $\eta$ ,  $\phi$ , and  $p_T$  of the leading muon

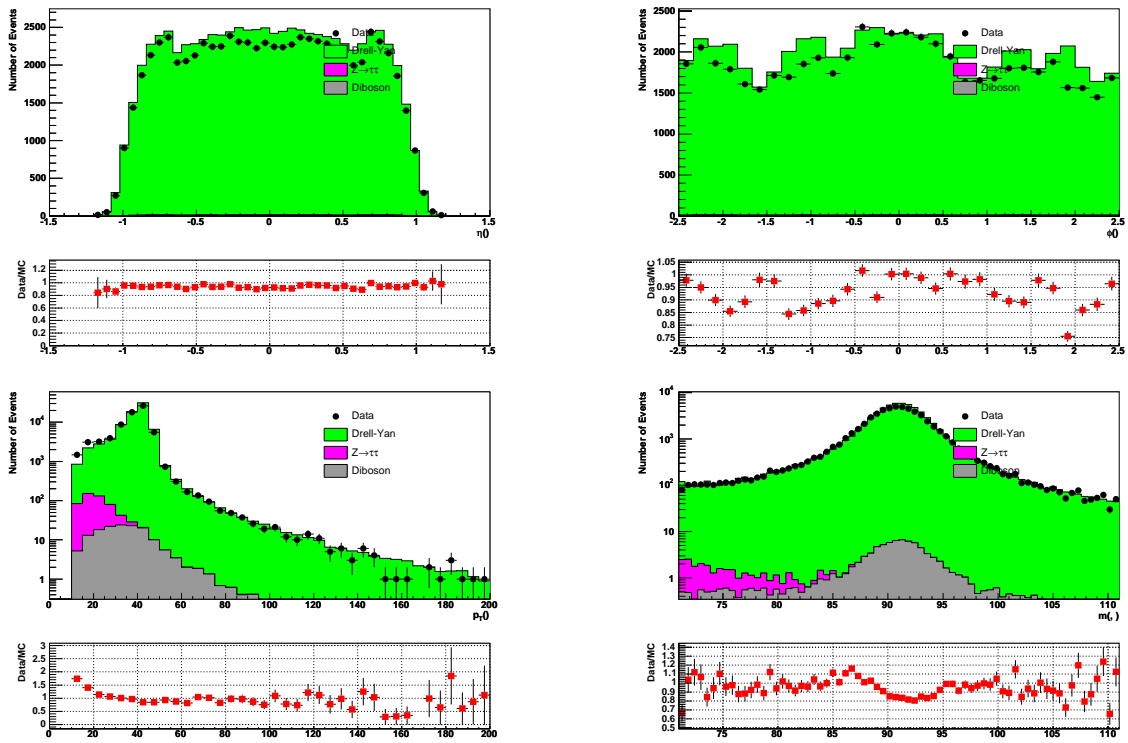


Figure 13: Plots of the second muon in  $Z$  events with two muons:  $\eta$ ,  $\phi$  and  $p_T$  of the second muon and the dilepton mass peak.

## 6.4 Summary of Observed and Predicted Numbers of W's and Z's

We summarize the number of  $W$ s and  $Z$  in the validation samples in Table 18. We find agreement with the MC in most of the categories. The biggest discrepancy (14%) is in the number of  $Z$ s found using loose electrons.

Selection	$W$ +jets	Drell-Yan	$Z \rightarrow \tau\tau$	Diboson	QCD	Expected	Observed
$W(e, \cancel{E}_T)$	660451	5742	797	821	1775	668338	636119
$W(\mu_{CMUP}, \cancel{E}_T)$	347830	34513	463	466	0	383578	365583
$W(\mu_{CMX}, \cancel{E}_T)$	216293	20696	283	251	0	237664	234823
$Z(e, e)$	0	23658	138	52	0	23883	24120
$Z(e, e_{loose})$	22	3953	39	8	0	4027	4706
$Z(\mu_{CMUP}, \mu_{CMU})$	0	1379	9	3	0	1393	1423
$Z(\mu_{CMUP}, \mu_{CMP})$	0	1774	10	4	0	1789	1870
$Z(\mu_{CMUP}, \mu_{CMUP})$	0	6781	40	16	0	6843	6729
$Z(\mu_{CMUP}, \mu_{CMX})$	0	4020	24	8	0	4057	4378
$Z(\mu_{CMUP}, \mu_{CMIO})$	0	1707	10	3	0	1723	1509
$Z(\mu_{CMX}, \mu_{CMU})$	0	875	5	1	0	883	893
$Z(\mu_{CMX}, \mu_{CMP})$	0	1080	6	2	0	1090	1075
$Z(\mu_{CMX}, \mu_{CMUP})$	0	3749	20	8	0	3781	4208
$Z(\mu_{CMX}, \mu_{CMX})$	0	2582	13	4	0	2602	3074
$Z(\mu_{CMX}, \mu_{CMIO})$	0	918	5	2	0	927	954

Selection	$W$ +jets	Drell-Yan	$Z \rightarrow \tau\tau$	Diboson	Expected	Observed
$W(e, \cancel{E}_T)$	2077655	19513	2619	2303	2103572	1833188
$W(\mu_{CMUP}, \cancel{E}_T)$	937785	94143	1279	1117	1035059	924564
$W(\mu_{CMX}, \cancel{E}_T)$	643062	62634	882	0	707582	640331
$Z(e, e)$	0	76968	403	154	77599	68965
$Z(e, e_{loose})$	64	12806	110	25	13015	13566
$Z(\mu_{CMUP}, \mu_{CMU})$	0	4206	25	8	4243	3974
$Z(\mu_{CMUP}, \mu_{CMP})$	0	5071	29	11	5115	5135
$Z(\mu_{CMUP}, \mu_{CMUP})$	0	20526	109	45	20696	18256
$Z(\mu_{CMUP}, \mu_{CMX})$	0	12538	70	24	12641	11951
$Z(\mu_{CMUP}, \mu_{CMIO})$	0	4700	27	10	4742	3946
$Z(\mu_{CMX}, \mu_{CMU})$	0	2763	14	4	2784	2604
$Z(\mu_{CMX}, \mu_{CMP})$	0	3217	17	6	3242	3002
$Z(\mu_{CMX}, \mu_{CMUP})$	0	11856	63	23	11949	11627
$Z(\mu_{CMX}, \mu_{CMX})$	0	8133	42	12	8192	8659
$Z(\mu_{CMX}, \mu_{CMIO})$	0	2684	16	5	2709	2577

Table 18: Event counts in  $W$  and  $Z$  selection samples. The predictions from  $W$  and  $DY$  + jets are from Alpgen. Contributions from  $WW$ ,  $WZ$ , and  $ZZ$ ,  $t\bar{t}$ , and single top are from Pythia.

## 6.5 Checking the Ratio of W to Z production ‘R’, as a Function of Time

We use the W and Z cross sections for each run period [2] to validate the datasets. We discuss the data samples selected on electron and muon triggers separately below.

17 and 18) The cross sections are in the correct range according to the theory, but there seem to be run-dependent effects: the cross sections go down with time. The ratio of  $\sigma_W/\sigma_Z$ , however, is correct and constant with respect to run number.

### 6.5.1 Electron-Triggered Modes

Figures 14 and Figures 15 show the W and Z cross-sections, respectively, as a function of run number [2]. There is a drop in the W cross section of approximately 10%. There is a similar, but smaller, drop in the Z cross section, suggesting that the problem is not in the luminosity calculation, but instead is in the electron trigger, ID, or met calculation.

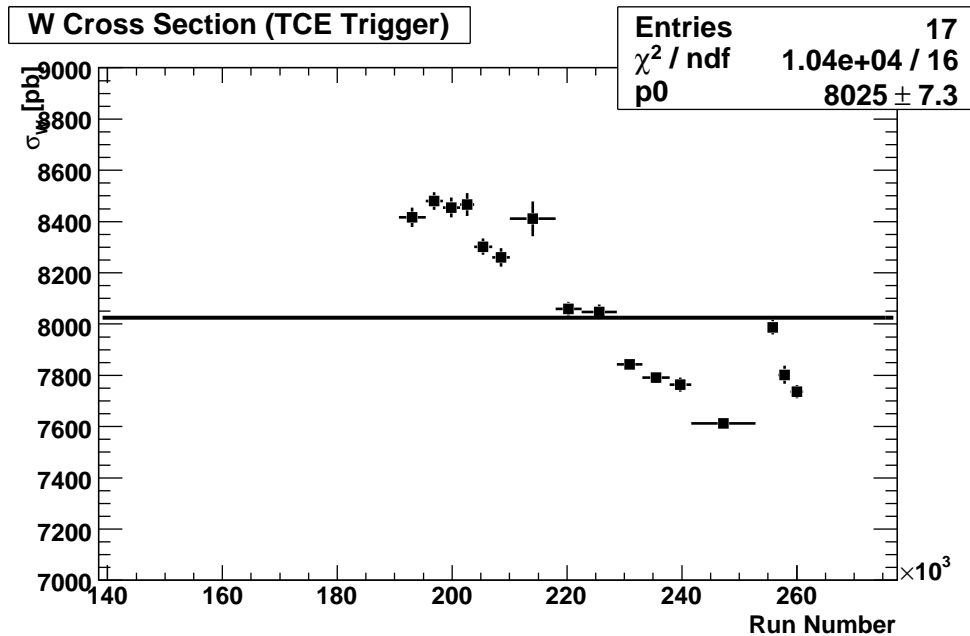


Figure 14: The W cross section for each run period, using the electron trigger. The cross sections are calculated using the efficiencies measured from the W+jets and Drell-Yan MC samples. It is unknown why the W cross section is measured to be decreasing with time.

The ratio of W to Z production is a very precise test of problems in lepton trigger efficiencies, lepton identification efficiencies, or problems with  $\cancel{E}_T$  [?]. Figure 16 shows  $R$  versus run number.

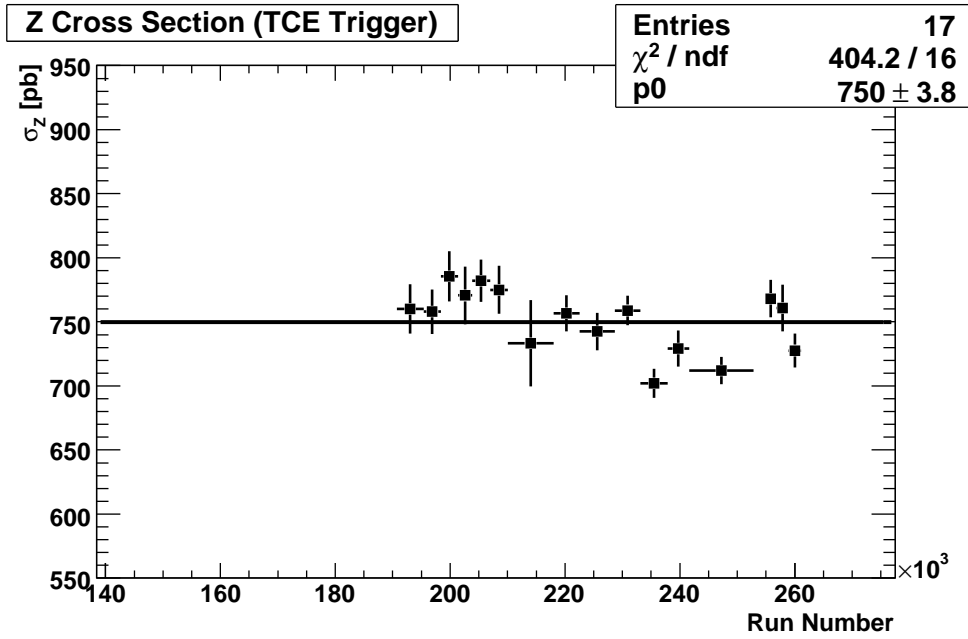


Figure 15: The Z cross section for each run period, using the electron trigger. The cross sections are calculated using the efficiencies measured from the Drell-Yan MC samples. The scatter and secular trend of the data are outside CDF estimates of point-to-point luminosity.

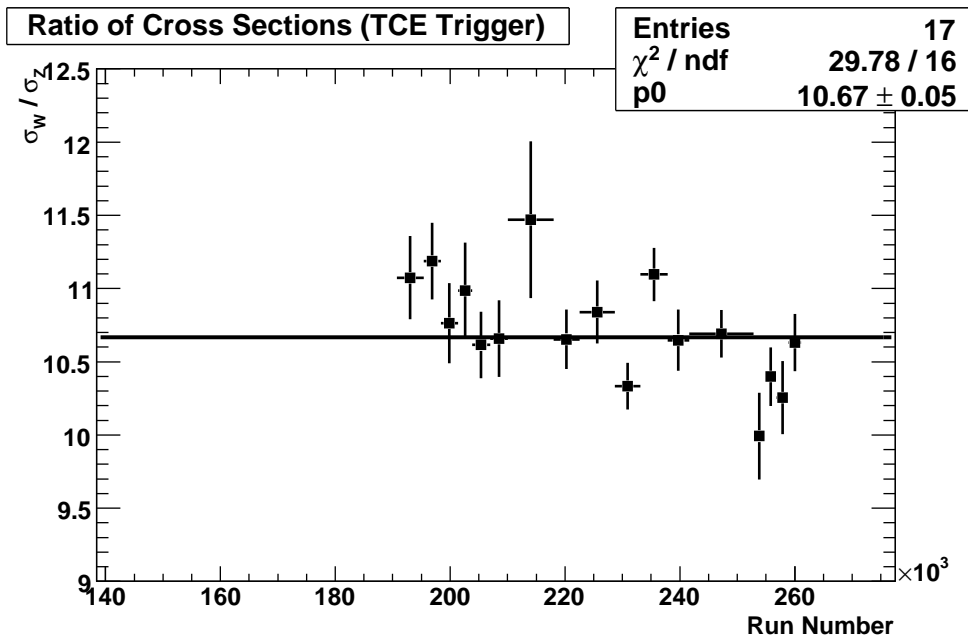


Figure 16: The ratio, R, of W and Z cross sections for each run period, using the electron trigger. The cross sections are calculated using the efficiencies measured from the W+jets and Drell-Yan MC samples. It is unknown why the W cross section is measured to be decreasing with time.

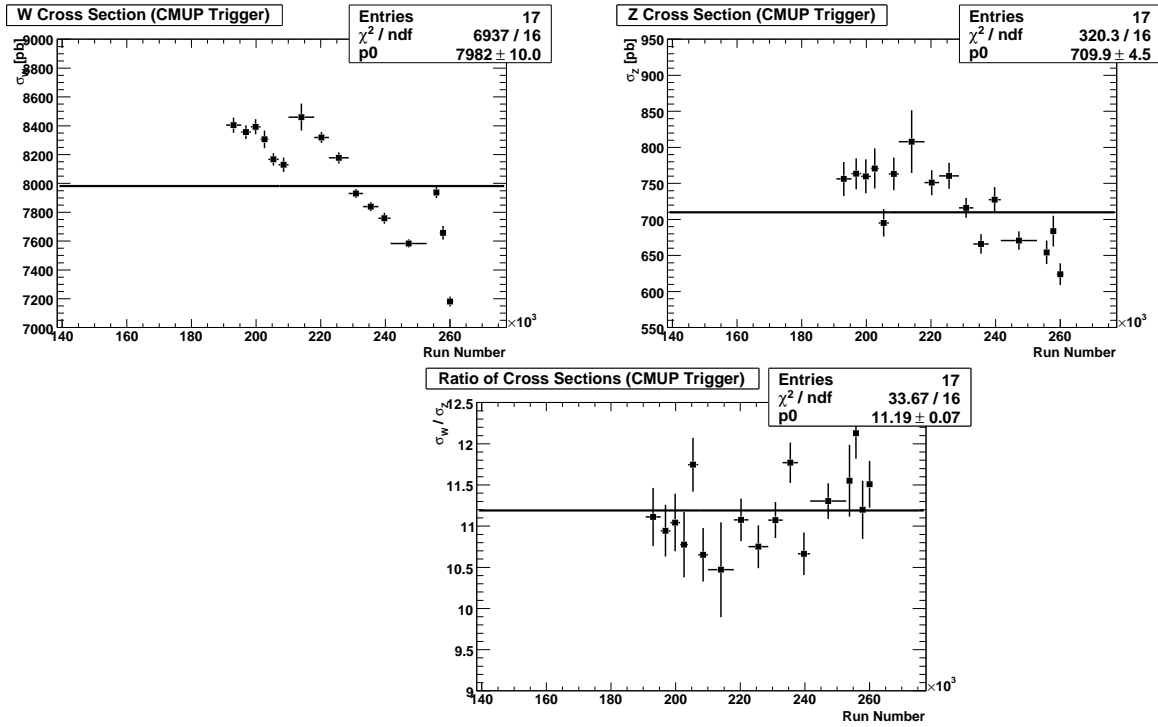


Figure 17: W and Z mass peaks and cross sections for each run period, using the CMUP trigger. The cross sections are calculated using the efficiencies measured from our W+jets and Drell-Yan MC samples.

### 6.5.2 Muon-Triggered Modes

This needs filling in and separating out the figures as in the electron section above...

## 6.6 Analysis Sample Selection

# 7 Soft Lepton Identification

## 7.1 Soft Electrons

We identify soft electrons using a method adapted from the B-Group's soft electron tagger. For every track in the event, we require that it passes track quality cuts and fiduciality cuts:

- 20 axial and 20 stereo COT hits
- At least 2 COT superlayers with 6 hits
- Track is fiducial to CES, CPR, and calorimeter
- Track  $|\eta| < 1$ .

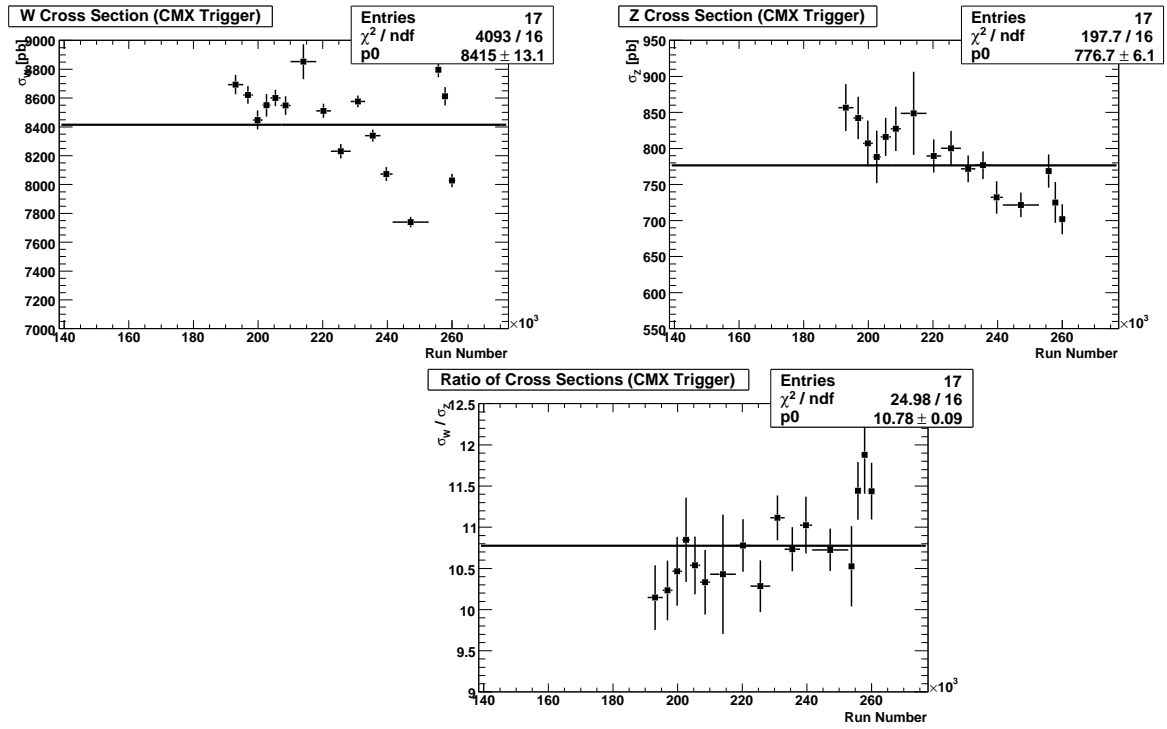


Figure 18: W and Z mass peaks and cross sections for each run period, using the CMX trigger. The cross sections are calculated using the efficiencies measured from our W+jets and Drell-Yan MC samples.

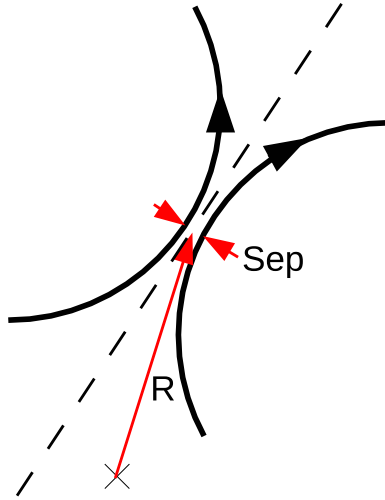


Figure 19: Conversion identification variables

After these cuts, we use a likelihood-based calculator to identify tracks that come from electrons. The likelihood calculator is described in Section 7.1.2, and is trained completely on data.

### 7.1.1 Training Sample Selection

We use identified conversions as a training sample of real electrons. Using the 8-GeV electron trigger, we obtain a pure sample of conversions by looking for pairs of tracks that match the following criteria: (see Figure 19 for an illustration of the variables)

- Opposite sign
- $|Sep| < 0.2$  cm
- $\Delta \cot(\theta) < 0.1$
- $R_{Conv} > 8$

After these cuts, we fit the  $\Delta \cot(\theta)$  distribution to determine the non-conversion background under the peak. We use the sideband of the distribution ( $0.06 < |\Delta \cot(\theta)| < 0.1$ ) to subtract out this background.

Since the higher- $p_T$  leg of the conversion will be trigger-biased, we train the likelihood function using the soft leg. We also don't use conversion pairs in which the hard leg extrapolates to the same calorimeter towers that are used for the soft leg, since those conversions have a very different  $E_{em}/p$  distribution.

We use generic tracks as the background sample for the likelihood. Using the 20-GeV muon trigger, we eliminate all tracks that pass a conversion filter: (with different cuts than above, since we're trying to eliminate all conversions instead of find a pure sample of them)

- Opposite sign

- $|Sep| < 0.2$  cm
- $\Delta \cot(\theta) < 0.03$
- $R_{Conv} > 0$

To further reduce the contamination of this ‘fake’ sample by real electrons, we ignore all events that have any identified hard electron (using standard CDF cuts) or any SecVtx tag. We only use tracks that are in jets, since those are the only tracks in our analysis for which we have a background expectation. (See Section ??)

### 7.1.2 Discriminating Variables

We use seven discriminating variables in our likelihood calculation:  $\frac{dE}{dx}$ ,  $E_{EM}/P$ ,  $\frac{E_{Had}}{E_{EM}}$ ,  $E_{CPR}$ ,  $E_{CES}$ ,  $\Delta X_{CES}$ , and  $\Delta Z_{CES}$ . The CES variables are calculated using the 2-dimensional CES shower algorithm (described in Section ??). The calorimeter variables are calculated using a narrow, two-tower wide section of the calorimeter.

We calculate a likelihood for each variable and multiply these likelihoods together to get the final likelihood.

$\frac{dE}{dx}$  is heavily dependent upon the momentum of the track, and is most useful as a discriminant for low-momentum tracks. Therefore, the likelihood function depends on  $p$ . If  $\frac{dE}{dx}$  wasn’t calculated for the track, either because there weren’t enough COT hits or because  $\frac{dE}{dx}$  was turned off in that run, this portion of the likelihood is set to 1. Plots of the fits are in Figure 23.

$$\begin{aligned}
 L_{dE/dx} &= \begin{cases} 1 & N_{dE/dx} = 0 \\ p_0 \cdot \text{landau}(dE/dx, p_1, p_2) & N_{dE/dx} > 0 \end{cases} \\
 p_0 &= q_0 + q_1 \cdot \exp(-p/q_2) + q_3 \cdot \exp(-p/q_4) \\
 p_1 &= r_0 + r_1 \cdot \exp(-p/r_2) \\
 p_2 &= s_0 + s_1 \cdot \exp(-p/s_2)
 \end{aligned}$$

For the calorimeter energies, we use a narrow, two-tower area of the detector. The first tower is the one to which the track extrapolates, and the the second is one tower to the East or West, depending on which one the track’s CES shower is nearer to. Plots of the fits are in Figure 21 and Figure!22.

$$\begin{aligned}
 L_{E_{em}/p} &= p_0 * \frac{\exp(p_1 \cdot (E_{em}/p - p_2))}{1 + \exp(p_1 \cdot (E_{em}/p - p_2))} + p_3 * \text{Gaus}(E_{em}/p, p_4, p_5) \\
 p_0 &= q_0 + q_1 \cdot \exp(-p/q_2) \\
 p_1 &= r_0 + r_1 \cdot \exp(-p/r_2) \\
 p_2 &= s_0 + s_1 \cdot \exp(-p/s_2) \\
 p_3 &= t_0 + t_1 \cdot p \\
 p_4 &= u_0 + u_1 \cdot p \\
 p_5 &= v_0
 \end{aligned}$$

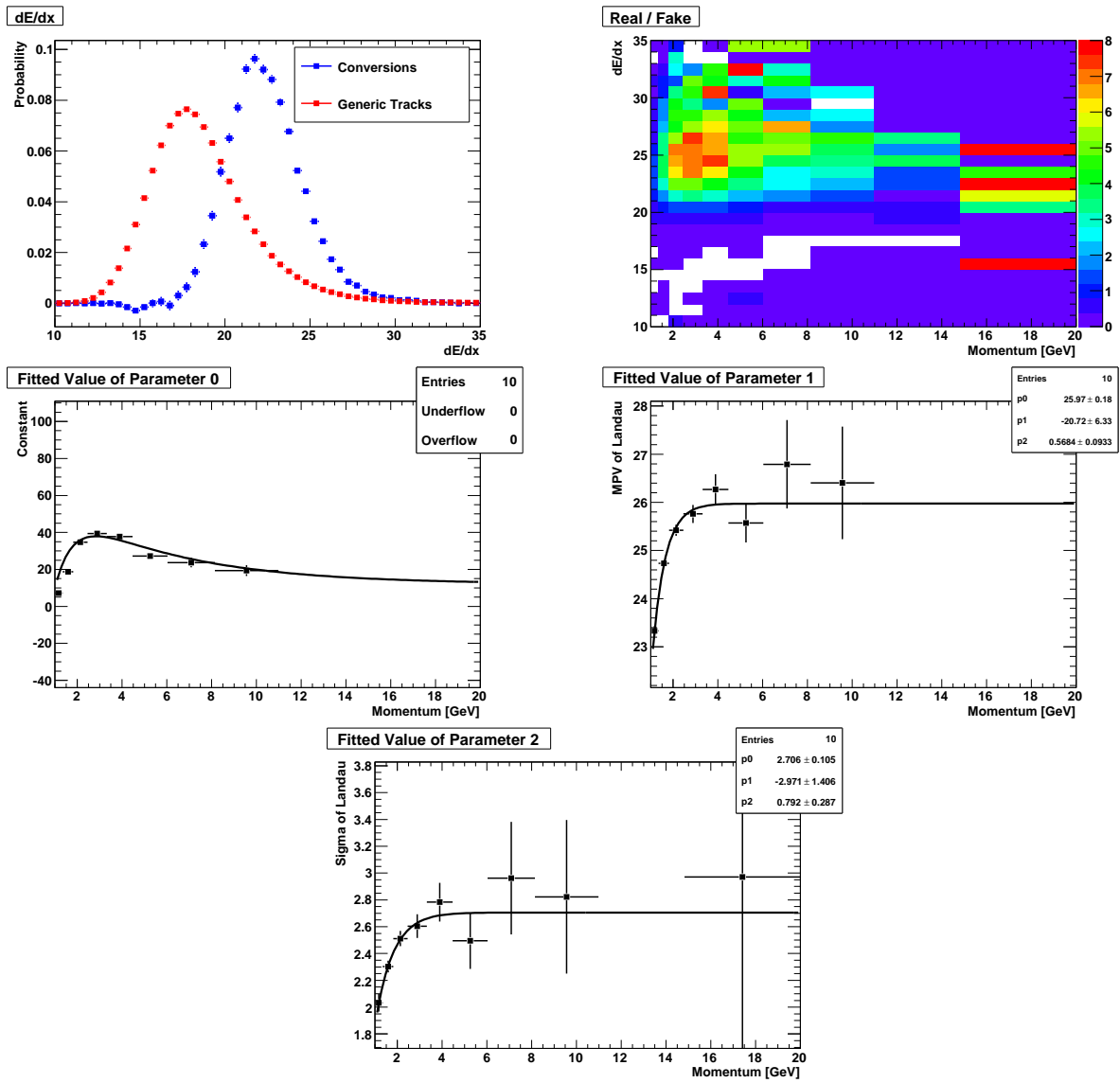


Figure 20: The fits to the  $\frac{dE}{dx}$  likelihood function.

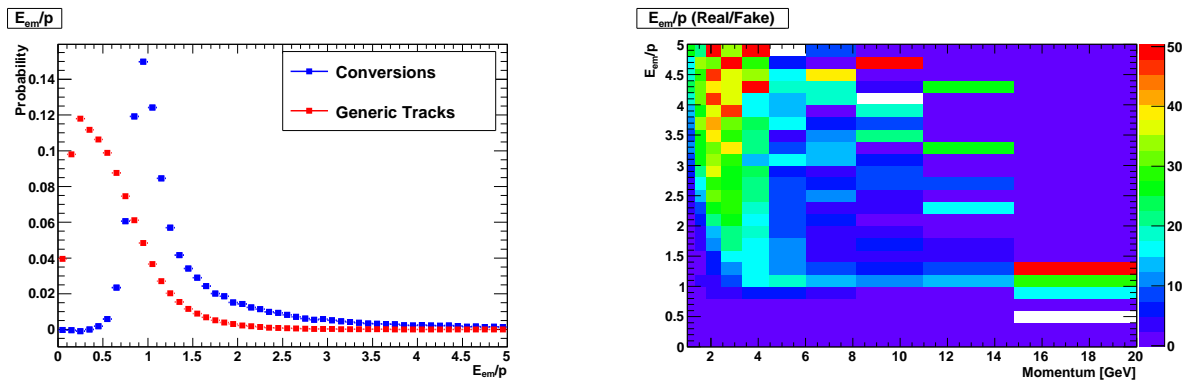


Figure 21: The plots for the  $\frac{EmE}{p-iso}$  likelihood function

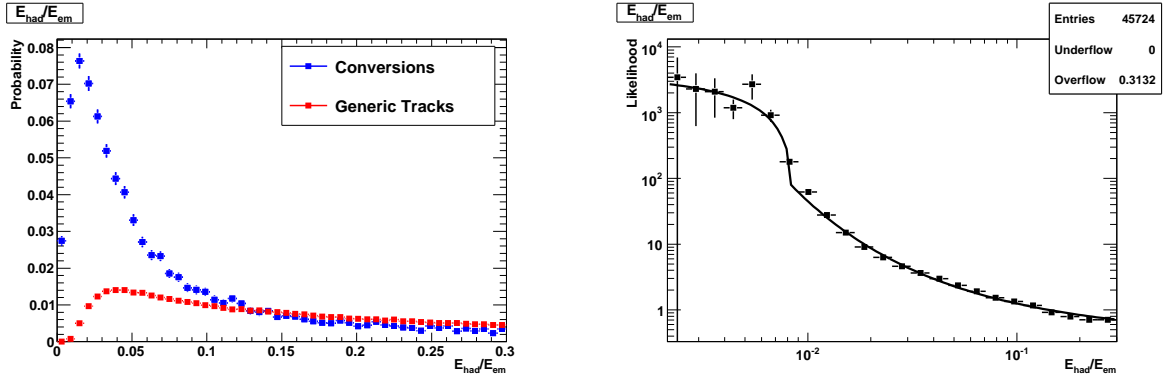


Figure 22: The plots for the  $\frac{E_{Had}}{E_{EM}}$  likelihood function

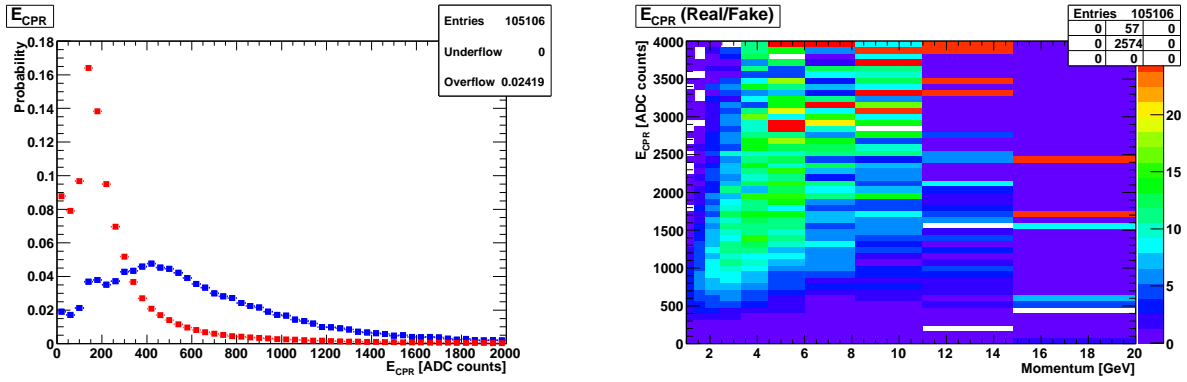


Figure 23: The plots for the  $E_{CPR}$  likelihood function

$$L_{Had/Em} = \begin{cases} 1 & N_{dE/dx} = 0 \\ p_0 \cdot \text{landau}(dE/dx, p_1, p_2) & N_{dE/dx} > 0 \end{cases}$$

The  $E_{CPR}$  distribution depends on the momentum of the track. Also, there is a factor of  $\sin(\theta)$  to correct for the fact that tracks at a larger angle have a longer path length through the CPR:

$$\begin{aligned} L_{CPR} &= p_0 \cdot \text{landau}(E_{CPR}/\sin(\theta), p_1, p_2) \\ p_0 &= q_0 + q_1 \cdot \text{landau}(p, q_2, q_3) \\ p_1 &= r_0 + r_1 \cdot \text{gaus}(p, r_2, r_3) \\ p_2 &= s_0 + s_1 \cdot \text{gaus}(p, s_2, s_3) \end{aligned}$$

The  $E_{CES}/p$  distribution depends on the track momentum. Plots of the fits are in Figure 24.

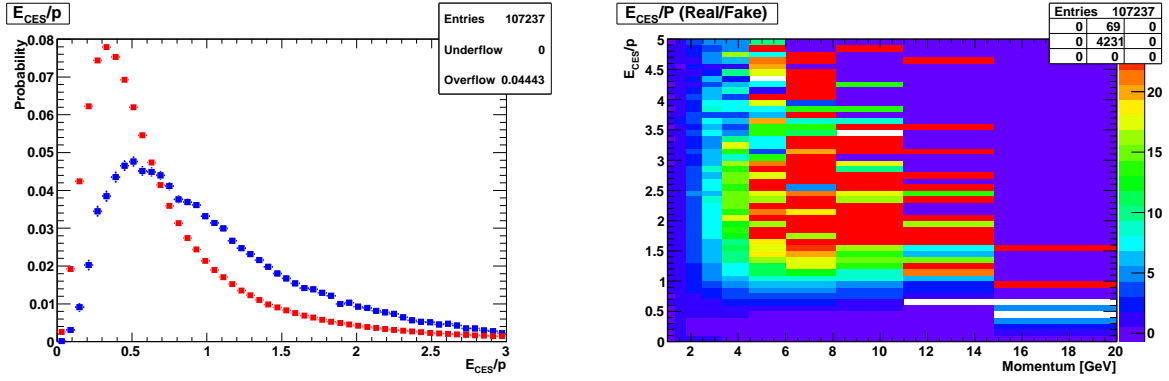


Figure 24: The fits to the parameters for the  $E_{CES}$  likelihood function

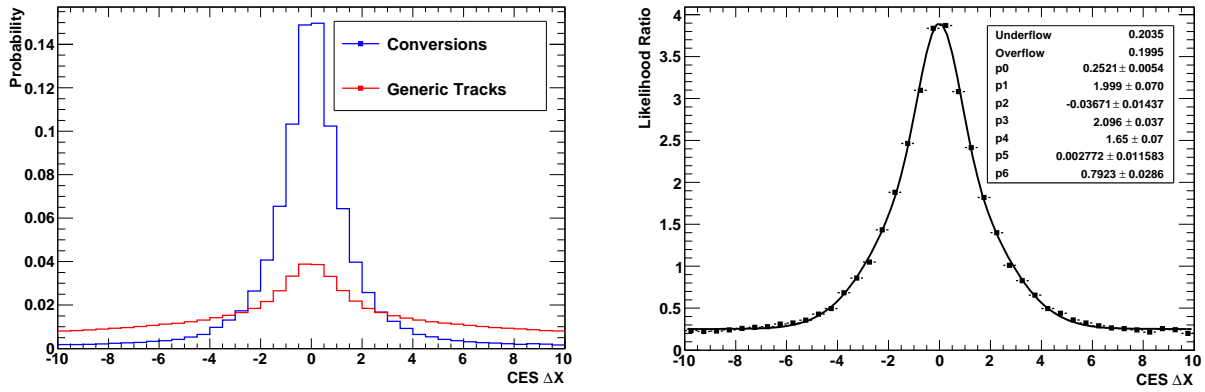


Figure 25: The plots for the CES  $\Delta X$  likelihood function

$$\begin{aligned}
 L_{CES} &= p_0 + p_1 \cdot \text{landau}(E_{CES}/p, p_2, p_3) \\
 p_0 &= q_0 \cdot \exp(-p/q_1) \\
 p_1 &= r_0 + r_1 \cdot p \\
 p_2 &= s_0 + s_1 \cdot \exp(-p/s_2) + s_3 \cdot \exp(-p/s_4) \\
 p_3 &= t_0 + t_1 \cdot \exp(-p/t_2) + t_3 \cdot \exp(-p/t_4)
 \end{aligned}$$

The distributions for the track-CES matching variables,  $\Delta X_{CES}$  and  $\Delta Z_{CES}$ , depend on both momentum and angle. The distributions are wider at lower momentum and at higher  $|\eta|$ . The plot of the fit for CES  $\Delta X$  is in Figure 25, and for CES  $\Delta Z$  is in Figure 26.

$$\begin{aligned}
 L_{\Delta X_{CES}} &= p_0 + p_1 \cdot \exp\left(\frac{-(\Delta X - p_2)^2}{2p_3^2}\right) + p_4 \cdot \exp\left(\frac{-(\Delta X - p_5)^2}{2p_6^2}\right) \\
 L_{\Delta Z_{CES}} &= p_0 + p_1 \cdot \exp\left(\frac{-(\Delta Z - p_2)^2}{2p_3^2}\right) + p_4 \cdot \exp\left(\frac{-(\Delta Z - p_5)^2}{2p_6^2}\right)
 \end{aligned}$$

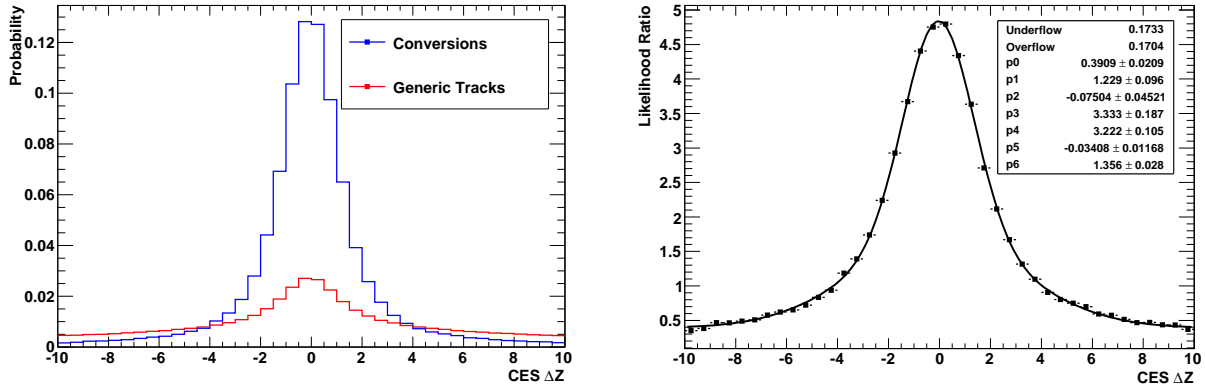


Figure 26: The plots for the CES  $\Delta Z$  likelihood function.

### 7.1.3 Efficiency and Fake Rate Matrices

The likelihood calculation doesn't work in the Monte Carlo, since it depends on the precise response of the calorimeters at low energy, which isn't modeled very well. Instead, we calculate the efficiency and fake rate of the soft electron identification as a function of  $p_T$ ,  $\eta$ , and isolation. We apply this efficiency or fake rate to each taggable candidate in the MC to find the predicted number of identified electrons.

We divide  $p_T$ ,  $\eta$ , and isolation into three bins each, and fit the efficiency (and fake rate) in each of these 27 bins to a second-degree polynomial in each variable.

We validate the fake matrix by testing it on a sample of generic tracks in jets, using the same electron removal that was used for the training sample of generic tracks: Conversion removal, hard electron removal, and heavy flavor removal.

Fitting the likelihood distribution of this jet sample, we find that it corresponds to 93.6% of the “fake” distribution plus 6.4% of the conversion distribution. Applying the efficiency matrices to the taggable tracks in this mixture gives good agreement with the data, as shown in Figure 27

### 7.1.4 Validation

\*\* Plots of SE kinematic variables in data compared with MC \*\*

## 7.2 Soft Muons

# 8 Track/ $\pi^0$ Clustering

The CDF  $\tau$  finder algorithm starts by finding a six tower calorimeter cluster with  $p_T > 6$  GeV and a seed track with  $p_T > 3$  GeV [need tauID ref]. It then requires that there be no tracks or  $\pi^0$ s inside an isolation annulus between  $10^\circ$  and  $30^\circ$  with respect to the seed track. Additional tracks and  $\pi^0$ s are added to the  $\tau$  candidate object if they are not in the isolation annulus and have  $p_T > 1$  GeV. We find that after requiring a tight lepton,  $\cancel{E}_T > 20$  GeV, and a single  $\tau$  finder object with visible  $p_T > 5$  GeV that the efficiency for the signal Monte

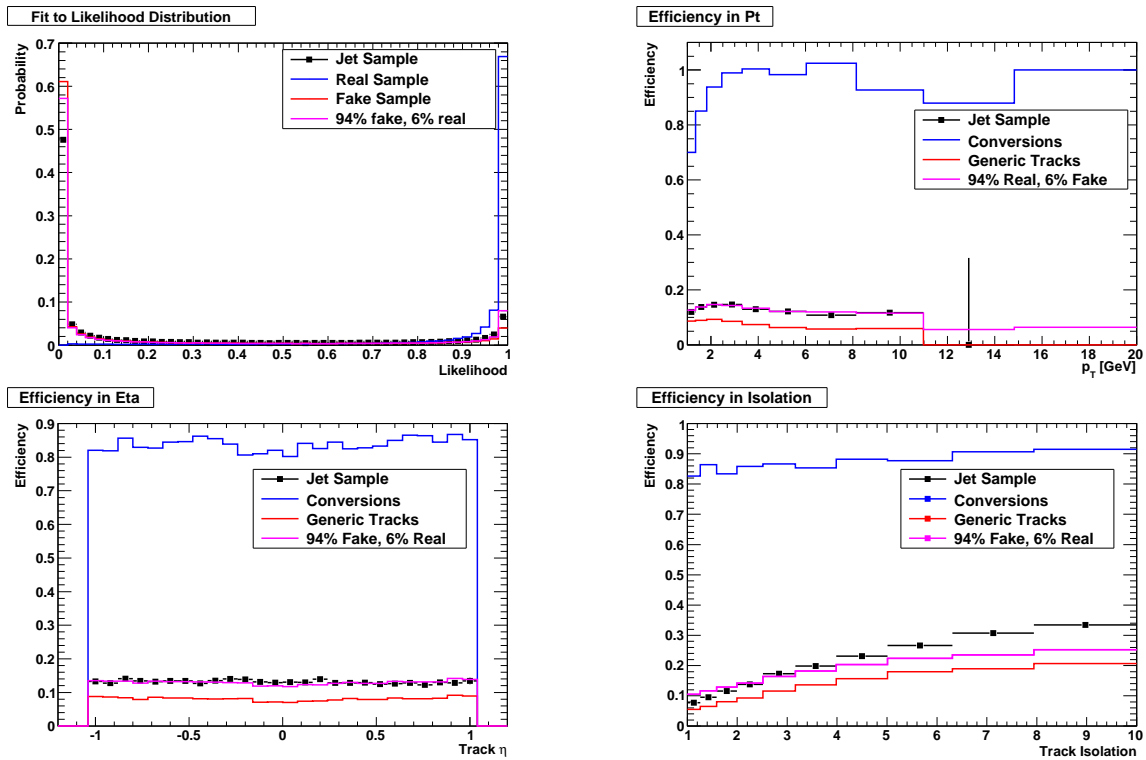


Figure 27: Validation plots of the soft electron fake matrix. Shown is the fit to the likelihood distribution in the jet sample, and the result of applying this mixture of “real” and “fake” to the  $p_T$ ,  $\eta$ , and isolation distributions.

Carlo sample to pass the selection is 3%. We can understand why this efficiency is low by examining the kinematics of the signal sample.

Figure 28 shows the angle between the leading generated  $\tau$  prong in an event and the prongs from a neighboring  $\tau$  that comes from the same  $a^0$ . The most probable value of the angle between these prongs is near  $20^\circ$ . Figure 29 shows the  $p_T$ -weighted  $\Delta\eta\Delta\phi$  distribution of reconstructed prongs from the neighboring  $\tau$  with respect to the leading  $\tau$  prong direction. The isolation annulus boundaries used by the  $\tau$  finder algorithm are superimposed. The collinearity of the  $\tau$  production in the signal mode often causes tracks from neighboring  $\tau$  decays to land inside the  $\tau$  finder isolation annulus and thereby causes the  $\tau$  candidate to be rejected.

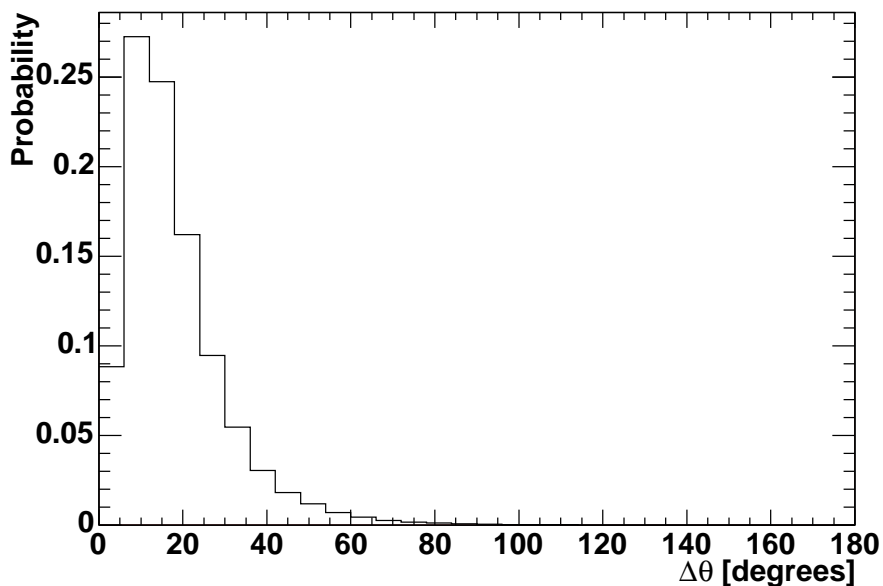


Figure 28: The angle between the leading generated prong and prongs from a neighboring  $\tau$  that comes from the same  $a^0$  in the NMSSM  $H \rightarrow 4\tau$  sample. Neighboring prongs tend to be within  $30^\circ$  of the leading prong.

Because the  $\tau$  finder is ill-suited to reconstruct collinear  $\tau$  decays, such as the one we are interested in, we must use some other kind of  $\tau$  identification. We have therefore developed a clustering algorithm based only on tracks and  $\pi^0$ s. The algorithm begins by selecting the highest  $p_T$  track above some  $p_T^{seed}$  threshold. This seed track is required to pass track quality cuts, have  $|z_0| \leq 60$  cm,  $|z_0 - z_{PV}| \leq 10$  cm,  $|d_0| \leq 0.2$  cm,  $9 \leq z_{CES} \leq 230$  cm, and to not be associated with a reconstructed high- $p_T$  electron or muon. Additional tracks are added to the cluster if they have pass track quality cuts,  $p_T \geq p_T^{shoulder}$ ,  $|z_{shoulder} - z_{seed}| \leq 5$  cm,  $\Delta R \leq \Delta R_{cluster}$  with respect to the seed track, and are not associated with a reconstructed high- $p_T$  electron or muon. Finally  $\pi^0$ s having  $p_T \geq p_T^{\pi^0}$  and with the  $\Delta R_{cluster}$  cone are added to the cluster. Thus, we have four parameters that define the cluster:  $p_T^{seed}$ ,  $p_T^{shoulder}$ ,  $p_T^{\pi^0}$ , and  $\Delta R_{cluster}$ .

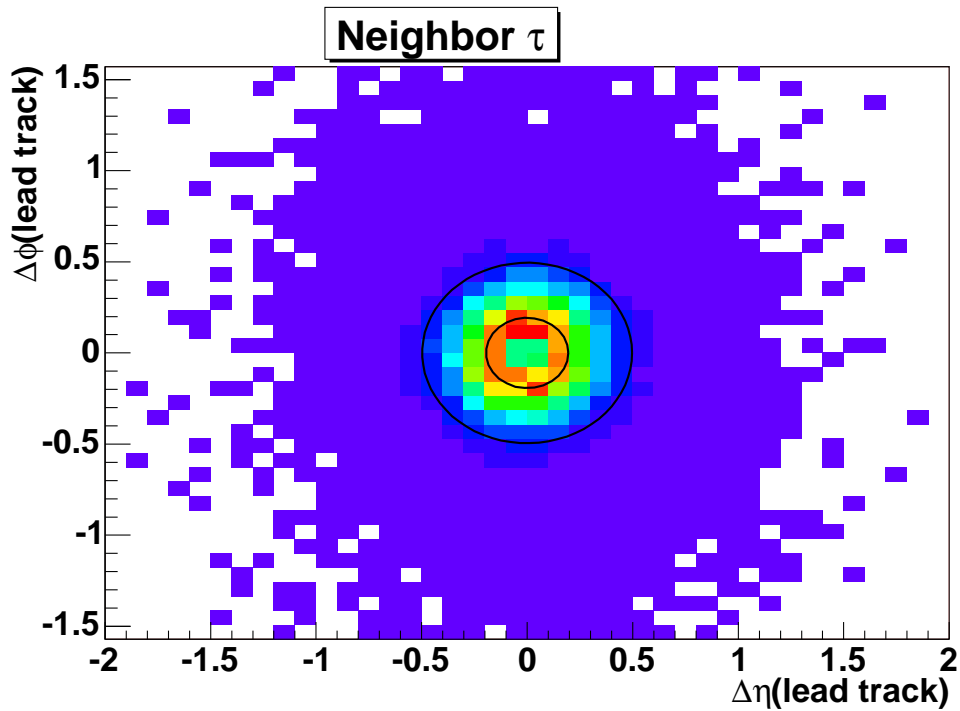


Figure 29: The  $p_T$ -weighted  $\Delta\eta\Delta\phi$  distribution of reconstructed prongs from neighboring  $\tau$  with respect to the leading  $\tau$  prong direction in the NMSSM  $H \rightarrow 4\tau$  Monte Carlo sample. The  $10^\circ$  inner and  $30^\circ$  outer isolation radii used by the  $\tau$ -finder algorithm are superimposed.

## 8.1 Clustering Parameters

To determine the optimal values of the clustering parameters, we investigate the effect of the parameters on the distributions of the mass of the cluster and the mass of di-clusters in our signal Monte Carlo sample. If the cluster contains two  $\tau$ 's, we expect to see a broad smeared peak corresponding to the  $a^0$  mass in the  $M(\text{cluster})$  distribution. Similarly, we expect to see a broad peak corresponding to  $M(H)$  in the di-cluster mass distribution. Figure 30 shows a comparison of the mass distributions for a variety of  $p_T^{\text{seed}}$  thresholds and two  $\Delta R_{\text{cluster}}$  thresholds, 0.3 and 0.5. We can see that the broad peak corresponding to  $M(a^0)$  grows when the cone size is increased to 0.5, as expected since a spatially larger cluster has a higher probability of containing decay products from both  $\tau$ 's arising from the same  $a^0$ .

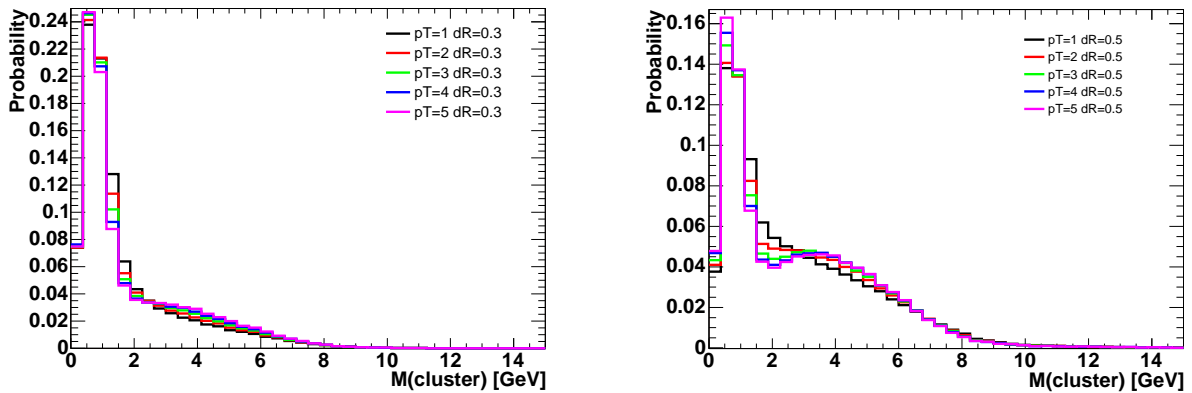


Figure 30: The track- $\pi^0$  cluster mass for different seed track  $p_T$  thresholds on the signal NMSSM Monte Carlo sample. On the left is the comparison using a cone radius of  $\Delta R = 0.3$  and on the right is the same for a cone radius of  $\Delta R = 0.5$ .

Figure 31 shows the distributions of the same clusters in the di-cluster mass. There are a number of interesting features in these plots. In the  $\Delta R = 0.3$  plot, notice that there is a narrow peak near  $M = 9$  GeV corresponding to  $M(a^0)$  for all  $p_T^{\text{seed}}$  thresholds. There is also a broad peak below 100 GeV corresponding to  $M(H)$  for all  $p_T^{\text{seed}}$  thresholds except  $p_T^{\text{seed}} = 1$  GeV. Similarly, in the  $\Delta R = 0.5$  plot the  $p_T^{\text{seed}} = 1$  GeV distribution does not show this peak while the others do. We conclude that the  $p_T^{\text{seed}} = 1$  GeV threshold is so low that tracks from the underlying event are being added to the cluster and washing out kinematic information that is helpful to identify the signal. By studying these distributions as well as the corresponding selection efficiencies, we determine that the optimal set of clustering parameters for our signal topology is  $p_T^{\text{seed}} = 2.5$  GeV,  $p_T^{\text{shoulder}} = 2$  GeV,  $p_T^{\pi^0} = 500$  MeV, and  $\Delta R_{\text{cluster}} = 0.5$ .

We can see a relative comparison of the effectiveness of clusters selected with these parameters by looking at fig. 32 which shows a comparison of the  $N(\tau)$  probability for clusters, cone 0.4 jets, and cone 0.7 jets when applied to the signal Monte Carlo sample. We see that clusters are less likely to contain no  $\tau$ 's than either kind of jet. They are also more likely to contain 1 or 2  $\tau$ 's than the other objects.

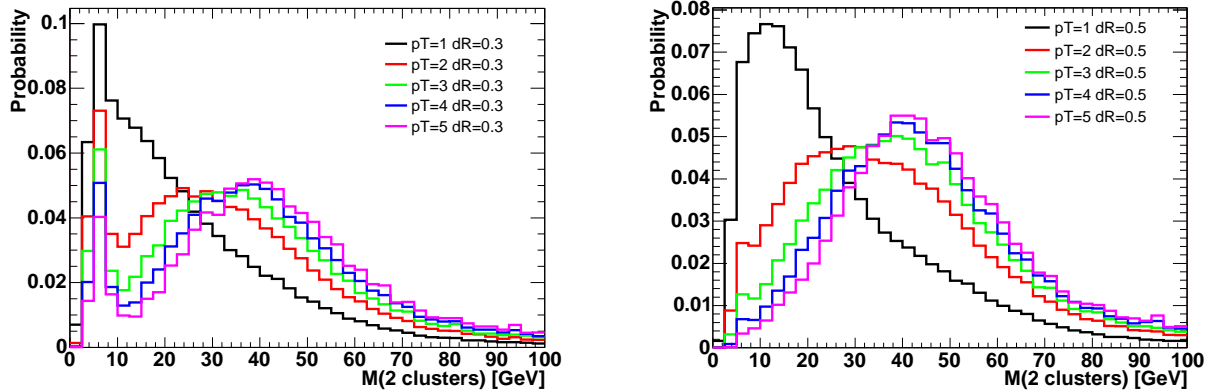


Figure 31: The mass of two track- $\pi^0$  clusters for different seed track  $p_T$  thresholds on the signal NMSSM Monte Carlo sample. On the left is the comparison using a cone radius of  $\Delta R = 0.3$  and on the right is the same for a cone radius of  $\Delta R = 0.5$ .

## 8.2 Single and Di-Tau Likelihoods Ratios

Clusters that contain two  $\tau$ 's have different kinematic properties than those that contain only one  $\tau$ . We therefore construct two separate likelihood ratios: one to discriminate single  $\tau$  clusters from jets and another to discriminate di- $\tau$  clusters from jets. The technique is the same for both ratios. We choose a set of discriminating variables  $x_1, \dots, x_n$  for each event and construct probability density functions (pdfs) for signal,  $f^s(x_i)$ , and background,  $f^b(x_i)$ . We then define the combined selection variable  $y$ , as,

$$y = \prod_{i=1}^n y_i; \quad y_i = \frac{f_i^b(x_i)}{f_i^s(x_i)}. \quad (1)$$

We can then calculate the combined likelihood as  $\mathcal{L} = 1/(1 + y)$ .

The following discriminating variables were used in the construction of the likelihood ratio probability density functions:

- The number of tracks in the cluster,  $N(\text{tracks})$ ;
- The number of  $\pi^0$ s in the cluster,  $N(\pi^0)$ .
- The number of “live” calorimeter towers,  $N(\text{live})$ , defined as the number of towers to which tracks in the cluster extrapolate.
- $p_T(\text{cluster})$ .
- $M(\text{cluster})$  where the mass is calculated using the tracks and  $\pi^0$ s contained in the cluster.
- The electromagnetic fraction of the cluster defined as  $f_{EM} = E_{EM}/(E_{EM} + E_{HAD})$  where  $E_{EM}$  and  $E_{HAD}$  refer to the electromagnetic and hadronic energies in the calorimeter, respectively.

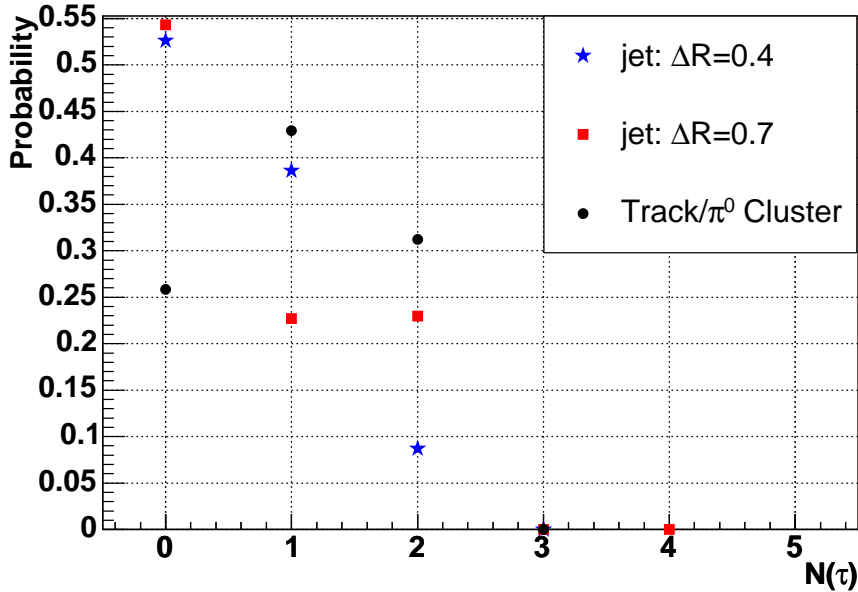


Figure 32: The probability that a cone 0.4 jet, a cone 0.7 jet, and a track- $\pi^0$  cluster contain 0, 1, 2, 3, or 4 generated  $\tau$ 's in the NMSSM  $H \rightarrow 4\tau$  Monte Carlo sample. The cluster is created using a seed track  $p_T$  threshold of 2.5 GeV, a shoulder track  $p_T$  threshold of 2 GeV, a  $\pi^0$   $p_T$  threshold of 500 MeV, and a cone radius of  $\Delta R = 0.5$ ,

- The 1<sup>st</sup> magnetic parallel moment of the cluster, defined as

$$k_{\parallel} = \sum_i^{\text{tracks}, \pi^0} \frac{\vec{v}_i \cdot \hat{v}_{\text{cluster}}}{M(\text{cluster})}, \quad (2)$$

where  $\vec{v}_i$  is the 3-vector of a track or  $\pi^0$  in the cluster,  $\hat{v}_{\text{cluster}}$  is the direction of the cluster 3-vector, and  $M(\text{cluster})$  is the mass of the cluster.

The efficiencies of a typical set of cut values for the above kinematic variables is given in Table 8.2 for data and signal Monte Carlo. Both the sequential and  $N - 1$  efficiencies are presented in this table. We see that the most effective selection is  $f_{EM}$  for data and  $k_{\parallel}$  for the Monte Carlo.

The probability density functions were constructed using signal Monte Carlo for the signal pdfs and Alpgen  $W$ +jets Monte Carlo for the background pdfs. For the discrete variables  $N(\text{tracks})$ ,  $N(\pi^0)$ , and  $N(\text{live})$ , the selection variable  $y_i$  is calculated for each bin. Figures 33, 34, and 35 show the distributions of  $N(\text{tracks})$ ,  $N(\pi^0)$ , and  $N(\text{live})$ , respectively, for single  $\tau$  clusters, di- $\tau$  clusters, and  $W$ + jets background.

For the continuous variables, the likelihood ratio calculation is implemented by fitting the distribution of each selection variable,  $y_i$ , and then multiplying these functions as per Eq. 1. The parameterization of  $y(p_T)$  for single(di-) $\tau$  discrimination is given in Eq. 4(5) and the corresponding signal and background pdfs and fits to  $y(p_T)$  are given in Figs. 36 and 37.

Cut	$\epsilon_{\text{seq}}(\text{Data})$	$\epsilon_{N-1}(\text{Data})$	$\epsilon_{\text{seq}}(\text{MC})$	$\epsilon_{N-1}(\text{MC})$
$N(\text{tracks}) < 6$	0.818	0.014	0.866	0.069
$N(\pi^0) < 6$	0.802	0.010	0.848	0.062
$N(\text{live}) < 5$	0.686	0.014	0.766	0.066
$k_{\parallel} \geq 10$	0.399	0.018	0.433	0.137
$f_{EM} \leq 0.35$ or $f_{EM} \geq 0.85$	0.097	0.030	0.187	0.130
$M(\text{cluster}) \geq 2.5$ GeV	0.010	0.027	0.061	0.112
$p_T(\text{cluster}) \geq 20$ GeV	0.009	0.010	0.061	0.061

Table 19: Cut-by-cut selection efficiencies for track- $\pi^0$  clusters applied to data and the signal Monte Carlo. Two efficiencies are given.  $\epsilon_{\text{seq}}$  is the sequential cut efficiency where the cuts are applied in the order listed in the table.  $\epsilon_{N-1}$  is the efficiency of all the selections except for the one given in the corresponding row.

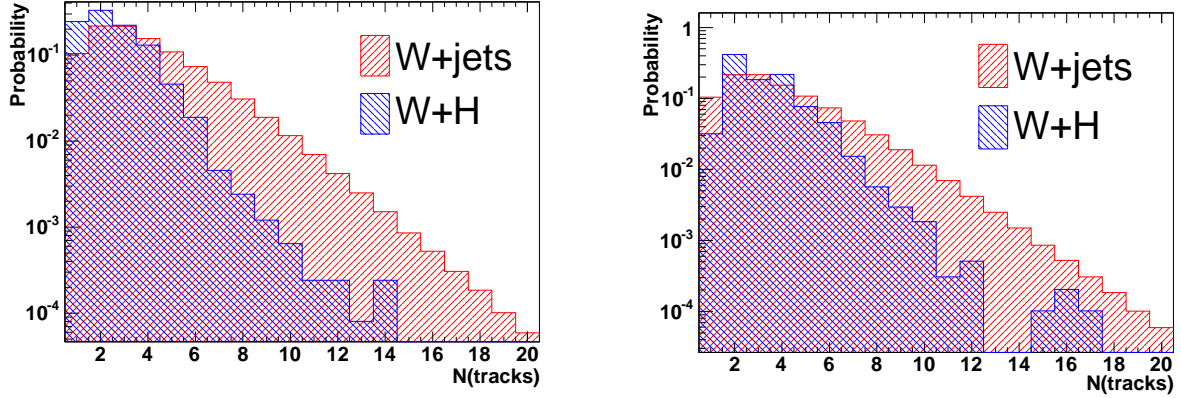


Figure 33: The signal and background distributions of the number of tracks in a cluster. On the left, the signal clusters are matched to a single generated MC  $\tau$ . On the right, the clusters are matched to two generated MC  $\tau$ 's.

$$y_{1\tau}(p_T) = \begin{cases} f_{low\ p_T}^{1\tau}(p_T) & p_T < 9.9 \text{ GeV} \\ f_{high\ p_T}^{1\tau}(p_T) & 9.9 \leq p_T < 70 \text{ GeV} \\ f_{high\ p_T}^{1\tau}(70) & p_T \geq 70 \text{ GeV} \end{cases} \quad (3)$$

$$f_{low\ p_T}^{1\tau}(x) = \exp(C + s \cdot x)$$

$$f_{high\ p_T}^{1\tau}(x) = p_2 + p_3 \cdot x + p_4 \cdot x^2 + p_5 \cdot x^3 + p_6 \cdot x^4$$

$$y_{2\tau}(p_T) = \begin{cases} f_{low}^{2\tau}(p_T) & p_T < 32 \text{ GeV} \\ f_{high}^{2\tau}(p_T) & 32 \leq p_T < 70 \text{ GeV} \\ f_{high}^{2\tau}(70) & p_T \geq 70 \text{ GeV} \end{cases} \quad (4)$$

$$f_{low}^{2\tau}(x) = p_0 + p_1 \cdot x + p_2 \cdot x^2 + p_3 \cdot x^3 + p_4 \cdot x^4 + p_5 \cdot x^5$$

$$f_{high}^{2\tau}(x) = p_6 + p_7 \cdot x + p_8 \cdot x^2$$

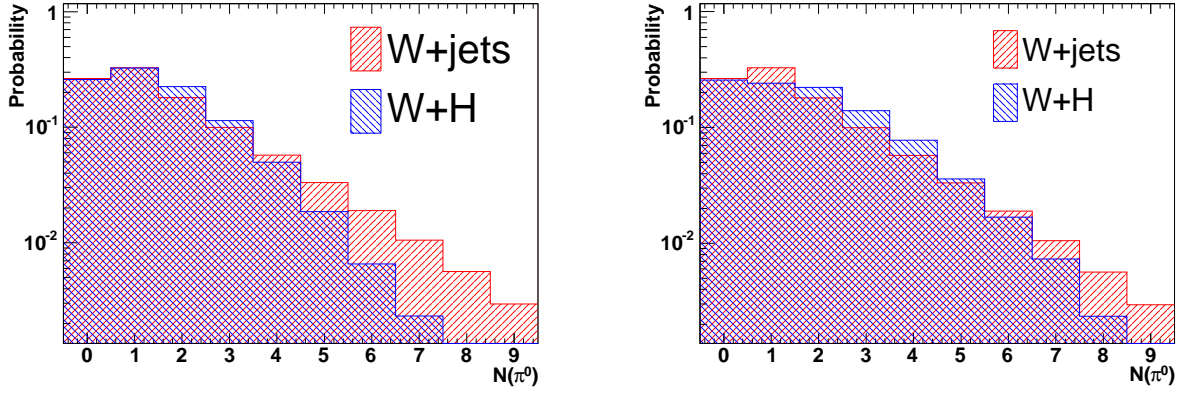


Figure 34: The signal and background distributions of the number of  $\pi^0$ s in a cluster. On the left, the signal clusters are matched to a single generated MC  $\tau$ . On the right, the clusters are matched to two generated MC  $\tau$ 's.

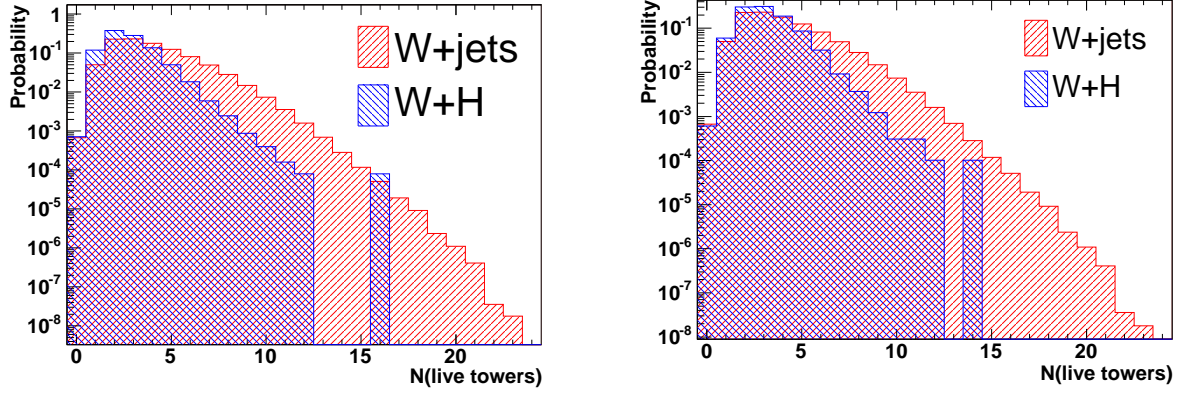


Figure 35: The signal and background distributions of the number of “live” calorimeter towers in a cluster. On the left, the signal clusters are matched to a single generated MC  $\tau$ . On the right, the clusters are matched to two generated MC  $\tau$ 's.

The parameterization of  $y(k_{||})$  for both single and di- $\tau$  discrimination is the same functional form and is given in Eq. 6. The corresponding signal and background pdfs and fits to  $y(k_{||})$  are given in Figs. 38 and 39.

$$y(k_{||}) = \begin{cases} f_{low\ k_{||}}(k_{||}) & k_{||} < 9.5\ \text{GeV} \\ f_{high\ k_{||}}(k_{||}) & 9.5 \leq k_{||} < 22\ \text{GeV} \\ f_{high\ k_{||}}(22) & k_{||} \geq 22\ \text{GeV} \end{cases} \quad (5)$$

$$f_{low\ k_{||}}(x) = p_0 + p_1 \cdot x + p_2 \cdot x^2 + p_3 \cdot x^3 + p_4 \cdot x^4$$

$$f_{high\ k_{||}}(x) = \exp(C + s \cdot x)$$

The likelihood ratio of the electromagnetic fraction is parameterized as a 6<sup>th</sup> order polynomial for both single and di- $\tau$  likelihoods. Figures 40 and 41 show the distributions of the signal and background pdfs and fits to  $y(f_{EM})$  for single and di- $\tau$  likelihoods, respectively.

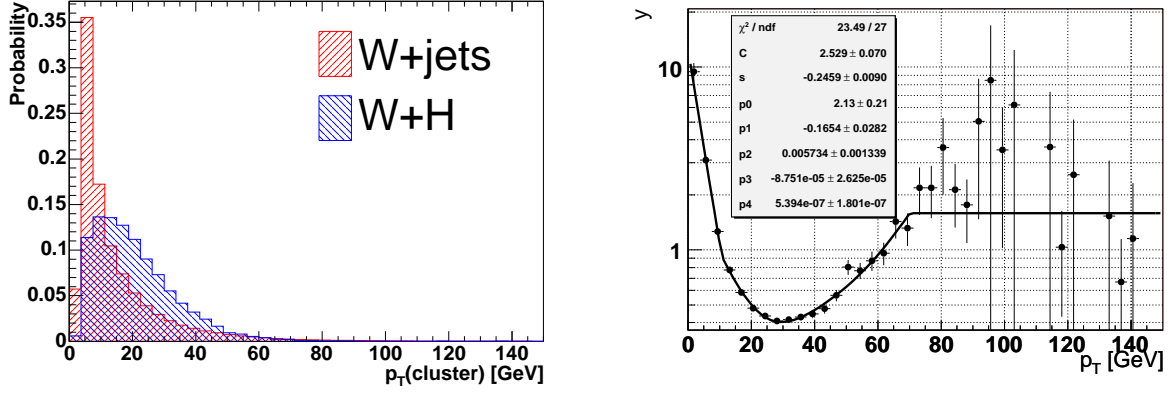


Figure 36: The signal and background distributions of cluster  $p_T$  on the left and the fit to the corresponding likelihood ratio on the right. Clusters are matched to a single MC signal  $\tau$ .

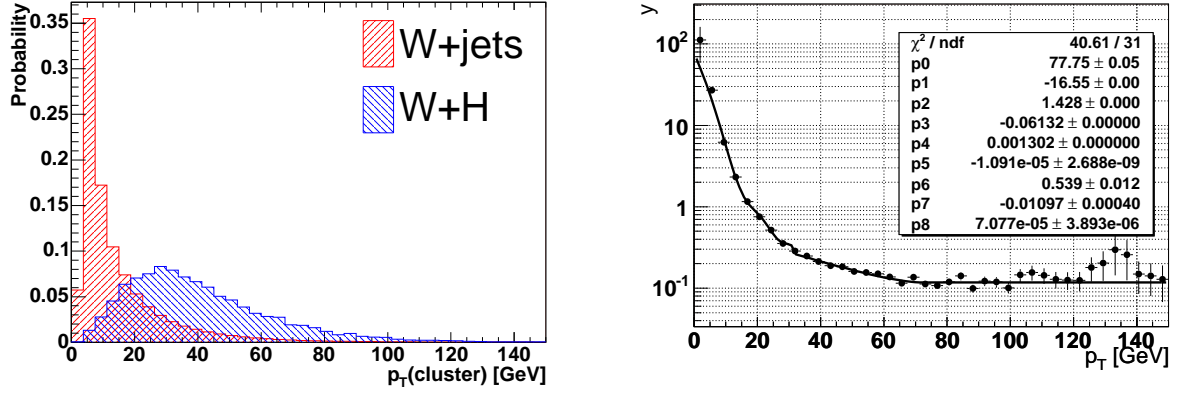


Figure 37: The signal and background distributions of cluster  $p_T$  on the left and the fit to the corresponding likelihood ratio on the right. Clusters are matched to two MC signal  $\tau$ 's.

$$y_{1\tau}(M) = \begin{cases} f_{low M}^{1\tau}(0.2) & M < 0.2 \text{ GeV} \\ f_{low M}^{1\tau}(M) & 0.2 \leq M < 2.6 \text{ GeV} \\ f_{high M}^{1\tau}(M) & 2.6 \leq M < 10.6 \text{ GeV} \\ f_{high M}^{1\tau}(10.6) & M \geq 10.6 \text{ GeV} \end{cases} \quad (6)$$

$$f_{low M}^{1\tau}(x) = p_0 + p_1 \cdot x + p_2 \cdot x^2 + p_3 \cdot x^3 + p_4 \cdot x^4 + p_5 \cdot x^5$$

$$f_{high M}^{1\tau}(x) = p_6 + p_7 \cdot x + p_8 \cdot x^2 + p_9 \cdot x^3 + p_{10} \cdot x^4 + p_{11} \cdot x^5$$

The parameterization of  $y(M)$  for single(di-) $\tau$  discrimination is given in Eq. 7(8) and the corresponding signal and background pdfs and fits to  $y(M)$  are given in Figs. 42 and 43.

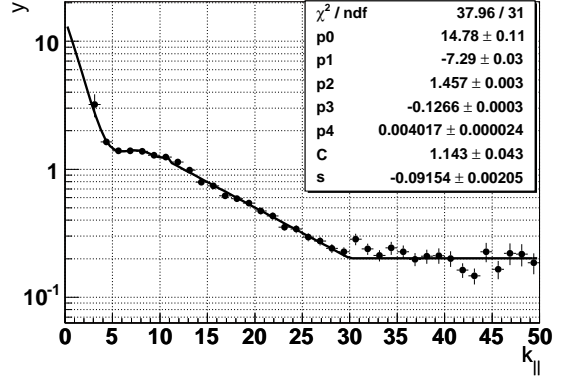
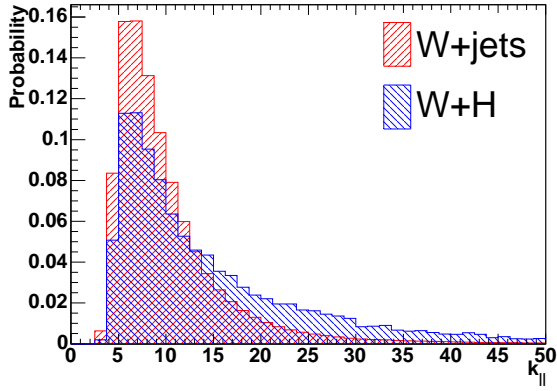


Figure 38: The signal and background distributions of cluster 1<sup>st</sup> magnetic moment on the left and the fit to the corresponding likelihood ratio on the right. Clusters are matched to a single MC signal  $\tau$ .

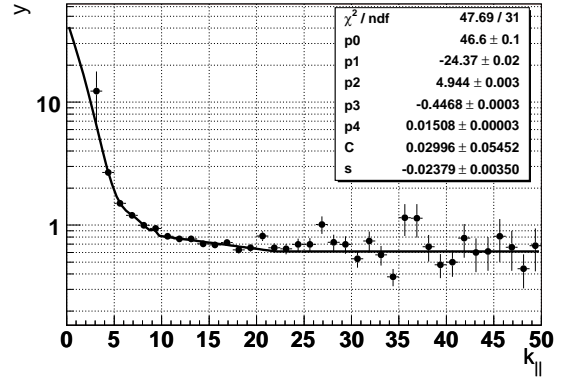
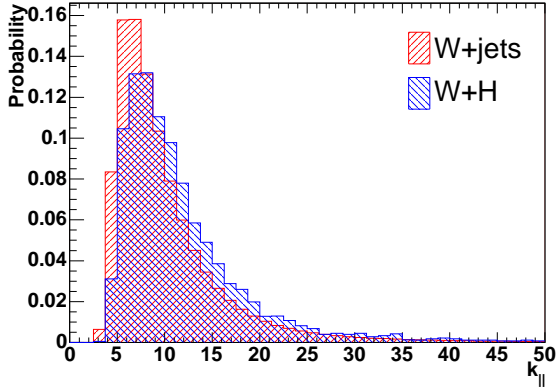


Figure 39: The signal and background distributions of cluster 1<sup>st</sup> magnetic moment on the left and the fit to the corresponding likelihood ratio on the right. Clusters are matched to two MC signal  $\tau$ 's.

$$y_{2\tau}(M) = \begin{cases} f_{low\ M}^{2\tau}(0.2) & M < 0.2 \text{ GeV} \\ f_{low\ M}^{2\tau}(M) & 0.2 \leq M < 2.85 \text{ GeV} \\ f_{high\ M}^{2\tau}(M) & 2.6 \leq M < 10 \text{ GeV} \\ f_{high\ M}^{2\tau}(10) & M \geq 10 \text{ GeV} \end{cases} \quad (7)$$

$$f_{low\ M}^{2\tau}(x) = \text{Landau Distribution}$$

$$f_{high\ M}^{2\tau}(x) = p_0 + p_1 \cdot x + p_2 \cdot x^2 + p_3 \cdot x^3 + p_4 \cdot x^4$$

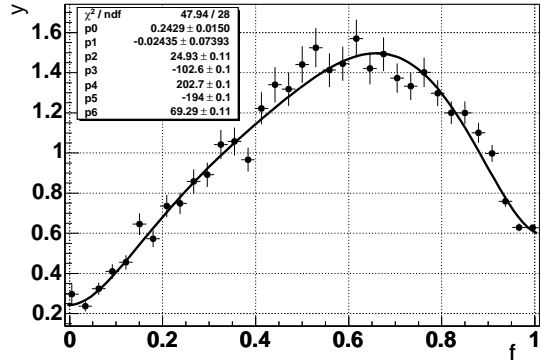
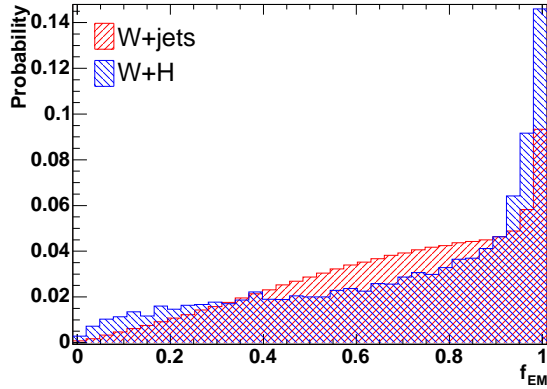


Figure 40: The signal and background distributions of cluster electromagnetic fraction on the left and the fit to the corresponding likelihood ratio on the right. Clusters are matched to a single MC signal  $\tau$ .

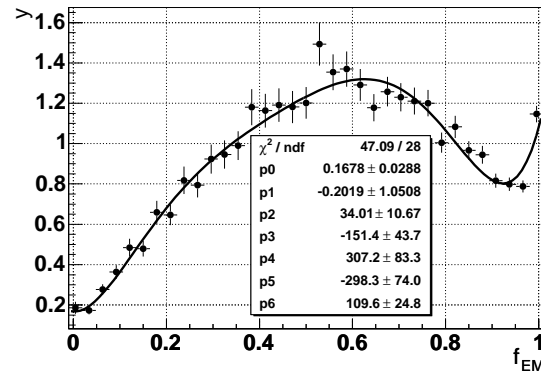
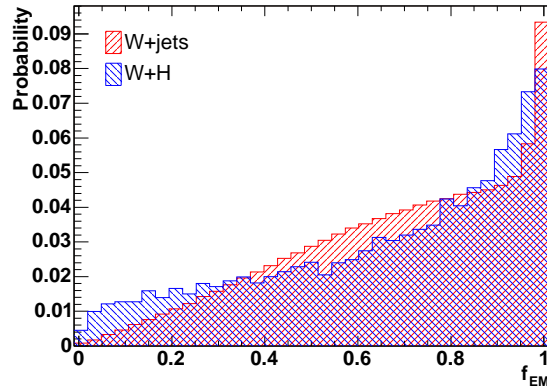


Figure 41: The signal and background distributions of cluster electromagnetic fraction on the left and the fit to the corresponding likelihood ratio on the right. Clusters are matched to two MC signal  $\tau$ 's.

## 9 Results

### 9.1 Object Counts

We expect very few signal events in our data, and those that do exist will be spread out between multiple signatures. For example, the four taus could decay into two clusters, an electron and a muon, or into four clusters, etc. Therefore, we examine multiple signatures for our signal. We create a list of possible signatures and only use those that are expected to have the highest signal-to-background ratio.

We sort events into bins based on the number of soft objects we find: Track/ $\pi^0$  clusters, soft  $e^+$ , soft  $e^-$ , soft  $\mu^+$ , and soft  $\mu^-$ . We further divide these bins by the charge correlation of overlapping objects: For example, a positive lepton overlapping a positive cluster will contribute +1 to the charge correlation, while a negative lepton overlapping a positive cluster will contribute -1. (see table 20)

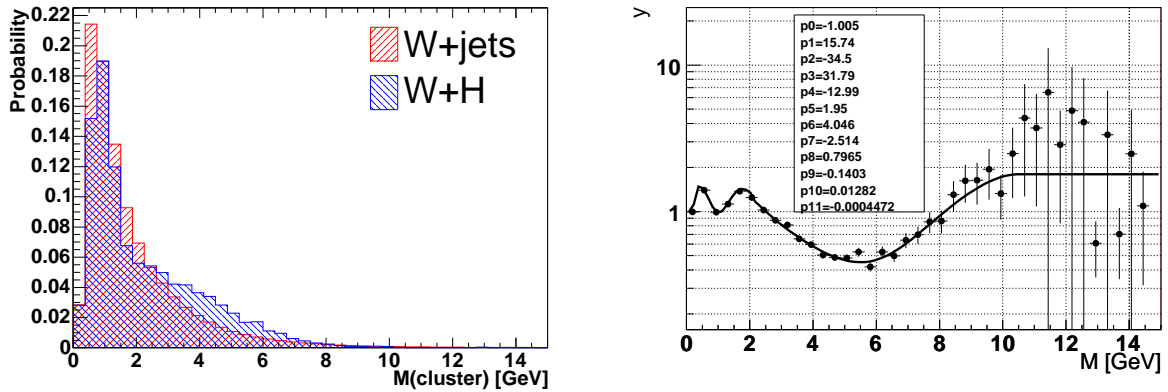


Figure 42: The signal and background distributions of cluster mass on the left and the fit to the corresponding likelihood ratio on the right. Clusters are matched to a single MC signal  $\tau$ .

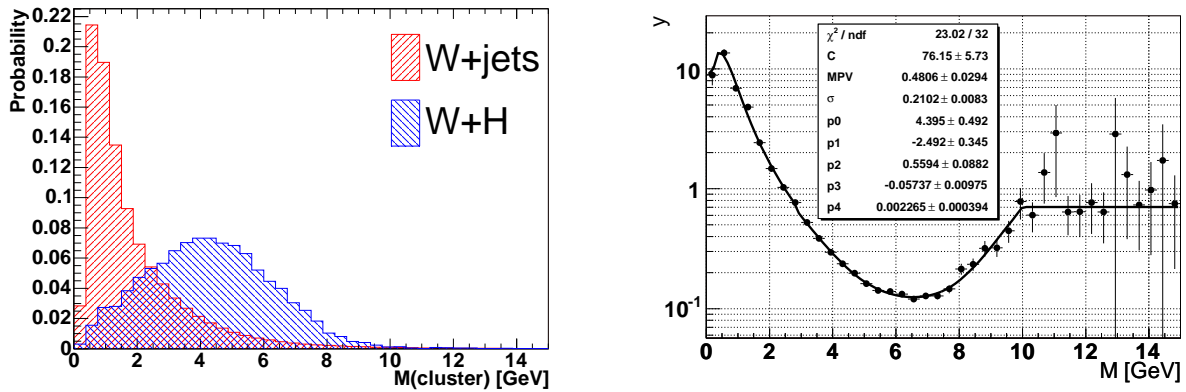


Figure 43: The signal and background distributions of cluster mass on the left and the fit to the corresponding likelihood ratio on the right. Clusters are matched to two MC signal  $\tau$ 's.

We expect that the bins with the best signal-to-background ratio will be those with both an electron and a muon.

Once we have determined the best set of bins to use, we will calculate a  $\chi^2$  of the actual event counts in each bin compared to the expected count. Additionally, we will make the kinematic plots using only those bins that are useful.

## References

- [1] Stirling NNLO W and Z ref
- [2] URL of list of run periods
- [3] B. Cooper and A. Messina, “Estimation of the Background to  $W^\pm \rightarrow e^\pm \nu + n$  Jet Events”, CDF Note 6636.

Clusters	$e^+$	$e^-$	$\mu^+$	$\mu^-$	QCorr	Background	Signal	$S/B$	Cumulative $S/\sqrt{B}$
3	0	1	1	0	2	0	0.0012		
2	1	1	1	0	-1	0	0.0012		
2	0	1	0	1	-2	0	0.0012		
2	0	0	1	1	-1	0	0.0012		
1	1	1	0	1	-2	0	0.0012		
1	0	1	1	1	-1	0	0.0012		
1	0	0	2	0	-1	0	0.0012		
1	0	0	0	2	-1	0	0.0012		
1	1	1	1	0	-1	0.0018	0.0023	1.3	0.280486
1	0	0	1	1	1	0.0018	0.0012	0.66	0.218333
1	0	1	1	0	-2	0.0062	0.0023	0.38	0.155563
2	0	1	1	0	-2	0.0047	0.0012	0.25	0.137855
1	0	3	0	0	-1	0.0053	0.0012	0.22	0.126499
2	1	0	1	0	-1	0.007	0.0012	0.17	0.116061
3	0	0	0	1	-1	0.022	0.0023	0.1	0.0964206
1	1	0	0	1	-2	0.013	0.0012	0.09	0.0905083
1	0	1	0	1	1	0.016	0.0012	0.075	0.0849686
2	0	1	0	1	0	0.016	0.0012	0.074	0.0813014
2	0	1	0	1	-1	0.036	0.0023	0.065	0.0754973
1	0	1	1	0	-1	0.02	0.0012	0.059	0.0733774
2	2	1	0	0	-2	0.021	0.0012	0.055	0.0716222
3	2	0	0	0	-2	0.022	0.0012	0.054	0.0701451
2	2	0	0	0	-2	0.065	0.0035	0.054	0.0675542

Table 20: Expected event counts in bins by number of clusters, number of leptons, and charge correlation between clusters and leptons. This count is from the control region, which uses an inverted likelihood selection on the clusters. This table continues for more rows, but the cumulative  $S/\sqrt{B}$  just gets worse.

[4] [http://www-cdf.fnal.gov/internal/physics/joint\\_physics/instructions/electron\\_cuts\\_gen6.html](http://www-cdf.fnal.gov/internal/physics/joint_physics/instructions/electron_cuts_gen6.html)