THE UNIVERSITY OF CHICAGO

A SEARCH FOR ANOMALOUS PRODUCTION OF EVENTS WITH MULTIPLE LEPTONS AND A W OR Z BOSON

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
SCOTT WILBUR

CHICAGO, ILLINOIS
DECEMBER 2011

TABLE OF CONTENTS

LIST OF FIGURES v

LIST OF TABLES viii

ACKNOWLEDGMENTS xi

ABSTRACT xiii

1 INTRODUCTION 1

2 ANALYSIS STRATEGY 3

3 THE CDF II DETECTOR 4

4 W AND Z BOSON SAMPLE SELECTION 7
- 4.1 W Selection 8
- 4.2 Z Selection 8

5 QCD BACKGROUND ESTIMATION TECHNIQUE 11
- 5.1 Anti-Selected Electron Definition 11
- 5.2 Fits to the Missing Transverse Energy Distribution 11

6 W AND Z SAMPLE VALIDATION 14
- 6.1 W Boson Sample Validation 14
  - 6.1.1 $W \to e\nu$ events 14
  - 6.1.2 $W \to \mu\nu$ events 14
- 6.2 Z Sample Validation 18
  - 6.2.1 $Z \to e^+e^-$ 18
  - 6.2.2 $Z \to \mu^+\mu^-$ 19
- 6.3 The Ratio of W to Z Production 'R' as a Precision Check 22

7 SOFT LEPTON IDENTIFICATION 25
- 7.1 Soft Electrons 25
- 7.2 Identification Algorithm and Candidate Selections 25
  - 7.2.1 Training Samples and Efficiency and Misidentification Rate Measurements 26
  - 7.2.2 Validation and Systematic Uncertainty Determination 30
- 7.3 Soft Muons 31
  - 7.3.1 Identification Algorithm and Candidate Selections 31
  - 7.3.2 Efficiency and Misidentification Rate Measurements 33
  - 7.3.3 Soft Muon Systematic Uncertainty Determination 38
- 7.4 Application of Soft Lepton Identification to W/Z Samples 43

8 BACKGROUND PREDICTION 44
8.1 Heavy Flavor Background Fraction 44
8.2 Normalization of Soft Electron Multiplicities 46

9 RESULTS 47
9.1 Soft Lepton Multiplicity 47
9.2 Soft Lepton Kinematics 52
9.3 Benchmark Model 52
9.4 Application to Other Models 54

10 CONCLUSIONS 56

BIBLIOGRAPHY 57

A DATA SAMPLES 61
A.1 Signal Region Data Samples 61
A.2 Soft Electron Identification Data Samples 61
A.3 Soft Muon Identification Data Samples 63

B MONTE CARLO SAMPLES 64

C SOFT ELECTRON LIKELIHOOD VARIABLES 69

D SUMMARY OF MULTI-LEPTON EVENTS 73

LIST OF FIGURES

1.1 A diagram of a Higgs decay to a pair of lightest supersymmetric neutralinos ($\tilde{N}_1$) which then cascade through a dark sector to a lightest dark sector particle ($\tilde{n}_d$ and a number of dark photons (γ_d). The dark photons then decay back into the SM in the form of leptons. This model is adapted from Ref. [11]. 2

3.1 An elevation view of the CDF II detector. The detector is approximately symmetric around the collision point. 6

5.1 The fits to the $\not{E}_T$ distribution of events with $m_T > 20$ GeV and $\Delta\phi(\not{E}_T, l) > 0.5$, in each of the three trigger lepton categories. The "electroweak" template is obtained from Monte Carlo and the "QCD" template is obtained from the anti-selected electron data sample. The systematic uncertainty of 26% found in [30] is shown. 13

6.1 W validation distributions, electron trigger: p_T of the highest-p_T good electron, $\not{E}_T$ and H_T in the event, M_T of electron and $\not{E}_T$. The W selections of $\not{E}_T > 25$ GeV and $m_T > 20$ GeV are included. 15
6.2 W validation distributions, CMUP trigger: p_T of the highest-p_T good muon, $\not{E}_T$ and H_T in the event, M_T of the good highest-p_T muon and $\not{E}_T$. The W selections of $\not{E}_T > 25$ GeV and $m_T > 20$ GeV are included. 16
6.3 W validation distributions, CMX trigger: p_T of the highest-p_T good muon, $\not{E}_T$ and H_T in the event, M_T of the good highest-p_T muon and $\not{E}_T$. The W selections of $\not{E}_T > 25$ GeV and $m_T > 20$ GeV are included. 17
6.4 Distributions of e-triggered Z events: the p_T of the two leading electrons, the H_T in the event, and the dilepton mass. In all the distributions except the mass distribution, there is a requirement that the dilepton mass is between 76 and 106 GeV. 18
6.5 Distributions of CMUP μ-triggered Z events: the p_T of the two leading electrons, the H_T in the event, and the dilepton mass. In all the distributions except the mass distribution, there is a requirement that the dilepton mass is between 76 and 106 GeV. 20
6.6 Distributions of CMX μ-triggered Z events: the p_T of the two leading electrons, the H_T in the event, and the dilepton mass. In all the distributions except the mass distribution, there is a requirement that the dilepton mass is between 76 and 106 GeV. 21
6.7 The ratio, R, of W and Z cross sections for each run period, using the electron trigger. 23
6.8 The ratio, R, of W and Z cross sections for each run period, using the CMUP muon trigger. 23
6.9 The ratio, R, of W and Z cross sections for each run period, using the CMX muon trigger. 24

7.1 Variables used to identify photon conversions to electron-positron pairs. On the left are the variables defined in the plane transverse to the beam. The beam position is denoted by an "x". R is the distance between the beam position and the point at which the two tracks are tangential or parallel to each other and δ is the distance between the two tracks at that point. On the right is the distribution of $\Delta\cot(\theta)$, where θ is the polar angle of the track in the $r-z$ plane. A fit is performed to find the signal (solid) and background (dashed) to estimate the sample composition under the peak. 27
7.2 Left: The likelihood distributions for electrons (closed squares) and non-electrons (open circles) after all preselection criteria. Only those candidates with a likelihood > 0.99 are identified as electrons. Right: The efficiency as a function of p_T for the identification of electrons and tracks misidentified as electrons after the likelihood selection. 28
7.3 Examples of fits to the soft electron efficiency (left) and fake rate (right) as a function of track isolation in a particular p_T and η bin. 28
7.4 A comparison of the measured and predicted efficiencies (left) and fake rates (right). The kinematic variables shown, from top to bottom, transverse momentum p_T, pseudo-rapidity η, and track isolation are those used to characterize the response of the soft electron algorithm. 29
7.5 Predicted and observed soft electron misidentification rates obtained from a QCD (jet) sample. On the left are the identification rates as function of p_T, and on the right the the identification rates as functions of η. 30
7.6 Results of the J/ψ mass fits in bins of the p_T of the softer, not-triggered, candidate leg of the J/ψ. 34
7.7 Results of the D^* mass fits in bins of the p_T of candidate π's coming from $D^0 \to K\pi$. 36
7.8 Results of the Λ mass fits in bins of the p_T of candidate p's. 37
7.9 A comparison of the soft muon likelihood distributions for μ, π, K, and p. . . . 38
7.10 Identification efficiency as a function of p_T for μ, π, K, and p. For the case of the μ, this is the rate at which real muons are identified. For the other species, it is the rate that the particle is misidentified as a muon. 39
7.11 Fit for soft muon efficiency function over J/ψ and Z events in the η range $|\eta| < 0.15$. 40
7.12 Soft muon ID rate matrix in bins of p_T and η. 41
7.13 Soft muon identification efficiency as a function of the fractional isolation of the muon obtained from the J/ψ sample. 41
7.14 Observed and predicted soft muon rate in the JET100 sample with the selections as described in the text as a function of p_T. 42

8.1 The result of the simultaneous fit of the $W+1$ soft muon sample in the p_T^{rel} and d_0 significance of the soft muon. The data distribution is fit to the sum of three components: W+heavy quark, W+light quark/gluon, and Drell-Yan. 45

9.1 Multiplicity of additional electrons and muons after the W boson selection. The two-dimensional histogram of N_μ vs. N_e is presented in slices of N_e for ease of viewing. Both hard and soft leptons (but not the initial leptons used for the W or Z identification) are counted. Note that the distributions combine the electron- and muon-triggered events. 48
9.2 Multiplicity of additional electrons and muons after the Z selection. The two-dimensional histogram of N_μ vs. N_e is presented in slices of N_e for ease of viewing. Both hard and soft leptons (but not the initial leptons used for the W or Z identification) are counted. Note that the distributions combine the electron- and muon-triggered events. 49
9.3 Distributions of the invariant mass of each pair of soft leptons ee (top), $e\mu$ (center), and $\mu\mu$ (bottom). The W-selected events are on the left and the Z-selected events are on the right. Note that the distributions combine the electron-and muon-triggered events. The contribution from conversions swamps any new physics signal in the $m(e, e)$ and $m(e, \mu)$ distributions, but the $m(\mu, \mu)$ distribution (bottom) is sensitive to the benchmark model as well as to other new physics models. 53

C.1 A comparison of dE/dx between electrons and non-electrons. The lowest-momentum tracks, with $p_T < 1.5$ GeV, are shown on the left and the higher-momentum tracks, with $p_T > 6$ GeV, are shown on the right. 70
C.2 A comparison of the calorimeter energy distributions ($\frac{EmE}{p \cdot iso}$ on the left $\frac{E_{Had}}{E_{EM}}$ on the right) between electrons and non-electrons. 70
C.3 A comparison of E_{CES}/p on the left and E_{CPR} on the right for electrons and non-electrons. 71
C.4 A comparison of CES ΔX and ΔZ for electrons and non-electrons. 72

LIST OF TABLES

4.1 Selections to identify tight central electrons 8
4.2 Selections to identify muons. 9
4.3 Selections to identify loose central electrons 9
4.4 Event counts in W and Z boson samples, split up by categories of the leading and subleading leptons. 10

5.1 Central tight electron identification selections [26] divided into two categories: those that shape the kinematics of the event, referred to as "kinematic", and those that discriminate between electrons and misidentified jets, referred to as "identification". 12

6.1 Summary of the predictions for $\mathrm{W}^{\pm} \to e^{\pm}\nu$. The QCD contribution is estimated as in Sec. 5. The other contributions are calculated from the MC, scaled according to the fit described in Section 5.2. 15
6.2 Summary of the predictions for $\mathrm{W}^{\pm} \to \mu^{\pm}\nu$ with a CMUP trigger. The QCD contribution is estimated as in Sec. 5. The other contributions are calculated from the MC, scaled according to the fit described in Section 5.2. 16
6.3 Summary of the predictions for $\mathrm{W}^{\pm} \to \mu^{\pm}\nu$ with a CMX trigger. The QCD contribution is estimated as in Sec. 5. The other contributions are calculated from the MC, scaled according to the fit described in Section 5.2. 17
6.4 Summary of the predictions and observations for $Z/\gamma^* \to e^+e^-$ where the trigger is TCE. The prediction is shown over the entire range of $m(e^+e^-)$ as well as in the Z mass window, $76 < m(e^+e^-) < 106$ GeV. 19
6.5 Summary of the predictions and observations for $Z/\gamma^* \to \mu^+\mu^-$ where the trigger is CMUP muon. The prediction is shown over the entire range of $m(\mu^+\mu^-)$ as well as in the Z mass window, $76 < m(\mu^+\mu^-) < 106$ GeV. 20
6.6 Summary of the predictions and observations for $Z/\gamma^* \to \mu^+\mu^-$ where the trigger is CMX muon. The prediction is shown over the entire range of $m(\mu^+\mu^-)$ as well as in the Z mass window, $76 < m(\mu^+\mu^-) < 106$ GeV. 21
6.7 The measured ratio R of W boson to Z boson production, and the trigger rate uncertainty calculated by comparing it to the NNLO calculated value of 10.67. . 22

7.1 Soft muon candidate selection criteria. 32
7.2 Number of events predicted by applying the soft muon misidentification matrix and observed in JET50 and JET100 data. 40

8.1 The result of the fit to correct the heavy flavor fraction. 45

9.1 Summary of predicted and observed event counts by number of additional electrons (N_e) and muons (N_μ) after the W boson selection. The prediction of a model described in Section 9.3 is also shown for comparison. Bins with less than 0.25 expected events in both signal and background and 0 observed events are not shown. 50

9.2 Summary of predicted and observed event counts by number of additional electrons (N_e) and muons (N_μ) after the Z selection. The prediction of a model described in Section 9.3 is also shown for comparison. Bins with less than 0.25 expected events in both signal and background and 0 observed events are not shown. 51
9.3 Sources of systematic uncertainties. Their size is measured both as a percentage and as the number of events in a benchmark-signal-rich region, defined as a W or Z boson plus at least 3 additional muons with $p_T > 3$ GeV. Note that, although some of the systematics are large, they have little effect in the signal region due to there being negligible SM background. 51
9.4 Parameters used for the benchmark model based on that in Ref. [11]. The first five parameters are the inputs to the MSSM including the branching fractions for $\chi_0 \to \chi_d + N\gamma_d$ [38]. 54
9.5 Summary of kinematic requirements to find various objects. These numbers can be used to set limits on many models that predict production of additional leptons. 55

A.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 the triggers are unprescaled and the same good run list is used for both triggers. 62

B.1 Alpgen $W+$ partons dataset names, processes, cross sections, and number of events generated. Note that the cross-sections listed include a K-factor of 1.4. . 65
B.2 Alpgen $Z+$ partons dataset names, processes, cross sections, and number of events generated. The mass range in the process name refers to the generated mass range of the dilepton pair. Note that the cross-sections listed do not include a K-factor. 65
B.3 Alpgen Drell-Yan + partons dataset names, processes, cross sections, and number of events generated. The mass range in the process name refers to the generated mass range of the dilepton pair. Note that the cross-sections listed do not include a K-factor. 66
B.4 Alpgen $W+$ heavy quark dataset names, processes, cross sections, and number of events generated. Note that the cross-sections listed include a K-factor of 1.4. 67
B.5 Alpgen Drell-Yan + heavy quark dataset names, processes, cross sections, and number of events generated. The mass range in the process name refers to the generated mass range of the dilepton pair. Note that the cross-sections listed do not include a K-factor. 68
B.6 Pythia dataset names, processes, cross sections, and number of events generated. The cross-sections listed include a K-factor of 1.4. 68

D.1 Summary of run 211441, event 7455273. This event passed the CMUP-triggered W boson selection. 73
D.2 Summary of run 222885, event 4745161. This event passed the TCE-triggered W boson selection. 74
D.3 Summary of run 228735, event 17076. This event passed the CMX-triggered W boson selection. 74

D.4 Summary of run 231693, event 2501991. This event passed the TCE-triggered W boson selection. 74
D.5 Summary of run 273941, event 2792032. This event passed the CMX-triggered W boson selection. 75
D.6 Summary of run 275385, event 9123310. This event passed the TCE-triggered W boson selection. 75

ACKNOWLEDGMENTS

This work wouldn't have been possible without the help and support of a large number of people. I am indebted to them all for their help, whether it was with physics or with other aspects of the work.

First of all, I would like to thank my advisor, Professor Henry Frisch. This work wouldn't have been possible without his guidance, both in the actual physics work and in navigating the occasional mires of bureaucracy I found along the way. His instruction in experimental physics techniques was always instructive, and his instruction in the negotiation techniques of famous painters has proven to be just as useful.

Secondly, none of this would have happened without the support of my family. If my parents hadn't taught me a love of science early, I never would have gotten this far. They deserve most of the credit for saving me from the fate of being a humanities major.

I'd like to extend special thanks to my coworkers here at Chicago, Dan Krop and Carla Pilcher. They greatly assisted me with parts of this work, and also worked with me on other analyses. The soft muon section of this analysis, in particular, owes much of its content to Dan's work.

This analysis would never have been published without the assistance of the conveners and godparents at CDF. Ben Kilminster and Eric James, the conveners of the Higgs Discovery Group, guided me through the blessing process and made sure that my work was ready to be presented at various conferences. Eva Halkiadakis, Viviana Cavaliere and Aidan Robson, the godparents for this analysis, took my rough draft of a paper and helped me shape it into a manuscript of publishable quality.

I would also like to extend my gratitude to all of my professors here at the University of Chicago. The classes I took from Professors Kathy Levin, Jon Rosner, Bob Wald, Bob Geroch, Carlos Wagner and Henry Frisch were all enlightening, even those that were on topics far outside my field of study.

Finally, my sincere thanks goes to my entire thesis committee. Thanks to Henry Frisch,

Jim Pilcher, Bob Rosner, and Carlos Wagner for reviewing my work, and most importantly for signing their names to that all-important piece of paper.

ABSTRACT

A search is presented for anomalous production of events containing multiple low-energy leptons produced in association with a W or Z boson using the Collider Detector at Fermilab. The leptons are not required to be isolated, and the signature is therefore sensitive to a wide range of so-called “lepton jet” topologies. The search uses data corresponding to 5.1 fb^{-1} of integrated luminosity from $p\bar{p}$ collisions at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV. The observed events are compared with the standard model predictions in bins of additional electron and muon multiplicity. No indications are found of phenomena beyond the standard model. A 95% confidence level limit is presented on the production cross section for an example benchmark point of a dark-matter inspired model of supersymmetric hidden valley Higgs production. Object identification efficiencies are also provided in order to enable the calculation of limits on additional models.

CHAPTER 1

INTRODUCTION

The standard model (SM) of particle physics [1] has been very successful in predicting observations made at high-energy particle colliders. However, it is known to be incomplete. There must be new physics beyond the reach of the current generation of experiments, and there are a plethora of theories predicting what that new physics might be.

Testing all of these models individually would be impossible, due to the number of models and the number of free parameters in each model. Therefore, "signature-based" searches are performed: a signature is chosen that is common to many new models, but has a low rate of background production in the SM. A search for this signature can both act as a precision test of the SM (by looking for disagreement with the low predicted background) and check for hints of many different possible theories of new physics.

The signature of multiple leptons is common in many models of physics beyond the SM with light mass scales and couplings to the electroweak sector, such as the Next-to-Minimal Supersymmetric Model [2], little Higgs models [3], and R-parity violating MSSM models [4]. Some of these new physics scenarios propose explanations for the nature of dark matter [5] as well as the existence of other, yet-undiscovered particles in long decay chains. In addition to predicting large numbers of leptons, these models also often predict that clusters of leptons are produced spatially close to each other. These clusters are often referred to in the literature as "lepton jets" [6]. Due to the unique characteristics of these models, they could have evaded previous searches for an excess of leptons, such as diboson searches [7] and SUSY-inspired multi-lepton searches [8]. The high multiplicity of leptons can lead to low lepton momenta, well below the usual cutoff of 10-20 GeV. Additionally, collimated lepton jets will fail the standard requirement that leptons be isolated in the detector.

One of the recent promising proposals involves the phenomenology of light supersymmetric hidden sectors [9] where the lightest visible superpartner, the equivalent of the LSP in the MSSM, is allowed to cascade into a hidden sector. The existence of such sectors has been further motivated by recent observed astrophysical anomalies [10] which may be signatures of dark matter annihilation [5] or decays into a light hidden sector [6].

As an example, Figure 1.1 shows a typical decay chain in a model in which the Higgs decays to a light hidden sector resulting in events with a high multiplicity of leptons [11].

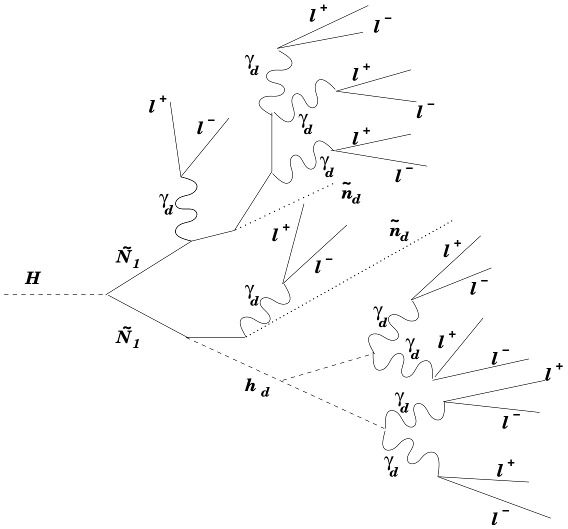

Figure 1.1: A diagram of a Higgs decay to a pair of lightest supersymmetric neutralinos ($\tilde{N}_1$) which then cascade through a dark sector to a lightest dark sector particle ($\tilde{n}_d$ and a number of dark photons (γ_d). The dark photons then decay back into the SM in the form of leptons. This model is adapted from Ref. [11].

This thesis presents a signature-based search for anomalous production of multiple leptons in association with W and Z bosons. Previous searches for lepton jets at the Tevatron [12] and at the LHC [13] have focused on searching for clusters of leptons, sometimes with very specific requirements on the size of the clusters. These searches have resulted in no evidence for lepton jets. A more general search, sensitive to a wide range of scenarios that predict multiple leptons, is performed here.

The data used here correspond to 5.1 fb^{-1} of integrated luminosity at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV collected using the CDF detector at Fermilab between December 2004 and January 2010. Within the events containing leptonically decaying W and Z bosons, a search is performed for additional 'soft' leptons with no isolation requirements and with momentum greater than 3 GeV for muons and 2 GeV for electrons [14].

CHAPTER 2

ANALYSIS STRATEGY

The baseline data sets for this analysis consist of leptonically decaying W and Z boson events selected with high-p_T leptons [26]. The selection of these events is described in Chapter 4. The kinematic distributions are used to validate the W and Z boson selections. A check of the ratio of cross-sections $\sigma(W)/\sigma(Z)$ is also used to validate the trigger efficiency and the lepton reconstruction efficiency. This ratio R of W to Z production is predicted to NNLO with a precision of a few percent [27], providing a very precise test of these efficiencies.

After the W or Z boson reconstruction, additional low-p_T electrons and muons are identified in the events with no isolation requirements. Purely data-driven techniques are used to develop the soft lepton identification algorithms. The selection of soft leptons is more fully described in Chapter 7.

The numbers of additional electrons and muons are counted in the inclusive W and Z data sets. The SM predicts very few W and Z boson events with multiple additional leptons, and so this is the analysis search region. We search for anomalous production of additional electrons and muons in association with W and Z bosons by comparing the observed events to the SM expectations in bins of additional lepton multiplicity. A limit is set on an example benchmark model based on the count of events with more than one additional lepton, and a table of efficiencies is provided to facilitate setting limits on other models. These results are described in Chapter 9.

CHAPTER 3

THE CDF II DETECTOR

The CDF II detector is a cylindrically-symmetric spectrometer designed to study $p\bar{p}$ collisions at the Fermilab Tevatron. The detector has been extensively described in detail elsewhere in the literature [15]. An elevation view of the detector is shown in Figure 3.1 Here the detector subsystems relevant for this analysis are described.

Tracking systems are used to measure the momenta of charged particles, to reconstruct primary and secondary vertices, and to trigger on and identify leptons with large transverse momentum [16]. Silicon strip detectors [17] and the central outer tracker (COT) [18] are contained in a superconducting solenoid that generates a magnetic field of 1.4 T. The silicon strip system provides up to 8 measurements in the $r - \phi$ and $r - z$ views and helps to reconstruct tracks in the region $|\eta| < 2$ [16]. The COT is an open-cell drift chamber that makes up to 96 measurements along the track of each charged particle in the region $|\eta| < 1$. Sense wires are arranged in 8 alternating axial and $\pm 2^\circ$ stereo super-layers. The resolution in p_T, σ_{p_T}/p_T, is $\approx 0.0015\, p_T/\mathrm{GeV}$ for tracks with only COT measurements, and $\approx 0.0007\, p_T/\mathrm{GeV}$ for tracks with both the silicon and COT measurements.

Calorimeters are segmented with towers arranged in a projective geometry. Each tower consists of an electromagnetic and a hadronic compartment [19, 20, 21]. The central electromagnetic calorimeter (CEM) and central hadronic calorimeter (CHA) cover the central region ($|\eta| < 1.1$), while the plug electromagnetic calorimeter (PEM) and plug hadronic calorimeter (PHA) cover the 'end plug' region ($1.1 < |\eta| < 3.6$). In this analysis, a high-E_T electron is required to be identified in the central region, where the CEM has a segmentation of 15° in ϕ and ≈ 0.1 in η [15], and an E_T resolution of $\sigma(E_T)/E_T \approx 13.5\%/\sqrt{E_T/\mathrm{GeV}} \oplus 2\%$ [19]. Two additional systems in the central region with finer spatial resolution are used for electron identification in this analysis. The central strip system (CES) uses a multi-wire proportional chamber to make profile measurements of electromagnetic showers at a depth of 6 radiation lengths (approximately shower maximum) [19]. The central preshower detector (CPR) is located just outside the solenoid coil on the front face of the CEM. In 2004 the CPR was upgraded from the Run I configuration of wire proportional chambers, similar to those used in the CES, to a fast scintillator system [21]. This analysis only uses data collected after the CPR upgrade.

Muons are identified using the central muon systems [22]: CMU and CMP for the pseudo-rapidity region of $|\eta| < 0.6$, and CMX for the pseudo-rapidity region of $0.6 < |\eta| < 1.0$. The CMU system uses four layers of planar drift chambers to detect muons with $p_T > 1.4$ GeV. The CMP system consists of an additional four layers of planar drift chambers located behind 0.6 m of steel outside the magnetic return yoke, and detects muons with $p_T > 2.2$ GeV. The CMX system detects muons with $p_T > 1.4$ GeV with four to eight layers of drift chambers, depending on the direction of the muon.

The luminosity is measured using two sets of gas Cerenkov counters [23], located in the region $3.7 < |\eta| < 4.7$. The total uncertainty on the luminosity is estimated to be 5.9%, where 4.4% comes from the acceptance and operation of the luminosity monitor and 4.0% from the calculation of the inelastic $p\bar{p}$ cross-section [24].

A three-level online event selection (trigger) system [25] selects events to be recorded for further analysis. The first two trigger levels consist of dedicated fast digital electronics analyzing a subset of the full detector data. The third level, applied to the full data of those events passing the first two levels, consists of a farm of computers that reconstruct the data and apply selection criteria consistent with the subsequent offline event processing.

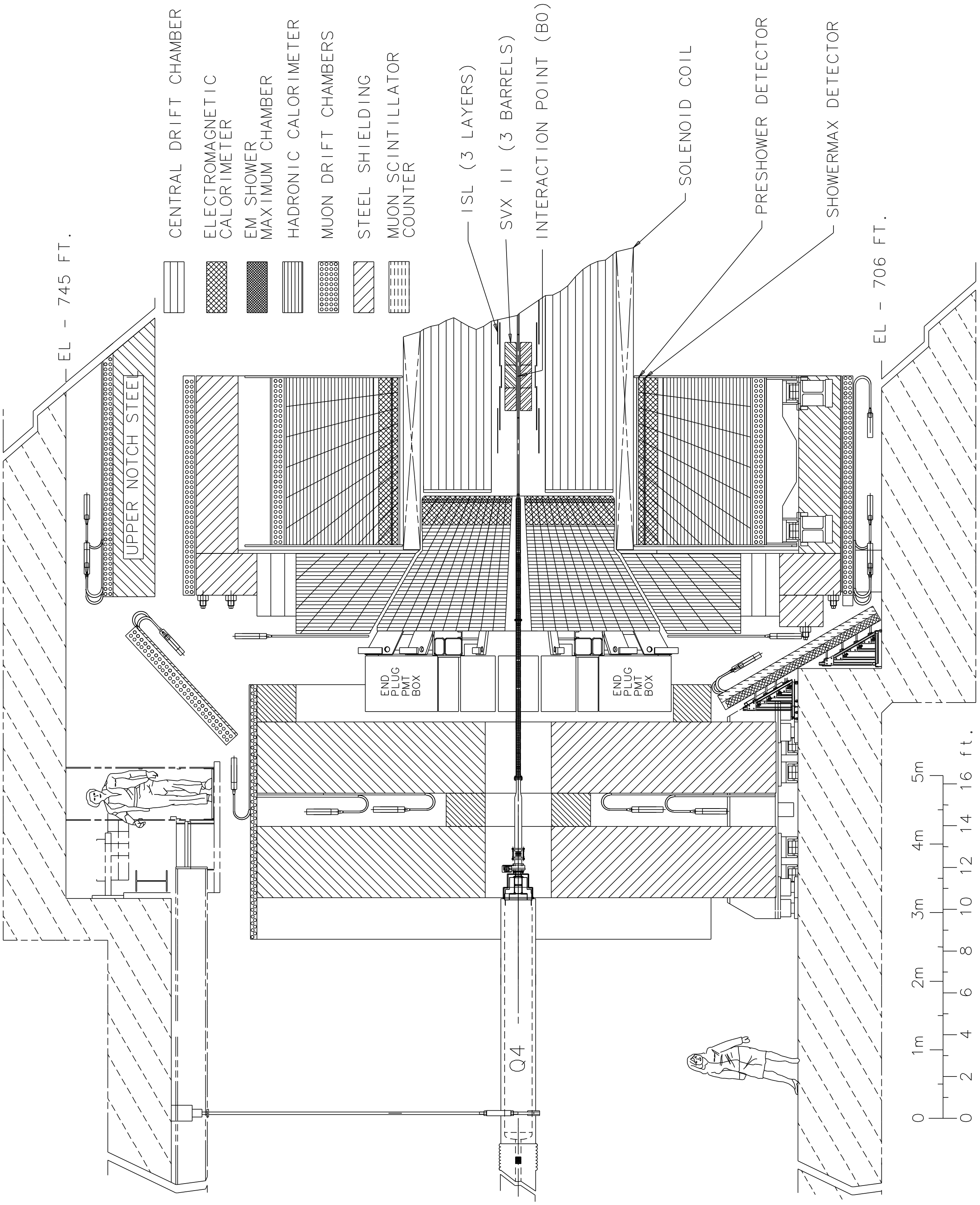

Figure 3.1: An elevation view of the CDF II detector. The detector is approximately symmetric around the collision point.

CHAPTER 4

W AND Z BOSON SAMPLE SELECTION

Events for this analysis are selected with three different triggers [25]. Around half the events are selected with a trigger requiring a high-p_T central electron in the CEM ($E_T > 18$ GeV, $|\eta| < 1.0$). In addition, two muon triggers, one requiring hits in both the CMP and CMU and the other requiring hits in the CMX, collect events with central muons ($p_T > 18$ GeV, $|\eta| < 1.0$). These datasets are described more fully in Appendix A.

Further selection criteria are imposed on triggered events offline. The standard CDF selections [26] are used to identify hard ($>$ 20 GeV) electrons and muons, with the additional requirement that hard muons have silicon hits. Two categories of hard leptons are considered: 'tight' leptons, which are described in Tables 4.1 and 4.2, are required as the trigger leg of the W and Z. A less stringent category of 'loose' leptons, described in Tables 4.2 and 4.3, are required for the non-trigger leg of Z reconstruction.

In order to reduce the electron background from photon conversions, the electron(s) from the W or Z boson decay are required to pass a conversion filter. If there is an electron candidate that has opposite charge, $\Delta \cot\theta < 0.04$ and $|\delta| < 0.2$, then the electron is called a conversion. (See Fig. 7.1 for an explanation and illustration of these variables.) However, the electron candidate is kept if its partner conversion track also has another partner track, since the three tracks are assumed to originate from an electron which radiates a photon which subsequently converts.

In order to reduce the background from mesons decaying to muons within the tracking chamber, the muon(s) from the W or Z boson decay must pass a decay-in-flight (DIF) removal algorithm. A track left by a pion or a kaon that decayed to a muon within the detector will leave real hits in the muon detectors, but the track in the tracking chamber will have a noticeable 'kink' in it where the decay occurred. The DIF algorithm removes these tracks by requiring the χ^2 per degree of freedom of the track fit to be less than 3 and the impact parameter of the track to be less than 0.02 cm. Additionally, for tracks with

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

Table 4.1: Selections to identify tight central electrons

$p_T > 300$ GeV, it requires $N_{transitions} > 30$, where $N_{transitions}$ is the number of times the pattern of track hits crosses the fitted track [28]. Muons consistent with cosmic rays are vetoed [29].

4.1 W Selection

The W boson selection requires a tight trigger lepton (defined above) with $\mathrm{E_T} > 20$ GeV, $\not\!\!E_T > 25$ GeV, and transverse mass of lepton+$\not\!\!E_T > 20$ GeV. In order to remove events where the $\not\!\!E_T$ arises from a mismeasured lepton, the difference in ϕ between the highest-energy lepton and the $\not\!\!E_T$ is required to be greater than 0.5 radians. The predicted and observed numbers of W boson events selected from each trigger are summarized in Table 4.4.

4.2 Z Selection

The Z boson selection requires two opposite-sign leptons: either two electrons or two muons. In addition, the invariant mass of the two leptons must be between 76 GeV and 106 GeV.

Muon Selections
Track must be fiducial to CMU, CMP, or CMX
$E_{EM} \leq 2$ GeV
$E_{HAD} \leq 6$ GeV
$E_{EM} + E_{HAD} \geq 0.1$ GeV for CMIO muons
$\mathrm{Iso}(R = 0.4)/E_T \leq 0.1$
≥ 3 COT axial segments with ≥ 5 hits
≥ 2 COT stereo segments with ≥ 5 hits
≥ 1 Si hit
$z_0^{\text{track}} \leq 60$ cm
if $p_T > 300$ GeV, $n_{transitions} \geq 30$
$d_0 < 0.02$ cm
$\Delta X_{CMU} \leq 3$ cm
$\Delta X_{CMP} \leq 5$ cm
$\Delta X_{CMX} \leq 6$ cm

Loose Muon Selections
$p_T \geq 10$ GeV

Tight Muon Selections
Muon must be of type CMUP or CMX
$p_T \geq 20$ GeV

Table 4.2: Selections to identify muons.

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

Table 4.3: Selections to identify loose central electrons

Predictions are made in two dielectron categories: One electron from the Z must be tight (Table 4.1), and the other may be tight or loose (Table 4.3). The predicted and observed numbers of events in both of these categories are summarized separately in Table 4.4.

Predictions are also made in ten dimuon categories: One muon must be tight, and therefore either a CMUP or a CMX muon. The other muon may be loose, and therefore may have any of the CMU, CMP, CMUP, CMX, or CMIO types. The predicted and observed numbers of events in all of these categories are summarized in Table 4.4.

Selection	Expected	Observed
W(e, $\not\!\!E_T$)	2571230	2548108
W(μ_{CMUP}, $\not\!\!E_T$)	1289610	1279001
W(μ_{CMX}, $\not\!\!E_T$)	904569	895257
Z(e, e)	156894	160251
Z(e, e_{loose})	25506	28896
Z(μ_{CMUP}, μ_{CMU})	8008	8391
Z(μ_{CMUP}, μ_{CMP})	9736	10433
Z(μ_{CMUP}, μ_{CMUP})	39620	36632
Z(μ_{CMUP}, μ_{CMX})	12893	13547
Z(μ_{CMUP}, μ_{CMIO})	9303	8489
Z(μ_{CMX}, μ_{CMU})	5860	6024
Z(μ_{CMX}, μ_{CMP})	6762	6863
Z(μ_{CMX}, μ_{CMUP})	14162	14467
Z(μ_{CMX}, μ_{CMX})	17245	17906
Z(μ_{CMX}, μ_{CMIO})	5852	5967

Table 4.4: Event counts in W and Z boson samples, split up by categories of the leading and subleading leptons.

CHAPTER 5

QCD 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 $\not\!\!E_T$ 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. The contribution from these QCD processes is estimated by using a data-derived model for these kinds of events [30]. This is accomplished by defining an object that is similar to an electrons, but has a much larger rate of contamination from jets, labeled an "anti-selected electron".

5.1 Anti-Selected Electron Definition

Following the technique described in Ref.[30], the standard CDF high-p_T electron identification selections are modified to select objects that are mostly fake electrons. The selections are divided into two categories, as shown in Table 5.1. The selections that affect the kinematics of the event are labeled "Kinematic Selections," while those designed to discriminate electrons from misidentified jets are labeled "Identification Selections." The anti-selected electron sample is defined by requiring that prospective objects pass all kinematic selections while simultaneously failing at least two of the identification selections. 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

The number of events that arise from QCD is obtained by fitting the $\not\!\!E_T$ 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-selected electron sample.

Kinematic Selections
$\eta_{\text{Detector}} \leq 1.1$ (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
Identification Selections
$E_{HAD}/E_{EM} \leq 0.055 + 0.00045 \times E$
$\chi^2_{CES} \leq 10$
$L_{shr} \leq 0.2$
$-3.0 \leq Q \times \Delta X_{CES} \leq 1.5$ cm
$\lvert\Delta Z_{CES}\rvert < 3$ cm

Table 5.1: Central tight electron identification selections [26] divided into two categories: those that shape the kinematics of the event, referred to as “kinematic”, and those that discriminate between electrons and misidentified jets, referred to as “identification”.

The QCD template is obtained from the anti-selected electron sample after subtracting the expected W boson contamination using the Monte Carlo. This fit is performed using the sample without the $\not\!\!E_T$ selection, in order to get an accurate estimate of the number of QCD events, most of which fail this selection. After the fit is performed across the $\not\!\!E_T$ distribution, the number of QCD events in the W boson signal region is calculated by applying the selection of $\not\!\!E_T > 25$ GeV. The Monte Carlo electroweak contribution and the data-derived QCD template are scaled to the result obtained from this $\not\!\!E_T$ fit. Figure 5.1 shows the fits to the electron dataset and to both muon-triggered datasets. A systematic uncertainty of 26% is applied to the QCD normalization, as found in [30].

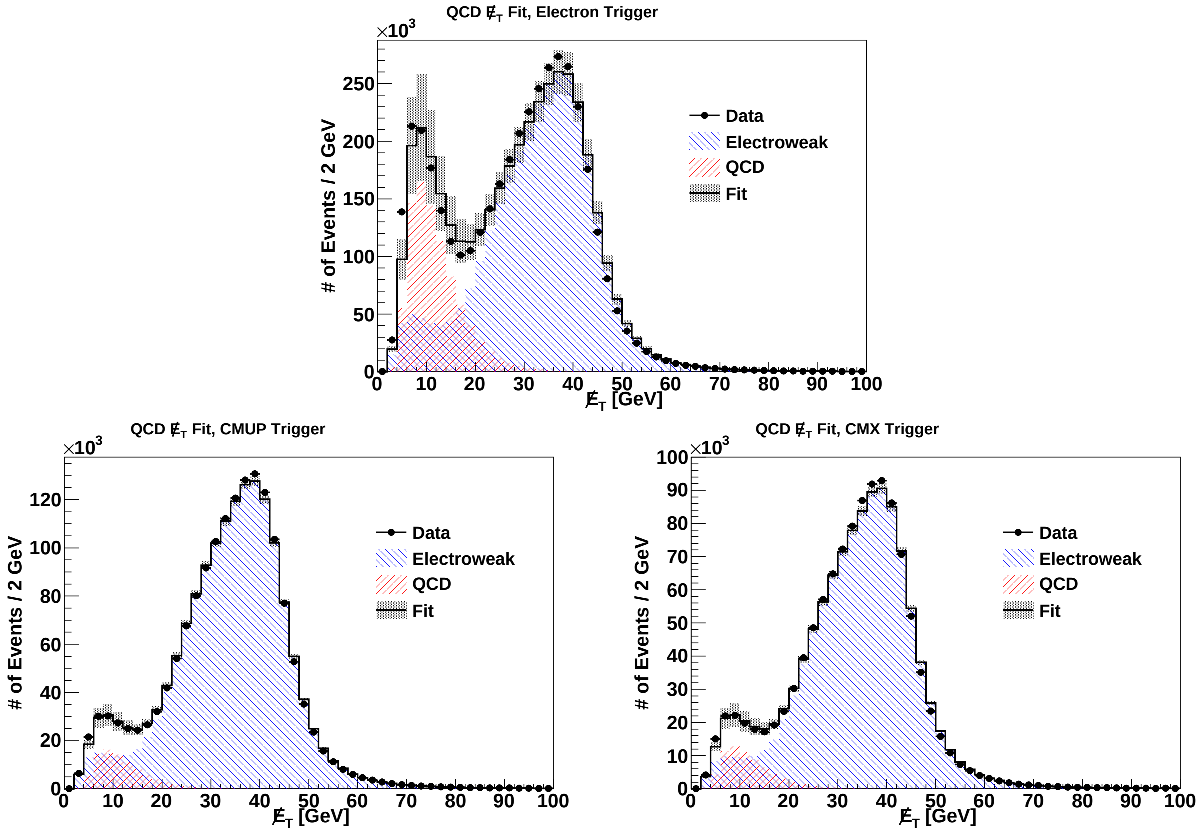

Figure 5.1: The fits to the $\not\!\!E_T$ distribution of events with $m_T > 20$ GeV and $\Delta\phi(\not\!\!E_T, l) > 0.5$, in each of the three trigger lepton categories. The "electroweak" template is obtained from Monte Carlo and the "QCD" template is obtained from the anti-selected electron data sample. The systematic uncertainty of 26% found in [30] is shown.

CHAPTER 6

W AND Z SAMPLE VALIDATION

The W and Z boson samples are validated against a large sample of SM Monte Carlo events. The main SM backgrounds in this analysis are from W + jets, Drell-Yan, QCD multijet, top quark, and diboson production processes. The matrix element portion of the electroweak background events has been modeled using the ALPGEN [31] Monte Carlo (MC) program, except for the top production and diboson production backgrounds, which were modeled by PYTHIA [32]. PYTHIA was used to model the parton showering in all samples. These MC events are analyzed using a GEANT based detector simulation, CDFSim [33]. All of the MC datasets include the relevant K-factors in their normalization. In order to improve the statistics for the heavy flavor background, specialized MC datasets are used to model the W and Z boson plus heavy-flavor jet processes. Heavy flavor events are removed from the more general samples in order to avoid double counting these heavy flavor events. The MC datasets are described in Appendix B.

6.1 W Boson Sample Validation

6.1.1 $W \rightarrow e\nu$ events

In the electron-triggered W boson sample, 2,571,230 events are predicted versus the observed 2,548,108 events. Table 6.1 summarizes the predicted numbers of events, and Figure 6.1 shows the validation distributions of kinematic variables in the $W \rightarrow e\nu$ sample.

6.1.2 $W \rightarrow \mu\nu$ events

In CMUP-triggered events, 1,289,610 $W \rightarrow \mu\nu$ events are expected versus the observed 1,279,011 events, while in CMX-triggered events, 904,569 are expected and 895,257 are observed. Tables 6.2 and 6.3 summarize the predicted numbers of events, and Figures 6.2 and 6.3 show the validation distributions of kinematic variables in the $W \rightarrow \mu\nu$ sample.

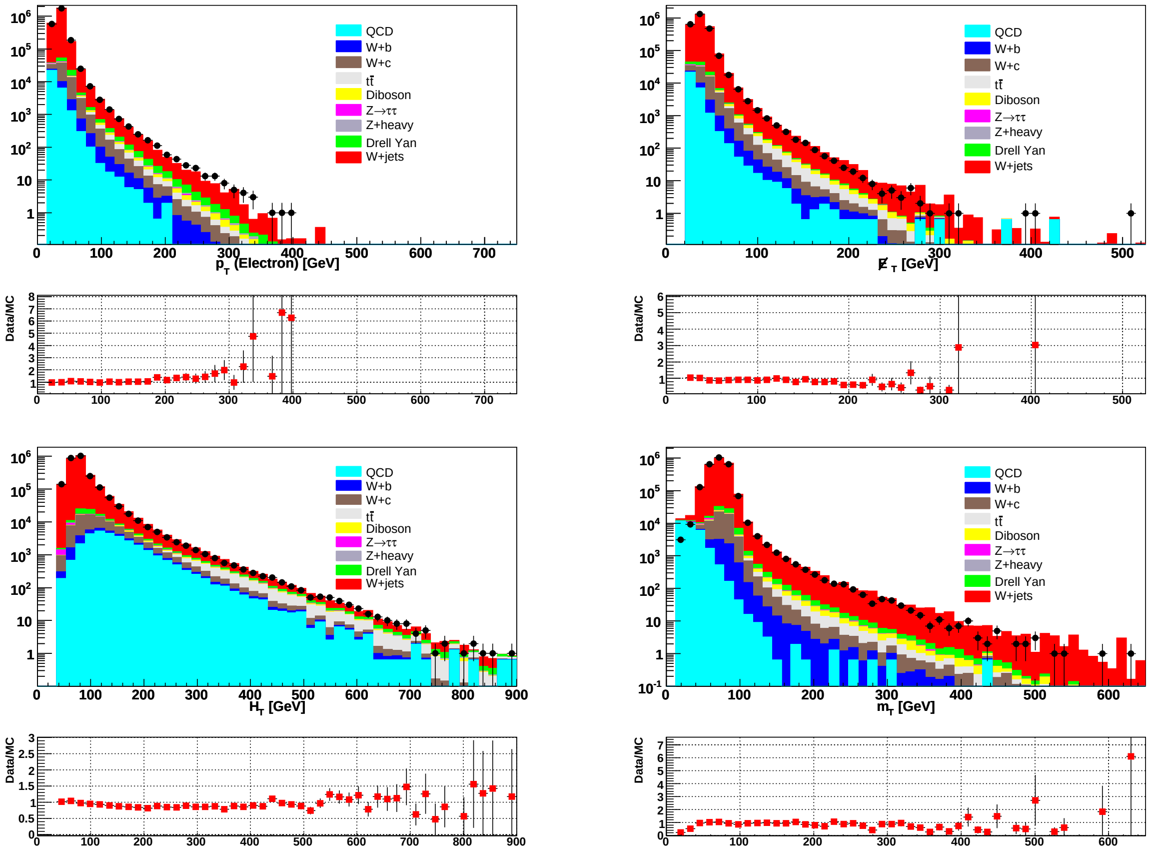

Figure 6.1: W validation distributions, electron trigger: p_T of the highest-p_T good electron, $\not{E}_T$ and H_T in the event, M_T of electron and $\not{E}_T$. The W selections of $\not{E}_T > 25$ GeV and $m_T > 20$ GeV are included.

Source	Number of Events
W+light jets	2450327
W+b	7573
W+c	50493
Drell-Yan	23095
$Z \to \tau\tau$	3400
Z+heavy	499
$t\bar{t}$	2113
Diboson	3456
QCD	30277
Expected total	2571230
Observed	2548108

Table 6.1: Summary of the predictions for $W^{\pm} \to e^{\pm}\nu$. The QCD contribution is estimated as in Sec. 5. The other contributions are calculated from the MC, scaled according to the fit described in Section 5.2.

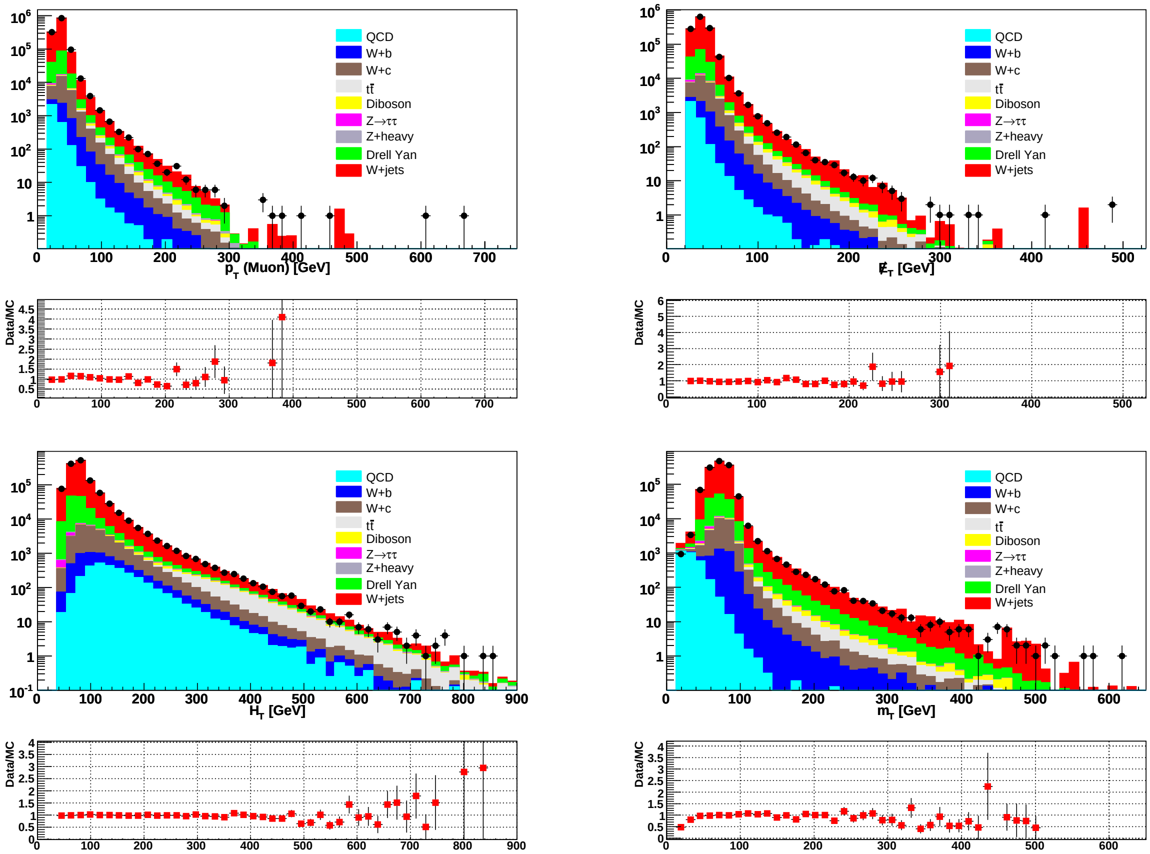

Figure 6.2: W validation distributions, CMUP trigger: p_T of the highest-p_T good muon, $\not{E}_T$ and H_T in the event, M_T of the good highest-p_T muon and $\not{E}_T$. The W selections of $\not{E}_T > 25$ GeV and $m_T > 20$ GeV are included.

Source	Number of Events
W+light jets	1139312
W+b	3589
W+c	23691
Drell-Yan	114194
$Z \rightarrow \tau\tau$	1708
Z+heavy	1337
$t\bar{t}$	1072
Diboson	1723
QCD	2986
Expected total	1289610
Observed	1279011

Table 6.2: Summary of the predictions for $W^{\pm} \rightarrow \mu^{\pm}\nu$ with a CMUP trigger. The QCD contribution is estimated as in Sec. 5. The other contributions are calculated from the MC, scaled according to the fit described in Section 5.2.

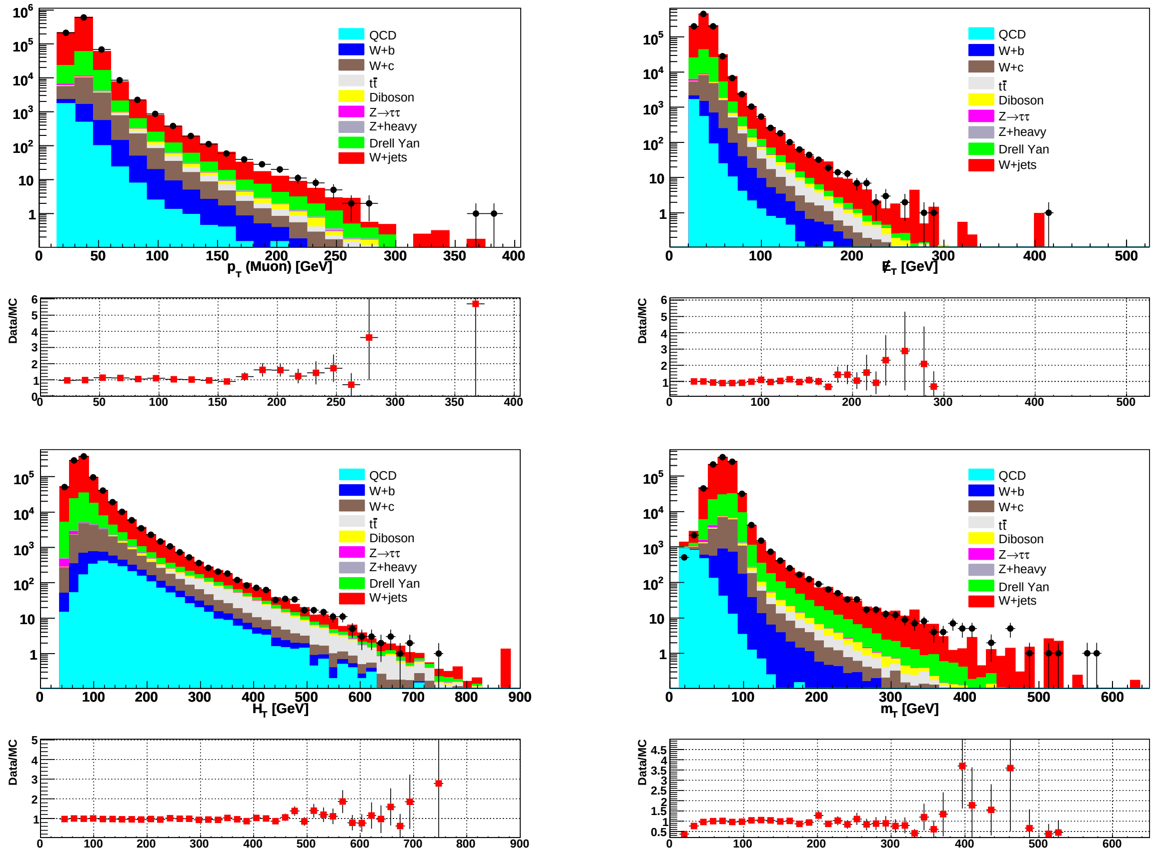

Figure 6.3: W validation distributions, CMX trigger: p_T of the highest-p_T good muon, $\not\!\!E_T$ and H_T in the event, M_T of the good highest-p_T muon and $\not\!\!E_T$. The W selections of $\not\!\!E_T > 25$ GeV and $m_T > 20$ GeV are included.

Source	Number of Events
W+light jets	802619
W+b	2337
W+c	15550
Drell-Yan	78109
$Z \to \tau\tau$	1204
Z+heavy	845
$t\bar{t}$	535
Diboson	1023
QCD	2348
Expected total	904569
Observed	895257

Table 6.3: Summary of the predictions for $W^{\pm} \to \mu^{\pm}\nu$ with a CMX trigger. The QCD contribution is estimated as in Sec. 5. The other contributions are calculated from the MC, scaled according to the fit described in Section 5.2.

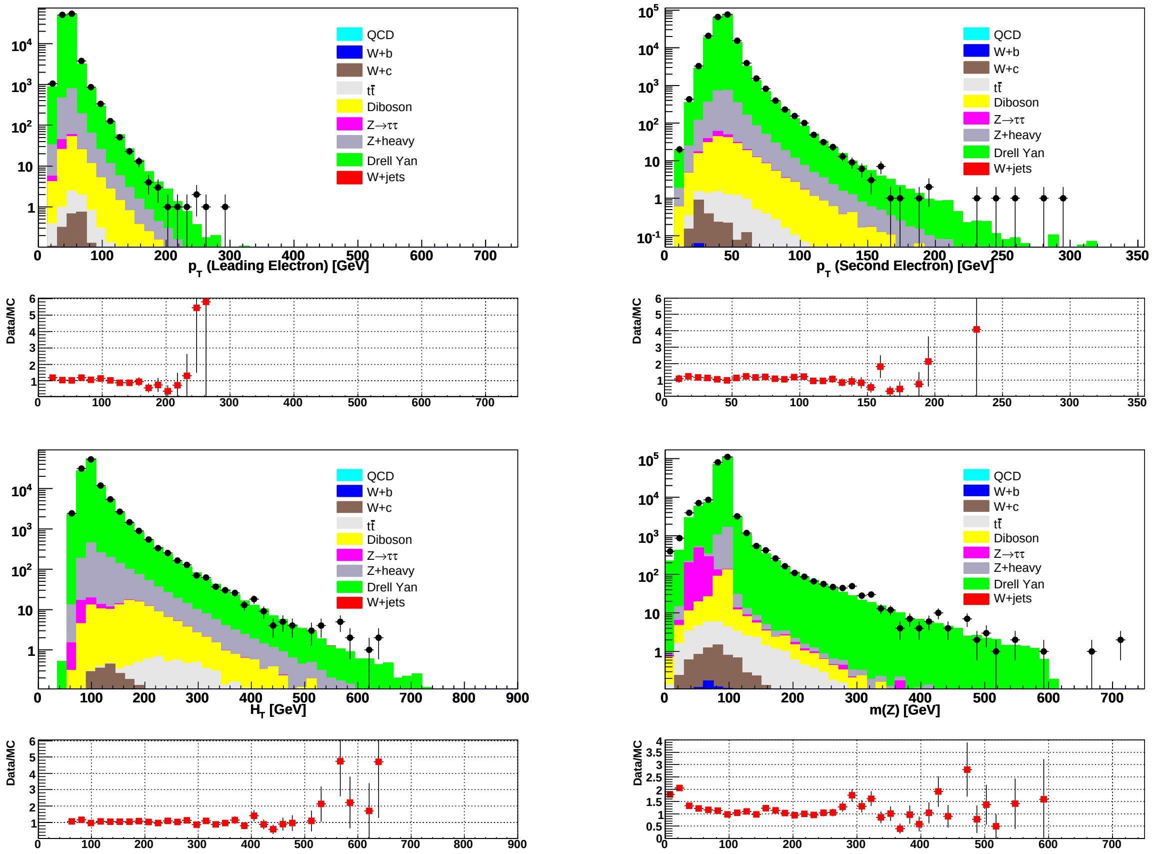

Figure 6.4: Distributions of e-triggered Z events: the p_T of the two leading electrons, the H_T in the event, and the dilepton mass. In all the distributions except the mass distribution, there is a requirement that the dilepton mass is between 76 and 106 GeV.

6.2 Z Sample Validation

6.2.1 $Z \to e^+e^-$

Predictions are made in two dielectron categories: One electron must be tight, and the second may be either tight or loose. The predicted and observed numbers of events in both of these categories is summarized separately in Table 4.4. Table 6.4 shows the predicted and observed event counts for electron-triggered Z events, with both categories of the second electron combined. Figure 6.4 shows validation distributions of the same sample.

Source	$N(Events)$ in full range	$N(Events)$ in Z window
W+light jets	63	16
W+b jets	0.9	0.2
W+c jets	5.7	1.8
Drell-Yan	120115	105801
$Z \to \tau\tau$	752	26
Z+heavy jets	1535	1438
$t\bar{t}$	22	5
Diboson	169	123
QCD	224	154
Expected total	122887	107567
Observed	129462	111063

Table 6.4: Summary of the predictions and observations for $Z/\gamma^* \to e^+e^-$ where the trigger is TCE. The prediction is shown over the entire range of $m(e^+e^-)$ as well as in the Z mass window, $76 < m(e^+e^-) < 106$ GeV.

6.2.2 $Z \to \mu^+\mu^-$

Predictions are made in ten dimuon categories: One muon must be selected from either a CMUP or a CMX trigger, and the second muon may come from any of the CMU, CMP, CMUP, CMX, or CMIO categories. The predicted and observed numbers of events in each of these categories is summarized separately in Table 4.4. Tables 6.5 and 6.6 show predicted and observed event counts for CMUP-triggered and CMX-triggered $Z \to \mu^+\mu^-$ events, with all five categories of the second muon combined. Figures 6.5 and 6.6 show validation distributions of the same sample.

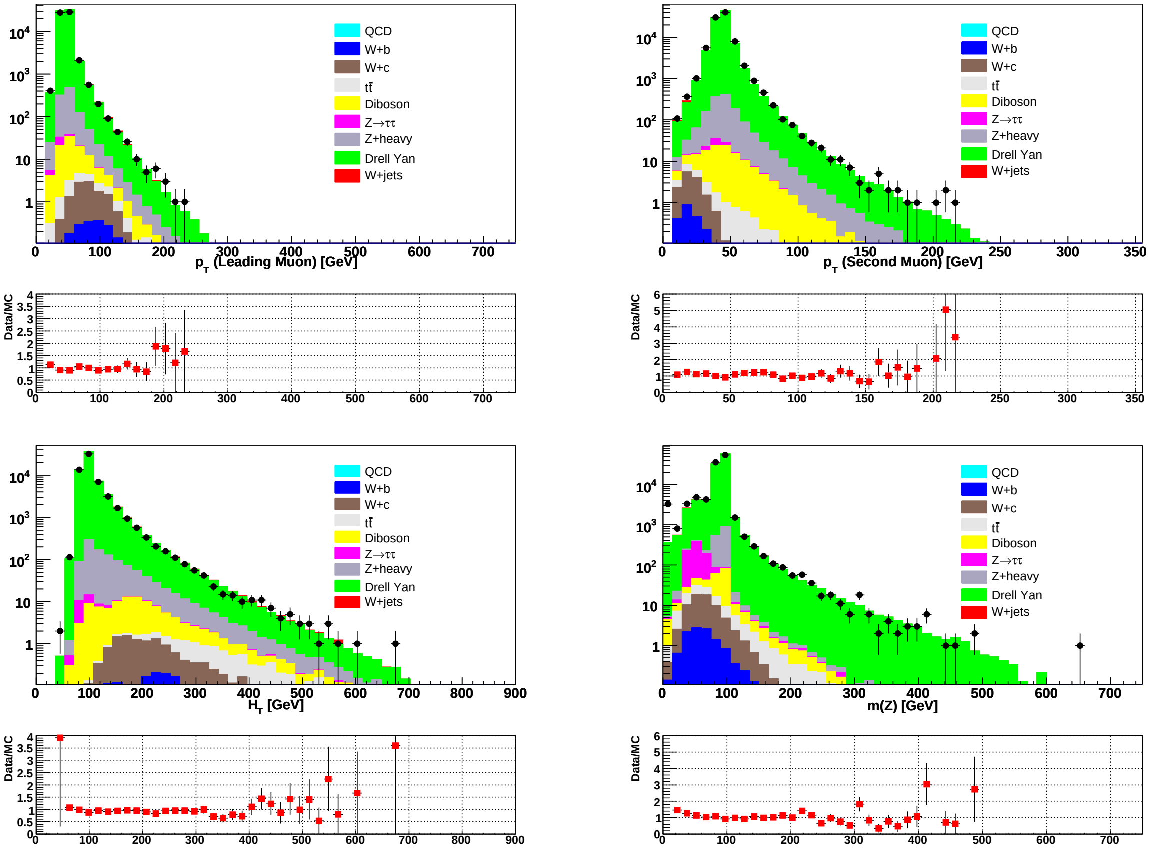

Figure 6.5: Distributions of CMUP μ-triggered Z events: the p_T of the two leading electrons, the H_T in the event, and the dilepton mass. In all the distributions except the mass distribution, there is a requirement that the dilepton mass is between 76 and 106 GeV.

Source	$N(Events)$ in full range	$N(Events)$ in Z window
W+light jets	366	73
W+b jets	11	1.8
W+c jets	53	10
Drell-Yan	75970	64783
$Z \to \tau\tau$	653	19
Z+heavy jets	1012	936
$t\bar{t}$	46	8.9
Diboson	131	86
QCD	22	15
Expected total	78263	65933
Observed	75674	60163

Table 6.5: Summary of the predictions and observations for $Z/\gamma^* \to \mu^+\mu^-$ where the trigger is CMUP muon. The prediction is shown over the entire range of $m(\mu^+\mu^-)$ as well as in the Z mass window, $76 < m(\mu^+\mu^-) < 106$ GeV.

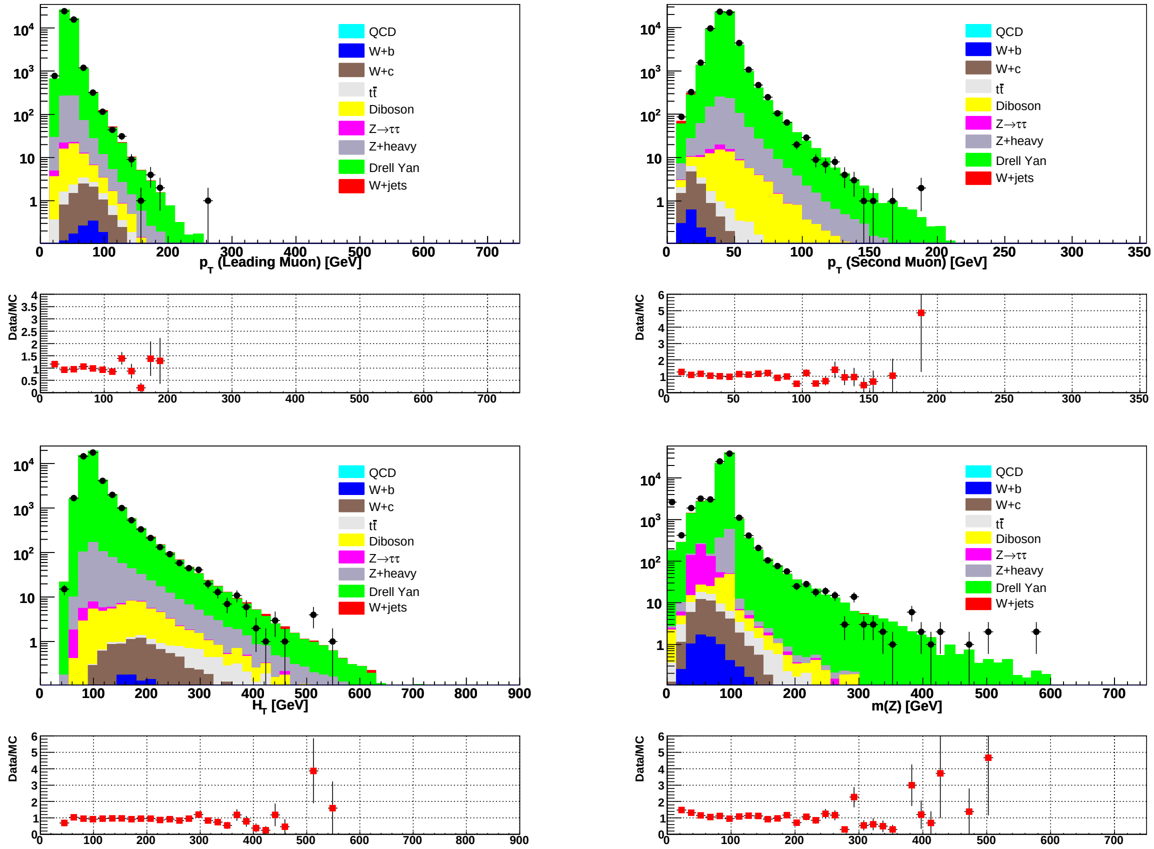

Figure 6.6: Distributions of CMX μ-triggered Z events: the p_T of the two leading electrons, the H_T in the event, and the dilepton mass. In all the distributions except the mass distribution, there is a requirement that the dilepton mass is between 76 and 106 GeV.

Source	$N(Events)$ in full range	$N(Events)$ in Z window
W+light jets	244	54
W+b jets	6	1.3
W+c jets	34	7.5
Drell-Yan	51634	44448
$Z\to\tau\tau$	417	9.6
Z+heavy jets	658	611
$t\bar{t}$	22	4.4
Diboson	78	52
QCD	17	12
Expected total	53112	45201
Observed	53254	42693

Table 6.6: Summary of the predictions and observations for $Z/\gamma^* \to \mu^+\mu^-$ where the trigger is CMX muon. The prediction is shown over the entire range of $m(\mu^+\mu^-)$ as well as in the Z mass window, $76 < m(\mu^+\mu^-) < 106$ GeV.

Trigger	Measured R	Systematic Uncertainty
TCE	10.84	1.6%
CMUP	11.30	5.9%
CMX	10.88	2.0%

Table 6.7: The measured ratio R of W boson to Z boson production, and the trigger rate uncertainty calculated by comparing it to the NNLO calculated value of 10.67.

6.3 The Ratio of W to Z Production 'R' as a Precision Check

The ratio of W to Z production is a very precise test of problems in lepton trigger efficiencies, lepton identification efficiencies, or problems with $\not\!\!E_T$ [28]. This ratio is calculated in each differently-triggered dataset, and the deviation from the theoretical value of $R = 10.67$ [27] is used as a systematic uncertainty on the trigger efficiency. Figures 6.7, 6.8, and 6.9 show R versus run number for electron-, CMUP muon-, and CMX muon-triggered events. The fluctuation of R with time is at least partially due to luminosity effects. At higher luminosity, leptons are less likely to pass isolation selections, decreasing the measured Z boson cross section more than that of the W boson. Table 6.7 shows the measured values of R and the corresponding systematic uncertainties.

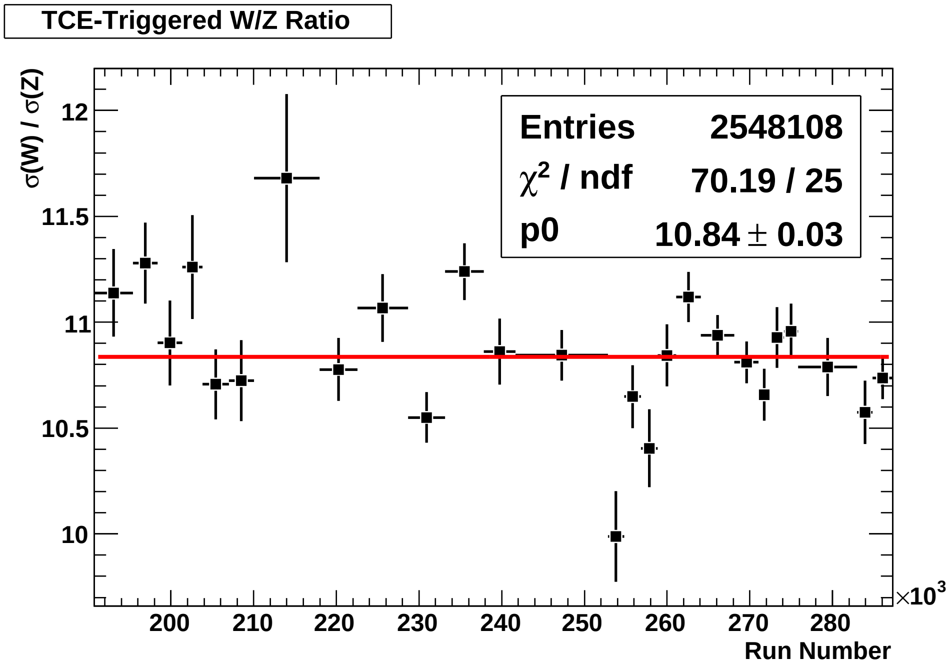

Figure 6.7: The ratio, R, of W and Z cross sections for each run period, using the electron trigger.

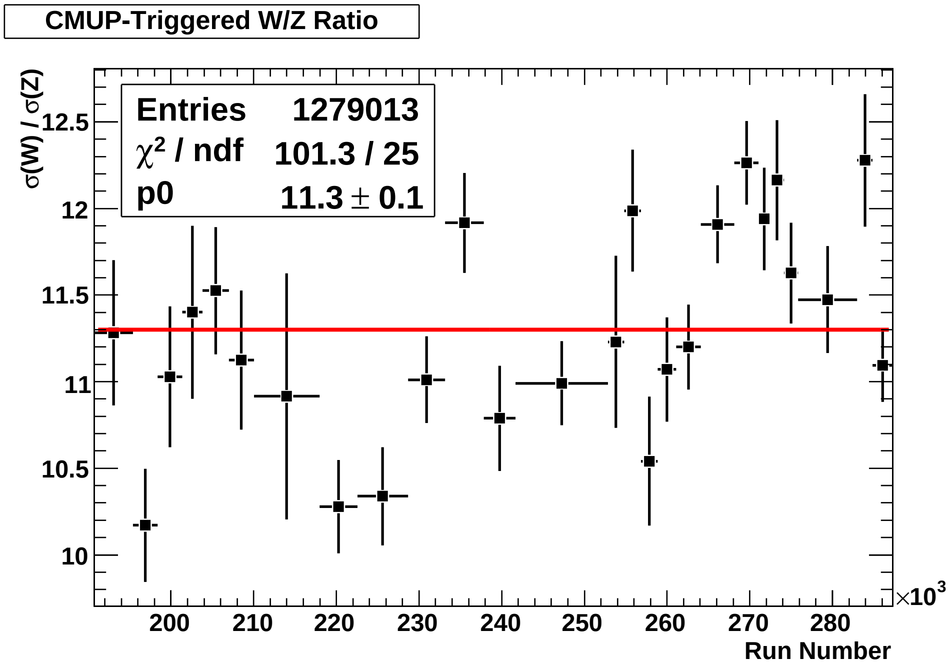

Figure 6.8: The ratio, R, of W and Z cross sections for each run period, using the CMUP muon trigger.

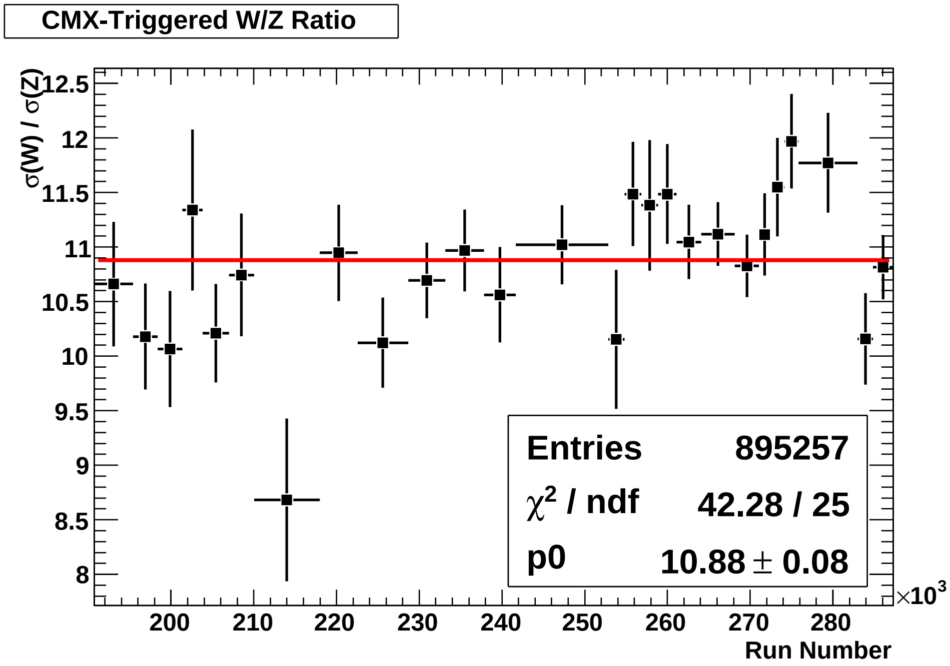

Figure 6.9: The ratio, R, of W and Z cross sections for each run period, using the CMX muon trigger.

CHAPTER 7

SOFT LEPTON IDENTIFICATION

The identification of low-p_T, or “soft”, leptons is a main focus of this analysis. Likelihood-based methods are used to identify soft electrons and muons. The identification algorithms are described here, along with the methods used to validate them and evaluate their systematic uncertainties.

7.1 Soft Electrons

Soft electrons are identified using a likelihood method trained on a signal sample from photon conversions and a background sample from other tracks with electron sources removed.

7.2 Identification Algorithm and Candidate Selections

Every track in the event is a soft electron candidate, provided that it passes track quality selections and fiduciality selections:

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

After this preselection, a likelihood-based calculator is used to identify electrons. The likelihood calculator uses seven discriminating variables: the energy loss as the track traverses the tracking chamber, the electromagnetic and hadronic calorimeter energies, the energies deposited in the preradiator and the showermax detector, and the two-dimensional distance $(\Delta x, \Delta z)$ between the extrapolated position of the track and the shower in the CES. The calorimeter variables are calculated using a narrow, two-tower-wide section of the calorimeter, as opposed to the larger integrated area used for the high-p_T electron selection [34]. Appendix C contains more information on these discriminating variables.

The likelihood is trained on data without resorting to the simulation. For each variable, a fit is performed to the ratio of the distribution in the real sample and the distribution in the fake sample. For each candidate, the value of each of these fit functions is multiplied together to get the final likelihood ($\mathcal{L}$):

$$\mathcal{L}_i = \frac{P(x_i|real)}{P(x_i|fake)}, \qquad Q = \prod_i \mathcal{L}_i, \qquad \mathcal{L} = \frac{Q}{1+Q}.$$

A candidate is identified as an electron if it passes the requirement $\mathcal{L} > 0.99$.

7.2.1 Training Samples and Efficiency and Misidentification Rate Measurements

Photon conversions are used as a pure sample of electrons to train the likelihood function. The 8 GeV electron trigger (see Appendix A.2) is used to obtain a pure sample of conversions by requiring that there be two tracks with opposite sign having $|\delta| < 0.2$ cm, $\Delta\cot(\theta) < 0.1$, and $R_{Conv} > 8$ cm. Figure 7.1a illustrates these variables.

After these selections, a fit is performed to the $\Delta\cot(\theta)$ distribution, shown in Figure 7.1b, to determine the non-conversion background under the peak. The sideband of the distribution ($0.06 < |\Delta\cot(\theta)| < 0.1$) is used to subtract out this background.

The likelihood distributions for the electron sample and for the non-electron background sample are shown in Figure 7.2 (left).

Since the higher-p_T leg of the conversion will be trigger-biased, the likelihood is trained using the softer leg. A conversion pair is not used if 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.

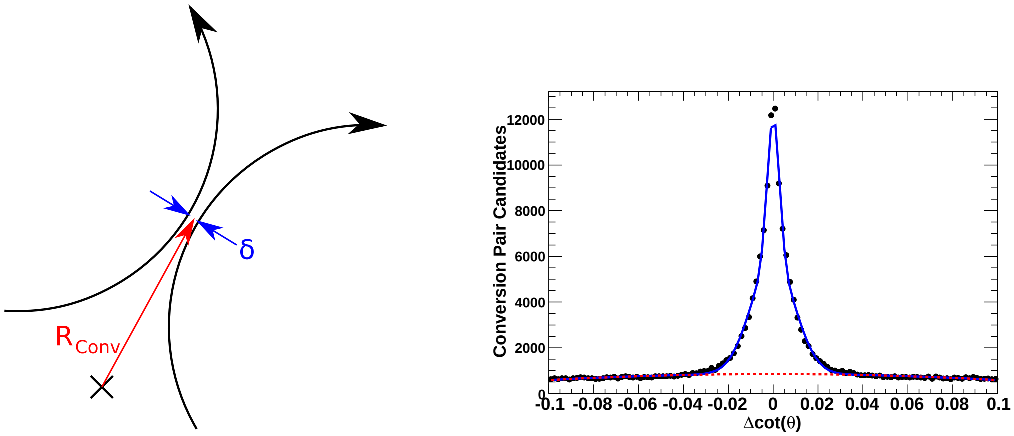

Figure 7.1: Variables used to identify photon conversions to electron-positron pairs. On the left are the variables defined in the plane transverse to the beam. The beam position is denoted by an "x". R is the distance between the beam position and the point at which the two tracks are tangential or parallel to each other and δ is the distance between the two tracks at that point. On the right is the distribution of $\Delta\cot(\theta)$, where θ is the polar angle of the track in the $r-z$ plane. A fit is performed to find the signal (solid) and background (dashed) to estimate the sample composition under the peak.

A sample of non-electron tracks with which to train the likelihood function is found using events from the 18 GeV muon trigger (See Appendix A.2). All tracks in the events that, along with another track, form a possible photon conversion are removed from the training sample. In addition, to reduce the contamination from real electrons, any event that contains an identified heavy quark decay or an identified electron is ignored.

The efficiency and fake rate are calculated in these training samples as functions of p_T, η, and track isolation. The separation in identification rate between electrons and non-electrons after the likelihood selection is shown in Figure 7.2 (right). This function is calculated by dividing the data into bins in p_T and η, and in each of these bins fitting the efficiency and fake rate to a linear function of the isolation.

This identification rate is applied to each candidate track in the MC to find the predicted number of identified electrons.

Some representative fits are shown in Figure 7.3. To show that this scheme takes into account any correlations between these kinematic variables, the results of applying these parameterizations to the training samples are shown in Figure 7.4.

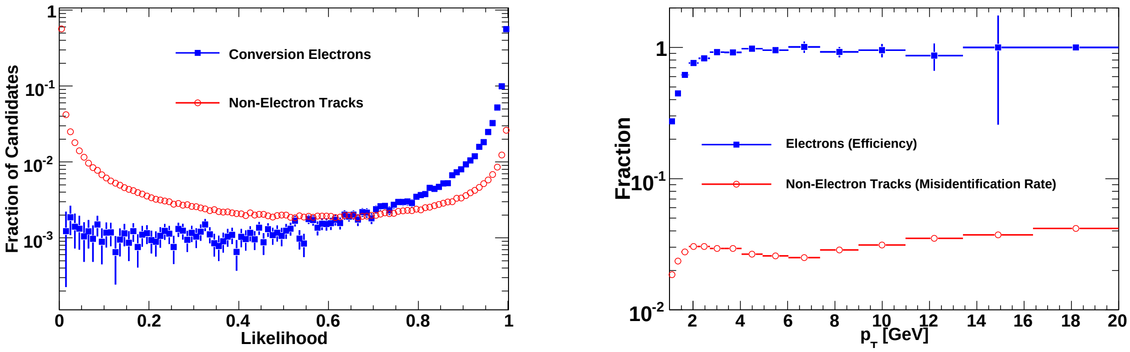

Figure 7.2: Left: The likelihood distributions for electrons (closed squares) and non-electrons (open circles) after all preselection criteria. Only those candidates with a likelihood > 0.99 are identified as electrons. Right: The efficiency as a function of p_T for the identification of electrons and tracks misidentified as electrons after the likelihood selection.

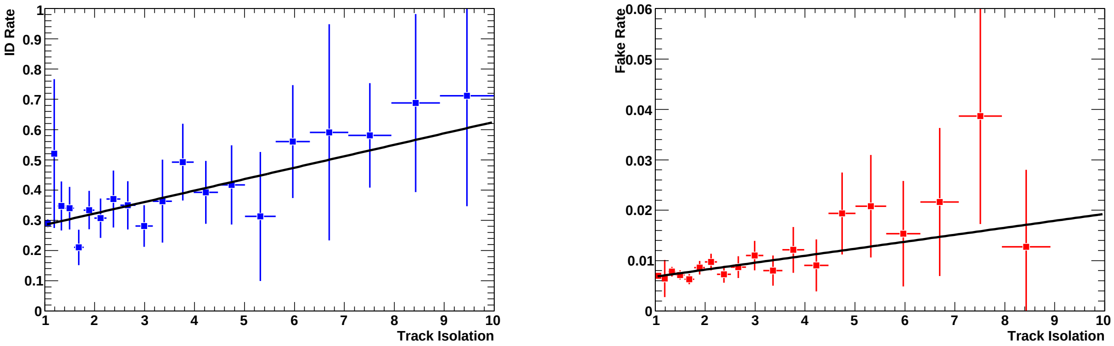

Figure 7.3: Examples of fits to the soft electron efficiency (left) and fake rate (right) as a function of track isolation in a particular p_T and η bin.

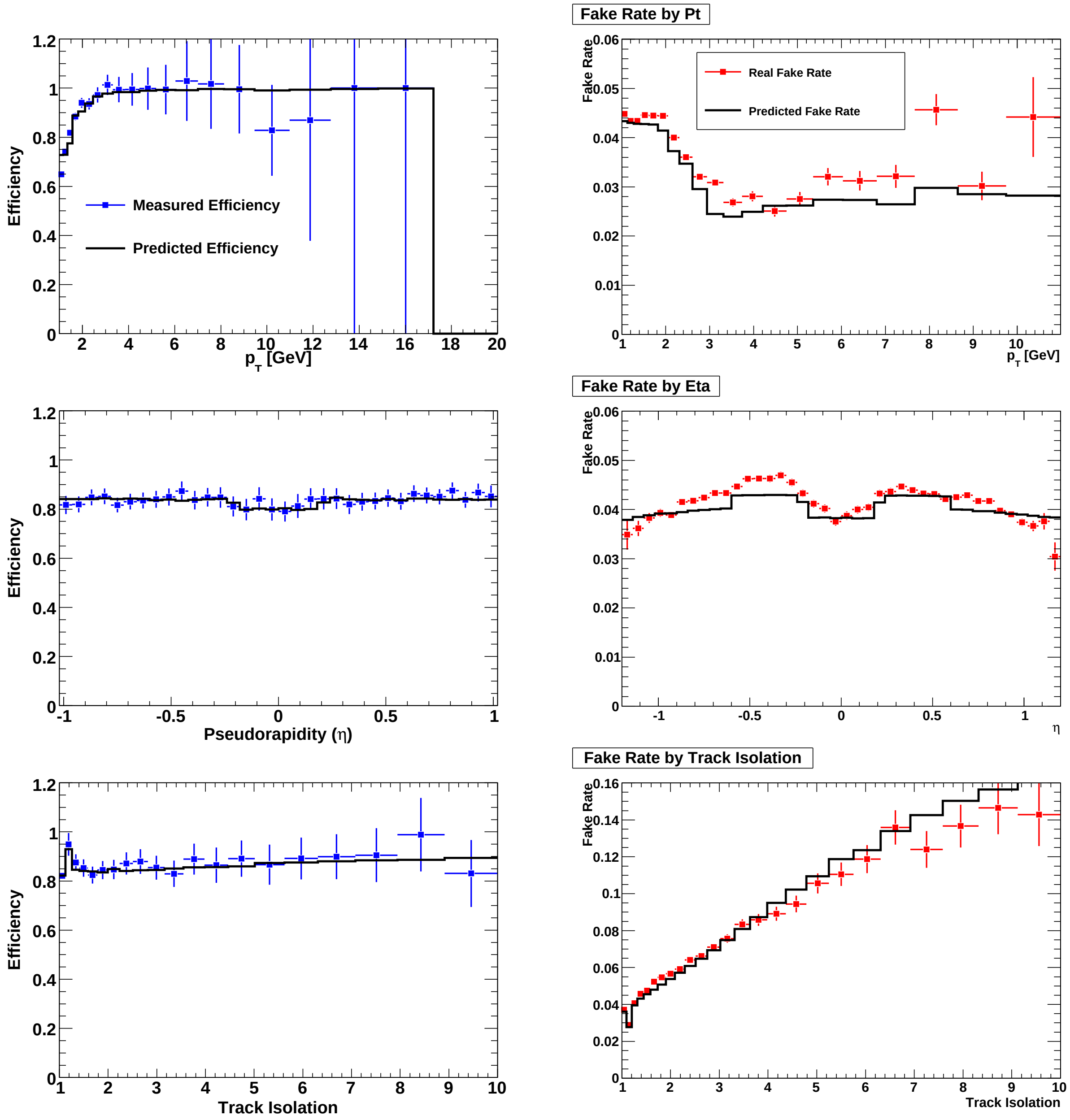

Figure 7.4: A comparison of the measured and predicted efficiencies (left) and fake rates (right). The kinematic variables shown, from top to bottom, transverse momentum p_T, pseudo-rapidity η, and track isolation are those used to characterize the response of the soft electron algorithm.

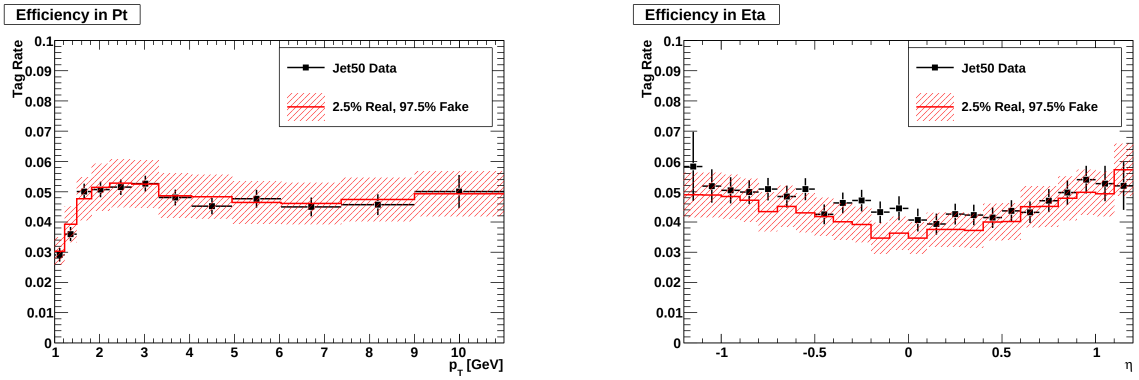

Figure 7.5: Predicted and observed soft electron misidentification rates obtained from a QCD (jet) sample. On the left are the identification rates as function of p_T, and on the right the the identification rates as functions of η.

7.2.2 Validation and Systematic Uncertainty Determination

The efficiency and fake rate parameterizations are checked on a data set triggered on jets having $E_T > 50$ GeV (See Appendix A.2). The parameterizations take into account the p_T, η and isolation of candidates in order to account for any kinematic differences between the training sample and the validation sample. First, the same electron removal that was used for the fake training sample (Section 7.2.1) is applied to the tracks in the jet sample. The likelihood distribution of all candidate tracks in the jet sample is then fit to templates from the real and fake likelihood training samples to obtain the fraction of real and fake electrons in the jet sample. The jet sample is found to consist of 2.5% real electrons, mostly coming from photon conversions from which only one electron was reconstructed. The predicted identification rate is then checked for agreement with the measured identification rate.

The disagreement between the calculated and observed identification rates is measured to be 1.6%. However, we observe larger disagreement in the shapes of the calculated and observed distributions in p_T and η, as seen in Figure 7.5. We assign a systematic uncertainty of 15%, which is sufficient to cover the observed disagreement. This systematic uncertainty is applied separately to the electron identification and misidentification rates.

7.3 Soft Muons

The soft muon identification algorithm described in Ref. [35] has been ported from the TopNtuple framework to the Stntuple framework with minimal modifications. The efficiency of the ported identification software is measured using reconstructed $J/\psi \rightarrow \mu^+\mu^-$ decays to obtain pure muon samples. The misidentification rates of pions and kaons are measured in $D^{*+} \rightarrow D^0\pi^+$ decays where D^0 decays as $D^0 \rightarrow K^-\pi^+$. Similarly, the proton misidentification rate is measured in $\Lambda^0 \rightarrow p\pi$. First, the algorithm is summarized.

7.3.1 Identification Algorithm and Candidate Selections

The soft muon identification algorithm relies on matching tracks identified in the COT to the track segments reconstructed in the muon chambers (muon stubs). Matching is done in the extrapolated position along the muon chamber drift direction (x), the longitudinal coordinate along the chamber wires (z) when available, and the difference in slope between the extrapolated COT track and the reconstructed muon chamber track segment (ϕ_L). Tracks are paired with muon chamber track segments based on the best match in x for those track segments within 50 cm of an extrapolated COT track.

The list of all such track-stub matching variables is:

- CMU dx,
- CMU dz,
- CMU $d\phi$,
- CMP dx,
- CMP $d\phi$,
- CMX dx,
- CMX dz,
- CMX $d\phi$.

Soft Muon Candidate Selections
$N(\text{COT}) \geq 48$
$N(\text{COT Axial}) \geq 24$
$N(\text{COT Stereo}) \geq 24$
$\|d_0\| < 0.3$ cm, where d_0 is the impact parameter with respect to the beamline
$\|z_0\| < 60$ cm
The track must extrapolate to within the physical boundary of a muon chamber.

Table 7.1: Soft muon candidate selection criteria.

Each of these variables can only be used if a stub exists in the corresponding system, i.e. CMP dx and $d\phi$ only have values if there is a CMP stub.

Table 7.1 lists the requirements placed on soft muon candidate tracks. The selections on the number of COT hits reduce the background from poorly measured tracks, while the selection on impact parameter removes some of the pion and kaon decay-in-flight background. Note that the fiducial definition described in the last bullet point differs from the one used in Ref. [35]. In that algorithm, candidates were declared fiducial if they extrapolated to within $3\sigma_{MS}$ outside of the physical chamber boundary, where σ_{MS} is the width of the multiple scattering distribution for a given p_T. This change was made due to the unavailability of the extrapolated track-to-chamber boundary distance in the Stntuple format.

The final likelihood, $\mathcal{L}$, is simply a sum Q of χ^2 terms built from each track-to-stub matching variable x_i described above:

$$Q = \sum_i^n \frac{(x_i - \mu_i)^2}{\sigma_i^2} = \sum_i^n y_i^2, \tag{7.1}$$

where μ_i and σ_i^2 are the expected mean and variance of the distribution of x_i. If the y_i are independent and normally distributed, the distribution of Q is χ^2 with a mean of n and $\sigma^2 = 2n$. The final likelihood is calculated by normalizing Q:

$$\mathcal{L} = \frac{Q - n}{\sigma(Q)}, \tag{7.2}$$

where $\sigma^2(Q)$ is the variance of Q.

The track-to-stub matching variance functions, σ_i in Eq. 7.1, are copied from the TopNtuple code directly into the Stntuple port with no modifications. The $\sigma(Q)$ term in the denominator of Eq. 7.2 is decomposed as in Ref. [35]

$$\sigma^2(Q) = 2n + \sum_{i,j} \rho(y_i^2, y_j^2), \tag{7.3}$$

and the ρ's are taken from the TopNtuple code.

In the final selection, we require that all identified soft muons must have a track segment in each muon chamber to which the track extrapolates and that $|\mathcal{L}| < 3.5$. For example, if a track should cross the physical volume of both the CMU and CMP detectors, there must be stubs in both detectors for it to be identified as a soft muon.

7.3.2 Efficiency and Misidentification Rate Measurements

The efficiency of the soft muon identification is measured using a pure sample of muons obtained from $J/\psi \to \mu\mu$ decays. These events are obtained using an online trigger requiring the presence of a muon with $p_T > 8$ GeV (See Appendix A.3), and the J/ψ is reconstructed by requiring that the trigger muon make a vertex with another track of opposite charge that has associated muon chamber hits. All track requirements listed in Sec. 7.3.1 are applied to both tracks. The J/Ψ candidate mass is required to be consistent with $2.9 < m(\mu\mu) < 3.3$ GeV, and signal and sideband regions are defined as follows:

- Left Sideband: $2.94 < m(\mu\mu) < 3.0$ GeV,
- Signal Region: $3.03 < m(\mu\mu) < 3.15$ GeV,
- Right Sideband: $3.18 < m(\mu\mu) < 3.24$ GeV,

The signal and sideband yields are fit in bins of p_T of the non-triggered leg and the fits are used to subtract out the background under the mass peak. These fits are shown in Fig. 7.6.

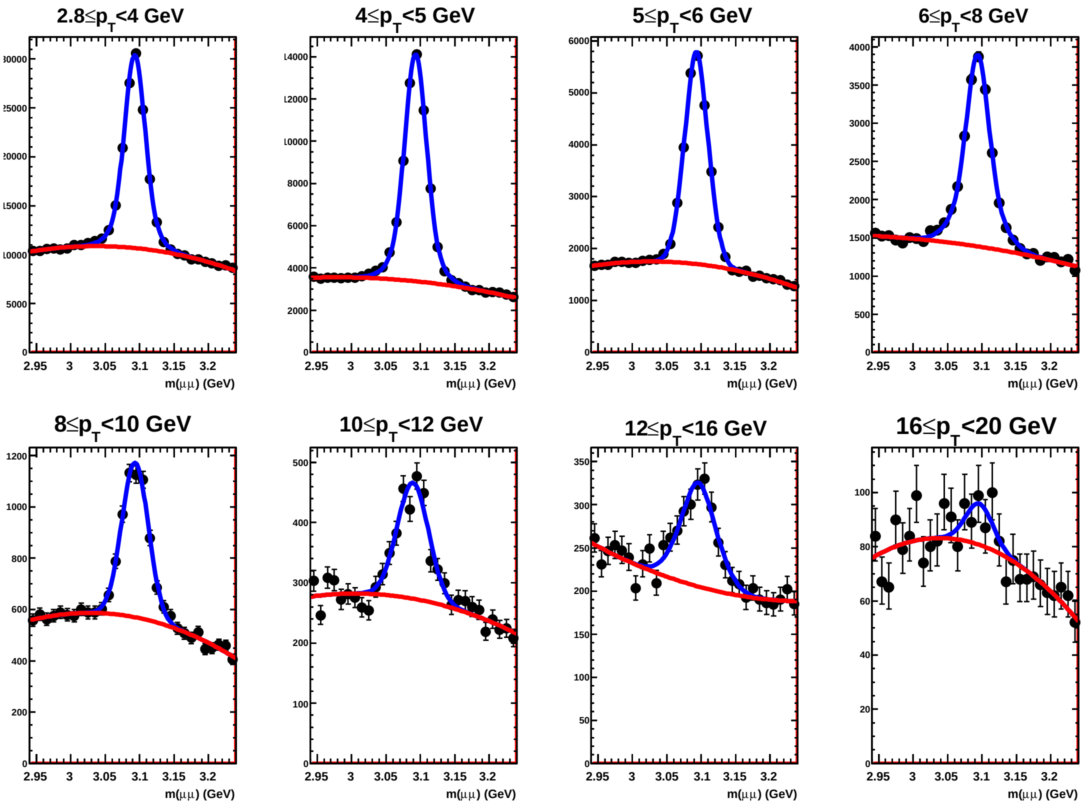
2.8≤p_T<4 GeV
4≤p_T<5 GeV
5≤p_T<6 GeV
6≤p_T<8 GeV
8≤p_T<10 GeV
10≤p_T<12 GeV
12≤p_T<16 GeV
16≤p_T<20 GeV

Figure 7.6: Results of the J/ψ mass fits in bins of the p_T of the softer, not-triggered, candidate leg of the J/ψ.

The misidentification rates of pions and kaons are measured in $D^{*+} \to D^0\pi^+$ decays where D^0 decays as $D^0 \to K^-\pi^+$. These events are obtained from the two track trigger (See Appendix A.3) and are reconstructed with the following selections:

- the K must have opposite charge to each of the two π's,
- $|z| \leq 5$ cm between any two tracks,
- the soft pion from the $D^* \to D^0$ decays must have $p_T \geq 500$ MeV,
- the kaon and pion from the D^0 decay must have $p_T \geq 2$ GeV,
- the kaon and pion from the D^0 decay must have $|d_0| \leq 0.2$ cm,
- $m(K\pi) - m(D^0) \leq 30$ MeV where $m(D^0)$ is the nominal D^0 mass,
- $p_T(D^0) \geq 5$ GeV,

- the impact parameter significance of the D^0 is required to be $d_0/\sigma(d_0) \geq 2$,
- $p_T(D^*) \geq 6$ GeV,
- $\Delta m = m(D^*) - m(D^0) \leq 170$ MeV.
- $\chi^2 \leq 100$ where χ^2 is from the vertex fit.

The π and K from the D^0 are required to form a vertex while the slow π from the D^* is attached to the primary vertex. The D^0 vertex is required to point back to the primary vertex. Signal and sideband regions are defined as follows:

- Left Sideband: $0.1396 < \Delta m < 0.141$ GeV,
- Signal Region: $0.14242 < \Delta m < 0.148421$ GeV,
- Right Sideband: $0.152 < \Delta m < 0.1625$ GeV,

The signal and sideband yields are fit in bins of p_T of the π or K from the D^0. These fits are shown in Fig. 7.7.

The misidentification rate of protons is measured using a sample of protons obtained from $\Lambda \rightarrow p\pi$ decays. These events are taken from the two track trigger (See Appendix A.3). The selections are as follows:

- the two tracks must pass the selections in Sec. 7.3.1,
- the two tracks are required to have opposite charge and fit to a vertex,
- $|\Delta z| \leq 2$ cm between the two tracks,
- the χ^2 of the vertex fit is required to be < 10,
- the decay length significance of the vertex is required to be $L_{xy}/\sigma(L_{xy}) \leq 20$,
- the Λ impact parameter is required to be $|d_0| < 0.02$ cm,
- $1.0 < m(\Lambda) < 1.16$ GeV.

Signal and Sideband regions are defined as follows:

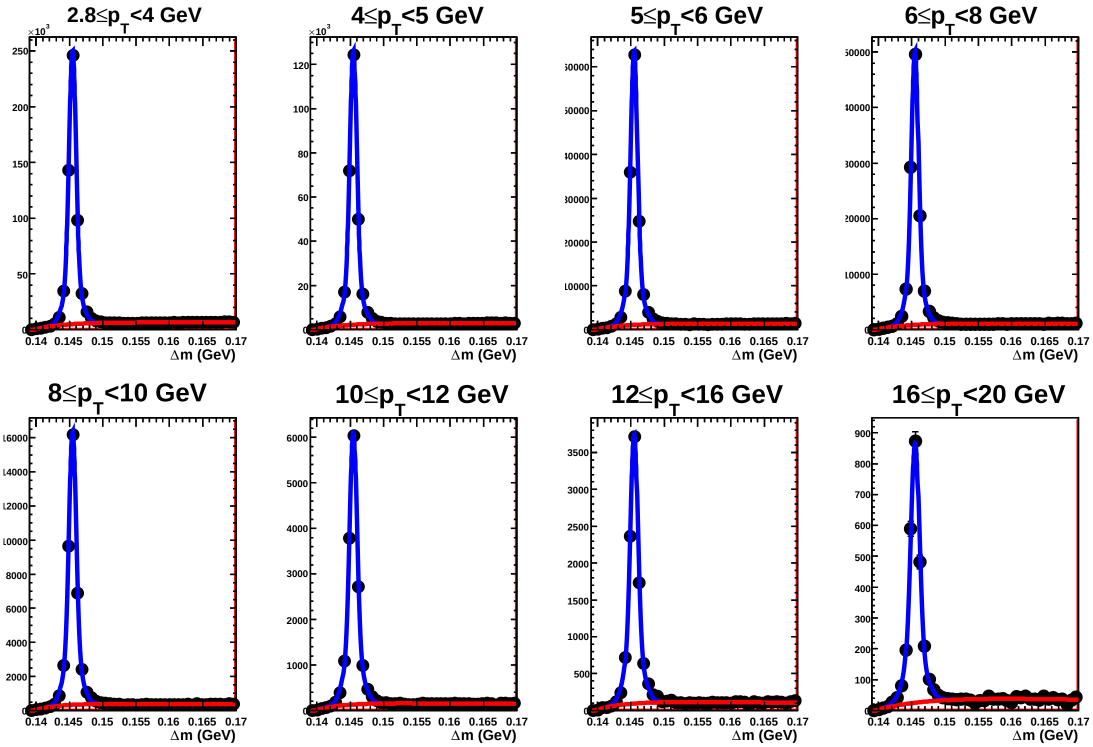

Figure 7.7: Results of the D^* mass fits in bins of the p_T of candidate π's coming from $D^0 \rightarrow K\pi$.

- Left Sideband: $1.101 < m(\Lambda) < 1.106$ GeV,
- Signal Region: $1.111 < m(\Lambda) < 1.121$ GeV,
- Right Sideband: $1.126 < m(\Lambda) < 1.131$ GeV,

The signal and sideband yields are fit in bins of proton p_T. These fits are shown in Fig. 7.8.

Figure 7.9 shows the distribution of muon likelihood $\mathcal{L}$ for μ, π, K, and p obtained from the signal regions in the samples described above. The μ likelihood peaks more strongly at small $\mathcal{L}$, as expected. The final selection, as described in Sec. 7.3.1, is $|\mathcal{L}| < 3.5$.

The technique described in Ref. [36] is used to obtain the efficiency and fake rates. The identification rate is determined as,

$$R_S = \frac{R_M - R_B \cdot f_B}{1 - f_B}, \tag{7.4}$$

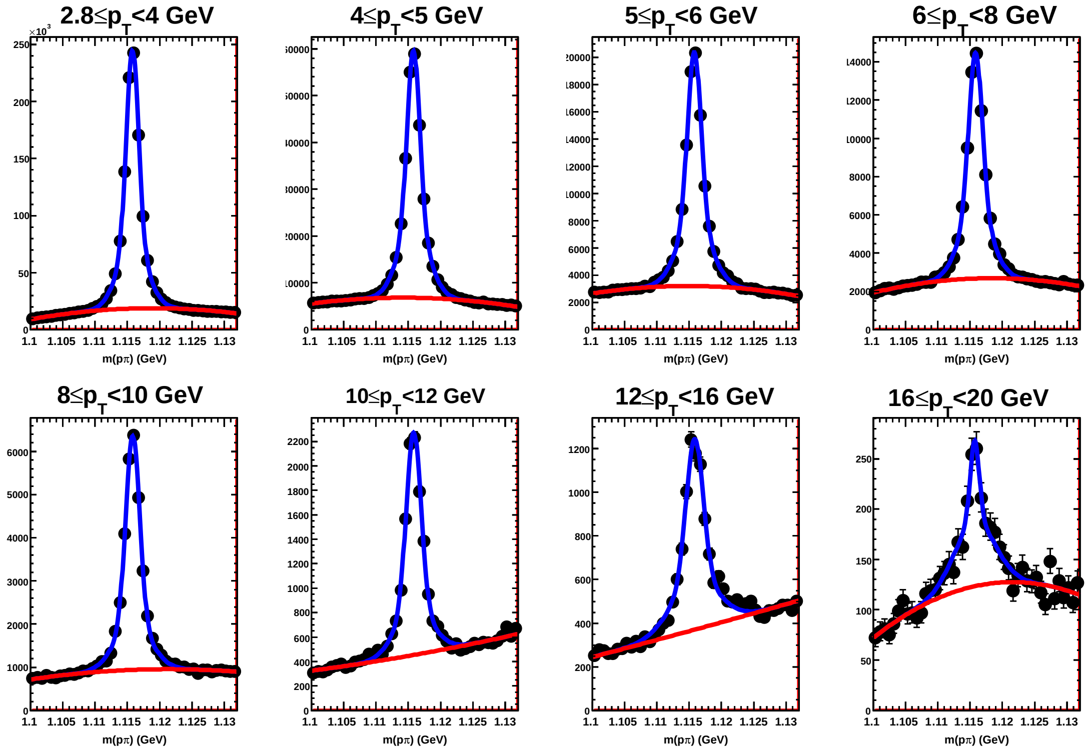

Figure 7.8: Results of the Λ mass fits in bins of the p_T of candidate p's.

where R_M and R_B are the identification rates measured in the signal and sideband regions, respectively, and f_B is the background fraction in the signal window. For fake rates, the measured misidentification rate, R_M, can be written in terms of the decay-in-flight rate as,

$$R_M = f_M \cdot R_{DIF} + (1 - f_M) \cdot R_{PT}, \tag{7.5}$$

where f_M is the decay-in-flight fraction and R_{DIF} and R_{PT} are the identification rates for decay-in-flight and punch-through, respectively. The R_{DIF} is used as presented in Ref. [36].

The identification efficiency is defined as $N(\text{identified})/N(\text{candidates})$, where the candidate requirements are shown in Table 7.1. The efficiencies for each particle type are plotted in Fig. 7.10 as a function of p_T. Note that for all particles except for muons, the "efficiency" is actually the rate that the particle is *misidentified* as a muon. Strong separation is observed between μ and backgrounds.

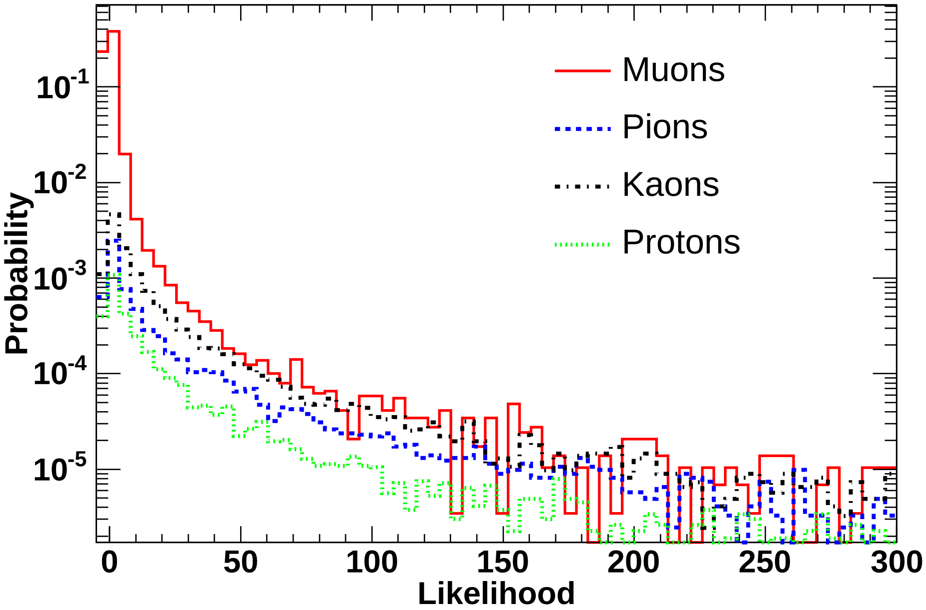

Figure 7.9: A comparison of the soft muon likelihood distributions for μ, π, K, and p.

An efficiency matrix is created in bins of p_T and η using the J/ψ sample. Because the sample is limited in statistics for $p_T > 12$ GeV, empty bins are filled in using interpolation between the low-p_T muons from J/Ψ decays and higher-p_T muons from Z decays. The soft muon identification is applied to Z events so that the region between the J/ψ and Z p_T may be correctly fitted. Figure 7.11 shows an example of these fits for candidates with $|\eta| < 0.15$. The final μ efficiency matrix is shown in Fig. 7.12.

For the corresponding binned misidentification matrix, the misidentification rate is measured in each of the three background samples. The π, K, and p matrices are then combined in the proportion found in W boson decays as presented in Table 3 of Ref. [36]. These weights are $f(\pi) = 0.719$, $f(K) = 0.156$, and $f(p) = 0.125$.

7.3.3 Soft Muon Systematic Uncertainty Determination

Separate systematic uncertainties are estimated for the true muon identification efficiency and the misidentification rate. The sideband subtraction technique used to obtain the muon

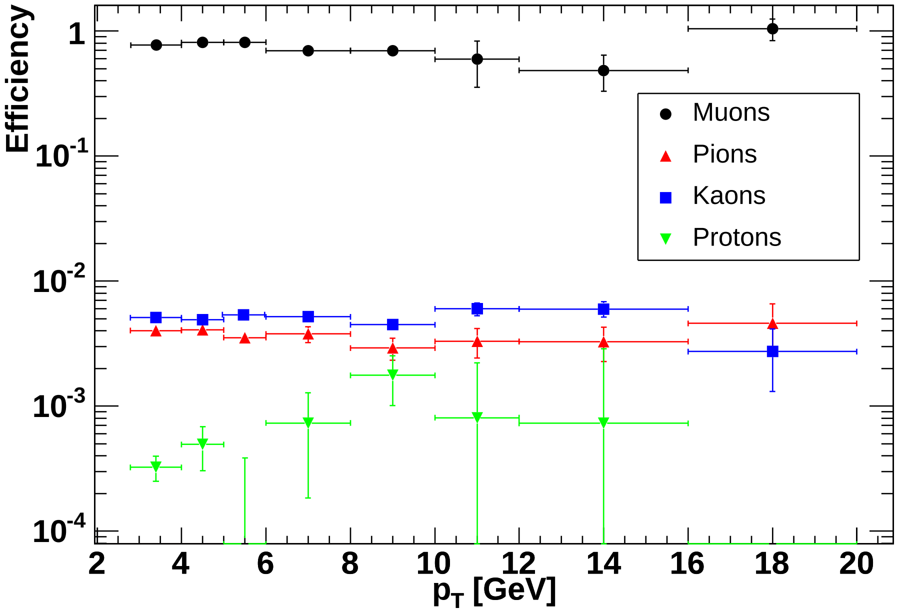

Figure 7.10: Identification efficiency as a function of p_T for μ, π, K, and p. For the case of the μ, this is the rate at which real muons are identified. For the other species, it is the rate that the particle is misidentified as a muon.

efficiency matrix introduces uncertainties arising from the statistics of the J/ψ sample. These uncertainties vary with p_T and η.

In addition, the maximum variation in efficiency of 8% arising from the difference between isolated and non-isolated candidates (See Figure 7.13) is used as an uncertainty representing the maximum possible difference between the J/ψ sample environment and the W/Z environment. This is added in quadrature to the statistical uncertainty arising from the sideband subtraction method (2% - 70%, depending on the bin) to obtain the final muon efficiency uncertainty.

The misidentification systematic uncertainty is obtained by selecting muon-free regions in JET samples and taking the difference between observed and predicted soft muon misidentification rates. The JET sample selections are as follows:

- At least 3 jets with $E_T > 15$ GeV and $|\eta| < 2.0$,
- Reject jets with positive SECVTX tag or negative SECVTX tag having $m(SV) > 0.3$

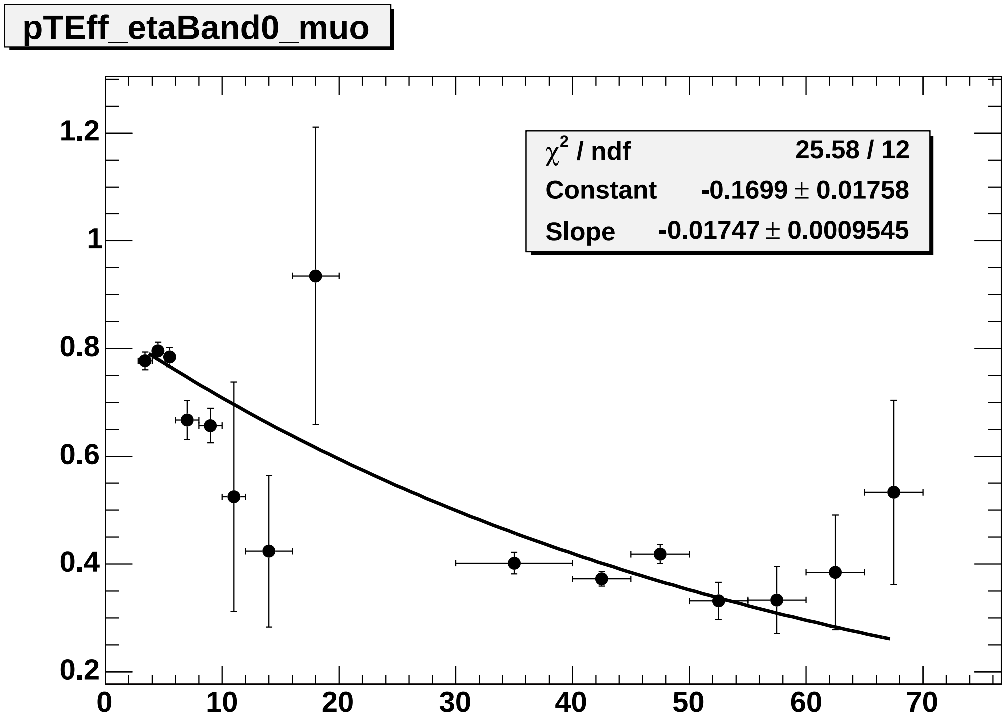

Figure 7.11: Fit for soft muon efficiency function over J/ψ and Z events in the η range $|\eta| < 0.15$.

Sample	**Identified**	**Predicted**	**Uncertainty**
JET50	517	505	2.3%
JET100	2331	2220	4.8%

Table 7.2: Number of events predicted by applying the soft muon misidentification matrix and observed in JET50 and JET100 data.

GeV,

- Reject candidate tracks in jets having $d_0/\sigma(d_0) > 2$.

Figure 7.14 shows the predicted and observed identification rate in JET100.

The differences between the predicted and observed number of events are shown in Table 7.2. Twice the largest error is used as the systematic uncertainty on the misidentification rate.

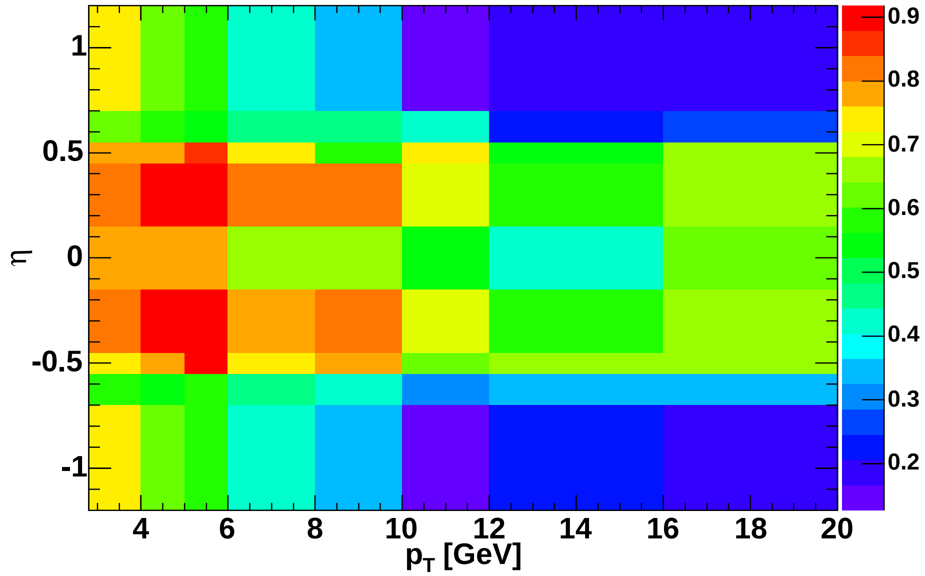

Figure 7.12: Soft muon ID rate matrix in bins of p_T and η.

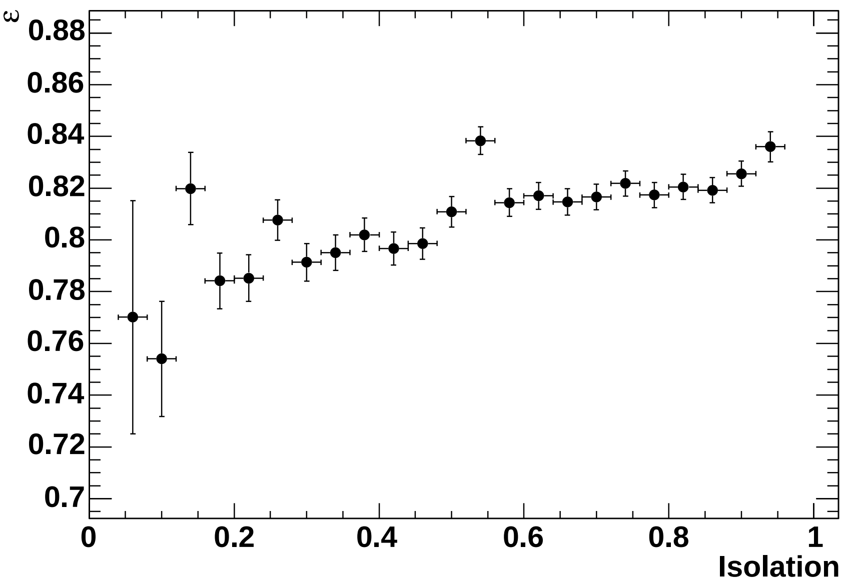

Figure 7.13: Soft muon identification efficiency as a function of the fractional isolation of the muon obtained from the J/ψ sample.

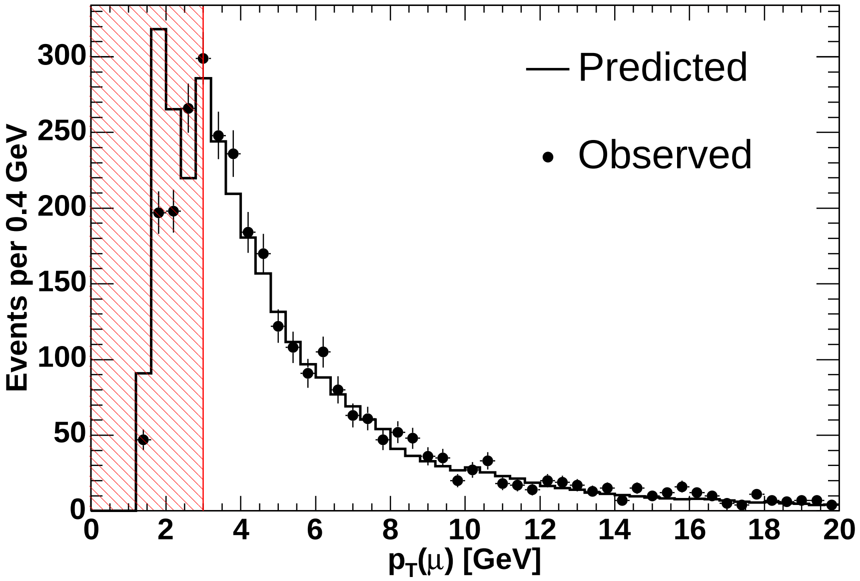

Figure 7.14: Observed and predicted soft muon rate in the JET100 sample with the selections as described in the text as a function of p_T.

7.4 Application of Soft Lepton Identification to W/Z Samples

Additional selection criteria are applied to soft lepton candidates in the high p_T W and Z boson data samples to reduce the amount of background in the search sample: Any track that is already identified as a high-p_T electron or muon in the W or Z boson selection is ineligible to be identified as a soft muon. To reject badly measured tracks, each track is required to have at least one hit in the silicon detector. For electron candidates, this hit must be in the first two layers of the silicon detector to help reject photon conversions. Each track is required to be inside of a reconstructed jet having $|\eta| < 2.0$ and transverse energy of $E_T > 5$ GeV, so that the heavy flavor fraction fit described later in Section 8.1 can be applied. Any track that is identified as a conversion partner is rejected. The track candidate must have a distance along the beamline $|\Delta z| < 5$ cm from the high p_T trigger lepton. If the trigger lepton is the same flavor as the soft lepton, the invariant mass M is calculated of the candidate + trigger, and the following mass ranges are rejected:

- $M < 5$ GeV to remove the J/ψ and $b\bar{b}$ backgrounds.
- $9 < M < 10$ GeV if the candidate track has opposite charge to the trigger lepton. This rejects Υ events.
- $80 < M < 100$ GeV if the candidate track has opposite charge to the trigger lepton. This rejects Z events.

CHAPTER 8

BACKGROUND PREDICTION

The main SM backgrounds in the signal region of this analysis are semileptonic heavy quark decays and photons converting to electron-positron pairs. The contributions from these backgrounds are estimated as described here.

8.1 Heavy Flavor Background Fraction

The leptonic decay of heavy flavor quarks is a significant background contribution to the soft leptons that are being counted in this analysis. This background is estimated using the data in the W/Z + exactly one soft muon bin, which should be dominated by SM processes. A fit is performed in two distributions of soft muons which are sensitive to the heavy flavor fraction: p_T^{rel}, which is the momentum of the muon transverse to the direction of the jet in which it is found, and $d_0/\sigma(d_0)$, which is the significance of the muon's impact parameter with respect to the beamline. A simultaneous fit is performed of these two distributions to a sum of templates from heavy flavor, light flavor, and Drell-Yan processes, as shown in Figure 8.1. These templates were acquired from the MC background samples. Only the contributions from light and heavy jets are allowed to vary, since the Drell-Yan cross section is already known to be well estimated from the Z boson sample validation.

The result of this fit, shown in Table 8.1, is used to normalize the contributions of the three types of processes in the higher-multiplicity sample. The results differ from the factor of 1.45 ± 0.17 found by previous analyses[37] due to the much lower energy requirement for the jets in this analysis. The uncertainty resulting from the fit, ranging from 5% to 34% in the various samples, is used as a systematic uncertainty on this normalization.

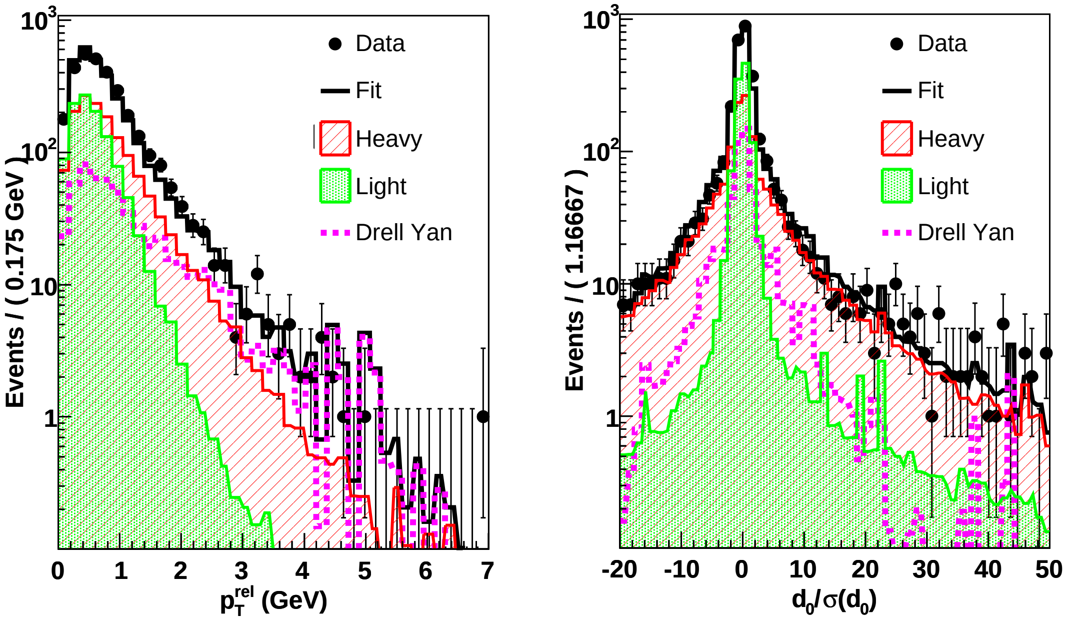

Figure 8.1: The result of the simultaneous fit of the $W + 1$ soft muon sample in the p_T^{rel} and d_0 significance of the soft muon. The data distribution is fit to the sum of three components: W+heavy quark, W+light quark/gluon, and Drell-Yan.

Selection	Component	Scale Factor
TCE W	Heavy	2.51 ± 0.20
	Light	0.93 ± 0.05
CMUP W	Heavy	3.20 ± 0.26
	Light	0.87 ± 0.10
CMX W	Heavy	3.38 ± 0.33
	Light	0.70 ± 0.13
TCE Z	Heavy	5.21 ± 0.71
	Light	0.75 ± 0.15
CMUP Z	Heavy	3.76 ± 0.97
	Light	1.33 ± 0.23
CMX Z	Heavy	4.07 ± 1.19
	Light	1.0 ± 0.27

Table 8.1: The result of the fit to correct the heavy flavor fraction.

8.2 Normalization of Soft Electron Multiplicities

The heavy flavor fit described in Section 8.1 normalizes all of the data to the $W/Z{+}1\mu$ bin. However, we find a mismatch in the $W/Z{+}1e$ bin, which is also expected to be dominated by SM processes. This mismatch is expected to be due to mismodeling of the number of photon conversions in the MC. The difference between the predicted and observed numbers in the W/Z plus exactly one electron bin is 34% in the W boson sample and 31% in the Z boson sample. This is used as a systematic uncertainty for the normalization of all other MC with at least one additional identified electron.

CHAPTER 9

RESULTS

We perform a broad search for additional electrons and muons in the previously identified W and Z boson events. This signature of multiple leptons is common in many models of new physics with light mass scales and couplings to the electroweak sector.

First, a sample of 4,722,370 W boson events and 342,291 Z boson events is obtained from 5.1 fb^{-1} of data. In these base samples, good agreement with predictions is observed in all kinematic distributions, as shown in Chapter 6.

Then, techniques are demonstrated for soft electron and muon identification with no requirement of isolation. For electrons, an efficiency of 80% at $p_T = 2$ GeV rising to 90%-100% for $4 < p_T < 20$ GeV is shown with a corresponding $2\% - 4\%$ misidentification rate. For muons, an efficiency between 80% at $p_T = 3$ GeV and 40% at $p_T = 20$ GeV is shown with a misidentification rate less than 1%. These efficiencies are shown in Figures 7.2 and 7.10.

9.1 Soft Lepton Multiplicity

Using the soft lepton identification techniques described in Section 7, the numbers of W and Z boson events with multiple additional leptons are counted. Figures 9.1 and 9.2 show the multiplicity of additional electrons (N_e) and muons (N_μ) in these events, with the SM expectation and observed data overlaid. The two-dimensional histograms of N_μ vs. N_e are presented in slices of N_e for ease of viewing. These expected and observed event counts are also presented in Tables 9.1 and 9.2 for ease of comparison with predictions from other models. The sources of systematic uncertainties are summarized in Table 9.3, with references to where in the thesis they are described and evaluated. Good agreement with the SM expectation is observed across the distributions.

In particular, very few multi-muon events are observed, which is the region where many

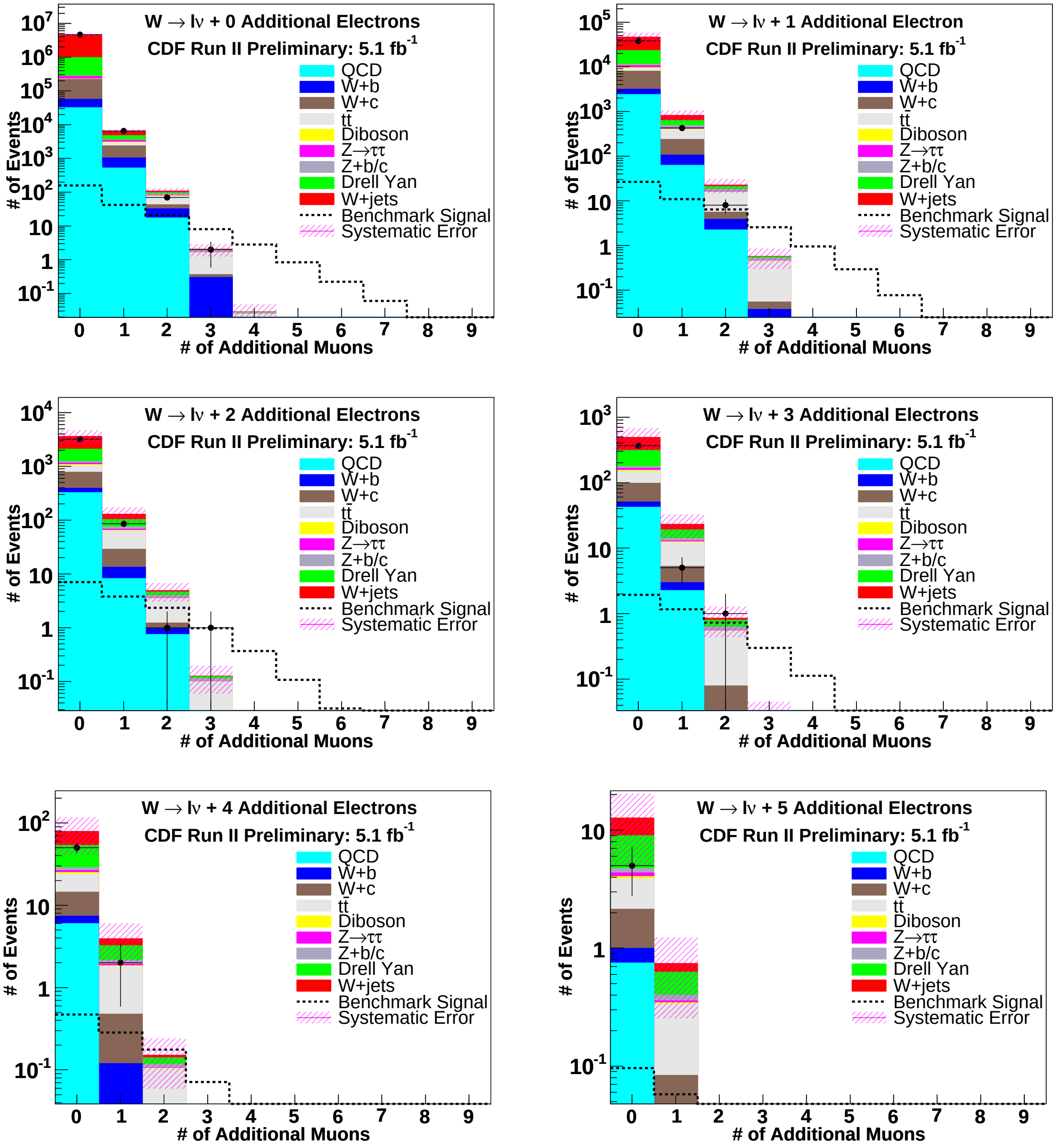

Figure 9.1: Multiplicity of additional electrons and muons after the W boson selection. The two-dimensional histogram of N_μ vs. N_e is presented in slices of N_e for ease of viewing. Both hard and soft leptons (but not the initial leptons used for the W or Z identification) are counted. Note that the distributions combine the electron- and muon-triggered events.

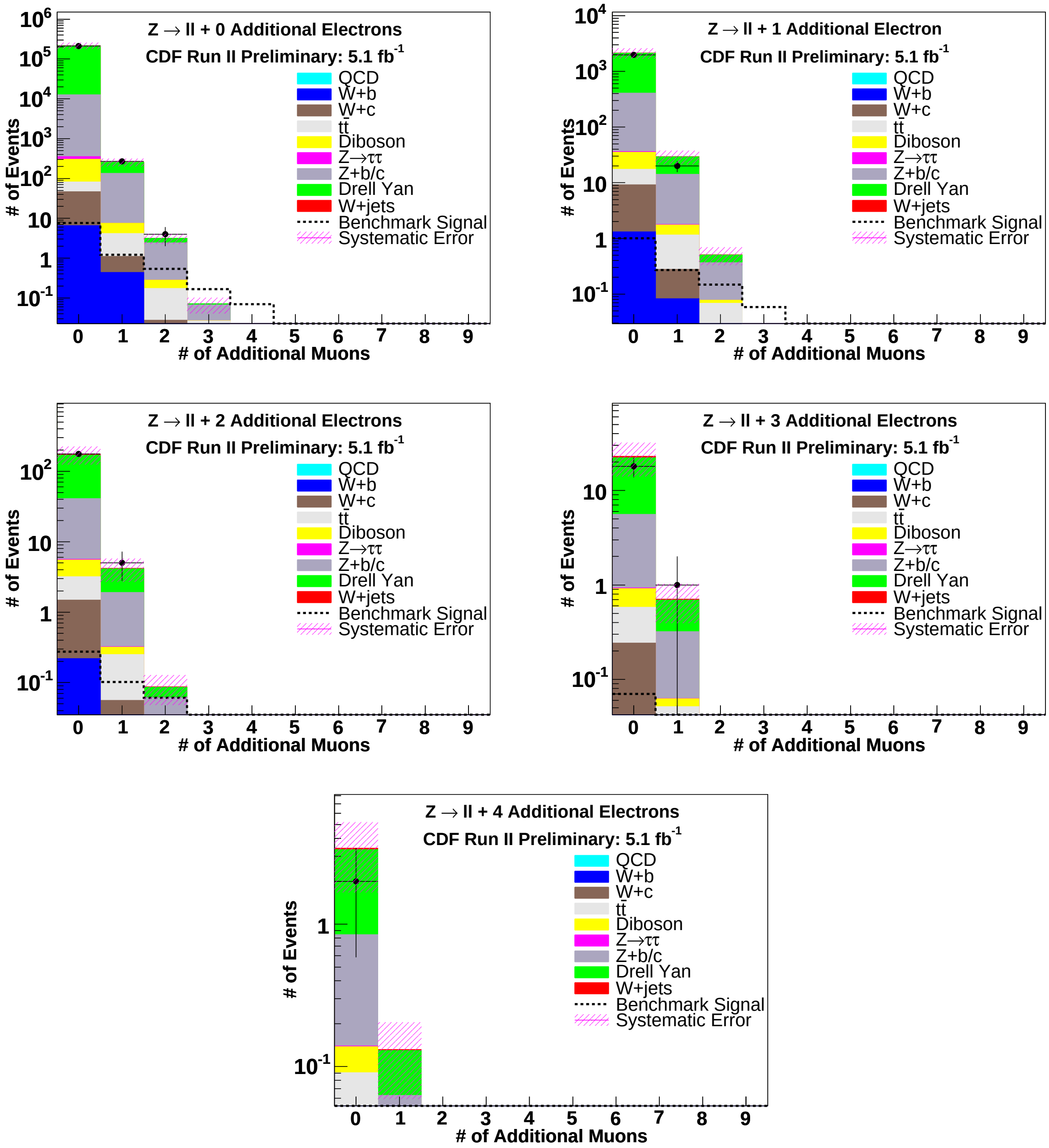

Figure 9.2: Multiplicity of additional electrons and muons after the Z selection. The two-dimensional histogram of N_μ vs. N_e is presented in slices of N_e for ease of viewing. Both hard and soft leptons (but not the initial leptons used for the W or Z identification) are counted. Note that the distributions combine the electron- and muon-triggered events.

N_e	N_μ	Predicted SM Background	Predicted Dark Higgs Signal	Observed
0	0	4623512 ± 315244	158	4673896
0	1	6463 ± 807	42	6498
0	2	109 ± 24	21	70
0	3	2.1 ± 0.79	8.0	2
0	4	0.029 ± 0.019	2.8	0
0	5	0.00026 ± 0.00023	0.83	0
1	0	46055 ± 11387	27	37778
1	1	824 ± 230	11	425
1	2	23 ± 7.8	6.4	8
1	3	0.58 ± 0.27	2.6	0
1	4	0.010 ± 0.0074	0.95	0
1	5	0.00011 ± 0.00011	0.29	0
2	0	3600 ± 1085	7.1	3184
2	1	129 ± 43	3.8	86
2	2	4.9 ± 1.8	2.3	1
2	3	0.13 ± 0.067	0.97	1
2	4	0.0031 ± 0.0024	0.37	0
3	0	491 ± 185	1.9	366
3	1	23 ± 9.3	1.2	5
3	2	0.85 ± 0.42	0.72	1
3	3	0.028 ± 0.017	0.30	0
4	0	79 ± 38	0.47	50
4	1	3.9 ± 2.1	0.28	2
5	0	13 ± 7.6	0.096	5
5	1	0.74 ± 0.49	0.058	0
6	0	2.0 ± 1.5	0.015	0

Table 9.1: Summary of predicted and observed event counts by number of additional electrons (N_e) and muons (N_μ) after the W boson selection. The prediction of a model described in Section 9.3 is also shown for comparison. Bins with less than 0.25 expected events in both signal and background and 0 observed events are not shown.

N_e	N_μ	Predicted SM Background	Predicted Dark Higgs Signal	Observed
0	0	215219 ± 36886	7.6	211448
0	1	255 ± 52	1.2	270
0	2	3.2 ± 0.89	0.54	4
1	0	2145 ± 447	1.0	1975
1	1	30 ± 8.1	0.27	20
1	2	0.51 ± 0.18	0.15	0
2	0	175 ± 50	0.28	176
2	1	4.2 ± 1.5	0.10	5
3	0	23 ± 9.0	0.070	18
3	1	0.71 ± 0.31	0.031	1
4	0	3.4 ± 1.8	0.019	2
5	0	0.52 ± 0.35	0.0044	0

Table 9.2: Summary of predicted and observed event counts by number of additional electrons (N_e) and muons (N_μ) after the Z selection. The prediction of a model described in Section 9.3 is also shown for comparison. Bins with less than 0.25 expected events in both signal and background and 0 observed events are not shown.

Systematic Source	Size (%)	Effect in Large S/B Region (Events)
Trigger Efficiency (Sec. 6.3)	±(1.6 - 5.9)%	±0.06
QCD fraction (Sec. 5)	±26%	0
Soft e real rate (Sec. 7.2.2)	±15%	±0.04
Soft e fake rate (Sec. 7.2.2)	±15%	±0.11
Soft μ real rate (Sec. 7.3.3)	±(8-70)%	±0.64
Soft μ fake rate (Sec. 7.3.3)	±10%	±0.34
Soft e normalization (Sec. 8.2)	±(31-39)%	±0.24
Heavy Flavor Fraction (Sec. 8.1)	±(5-34)%	±0.25

Table 9.3: Sources of systematic uncertainties. Their size is measured both as a percentage and as the number of events in a benchmark-signal-rich region, defined as a W or Z boson plus at least 3 additional muons with $p_T > 3$ GeV. Note that, although some of the systematics are large, they have little effect in the signal region due to there being negligible SM background.

lepton jet models would be expected to show an excess. Only three events containing 3 muons beyond the W or Z selection are observed, which is consistent with the SM expectation of 2.9 events. No events are observed containing four or more additional muons. A summary of the high lepton multiplicity events that are observed in the data is shown in Appendix D.

9.2 Soft Lepton Kinematics

Many new physics scenarios, in addition to creating excesses in the soft lepton multiplicity, would result in discrepancies in the soft lepton kinematics. For example, in the benchmark model described in Table 9.4, pairs of leptons are produced by decaying dark photons, which would create a mass peak at 300 MeV in the dilepton mass.

With no excess having been seen in the soft lepton multiplicity, the dilepton mass distributions are shown in Figure 9.3. All distribution are consistent with the SM predictions. The most discrepant is the $Z + \mu\mu$ mass distribution, but the discrepancy does not pass the 95% confidence level.

9.3 Benchmark Model

A series of representative lepton jet models is presented in Ref. [11]. The benchmark model chosen for this analysis is an adaptation of the 'Neutralino Benchmark Model,' in which the Higgs decays principally to a pair of the lightest supersymmetric particles, which then decay through a dark sector to lepton jets. A MC sample of signal events was generated from this model using Pythia. The signal from this model that this analysis is most sensitive to is associated production of a W or Z boson and a Higgs boson, which has a cross section of 389 fb^{-1}.

The particular parameters of the model [38] were chosen to create a 'typical' model of this class. The MSSM parameters (μ, m_1, m_2, $\tan(\beta)$ and $\sin(\alpha)$) avoid previous searches for supersymmetry while making the lightest supersymmetric partner (χ_0) the favored Higgs

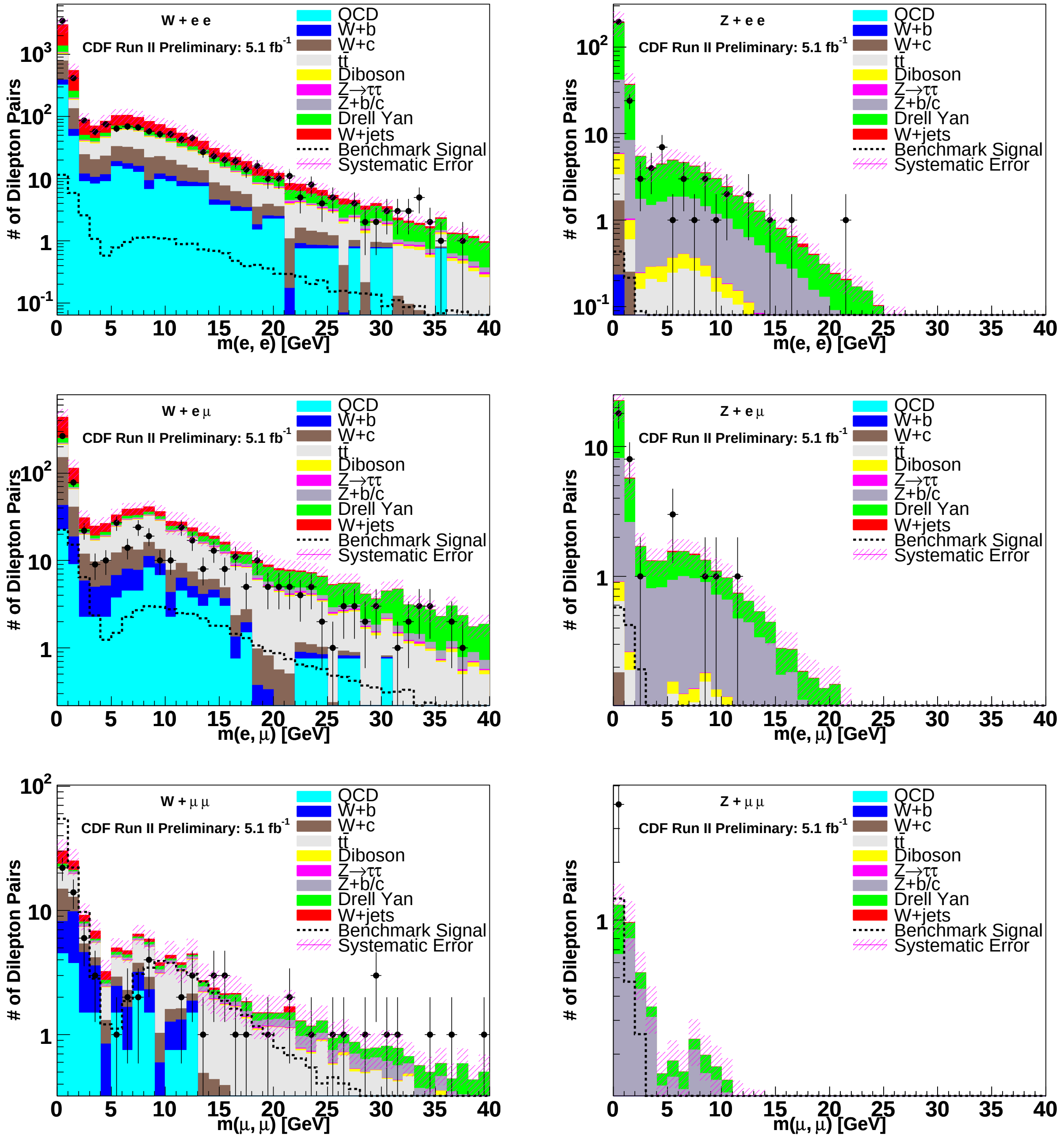

Figure 9.3: Distributions of the invariant mass of each pair of soft leptons ee (top), $e\mu$ (center), and $\mu\mu$ (bottom). The W-selected events are on the left and the Z-selected events are on the right. Note that the distributions combine the electron-and muon-triggered events. The contribution from conversions swamps any new physics signal in the $m(e,e)$ and $m(e,\mu)$ distributions, but the $m(\mu,\mu)$ distribution (bottom) is sensitive to the benchmark model as well as to other new physics models.

Parameter	Value
μ	149 GeV
m_1(bino)	13 GeV
m_2(wino)	286 GeV
$\tan(\beta)$	3.5
$\sin(\alpha)$	-0.28
m_{χ_0}	10 GeV
m_H	120 GeV
m_{χ_d}	1 GeV
m_{γ_d}	300 MeV
BR($\chi_0 \rightarrow \chi_d + 2\gamma_d$)	33%
BR($\chi_0 \rightarrow \chi_d + 3\gamma_d$)	33%
BR($\chi_0 \rightarrow \chi_d + 4\gamma_d$)	33%

Table 9.4: Parameters used for the benchmark model based on that in Ref. [11]. The first five parameters are the inputs to the MSSM including the branching fractions for $\chi_0 \rightarrow \chi_d + N\gamma_d$ [38].

decay channel. The Higgs has a mass near that favored by precision measurements. The branching fractions for χ_0 decaying into the dark neutralinos (χ_d) and dark photons (γ_d) simply model the sort of cascade decay illustrated in Figure 1.1. The mass of the dark photon is chosen in order to make the additional leptons that are produced approximately half muons and half electrons. These parameters are summarized in Table 9.4.

As an example, a limit is set on the production of this benchmark model. The limit is set at $0.312 \times \sigma$, or 112 fb, at 95% credibility. The model can be ruled out at the standard cross section at a confidence level of 99.7%.Both of these limits are set using the MCLIMIT tools [39] running over the combined W and Z channels in Figures 9.1 and 9.2.

9.4 Application to Other Models

In addition to the benchmark model discussed in Section 9.3, limits can be set on a wide range of alternate models. A rough estimate of the limit for a particular model can be made by normalizing its production to the W or Z boson cross section, applying the efficiencies in Figures 7.2 and 7.10 to the additional leptons, and comparing the result to the observed

Object	Requirements	Number Observed
W	$p_T(e/\mu) > 20$ GeV $\lvert\eta(e)\rvert < 1.1$, $\lvert\eta(\mu)\rvert < 1.5$ $\not\!\!E_T > 25$ GeV $m_T(l, \not\!\!E_T) > 20$ GeV $d\phi(l, \not\!\!E_T) > 0.5$	4,722,370
Z	$p_T(e/\mu) > 20$ GeV $p_T(e_2) > 12$ GeV, $p_T(\mu_2) > 10$ GeV $\lvert\eta(e)\rvert < 1.1$, $\lvert\eta(\mu)\rvert < 1.5$ 76 GeV $< m(l_1, l_2) <$ 106 GeV	342,291
soft e	$p_T(e) > 2$ GeV $\lvert\eta(e)\rvert < 1$ $\mathcal{L} > 0.99$ (Efficiency in Figure 7.2)	See Tables 9.1 and 9.2
soft μ	$p_T(\mu) > 3$ GeV $\lvert\eta(\mu)\rvert < 1.5$ $\lvert\mathcal{L}\rvert < 3.5$ (Efficiency in Figure 7.10)	See Tables 9.1 and 9.2

Table 9.5: Summary of kinematic requirements to find various objects. These numbers can be used to set limits on many models that predict production of additional leptons.

and predicted numbers of additional leptons in Tables 9.1 and 9.2. For ease of reference, a summary of the kinematic selections for identified objects is presented in Table 9.5.

In general, any model that predicts significant numbers of 3-muon events can be ruled out, since only three such events are observed in the sample, consistent with the SM background. However, models that produce multiple electrons can more easily be accommodated, since photon conversions result in a much higher background in that region.

CHAPTER 10

CONCLUSIONS

This analysis expands the reach of previous searches for additional leptons by allowing leptons to be reconstructed from a much lower p_T threshold and with no requirement of isolation. This greatly increases the acceptance to find lepton jets or similar excesses of leptons from effects beyond the SM. No indication of such new effects is seen in the data sample. A 95% confidence level limit is set on an example benchmark model of supersymmetric Higgs production, and a framework is provided to set limits on a class of other models.

REFERENCES

[1] S.L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. D **2**, 1285 (1970);

[2] R. Dermisek and J. F. Gunion, Phys. Rev. Lett. **95**, 041801 (2005); R. Dermisek and J. F. Gunion, Phys. Rev. D. **75**, 075019 (2007).

[3] C. T. Hill and R. J. Hill, Phys. Rev. D **75**, 115009 (2007); Phys. Rev. D **76**, 115014 (2007).

[4] C. Cheung, J. T. Ruderman, L. Wang, and I. Yavin, JHEP **1004**, 116 (2010)

[5] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slayter and N. Weiner, Phys. Rev. D **79**, 015014 (2009); P. Meade, M. Papucci, A. Strumia, and T. Volansky, arXiv:0905.0480 [hep-ph]; M. Pospelov and A. Ritz, Phys. Lett. B **671**, 391 (2009); C. Cheung, J. T. Ruderman, L. T. Wang and I. Yavin, Phys. Rev. D **80**, 035008 (2009); A. Katz and R. Sundrum, JHEP **0906**, 003 (2009).

[6] J. T. Ruderman and T. Volansky, arXiv:0908.1570 [hep-ph]; X. Chen, JCAP **0909**, 029 (2009); J. Mardon, Y. Normura and J. Thaler, Phys. Rev. D **80** 035013 (2009).

[7] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. D **83**, 112008 (2011)

[8] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **101**, 251801 (2008)

[9] M. J. Strassler and K. M. Zurek, Phys. Rev. Lett. B **651**, 374 (2007); T. Han, Z. Si, K. M. Zurek, and M. J. Strassler JHEP **0807**, 008 (2008).

[10] O. Adriania *et al.* [PAMELA Collaboration], Nature **458**, 607 (2009); F. Aharonian *et al.* [H.E.S.S Collaboration], Phys. Rev. Lett. **101**, 261104 (2008); A. A. Abdo *et al.* [Fermi LAT Collaboration], Phys. Rev. Lett. **102**, 181101 (2009).

[11] A. Falkowski, J. T. Ruderman, T. Voalnsky and J. Zupan, JHEP **1005**, 077 (2010); A. Falkowski, J. T. Ruderman, T. Volansky and J. Zupan, Phys. Rev. Lett. **105**, 241801 (2010).

[12] V. M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **105**, 211802 (2010)

[13] CMS Collaboration, JHEP **1107** 098 (2011)

[14] We use the convention that "momentum" refers to pc and "mass" to mc^2.

[15] F. Abe *et al.*, Nucl. Instrum. Methods **271**, 387 (1988); D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **71**, 052003 (2005); The CDF-II Detector Technical Design Report, Fermilab-Pub-96/390-E; A. Abulencia *et al.* (CDF Collaboration), J. Phys. G **34**, 2457 (2007); T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. D **77**, 112001 (2008).

[16] The CDF II detector uses a cylindrical coordinate system in which ϕ is the azimuthal angle, r is the radius from the nominal beam line, and z points in the proton beam direction. The transverse $(r-\phi)$ plane is perpendicular to the z axis. Transverse momentum and energy are the respective projections of momentum measured in the tracking system and energy measured in the calorimeter system onto the $r-\phi$ plane, and are defined as $p_T = p\sin\theta$ and $E_T = E\sin\theta$. Here, θ is the polar angle measured with respect to the interaction vertex. Missing $\mathrm{E_T}$ ($\vec{\not{E}}_T$) is defined by $\vec{\not{E}}_T = -\sum_i E_T^i \hat{n}_i$, where i is the calorimeter tower number for $|\eta| < 3.6$, and $\hat{n}_i$ is a unit vector perpendicular to the beam axis and pointing at the i^{th} tower. The transverse mass m_T is defined as the invariant mass of the sum of the transverse projection of the momentum vector of one decay product and the missing energy vector $\vec{\not{E}}_T$. The pseudorapidity η is defined as $-\ln(\tan(\theta/2))$, where θ is measured with respect to the origin of the detector. We define the magnitude $\not{E}_T = |\vec{\not{E}}_T|$.

[17] A. Sill *et al.*, Nucl. Instrum. Methods **447**, 1 (2000); A. Affolder *et al.*, Nucl. Instrum. Methods **453**, 84 (2000); C.S. Hill *et al.*, Nucl. Instrum. Methods **530**, 1 (2000).

[18] A. Affolder *et al.*, Nucl. Instrum. Methods **526**, 249 (2004).

[19] L. Balka *et al.* Nucl. Instrum. Methods **267**, 272 (1988).

[20] S. Bertolucci *et al.* Nucl. Instrum. Methods **267**, 301 (1988).

[21] S. Kuhlmann *et al.*, Nucl. Instrum. Methods **518**, 39, 2004.

[22] G. Ascoli *et al.*, Nucl. Instrum. Methods **268**, 33 (1988); A. Artikov *et al.*, Nucl. Instrum. Methods **538**, 358 (2005).

[23] D. Acosta *et al.* (CDF Collaboration), Nucl. Instrum. Methods **461**, 540 (2001); D. Acosta *et al.*, Nucl. Instrum. Methods **494** 57 (2002).

[24] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. **94**, 091803 (2005).

[25] F. Abe *et al.* (CDF Collaboration), Nucl. Instrum. Methods **271**, 387 (1988).

[26] A. Abulencia *et al.*, J. Phys. G **34**, 2457 (2007).

[27] F. A. Berends, W. T. Giele, H. Kuijf, R. Kleiss, and W. James Stirling, Phys. Lett. B **224**, (1989).

[28] S. Paramonov, "A Limit on the Branching Ratio of the Flavor-Changing Top Quark Decay $t \to Zc$.", Ph.D. Thesis, University of Chicago (2009)

[29] T. Aaltonen *et al.* (CDF Collaboration), arXiv:hep-exp/1004.1181.

[30] B. Cooper, Ph.D. thesis, University Coll. London (2006), FERMILAB-THESIS-2006-61.

[31] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions", JHEP 0307:001 (2003)

[32] T. Sjostrand, S. Mrenna, P. Skands, "PYTHIA 6.4 Physics and Manual", JHEP 0605:026 (2006)

[33] E. Gerchtein and M. Paulini, Computing in High Energy and Nuclear Physics, p. 1 (2003)

[34] A. Abulencia *et al.* (CDF Collaboration), J. Phys.G Nucl. Part. Phys. **34** (2007)

[35] A. Ivanov, T. Schwarz, R. Erbacher, *"Calibration of Heavy Flavor Content in W+Jets Data"*, CDF Note 9403 (2008).

[36] A. Bridgeman, L. Cerrito, U. Grundler, T. Liss, and X. Zhang, *"A Heavy Flavor-Free Mistag Matrix for the Soft Muon Tagger"*, CDF Note 9083 (2008).

[37] A. Ivanov, T. Schwarz, R. Erbacher, *"Calibration of Heavy-Flavor Content in W + Jets Data"*, CDF Note 9403 (2008)

[38] M. Reece, Private Communication.

[39] T. Junk, NIM A434, p. 435-443 (1999)

[40] S. Fidele, P. Murat, *"2D Reconstruction of the CES Showers and Identification of High-p_T Electrons"*, CDF Note 7097 (2005)

APPENDIX A

DATA SAMPLES

Many differently-triggered data samples are used in this analysis. High-p_T lepton triggers are used to acquire the signal region data samples. Lower-p_T lepton triggers are used to acquire pure samples of leptons to train and test the soft lepton identification algorithms. Other triggers are used to select non-lepton tracks to check the misidentification rate of the soft lepton identification. All of these data samples are described more fully in this appendix.

A.1 Signal Region Data Samples

This analysis uses events selected with the "`ELECTRON_CENTRAL_18`", "`MUON_CMUP18`", and "`MUON_CMX18`" triggers as the search region. The good run list "`goodrun_v37_em_mu_si.list`", requiring good electrons, muons, and silicon, is used. The data encompass CDF Periods 1 to 27, covering the calendar period Dec. 7, 2004 to Jan. 6, 2010. The datasets, runs, run periods, and luminosities are listed in Table A.1. The total luminosity for this dataset is 5.1 fb^{-1}, and the numbers of electron- and muon-triggered events are 384,622,495 and 224,359,512, respectively.

A.2 Soft Electron Identification Data Samples

The real electron training sample for the soft electron identification algorithm is collected using the `ELECTRON_CENTRAL_8_L2_DPS` trigger. In order to be sure of having similar detector response in the training sample and the search sample, the same range of data is used in each sample. Datasets `blpcah`, `blpcai`, `blpcaj`, `blpcak` and `blpcam` are used.

The non-electron background sample is collected using the `MUON_CMUP18` and `MUON_CMX18` triggers. These are the same triggers used in the signal sample, and were chosen in order to provide in the training sample a similar environment to the signal sample. The differences

Dataset	Period	Runs	Luminosity	Dates
0h	1	190697 – 195408	363pb^{-1}	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^{-1}	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^{-1}	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^{-1}	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	2687pb^{-1}	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
	23	272470 – 274055		15 Feb 09 - 21 Mar 09
	24	274123 – 275848		22 Mar 09 - 4 May 09
	25	275873 – 277511		5 May 09 - 13 Jun 09
	26	282976 – 284843		15 Sep 09 - 25 Oct 09
	27	284858 – 287261		25 Nov 09 - 06 Jan 10

Table A.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 the triggers are unprescaled and the same good run list is used for both triggers.

are that no W or Z boson selection is applied and all sources of real electron contamination are removed in the training sample.

The systematic uncertainty in the soft electron identification is measured using non-electron tracks collected with the `JET_50` trigger. Datasets `gjt2ah`, `gjt2ai`, `gjt2bi`, `gjt2ci`, `gjt2bj`, `gjt2bk` and `gjt2bm` are used.

A.3 Soft Muon Identification Data Samples

The efficiency of the soft muon identification algorithm is measured using muons obtained from the `MUON_CMUP8` trigger. Periods `0h` through `0j` are used, corresponding to datasets `bmclah`, `bmclai` and `bmclaj`.

The misidentification rate of the soft muon identification is measured using sample of pions, kaons and protons obtained from D^* and Λ decays. These decays are found in events collected using the `EXO_TWO_TRACK` trigger. Datasets `hbhdah`, `hbhdai`, and `hbhdak` are used.

The systematic uncertainty in the soft muon misidentification rate is measured using non-muon tracks collected using the `JET_100` trigger. The dataset `gjt4ah` is used.

APPENDIX B

MONTE CARLO SAMPLES

This analysis uses the standard CDF Monte Carlo samples maintained by the CDF Top Group. Many of these samples consist of both low-luminosity files and high-luminosity files. In order to properly model the luminosity of the data sample used in this analysis, the low-luminosity and high-luminosity files are combined and the events are weighted according to the number of primary vertices in the event.

Because heavy quark (c and b) decays are an important background in this analysis, MC samples consisting specifically of W or Z bosons and heavy quarks are used. The W/Z plus jets samples contain some heavy quarks, but all events in those samples that produce a heavy quark are removed in order to avoid overlap with the heavy flavor samples.

The Alpgen W + jets and W + heavy quarks datasets are shown in Tables B.1 and B.4, and the Alpgen Drell-Yan and Drell-Yan + heavy quarks datasets are shown in Tables B.2, B.3 and B.5. The Pythia datasets are shown in Table B.6.

Process	σ	low $\mathcal{L}$ Dataset	low $\mathcal{L}$ N_{events}	high $\mathcal{L}$ Dataset	high $\mathcal{L}$ N_{events}
$W(e\nu) + 0p$	2.52 nb	pt0sw0	4929337	ut0s00	1985030
$W(e\nu) + 1p$	315 pb	pt0sw1	4909767	ut0s01	1984122
$W(e\nu) + 2p$	49.42 pb	pt0s2w	918835	ut0s02	400219
$W(e\nu) + 3p$	7.83 pb	pt0s3w	783415	ut0s03	396219
$W(e\nu) + \geq 4p$	1.44 pb	pt0s4w	453531	ut0s04	396219
$W(\mu\nu) + 0p$	2.52 nb	pt0sw5	5010637	ut0s05	1985030
$W(\mu\nu) + 1p$	315 pb	pt0sw6	4997783	ut0s06	1985030
$W(\mu\nu) + 2p$	49.42 pb	pt0s7w	877801	ut0s07	400219
$W(\mu\nu) + 3p$	7.83 pb	pt0s8w	817043	ut0s08	400219
$W(\mu\nu) + \geq 4p$	1.44 pb	pt0s9w	906265	ut0s09	400219
$W(\tau\nu) + 0p$	2.52 nb	ut0sw0	4868422	ut0s10	1984200
$W(\tau\nu) + 1p$	315 pb	ut0sw1	4981403	ut0s11	1985030
$W(\tau\nu) + 2p$	49.56 pb	ut0s2w	917094	ut0s12	400219
$W(\tau\nu) + 3p$	7.84 pb	ut0s3w	1008221	ut0s13	400219
$W(\tau\nu) + \geq 4p$	1.44 pb	ut0s4w	986494	ut0s14	396219

Table B.1: Alpgen $W+$ partons dataset names, processes, cross sections, and number of events generated. Note that the cross-sections listed include a K-factor of 1.4.

Process	σ	low $\mathcal{L}$ Dataset	low $\mathcal{L}$ N_{events}	high $\mathcal{L}$ Dataset	high $\mathcal{L}$ N_{events}
$Z(ee) + 0p$ [75 : 105]	158 pb	zt0sp0	2639520	bt0sz0	880438
$Z(ee) + 1p$ [75 : 105]	21.6 pb	zt0sp1	2624793	bt0sz1	1024551
$Z(ee) + 2p$ [75 : 105]	3.46 pb	zt0szb	4595453	bt0sz2	1793000
$Z(ee) + 3p$ [75 : 105]	0.55 pb	zt0s3p	524261	bt0sz3	192119
$Z(ee) + \geq 4p$ [75 : 105]	99.2 fb	zt0s4p	525065	bt0sz4	192119
$Z(\mu\mu) + 0p$ [75 : 105]	158 pb	zt0sp5	2659832	bt0sz5	1020551
$Z(\mu\mu) + 1p$ [75 : 105]	21.6 pb	zt0sp6	2652428	bt0sz6	1021555
$Z(\mu\mu) + 2p$ [75 : 105]	3.46 pb	zt0szt	4660506	bt0sz7	1793000
$Z(\mu\mu) + 3p$ [75 : 105]	0.55 pb	zt0s8p	536159	bt0sz8	192119
$Z(\mu\mu) + \geq 4p$ [75 : 105]	99.2 fb	zt0s9p	530242	bt0sz9	192119
$Z(\tau\tau) + 0p$ [75 : 105]	158 pb	zt0st3	5860164	bt0sza	2400920
$Z(\tau\tau) + 1p$ [75 : 105]	21.5 pb	zt0st4	5722772	bt0szb	2400920
$Z(\tau\tau) + \geq 2p$ [75 : 105]	4.14 pb	zt0st2	2263107	bt0szc	953280

Table B.2: Alpgen $Z+$ partons dataset names, processes, cross sections, and number of events generated. The mass range in the process name refers to the generated mass range of the dilepton pair. Note that the cross-sections listed do not include a K-factor.

Process	σ	low $\mathcal{L}$ Dataset	low $\mathcal{L}$ N_{events}	high $\mathcal{L}$ Dataset	high $\mathcal{L}$ N_{events}
$DY(ee) + 0p$ [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	zt0so6	192119
$DY(ee) + 1p$ [20 : 75]	8.39 pb	xt0s1p	530958	zt0so7	192119
$DY(ee) + 2p$ [20 : 75]	1.61 pb	xt0s2p	536159	zt0so9	1793000
$DY(ee) + 3p$ [20 : 75]	233 fb	xt0s3p	525670	zt0soa	192119
$DY(ee) + \geq 4p$ [20 : 75]	39.8 fb	xt0s4p	515638	zt0sob	192119
$DY(ee) + 0p$ [105 : 600]	4.07 pb	yt0s0p	519104	zt0sol	192119
$DY(ee) + 1p$ [105 : 600]	705 fb	yt0s1p	524895	zt0som	192119
$DY(ee) + 2p$ [105 : 600]	117 fb	yt0s2p	513214	zt0son	192119
$DY(ee) + 3p$ [105 : 600]	18.5 fb	yt0s3p	504749	zt0soo	192119
$DY(\mu\mu) + 0p$ [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	zt0soc	192119
$DY(\mu\mu) + 1p$ [20 : 75]	8.4 pb	xt0s6p	530696	zt0sod	192119
$DY(\mu\mu) + 2p$ [20 : 75]	1.6 pb	xt0s7p	520703	zt0sof	1792478
$DY(\mu\mu) + 3p$ [20 : 75]	233 fb	xt0s8p	509424	zt0sog	192119
$DY(\mu\mu) + \geq 4p$ [20 : 75]	39.8 fb	xt0s9p	523932	zt0soh	192119
$DY(\mu\mu) + 0p$ [105 : 600]	4.07 pb	yt0s5p	530941	zt0soq	192119
$DY(\mu\mu) + 1p$ [105 : 600]	706 fb	yt0s6p	529581	zt0sor	192119
$DY(\mu\mu) + 2p$ [105 : 600]	117 fb	yt0s7p	531006	zt0sos	192119
$DY(\mu\mu) + 3p$ [105 : 600]	18.5 fb	yt0s8p	510246	zt0sot	192119
$DY(\tau\tau) + 0p$ [20 : 75]	160 pb	xt0st0	1135920	-	-
$DY(\tau\tau) + 1p$ [20 : 75]	8.38 pb	xt0st1	1158902	-	-
$DY(\tau\tau) + \geq 2p$ [20 : 75]	1.82 pb	xt0st2	2270345	-	-
$DY(\tau\tau) + 0p$ [105 : 600]	4.07 pb	zt0s0h	268428	-	-
$DY(\tau\tau) + 1p$ [105 : 600]	707 fb	zt0s1h	268428	-	-
$DY(\tau\tau) + \geq 2p$ [105 : 600]	117 fb	zt0s2h	268428	-	-

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

Process	σ	low $\mathcal{L}$ Dataset	low $\mathcal{L}$ N_{events}	high $\mathcal{L}$ Dataset	high $\mathcal{L}$ N_{events}
$W(e\nu)+bb+0p$	4.17 pb	bt0s0w	1541069	bt0s00	593755
$W(e\nu)+bb+1p$	1.24 pb	bt0s1w	1545970	bt0s01	594426
$W(e\nu)+bb+\geq 2p$	402 fb	bt0s2w	1498550	bt0s02	604337
$W(\mu\nu)+bb+0p$	4.17 pb	bt0s5w	1539099	bt0s05	605333
$W(\mu\nu)+bb+1p$	1.24 pb	bt0s6w	1529300	bt0s06	601125
$W(\mu\nu)+bb+\geq 2p$	400 fb	bt0s7w	1501959	bt0s07	593788
$W(\tau\nu)+bb+0p$	4.17 pb	dt0sw0	769285	bt0s10	601802
$W(\tau\nu)+bb+1p$	1.24 pb	dt0sw1	1105495	bt0s11	596337
$W(\tau\nu)+bb+\geq 2p$	400 fb	dt0sw2	1468622	bt0s12	592389
$W(e\nu)+cc+0p$	7.00 pb	ct0s0w	2008023	bt0s15	796448
$W(e\nu)+cc+1p$	2.51 pb	ct0s1w	1983960	bt0s16	772496
$W(e\nu)+cc+\geq 2p$	879 fb	ct0s2w	2001927	bt0s17	777696
$W(\mu\nu)+cc+0p$	7.00 pb	ct0s5w	2018429	bt0s21	800448
$W(\mu\nu)+cc+1p$	2.51 pb	ct0s6w	2025229	bt0s22	792448
$W(\mu\nu)+cc+\geq 2p$	879 fb	ct0s7w	1990504	bt0s23	788236
$W(\tau\nu)+cc+0p$	7.00 pb	et0sw0	1973192	bt0s25	788448
$W(\tau\nu)+cc+1p$	2.51 pb	et0sw1	1985097	bt0s26	796690
$W(\tau\nu)+cc+\geq 2p$	879 fb	et0sw2	1921088	bt0s27	786908
$W(e\nu)+c+0p$	23.9 pb	st0sw0	1960065	ot0swd	800448
$W(e\nu)+c+1p$	4.75 pb	st0sw1	1964891	ot0swe	800448
$W(e\nu)+c+2p$	710 fb	st0sw2	1978900	ot0swf	799773
$W(\mu\nu)+c+0p$	23.9 pb	st0sw5	1992335	ot0swh	800448
$W(\mu\nu)+c+1p$	4.75 pb	st0sw6	1984842	ot0swi	800448
$W(\mu\nu)+c+2p$	710 fb	st0sw7	1974052	ot0swj	799678
$W(\tau\nu)+c+0p$	23.9 pb	st0swa	1532572	ot0swl	800448
$W(\tau\nu)+c+1p$	4.75 pb	st0swb	1532908	ot0swm	800448
$W(\tau\nu)+c+2p$	710 fb	st0swc	1504501	ot0swn	800448
$W(\tau\nu)+c+\geq 3p$	116 fb	st0swd	1510193	ot0swo	800448

Table B.4: Alpgen $W+$ heavy quark dataset names, processes, cross sections, and number of events generated. Note that the cross-sections listed include a K-factor of 1.4.

Process	σ	low $\mathcal{L}$ Dataset	low $\mathcal{L}$ N_{events}	high $\mathcal{L}$ Dataset	high $\mathcal{L}$ N_{events}
$Z(ee) + bb + 0p$ [75 : 105]	511 fb	zt0sb0		bt0szd	1544133
$Z(ee) + bb + 1p$ [75 : 105]	134 fb	zt0sb1	536159	bt0sze	192056
$Z(ee) + bb + \geq 2p$ [75 : 105]	38.5 fb	zt0sb2	525955	bt0szf	192119
$Z(\mu\mu) + bb + 0p$ [75 : 105]	511 fb	zt0sb5	529635	bt0szg	109368
$Z(\mu\mu) + bb + 1p$ [75 : 105]	134 fb	zt0sb6	530793	bt0szh	87944
$Z(\mu\mu) + bb + \geq 2p$ [75 : 105]	38.5 fb	zt0sb7	525695	bt0szi	192119
$Z(\tau\tau) + bb + \geq 0p$ [75 : 105]	625 fb	zt0sc0	536159	bt0szj	608337
$Z(ee) + cc + 0p$ [75 : 105]	511 fb	zt0sc0	1544133	bt0szk	284167
$Z(ee) + cc + 1p$ [75 : 105]	134 fb	zt0sc1	690239	bt0szl	288167
$Z(ee) + cc + \geq 2p$ [75 : 105]	38.5 fb	zt0sc2	663518	bt0szm	288167
$Z(\mu\mu) + cc + 0p$ [75 : 105]	511 fb	zt0sc5	671375	bt0szn	288167
$Z(\mu\mu) + cc + 1p$ [75 : 105]	134 fb	zt0sc6	663431	bt0szo	288167
$Z(\mu\mu) + cc + \geq 2p$ [75 : 105]	38.5 fb	zt0sc7	648338	bt0szp	288167
$Z(\tau\tau) + cc + \geq 0p$ [75 : 105]	625 fb	zt0sct	2056891	bt0szq	800448
$DY(ee) + bb + 0p$ [20 : 75]	293 fb	xt0sb0	529488	-	-
$DY(ee) + bb + 1p$ [20 : 75]	58.5 fb	xt0sb1	534522	-	-
$DY(ee) + bb + \geq 2p$ [20 : 75]	15.9 fb	xt0sb2	529502	-	-
$DY(\mu\mu) + bb + 0p$ [20 : 75]	293 fb	xt0sb5	529304	-	-
$DY(\mu\mu) + bb + 1p$ [20 : 75]	58.5 fb	xt0sb6	367279	-	-
$DY(\mu\mu) + bb + \geq 2p$ [20 : 75]	15.8 fb	xt0sb7	528903	-	-
$DY(\tau\tau) + bb + \geq 0p$ [20 : 75]	313 fb	xt0sbt	1510091	-	-
$DY(ee) + bb + 0p$ [105 : 600]	14.4 fb	yt0s0b	513872	-	-
$DY(ee) + bb + 1p$ [105 : 600]	4.2 fb	yt0s1b	529304	-	-
$DY(ee) + bb + \geq 2p$ [105 : 600]	1.2 fb	yt0s2b	523926	-	-
$DY(\mu\mu) + bb + 0p$ [105 : 600]	14.4 fb	yt0s5b	534522	-	-
$DY(\mu\mu) + bb + 1p$ [105 : 600]	4.2 fb	yt0s6b	529385	-	-
$DY(\mu\mu) + bb + \geq 2p$ [105 : 600]	1.2 fb	yt0s7b	529458	-	-
$DY(\tau\tau) + bb + \geq 0p$ [105 : 600]	18.1 fb	yt0stb	1515347	-	-

Table B.5: Alpgen Drell-Yan + heavy quark dataset names, processes, cross sections, and number of events generated. The mass range in the process name refers to the generated mass range of the dilepton pair. Note that the cross-sections listed do not include a K-factor.

Dataset	Process	σ	N_{events}
tt0s75	$t\bar{t}$	7.0 pb	4730477
it0sww	Diboson (WW)	12.4 pb	2291309
it0swz	Diboson (WZ)	3.65 pb	2328823
it0szz	Diboson (ZZ)	3.8 pb	2319470

Table B.6: Pythia dataset names, processes, cross sections, and number of events generated. The cross-sections listed include a K-factor of 1.4.

APPENDIX C

SOFT ELECTRON LIKELIHOOD VARIABLES

Seven discriminating variables are used in the soft electron likelihood calculation: $\frac{dE}{dx}$, E_{EM}/P, $\frac{E_{Had}}{E_{EM}}$, E_{CPR}, E_{CES}/p, ΔX_{CES}, and ΔZ_{CES}. Each of these variables is described below.

$\frac{dE}{dx}$ is the energy that a particle loses as it moves through the tracking chamber. This is both lower and more momentum-dependent for non-electrons than for electrons, and is therefore 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. A comparison between electrons and non-electrons is shown in Figure C.1.

To measure the electromagnetic and hadronic energies, the calorimeter responses are summed in 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. Electrons tend to leave most of their energy in the CEM, and therefore have E_{EM}/p near 1 and very low E_{Had}/E_{EM}. However, nearby tracks (for non-isolated electron tracks) will increase E_{EM}/p significantly, and so that distribution is corrected for the track isolation. Comparisons between electrons and non-electrons in these variables are in Figure C.2.

E_{CPR} is the energy deposited by the electron candidate in the CPR. This energy is measured using the 1+2+4 pad algorithm, which sums the energy from a single pad if the track extrapolates to the center of a pad, from two adjacent pads if the track extrapolates to near the boundary of two CPR pads, or from four pads if it extrapolates to near a corner.

The E_{CPR} distribution depends on the momentum of the track, as higher-momentum electrons will tend to deposit more energy in the CPR. However, non-electron background tracks are usually minimum ionizing particles, and will therefore leave only a small signature in the CPR regardless of their momenta. To account for the fact that particles traveling at

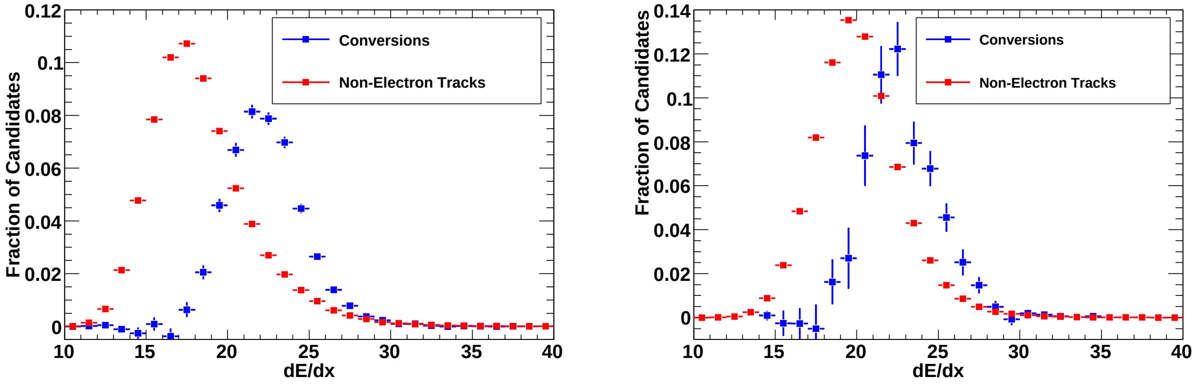

Figure C.1: A comparison of dE/dx between electrons and non-electrons. The lowest-momentum tracks, with $p_T < 1.5$ GeV, are shown on the left and the higher-momentum tracks, with $p_T > 6$ GeV, are shown on the right.

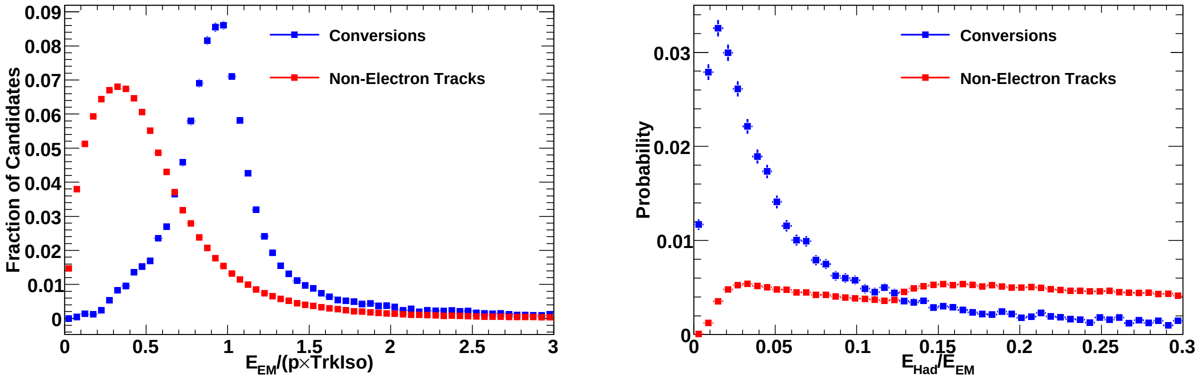

Figure C.2: A comparison of the calorimeter energy distributions ($\frac{EmE}{p \cdot iso}$ on the left $\frac{E_{Had}}{E_{EM}}$ on the right) between electrons and non-electrons.

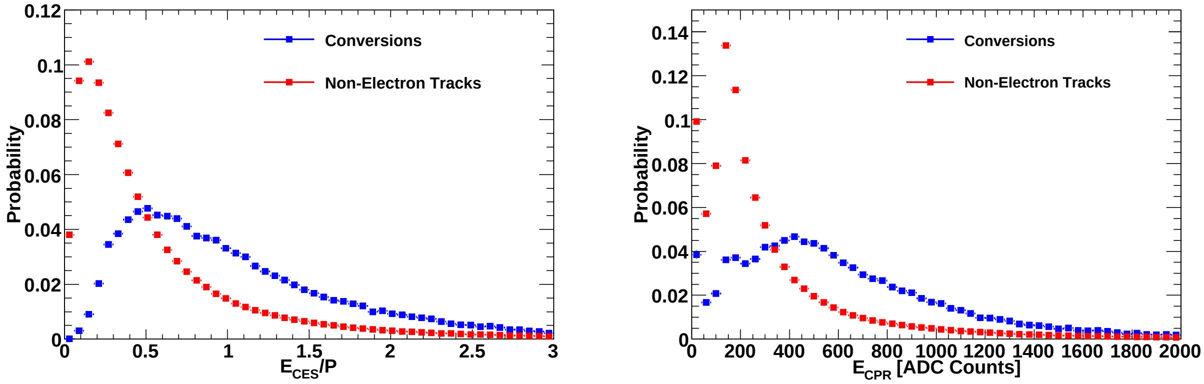

Figure C.3: A comparison of E_{CES}/p on the left and E_{CPR} on the right for electrons and non-electrons.

different angles have a different path length in the CPR, the energy is multiplied by a factor of $\sin(\theta)$.

E_{CES} is the energy of the nearest cluster in the CES to the point to which the track extrapolates. The standard CES measurement technique finds the nearest strip cluster in the x coordinate and the nearest wire cluster in the z coordinate separately. However, this analysis uses the 2D CES algorithm[40], which uses a more precise calibration to match the energies of strip and wire clusters and thereby locates the showers in two dimensions.

The E_{CES} distribution tends to be proportional to the track momentum, like the E_{EM} distribution. However, the CES has better spatial resolution than the CEM, particularly when using the 2D CES shower algorithm. Therefore, E_{CES} is a more useful discriminant for non-isolated tracks. A comparison of E_{CES} and E_{CPR} for electrons and non-electrons is shown in Figure C.3.

ΔX_{CES} and ΔZ_{CES} are the distances from the point to which the track extrapolates to the nearest cluster in the CES. The distributions for thees track-CES matching variables, depend on both momentum and angle. For electrons, the distributions are wider at lower momentum and at higher $|\eta|$. For non-electrons, the distributions are always very wide since hadrons don't shower well in the electromagnetic calorimeter. A comparison of the track-CES matching variables between electrons and non-electrons is shown in Figure C.4.

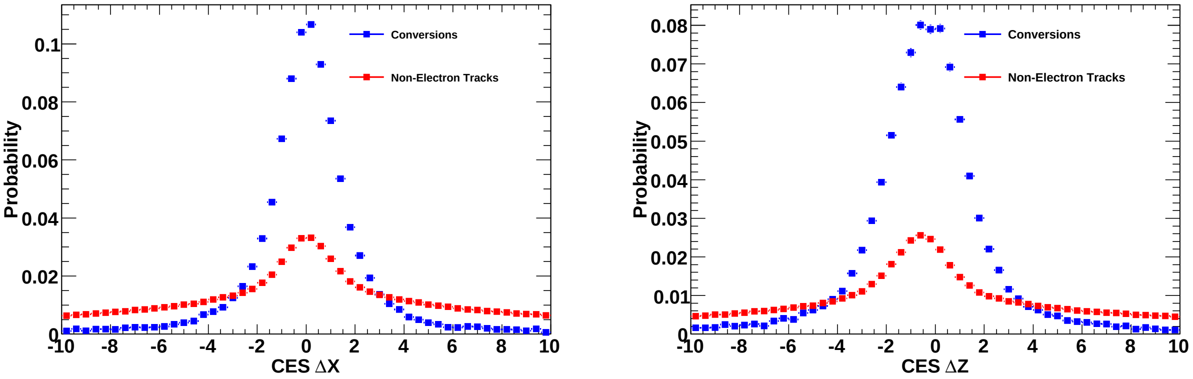

Figure C.4: A comparison of CES ΔX and ΔZ for electrons and non-electrons.

APPENDIX D

SUMMARY OF MULTI-LEPTON EVENTS

Each of the following tables is a summary of one of the multiple-lepton events found in the data. Any event that contains a W or Z boson and at least 3 additional muons is listed here. Also listed is any event that contains a W or Z boson plus at least 5 leptons if at least one of the additional leptons is a muon. These requirements are asymmetric in electrons and muons due to the conversion background swamping out parts of the multi-electron region.

Identified Object	p_T	η	ϕ
Tight μ^-	32.26	0.21	1.94
Tight μ^+	20.28	-0.14	4.52
Soft μ^-	10.28	-0.32	5.07
Soft μ^+	5.56	-0.29	5.11
$\not{E}_T$	56.63		1.02
H_T	183.14		

Table D.1: Summary of run 211441, event 7455273. This event passed the CMUP-triggered W boson selection.

Identified Object	p_T	η	ϕ
Tight e^-	99.53	-0.67	5.64
Soft μ^+	7.71	0.26	1.96
Soft e^+	7.68	0.26	1.96
Soft μ^+	7.21	0.25	1.99
Soft e^+	5.57	-0.34	3.96
Soft e^-	4.35	-0.29	4.20
$\not\!\!E_T$	52.35		2.50
H_T	252.44		

Table D.2: Summary of run 222885, event 4745161. This event passed the TCE-triggered W boson selection.

Identified Object	p_T	η	ϕ
Tight μ^+	41.75	-0.73	4.59
Soft μ^+	18.50	0.33	1.63
Soft μ^-	17.47	0.07	1.20
Soft μ^-	6.63	0.22	1.50
$\not\!\!E_T$	41.27		1.46
H_T	83.53		

Table D.3: Summary of run 228735, event 17076. This event passed the CMX-triggered W boson selection.

Identified Object	p_T	η	ϕ
Tight e^-	61.73	-0.78	5.89
Soft e^+	17.45	0.40	1.36
Soft e^+	15.08	-0.06	4.38
Soft μ^+	5.59	0.40	1.36
Soft e^+	6.82	0.44	1.39
Soft e^+	6.64	-0.09	4.34
$\not\!\!E_T$	89.60		1.66
H_T	533.87		

Table D.4: Summary of run 231693, event 2501991. This event passed the TCE-triggered W boson selection.

Identified Object	p_T	η	ϕ
Tight μ^-	31.22	0.79	3.28
Soft μ^-	5.33	-0.23	4.98
Soft e^-	3.74	0.16	2.74
Soft e^+	3.55	0.18	2.68
Soft e^+	2.74	-0.31	4.88
Soft e^+	2.53	-0.38	1.66
$\not{E}_T$	55.65		0.69
H_T	334.30		

Table D.5: Summary of run 273941, event 2792032. This event passed the CMX-triggered W boson selection.

Identified Object	p_T	η	ϕ
Tight e^+	21.44	-0.67	4.12
Soft μ^-	19.18	0.36	1.54
Soft e^+	16.06	0.66	4.33
Soft μ^-	14.64	-0.33	3.12
Soft μ^-	7.33	0.35	1.61
Soft e^+	2.92	0.26	1.51
$\not{E}_T$	136.51		1.65
H_T	463.55		

Table D.6: Summary of run 275385, event 9123310. This event passed the TCE-triggered W boson selection.