MEASUREMENT OF THE STANDARD MODEL ZZ CROSS-SECTION IN THE $ZZ \to \ell\ell\ell\ell$ CHANNEL

Tatiana Isabel Rodriguez

A DISSERTATION
in
Physics and Astronomy

Presented to the Faculties of the University of Pennsylvania
in
Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

2011

Supervisor of Dissertation

Elliot Lipeles
Assistant Professor, Physics and Astronomy

Co-supervisor of Dissertation

Nigel Lockyer
Adjunct Professor, Physics and Astronomy

Graduate Group Chairperson

Alan T. Johnson
Professor, Physics and Astronomy

Dissertation Committee

I. Joseph Kroll, Professor of Physics and Astronomy
Mark Trodden, Professor of Physics and Astronomy
Hugh H. Williams, Professor of Physics and Astronomy

Acknowledgements

I'd like to take a moment to acknowledge the help of all those people without whom I could not be handing in this doctoral thesis. I have never been a cynic, but even in my least optimistic mood I would never have predicted the longness and the slogness of the path I took to get here. There is no way I could possibly list, or even remember, all the people probably numbering into the hundreds to whom I laid bare all my grad school woes and who in return gently and kindly encouraged me to stay on the path. First and foremost on that incomplete list, however, standing out in bright relief, would be the name of Elliot Lipeles, my advisor. His kindness, intelligence, efficient pace, and overall helpfulness cannot, I repeat and underline, cannot be overstated. Thank you, Elliot, for finally guiding me to the finish line. Without your miraculous move to Penn I certainly would have left by now out of frustration.

Elliot may be the last, but he is no means the first, ray of hope that came at the last possible moment to keep me from veering off the track. There is also Joel Heinrich, who started working with me a week before I had trouble with my first

supervisor, and who eventually guided me towards working with my other, sweet, kind, co-advisor, Nigel Lockyer. Joel served as a mentor, a guide for my passion for physics and encouragement for me to pursue it. Nigel always made sure I had sight of the big picture and gave me my thesis topic, and always was on my side. After Nigel moved to Canada I had help from the kind and wonderful people at Fermilab, starting with Pasha Murat, then with Larry Nodulman from the Electroweak group, and finally Roman Lysak, Eric James and Sergo Jindariani from the $H \to WW$ group.

To the rest of my grad school committee, thank you too, for graciously hearing my defense and considering me to be among your scientific peers. To Joe Kroll especially, for being concerned with my progress and checking in with me periodically to show that someone cared.

Also, why not, I would like to thank Jorge Cham and Stephen Colbert for being my unofficial humor therapists.

Then there were all the great people who were and are my friends who I met at Penn, in Philly, at Fermilab, in Chicago, and at various conferences and independent travels around the world during the time of my graduate career. All of you helped sustain me while I went through this process and I thank you so much for it. I would in particular like to thank my parents, Katie, Josh, Teppei, Diane, Nate, Christina, Carolyn, Tamara, and Cynthia for their unfailing, unyielding support, and Chris Cox, Cesar, Anadi, Juan, Diana, and Derek for their human

understanding and encouragement to graduate. There are so many others whose love and expansion of my world I hold dear: Wes and Gina, everyone at the Metrology group at Fermi, Farrukh, Betta, Rob, my improv girls Carlie, Marissa, Lydia and Naomi, Anton, Abby, Rachel, Adrian, Kendrick, J.B., Malini, Joanna, Fen, Gosia, Dave, Almudena, Araceli, Danial, Alejandro, Alex Borisov, Karolina, Emily, Teresa, Malini, and so many others who have touched my heart and given me joy. I hope you know who you are. I hope in the future I find something wonderful to do and rely on your extended arms again.

ABSTRACT

MEASUREMENT OF THE STANDARD MODEL ZZ CROSS-SECTION IN THE $ZZ \to \ell\ell\ell\ell$ CHANNEL

Tatiana Isabel Rodriguez

Elliot Lipeles and Nigel Lockyer

In this thesis we study one of the last corners of the Standard Model to be thoroughly investigated in a hadron collider, the production of two simultaneous Z bosons. We analyze 6.1 fb^{-1} of data produced at Fermilab at a center-of-mass energy $\sqrt{s} = 1.96$ TeV and recorded by the CDF experiment. The predicted cross-section is 1.4 pb and we measured $2.18^{+0.64}_{-0.63}$ pb (stat)$\pm$0.30 (syst) using 14 observed events. This is the largest set of candidate events in this channel yet found and with our estimated signal in the sample of 9.54 events provides the smallest percentage uncertainty on the ZZ cross-section to date. We also use this large set of events to yield kinematic plots and measure ZZ properties that will be of use in probing for new physics in the future.

Contents

List of Tables

List of Figures

Introduction

I have divided this thesis into eight chapters. Chapter 1 explains in brief the Standard Model, the processes in the Standard Model underlying the process at the heart of this analysis, and the possible alternatives to that process that this analysis could probe. Chapter 2 goes through in brief the CDF architecture and how the design of that architecture contributes to finding the particles we are most concerned with here. Chapter 3 goes through lepton identification as matched to the component parts of the CDF detector. Chapter 4 describes in more detail the criteria we use to identify our four leptons and why we used them. Chapter 5 presents the signal and background gained from using these new lepton criteria. Chapter 6 presents the analysis at the high-level stage of reconstructing the ZZ from our leptons. Chapter 7 gives the cross-checks we used to make sure our new lepton types were reliable and the systematics on the analysis. Finally Chapter 8 presents the results and conclusion.

Chapter 1

Theoretical underpinnings

The branch of physics known as high-energy physics seeks, as its ultimate goal, to fundamentally understand our universe. Experimental particle physics has developed two methods of doing so: more precisely measuring the properties of particles that we already know; and searching for particles that no one has yet found. Most analyses fall in one camp or another. This analysis focuses primarily on the first, but has direct bearing on several possible searches.

1.1 The Standard Model

The Standard Model is the currently accepted framework for fundamental physics. Within its group-theoretical structure of $SU(3)_C \times SU(2)_L \times U(1)_Y$ it describes the interactions between all known fermions and bosons carrying the electromagnetic,

weak, and strong forces. Since its inception over thirty years ago it has time and again been confirmed by ever-increasingly precise experiments at ever-higher energy regimes. Its predictions have been verified to the 1% level, in some cases significantly better, and there is no clear, nondisputed measurement that differs from the theoretical value by more than 3σ significance.

However, there are several reasons why the Standard Model cannot be the complete underlying description of the universe. There are some major ones, such as gravity not being included in its scope, nor dark matter. The Standard Model has no satisfying explanation for why the neutrino masses are so very, very much tinier than the scale of the field that we believe gives them their mass, nor for its 19 or so free parameters. And to explain the masses of the gauge bosons, it relies on the existence of a Higgs particle, which has not yet been discovered. This Higgs, moreover, has quadratically divergent radiative corrections to its mass; according to its definition it must be light enough to be detected on the electroweak scale, but it must then suffer from an enormously unnatural fine-tuning of those massive radiative corrections canceling each other out. These are all problems requiring new physics answers, and hopefully we can find new data to challenge our existing understanding.

We emphasize this point about the contradictory nature of the Standard Model because of the nature of this analysis. This analysis will more precisely measure the cross-section of the ZZ as it decays to four leptons. However ZZ production is rare,

and in the four lepton channel this analysis is exploring, the resulting sample is also exceptionally pure. The original plan to search for Higgs$\rightarrow ZZ \rightarrow 4\ell$ was not completed, because the LHC sensitivity significantly exceeds what we could achieve, and they are rapidly closing in on Fermilab with respect to their luminosity. This analysis however does have a direct connection to that search because confirmation of our understanding of that process is part of a meaningful search for the Higgs.

So let us expound for a bit on the nature of these particles.

1.2 Electroweak theory

The Z boson is the neutral carrier of the weak force, but unlike the photon, it is massive. In the standard gauge description of the electroweak force, the gauge bosons cannot have explicit mass terms because then the theory becomes unrenormalizable. The Higgs mechanism was introduced to solve this problem.

1.3 The Higgs mechanism

In the unbroken electroweak symmetry, the gauge bosons obey $SU(2)$ symmetry and are massless, and the strong, weak, and electromagnetic forces have roughly equal coupling strength. The four spin-1 massless gauge bosons (W^1, W^2, W^3, B_Y) are generated as a result of the electroweak $SU(2) \times U(1)$ symmetry. However, we know that currently this symmetry is broken; the electromagnetic and weak forces

have different strengths and their respective carriers have different masses. To resolve this issue theorists introduced the concept of electroweak symmetry breaking through the Higgs mechanism.

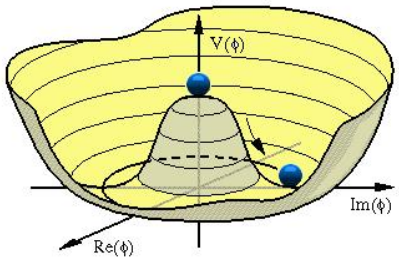


Figure 1.1: The Higgs quartic potential, and the Higgs field represented as a rolling ball.

This mechanism depends on the existence of the Higgs potential, a quartic potential shaped like a Mexican hat, at every point in space (as shown in Figure 1.1). At every point in space, a complex scalar field called the Higgs field takes a value. As the universe cooled, the field became less excited until it slipped into a minimum of the potential. Now the center value of the potential is a local maximum; if the Higgs field had rested at that point, the symmetry of the field would be preserved and the expectation value of the Higgs field in the vacuum of space would be zero. However, the true minimum lies at a lower potential, so the Higgs field must make a choice of direction and break the symmetry. Since the Higgs is a complex scalar doublet, it has four degrees of freedom; three of the four real field components become Goldstone bosons as a result of the symmetry breaking, and the fourth becomes the

Higgs boson. The Higgs field acquires a non-zero vacuum expectation value.

The electroweak gauge bosons, which couple to the Higgs, absorb the Goldstone bosons as longitudinal components of their polarization; this is equivalent to gaining mass. The W^1 and W^2 bosons become the massive W^+ and W^-, while the W^3 and B_Y rotate into each other through the weak mixing angle, and become the Z and γ, or photon. The Z acquires mass, while the photon does not. And thus the problem of massive gauge bosons has been resolved.

1.3.1 The Z boson

The Z was discovered at the UA1 and UA2 experiment at CERN in 1983 and has since been measured so precisely and understood so well that it is used for lepton calibration in new detectors. The fit accuracy can be seen for example in this $Z \rightarrow ee$ plot (Figure 1.2). Its measured production cross section is 256.0 ± 0.7(statistical error)$\pm$2.0(systematic error) pb at a center-of-mass energy of the colliding particles $\sqrt{s} = 1.96$ TeV at CDF. It decays with probability 10% to charged leptons, 20% to neutrinos, and 70% to hadronic particles, meaning almost entirely quarks.

The reason that the Z decays preferentially into quarks can partly be explained by the color factor; the Z has three more chances to decay into quarks than to leptons because there are three different colors for each quark. Furthermore, the Z couples to weak isospin and has both vector and axial-vector coupling components.

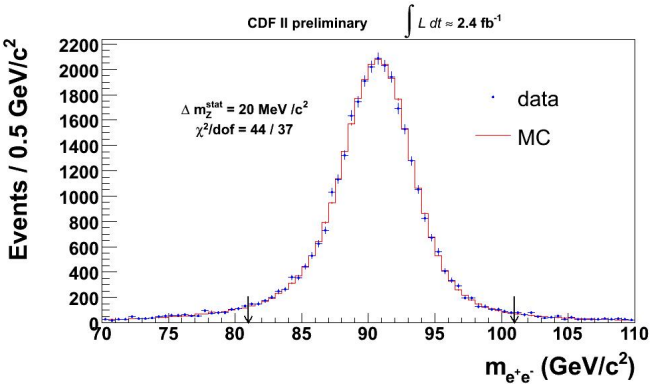


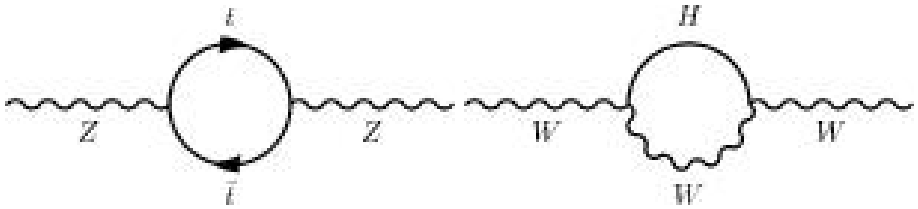
Figure 1.2: A very good fit of di-electrons to the Z mass peak.

Neutrinos have a higher weak isospin charge than leptons, and this difference acts to suppress Z decay into charged leptons by nearly a factor of 2.

The Z peak follows a Breit-Wigner distribution; the mass of any individual Z particle can be a long way from the peak. The mass of the Z at that peak has been experimentally determined to be 91.1876 ± 0.0021GeV at the LEP experiment. It is one of the free parameters of the Standard Model.

1.3.2 The Higgs boson

The Higgs has not yet been discovered, but there are many ongoing searches for it and already limits have been placed on its possible mass. The LEP collider at CERN was able to exclude the Higgs at 95% confidence level below 114 GeV through a direct search. Recently the H→WW groups at Fermilab managed to combine their results to set a Higgs mass exclusion at 95% confidence level between 160 GeV and 170 GeV.



(a) Top-loop correction to the W mass

(b) Higgs one-loop correction to the W mass

Figure 1.3: One-loop corrections used to constrain search for Higgs mass

Furthermore, by using indirect evidence from electroweak constraints, it is possible to restrict the area of search for the Higgs mass. The one-loop corrections to the W/Z and top propagators, as shown in the Feynman diagrams in Figure 1.3, affect the mass of the W and top, so knowing these masses we can set a constraint on the Higgs mass. There are perhaps a dozen other electroweak variables that constrain the Higgs, too. The current limit at 95% confidence level is $m_H < 157$ GeV . The $\Delta\chi^2$ fit of the Higgs mass based on these constraints is 87^{+35}_{-26} GeV–the combination of all these limits are in Figure 1.4. However, these calculations are very sensitive to the presence of new physics, so we cannot entirely rule out searching for the Higgs at higher masses.

Since the Higgs boson couples to mass, and the natural scale of the coupling is on the order of the vacuum expectation value of the Higgs, 246 GeV, the production cross-section of the Higgs via light quark-antiquark collisions would be far too small to detect. However, we can produce the Higgs at the Tevatron at next-to-leading order via gluon-gluon interaction; this means that our production Feynman diagram

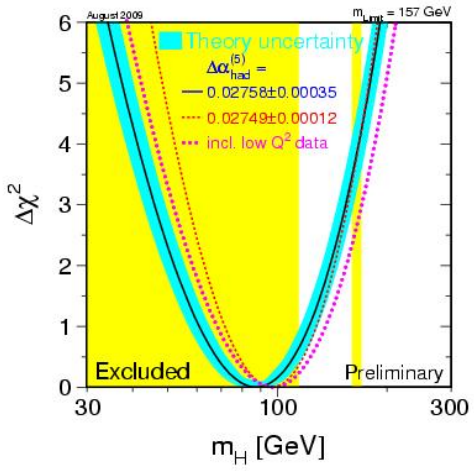


Figure 1.4: Log plot of most likely Higgs mass from electroweak constraints plus experimental constraints at 95% confidence level.

would look like the diagram at the top left of Figure 1.5. This production mechanism corresponds to the top curve in the Higgs production cross-section plot in Figure 1.6. The other Feynman diagrams correspond to the other curves in that plot, with lower production cross-section. While they do not contribute as much to the Higgs production rate, some channels such as WH associated production actually provide better detection sensitivity in the mass range under 130 GeV.

Although $H \to WW$ has gained some success as an analysis channel for having the sensitivity to exclude the Higgs in a 10 GeV area at 95% confidence, we can see clearly from Figure 1.6 that the branching fraction for $H \to ZZ$ is always smaller than for $H \to WW$. To understand how to improve the search for $H \to ZZ$, we need to understand the ZZ cross-section itself.

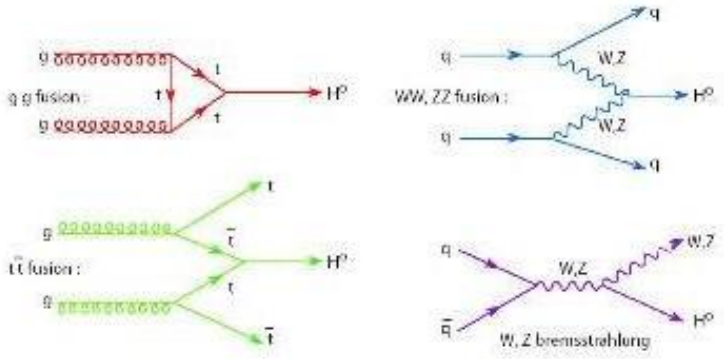


Figure 1.5: The four main production modes of the Higgs at hadronic colliders. gg is the most likely at the Tevatron and the LHC.

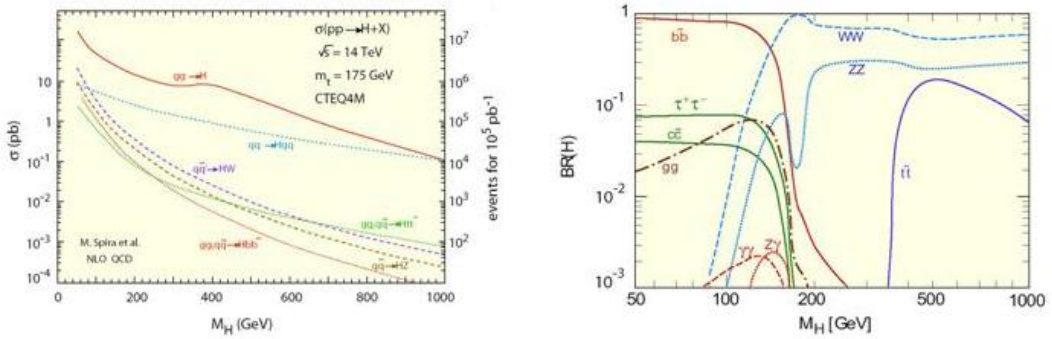


Figure 1.6: On the left, the main Higgs production cross-sections as a function of Higgs mass; on the right, the branching ratio of the Higgs to various decay modes.

1.4 ZZ in this analysis

The four lepton final state of the ZZ channel is important because it is the only ZZ final state in which there is no irreducible Standard Model background, besides the Standard Model ZZ itself. It is therefore the channel that is the most capable eventually of providing unambiguous evidence for a Higgs. There are also many other possible new physics contributions to the 4 lepton final state. For example, theorists have postulated a heavy Higgs, anomalous triple-gauge coupling, Z', or a Randall-Sundrum graviton, each with cross-section several orders of magnitude lower than the Standard Model ZZ cross-section. There exists a Littlest Higgs model in which the heavy photon (A_H) can decay to a WW pair, so pair-produced heavy quarks can lead to 4 W's.

ZZ can decay to other final states as well. The relative branching fractions between all the ZZ decay channels can be seen in Figure 1.7. The 4 quark channel is too difficult to distinguish from the normal jet background at CDF, which overwhelms every other process as can be seen in Figure 1.8. Even in the 2 quark-2 lepton channel it is very difficult to separate signal from background. The 4 neutrino channel is completely invisible, and either channel with 2 neutrinos suffers from both a difficulty in detecting neutrino multiplicity and a confusion with WW decay.

The $ZZ \rightarrow 4\ell$ channel, at 0.45%, has the smallest branching fraction of any ZZ channel. However, in a hadron collider such as Fermilab this channel has the largest signal significance of any other. The only background in this channel, given the

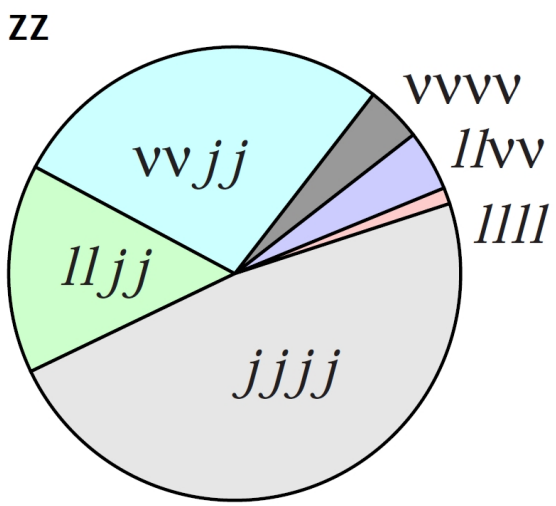


Figure 1.7: Relative branching fractions of all the decay channels from ZZ in the Standard Model. The 4 lepton channel is the smallest.

absence of any other Standard Model processes with a 4 lepton signature, is the rare, but for this analysis significant, misidentification of jets in the detector as leptons. As such the 4 lepton channel is our choice for measuring the cross-section of the ZZ.

1.5 Previous work

The ZZ was seen in several channels at the LEP collider in Switzerland before it shut down in 2001. In the $ZZ \rightarrow 4\ell$ channel only 1 event was observed. However, prior to 2008 it had never been observed at a hadron collider. Two groups from Fermilab made matching observations of the process each claiming observation, due to the tiny background involved, with only three signal candidate events in the

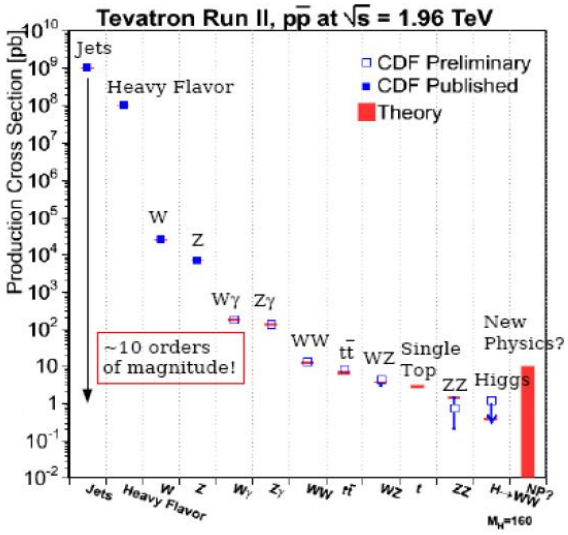


Figure 1.8: Log-plot comparison of cross-sections of all major CDF processes, including jets and ZZ.

data. These groups sought to maximize the signal over background significance of their measurements in order to cleanly claim observation of the process.

In the intervening time several more candidate events have been found by groups on both experiments at the Tevatron. Since ZZ has been observed, the next most pressing item on the list is to measure its cross-section as accurately as possible. And because it is such a rare process, as displayed in Figure 1.8, and the best way to see it is through the golden channel of $ZZ \rightarrow 4\ell$, it became clear that the best way to do so is to trade signal signficance for increased acceptance of leptons. We have done this in the present analysis, and our estimated signal in the sample is the highest to date, and the result of 14 observed candidate events in this channel is the largest sample ever found, to yield the smallest percentage error on the ZZ

cross-section measured in a single channel.

We are also grievously in debt to the original ZZ analysis done at CDF, on which this work is entirely based, as well as the $H \rightarrow WW$ group at CDF, which improves on the lepton-finding algorithms of the former. The code framework that we borrowed from these two groups was invaluable in running our analysis in a much shorter period than would have been possible otherwise. Rather than repeating their work, though, we were able to build on it and achieve a better result.

Chapter 2

Our detector

2.1 Fermilab

The detector used to conduct this analysis is one of the two detectors located on the main ring at Fermi National Accelerator Laboratory. This laboratory is located on a 6800 square acre patch of land approximately 35 miles west of Chicago.

The approximately 4-mile main ring, also known as the Tevatron, is buried about 20 feet underground. Within it are circulating 36 bunches of protons and antiprotons, controlled by superconducting magnets along the ring and traveling clockwise and anticlockwise, respectively. These bunches pass each other every 396 nanoseconds and are focused by focusing magnets on either side of the two main experiments to collide at a center-of-mass energy $\sqrt{s} = 1.96$ TeV at those points. After about 18 hours or so the rate of collisions falls, these particles are removed and a fresh batch

is injected.

2.1.1 Particle acceleration

To produce protons, hydrogen atoms from a normal hydrogen gas canister are ionized by adding an extra electron. These ions are then accelerated to 75 KeV, about the energy found in an old television tube in a structure called the Cockroft-Walton. They then pass through a linear accelerator, the Linac, to get their energy up to 400 MeV. At this point they pass through a carbon foil that strips away the electrons and permits only protons to pass. They are then accelerated further in the Booster and the Main Injector, shown in Figure 2.1, until they have sufficient energy to be injected into the Tevatron. Each of these accelerators is capable of covering only a limited range of energies.

To produce antiprotons, some energetic protons from the Main Injector are diverted to a nickel target in the Antiproton Source. Out of the multitude of particles created by the collision, antiprotons are then collected and focused into an accumulator ring. There they are cooled in order to compress their spatial resolution and stacked in bunches. Finally they are sent through the Main Injector to be accelerated and collected in the Recycler to wait for their next injection into the Tevatron, in the opposite direction as protons.

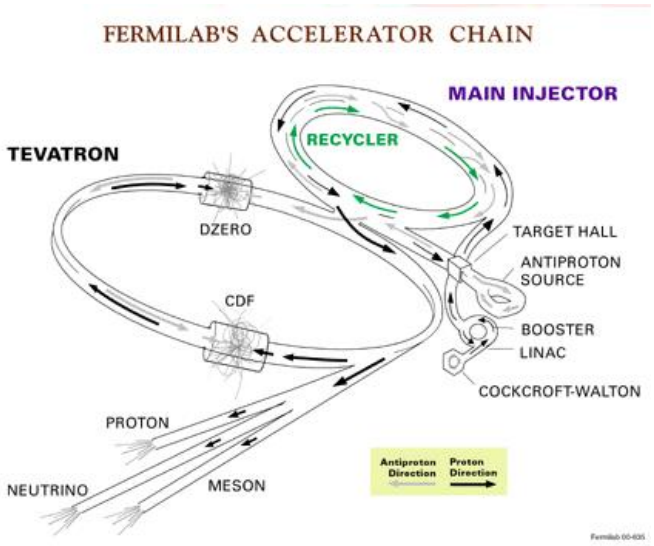


Figure 2.1: Schematic of the Fermilab accelerator chain.

2.1.2 Data collection

Besides the center-of-mass energy of 1.96 TeV, currently the second-highest in the world, the other important quantity provided our analysis by the Tevatron is the luminosity of collisions. This is a measure of the rate of proton-antiproton collision in the detector. The more highly overlapping the surface area of both bunches in the crossing, and the more frequent the bunch crossing, the higher the rate of data taking and the more likely to see interesting and rare processes over time. The amount of data produced at any given moment is called the instantaneous luminosity ($\mathcal{L}$). The total amount of data collected over the life of the detector is called the integrated luminosity ($\mathcal{L}$) and at CDF is measured in pb^{-1}. The unit of measurement, the barn (b), is equivalent to 10^{-24} cm^2, about the cross-sectional area of a uranium nucleus, and is so called because it was thought to be as large a

surface area to hit for a proton as a barn for a person. 1 pb^{-1} of integrated luminosity is 10^{-36}cm^2, and towards the end of running at CDF this amount of data can be collected in a matter of hours.

The number of events produced by a given process is related to the integrated luminosity by the total cross-section of the process, σ, by the relation

$$N = \sigma \int_T \mathcal{L} dt$$

The ZZ in our analysis is a very rare process, so the number of events seen in our analysis matters very much. So long as the measurement matches the prediction, then the higher the number of events, the better the measured ZZ cross-section. And if the two do not match by a significant enough margin, then the higher the likelihood that we have discovered some non-Standard Model process. So N is incredibly important.

2.2 CDF

CDF (Collider Detector at Fermilab) is one of the two particle detectors sitting along the ring. Many particle colliders are designed with at least two detectors, partly to compete with each other to stimulate more timely results, and partly to verify each other's results if something new and interesting is found.

The detector is also sometimes referred to as "the experiment". CDF is a very large experiment, measuring 12 meters on each side of a cube and weighing 5000 tons.

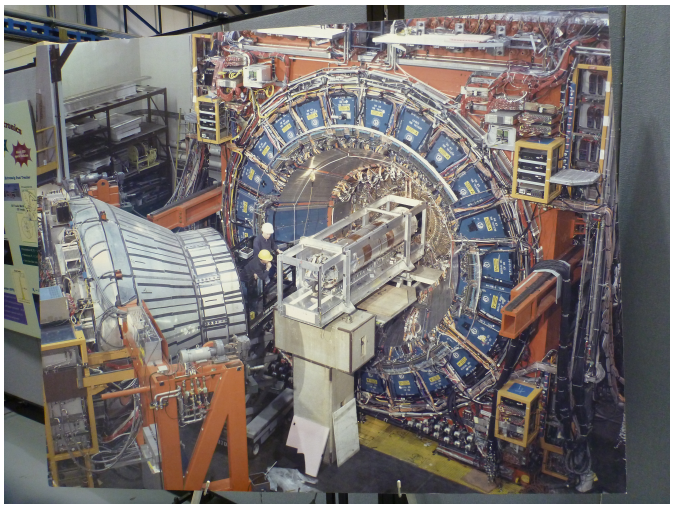


Figure 2.2: Photograph of CDF during detector installation.

It has been in operation since 1985. From 1992 to 1996 it collected 110 pb^{-1} of data during the phase of operation called Run I. Between 1996 and 2001 it underwent a massive upgrade and entered phase Run II. The data taken during Run II is used for this thesis. By the end of operations of the Tevatron this September 2011 CDF expects to collect 10 fb^{-1} of data over the course of Run II, or nearly 100 times more than collected in Run I. Along with the other detector, DØ, CDF discovered the top quark in 1995 during Run I. Since then it has observed all the Standard Model dibosons (WW,WZ,ZZ), many heavy baryons such as the Ξ_b, and singly produced top; measured CP violation in B mesons and B_s matter-antimatter oscillation; and made precision Standard Model measurements for nearly every predicted decay path. All so far have conformed to our expectation from the Standard Model, although there are a few hints of new phenomena. For instance, the measured asymmetry in $t\bar{t}$ production, meaning whether the top quark travels

along or against the direction of the incoming proton, has been found to deviate from the Standard Model prediction at a level of close to 2σ. Or, to take a more recent example, the intriguing disagreement between CDF and DØ about the bump in the $W + 2$jets mass distribution.

The detector is arranged in the onion-like design that has become standard for almost all major detectors on high-energy rings. It is essentially a cylinder that is wrapped around the collision point thickly enough to contain almost all the particles emitted from it. This allows for a wide variety of measurements to be made with the same detector.

We refer to its geometry using polar coordinates. The z-axis lies along the beampipe running through the center of the detector cylinder. The ϕ-axis surrounds the beampipe. The θ-axis rotates between the z-axis and the vertical. However, we typically do not refer to the θ location but rather to the pseudorapidity η, where

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

We use η in our coordinate description because differences in this quantity are not dependent on the z-boost of the original hard scattering. In this framework $\eta = 0 \equiv \theta = 90°$, $\eta = 1 \equiv \theta = 45°$, and at the beamline $\eta = \infty \equiv \theta = 0°$. We divide the detector into central and forward regions using $\eta = 1$. The forward region is also sometimes called the "plug" because of the design of the endcap calorimeters used to plug the forward detecting hole.

Another useful quantity that is described in CDF polar coordinates is ΔR, where

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

This quantity is often used to measure the number of tracks in a cone around a central track, or the amount of energy in a certain area around a central point. The use of η is valuable here as well because ΔR will have a circular base if η and ϕ have the same displacement, i.e. they measure about the same.

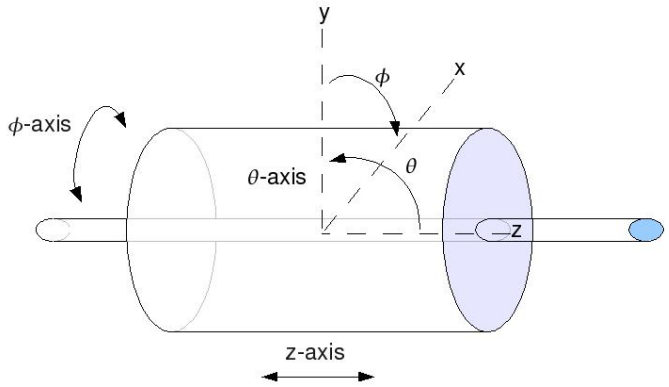


Figure 2.3: Graphical depiction of the polar coordinates of the CDF detector.

Magnets focus the proton and antiproton bunches to collide with each other at the center of the detector, and they are timed to interact exactly as they pass each other there. This way the resulting spray of particles can take advantage of the entire geometric detection of the experiment. However, the envelope of the particle bunch, while squeezed vertically and horizontally to maximize instantaneous luminosity, is spread out along approximately 60 cm in the z direction. So we cannot guarantee that the primary collision vertex is always at exactly $z = 0$ in the center of the detector, but on average this is the case.

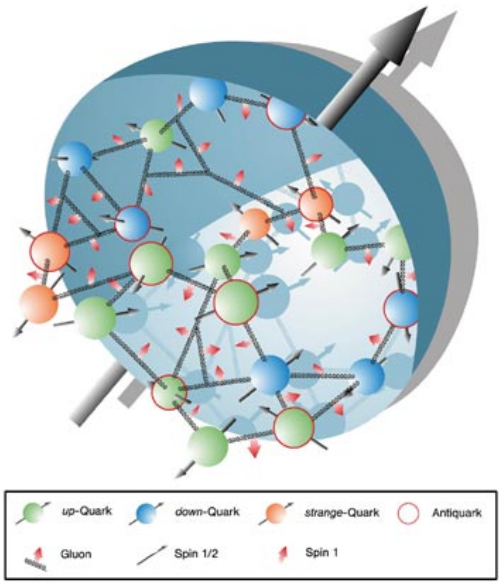
Within CDF is another magnet, a liquid-helium-cooled, superconducting solenoid. This magnet is what allows us to measure the momentum of a charged particle. The trajectory of a charged particle measured using tracking chambers, described in the next section, is bent by the magnetic field. The radius of the curvature of this trajectory can be measured and translated into the momentum of the particle transverse to the beamline via the formula

$$p_T = Bqr$$

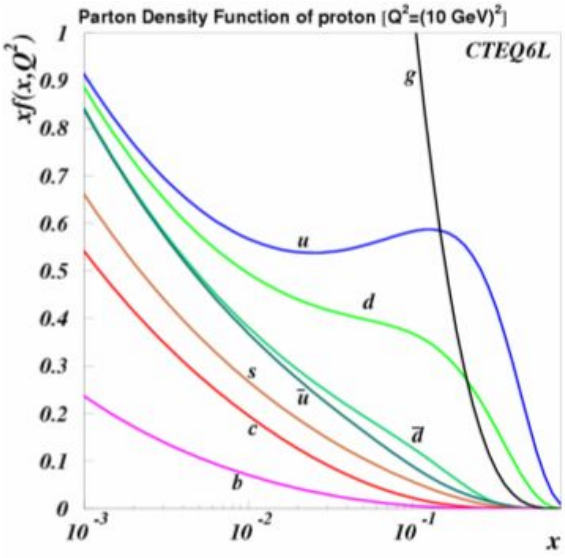
where B is the magnetic field of the solenoid, q is the charge of the particle passing through, and r is the radius of the curved trajectory. The magnetic field (of 1.5 Tesla) is approximately uniform within the solenoid. The Standard Model particles with a long enough lifetime to reach the tracking chambers will always have a charge of either 1 or -1 in units of the electron charge.

When we collide protons and antiprotons we are not actually colliding elementary particles. These hadrons are really bags of valence quarks (q), sea quarks (also q), and gluons (g). Each of these particles, or partons within the proton, carries a certain fraction of the proton's total momentum.

Most of the time when a proton and antiproton collide, which over the span of CDF Run II has been on average 2 to 6 times per bunch crossing, the quarks or gluons in each hadron experience a glancing collision, or else on average one particle carries a much larger fraction of the total momentum of the hadron than its opponent, so that the collision is boosted in the z direction. Only when there is a head-on



(a) QCD view of interior of a proton



(b) Parton distribution functions for particles inside a proton. $f(x, Q^2)$ on the y-axis is the probability that a certain parton carries a fraction of the total momentum of the proton, x, on the x-axis

collision of $g-g$, $g-q$, or $q-q$ each with high momentum fraction, then we have a chance to produce the high-mass, high-transverse energy particles that are so vital to yielding possible clues about the state of physics beyond the Standard Model. In our case we are concerned about the production of ZZ, expected to be well described theoretically, but is as of yet poorly measured experimentally, and could well be modified by interesting physics as described in Chapter 1.

Here I will list the major components of the design, and show how they are relevant to this analysis.

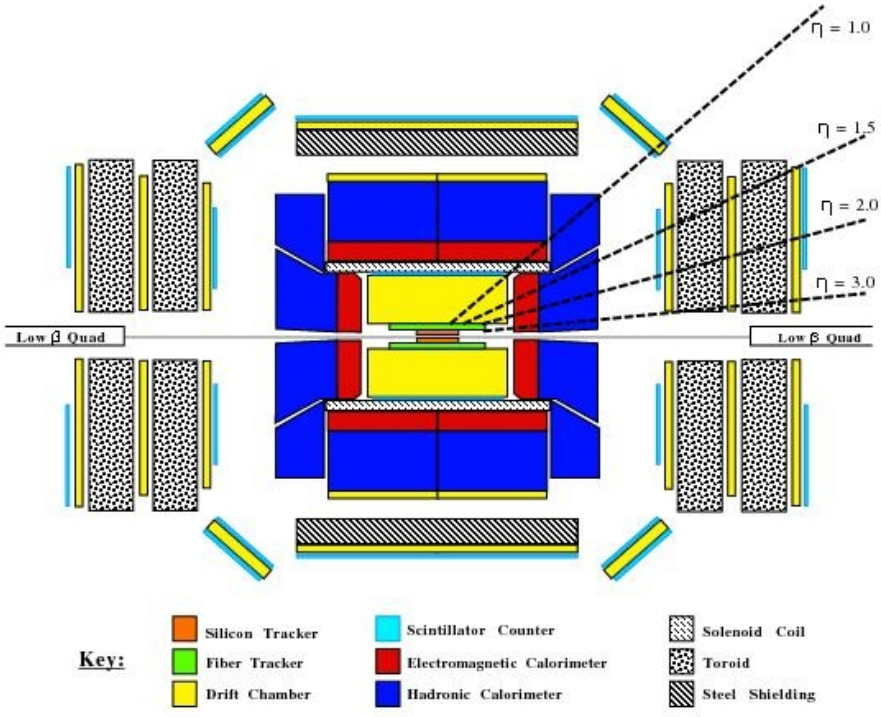


Figure 2.4: Schematic of CDF detector and components

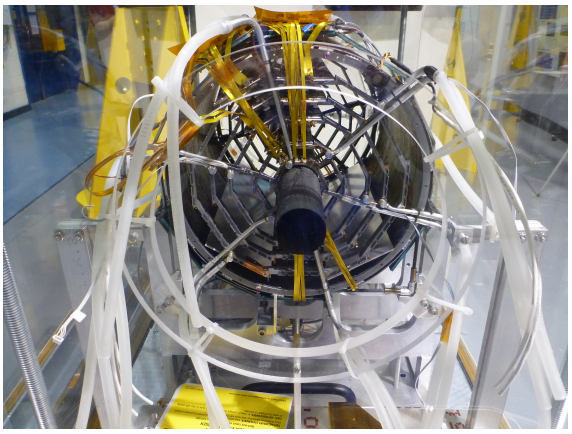
2.2.1 Silicon

We begin at the center of CDF. The tube carrying the particles to be collided is a 2.7 cm diameter beryllium pipe, chosen for being the metal with the smallest nuclear cross-section, in order to minimize particle interactions with non-detecting material close to the primary interaction point.

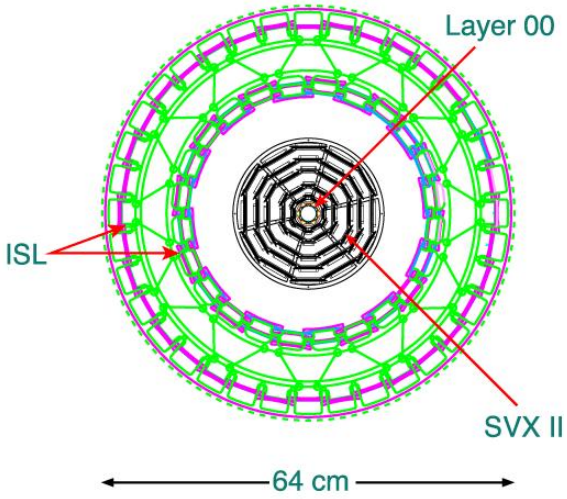
The first detecting layer is the Silicon Vertex Detector. This is a series of 6 layers of silicon strips that radiate outwards from the center in concentric dodecagons beginning from a layer sitting directly on the beampipe and ending at $r = 10$ cm. They are arranged along the z-axis in three barrels which extend for 90 cm total symmetrically around $z = 0$. The innermost layer, dubbed Layer 00 and sitting directly on the beampipe, was installed partway through Run II, and contributed greatly to the success in measuring b-quark properties such as the B_s mixing sensitivity in the recent years.

Silicon is a valuable detecting material because of its semiconductive properties. A voltage is held across each strip. When particles passing through the silicon interact with it electromagnetically, they knock out electrons in the material into the conducting band leaving positively-charged "holes". Now surrounding electrons and holes both drift in the electric field towards the edges of the silicon strip. This stimulates a pulse which is then read out as a hit. The advantage of silicon is its sensitivity to such pulses and its tiny spatial resolution of same. More about this and other detectors can be found in the PDG 2010 Review of Particle Physics,

Particle Detectors chapter [20].



(a) End view of the silicon detector. All the photographs in this section are taken from models and versions of the subdetectors from Run I displayed in the Assembly Hall in the main building of CDF.



(b) Schematic end view of the silicon detector.

There are 3 more layers of silicon in what we call the Intermediate Silicon Layer. This extends out to $r = 32$ cm and $|\eta| = 2$ and allows even greater precision tracking for particles that do not fall in the central region. Altogether the resolution on the silicon system is approximately 15μm on the impact parameter of the track and 0.3 mrad on the angle of the track with the x-axis.

2.2.2 COT

Outside the semiconductor tracking subdetector lies the next layer, the ionized gas multiwire chamber, which we call the COT, or Central Outer Tracker. This is the main subdetector that gives us the measurement of the charged particle momentum using the solenoid B-field. It is a cylinder measuring 3 meters along the beampipe and 3 meters in diameter. The COT is filled with lengthwise-strung, charged wires and 50%/50% ionized argon-ethane gas, which is occasionally circulated with oxygen. The mixture was chosen because of the high drift velocity of the electrons, about 100μm per ns, to ensure that high-transverse-momentum events occurring in successive bunch crossings would be recorded without overlap.

When a charged particle flies through this gaseous matrix, it knocks electrons off the ions in the gas and causes them to propagate in the electric field towards the wire. As these electrons drift towards the wire they knock off other electrons in the gas causing an electron shower. When the shower of electrons hits the wire, a pulse with amplitude proportional to the amount of charge propagates outwards towards the electronics on the end of the wire. If the pulse trips a certain threshold, it is recorded as a hit on the wire.

Along the radius there are 12 voltage-controlled wires in 8 layers, or superlayers as we call them. This gives us a total of 96 points radially with which we can determine to substantial accuracy the curvature of the track of the charged particle. In total the COT has 63,000 wires.

Figure 2.5: View of the COT along the beampipe. The wires within are reflecting the ambient light.

The resolution of the COT on the impact parameter is approximately 600μm, and the resolution on the angle of the track with the x-axis is 6 mrad. This is appreciably worse than that of the silicon, but since it is radially further away from the beampoint the precision is not as crucial, and as is always necessary to consider, it is much cheaper than silicon.

However since the COT is shaped cylindrically and does not extend out much further than the silicon along the z axis, the full tracking capability of the subdetector only extends to an $|\eta|$ of 1.0. After this we hit the end of the cylinder and the number of superlayers the particle passes through decreases along with its tracking accuracy, so in our analysis for such particles we see a decrease in efficiency and in momentum resolution. Fortunately for forward charged particles the silicon tracking extends out to $|\eta| = 2.0$.

2.2.3 Calorimetry

Once the charged particle leaves the tracking chamber it hits the second major component of the detector, the particle energy detector, or the calorimeter. The calorimeter extends out farther in η than the tracking chamber, allowing us to collect information from very forward particles, in other words particles with low transverse momenta, or those resulting from glancing proton-antiproton collisions.

The calorimeter is divided into 48 wedges, 24 on each side of the center, each radially about 8 meters long and spanning 15 degrees in ϕ space. Each wedge is divided into 9 separate towers, each with a roughly square radially presented surface area. There are some gaps in detection material at the cracks between towers, which are necessary for feeding out electronics and photodetector tubes. This geometry will be important when we describe our lepton identification.

There are two parts to each calorimeter tower corresponding to the two kinds of particles it can detect: electromagnetic energy from electrons and photons; and hadronic energy from hadrons and jets. The hadronic component takes up most of the tower and has radially alternating steel and photodetector layers; the EM component has radially alternating lead and photodetector.

Electrons and photons interact electromagnetically with the electrons and protons in the lead in the EM calorimeter. When attracted by the electromagnetic fields in the metal electrons also undergo bremsstrahlung and radiate some of their energy in the form of photons. As these electrons and photons propagate, they also knock off

Figure 2.6: Cross-section of single calorimeter wedge. The phototubes that collect and amplify the photons can be seen on top of each tower.

electrons from the atoms in the material, which carry another fraction of the energy of the original incoming particle. Each of these daughter electrons then knock out other electrons and creates a larger shower in the photodetector, until the average energy of each decreases and the shower stops entirely somewhere within the detector. As the high-energy electrons enter the photodetector, electrons within the atoms of the material get excited to a higher energy state. As they de-excite, photons are produced, which are then propagated on to the phototubes at the end of the calorimeter, which amplify the signal. The amplitude of the signal, and therefore the measured energy per photodetecting layer, will be proportional to the amount of light created by the shower.

In contrast, if the incoming particle is a hadron, its relatively larger mass will cause it to barrel through the electromagnetic fields of the atoms of the metal and interact only through nuclear collisions in the lead or steel, making a larger and more diffuse shower. This means the particle will penetrate much more deeply into the detector before it hits enough material to stop showering completely.

A small proportional-wire chamber, called the CES, is embedded about 6 radiation lengths deep within the EM calorimeter in the central region ($|\eta \leq 1.0|$). A radiation length is the mean distance that an electron will travel in the material before converting to $\gamma\gamma$ at 1/e, and is denoted by X_0. This quantity is proportional to the square of the charge of the electrons in the material, which is why lead is used in the EM calorimeter and is such an effective stopping material. The depth of

the central EM calorimeter is 19 X_0.

In the forward, or plug region ($1.1 \leq |\eta| \leq 3.6$), this detector is called the PES and consists of 2-D scintillating strips, and is also about 6 radiation lengths deep within the EM calorimeter. This is approximately the point of maximum electromagnetic shower, so the CES and PES can more precisely determine the position of objects passing through them, as well as distinguish between single photons, double photons from π^0 decay, and leptons. The depth of the plug EM calorimeter is 21 X_0.

The hadronic calorimeter is characterized not by the radiation length but by the interaction length (λ), which is proportional instead to the atomic number A since only nuclear interactions can stop a hadron. The central hadronic calorimeter is 4.5 interaction lengths deep. The plug hadronic calorimeter is 7λ deep. Both numbers indicate that in nearly every case all hadrons will be stopped in the calorimeter.

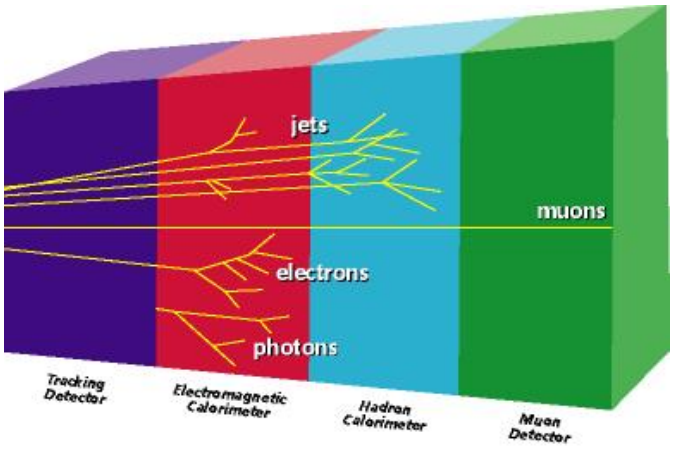


Figure 2.7: Schematic of an EM shower in the calorimeter, compared with a hadronic shower.

By looking at the amount of energy deposited in the calorimeter relative to the momentum of the particle, we can determine whether it was an electron, a photon (which, being chargeless, would leave no track), or a hadronic jet. We will elaborate on this topic in the section on particle identification.

2.2.4 Muon chambers

The muon detectors form the outermost layer of the onion, and are a bit more fragmented than the other subdetectors. There are five or six different muon chambers, some of them overlapping radially, but taken together covering the central region almost entirely and extending out to $|\eta| = 1.5$.

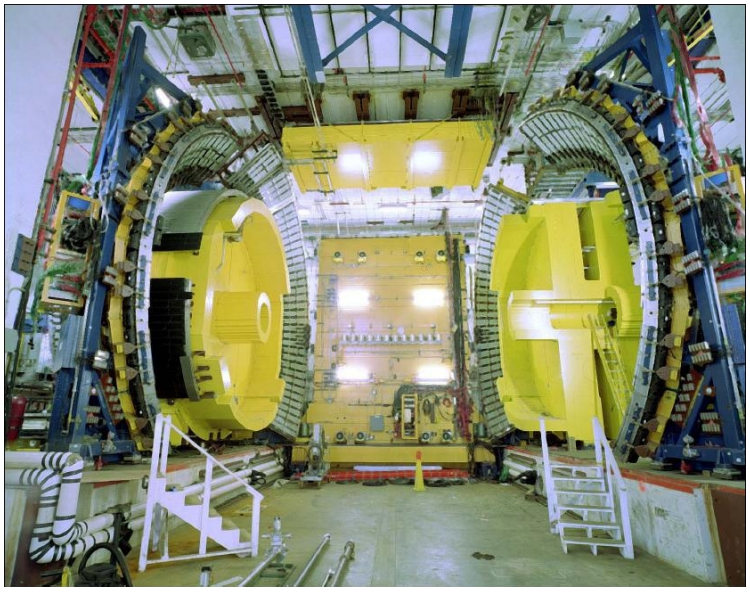


Figure 2.8: Some of the muon detectors prior to installation.

The point of these chambers are to detect muons alone. While electrons and jets

will leave tracks in the COT and be absorbed by the material in the calorimeter, muons will leave a track and fly straight through the calorimeter leaving very little energy within.

The muon chambers therefore are simply additional tracking chambers positioned radially outside the calorimeter and shielded by the calorimeter and layers of solid steel such that only muons are supposed to be recorded in them. Since they lie beyond the reach of the solenoid magnet, the hits on these detectors cannot be added to the momentum tracking from the COT and silicon. There are very few hits in these chambers anyway; the CMU and CMP–Central MUon detector and Central Muon uPgrade–have four wires each, and many muons that pass directly through do not always get every hit recorded. Any collection of two or more hits in any of these detectors is called a stub.

The CMU and CMP overlap in $\eta - \phi$ space to form the combined CMUP detector, which yields very solid muon detection. The CMX–Central Muon-eXtension–is another centrally located muon chamber which does well at detecting muons. The CMXMsKs is the Miniskirt/Keystone add-on to the CMX, at the position of underneath CDF–the miniskirt–and at the keystone of the arches halfway through Run II. Finally, the IMU or BMU–Intermediate MUon detector or Barrel MUon detector–is a forward muon detector, which means that it detects entirely past $|\eta| = 1.0$ and extends to $|\eta| = 1.5$.

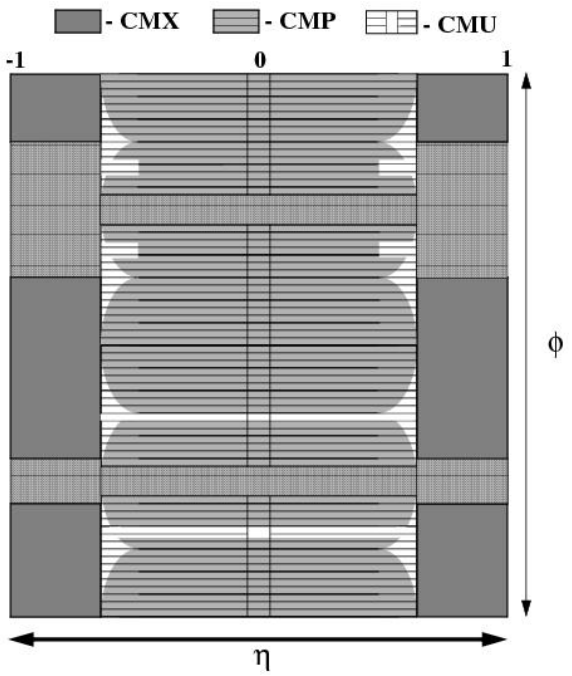


Figure 2.9: Schematic of central muon detector position in unrolled η-ϕ space.

2.3 Trigger system

Now that we know how the components of CDF fit together, we must take a second to discuss how they unite to record the particle physics data on which this and every other analysis is based.

As mentioned above there are between 2 and 6 collisions every 396 nanoseconds at CDF. Were we to record every single event, or bunch crossing, we would quickly become overwhelmed with data and run out of tape immediately. Therefore, we must filter out the unimportant, uninteresting events and try to keep only those that trip the highest energy and momentum threshholds, and represent the most energetic and transversely traveling particles from a huge collision.

In order to do that we institute three levels of triggering at CDF, each of which filters out a huge level of uninteresting data and allows us to record only petabytes, instead of hellabytes, of data.

The nominal output level at a bunch crossing rate of 396 nanoseconds is 2.5 MHz, 2.5 million events per second. The Level 1 trigger only records events that pass very basic cuts with minimum processing effort coming from the calorimeter, muon detectors and COT, with only 4 μs to make its decision. A large fraction of total events are cut out at this stage–the output rate of trigger Level 1 is 40 KHz.

Once an event passes Level 1 it is sent to one of the four storage containers in the buffer of the Level 2 trigger. Level 2 has a bit more time to accept more complex

information from the silicon and CES, as well as running more detailed algorithms such as clustering several calorimeter towers together to measure an overall jet energy. The output rate of Level 2 is only 350 Hz.

Finally an event that passes both these triggers gets read into the Level 3 trigger, which processes the entire event record using offline algorithms. The output rate of Level 3 is 100 Hz and these events are what gets stored as the data we run over to do our analysis.

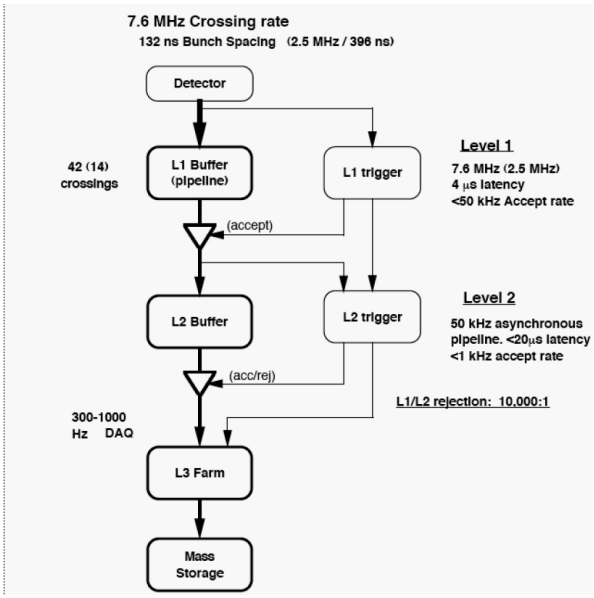


Figure 2.10: Schematic of the three-level trigger system.

Some of this data is collected through trigger paths, which are the algorithms employed at each level to search for specific objects in the event. Data collected through these trigger paths are then recorded as separate datasets on tape.

For our analysis we use only three trigger paths:

- ELECTRON_CENTRAL_18, which looks for a central electron with recorded transverse energy $\geq$ 18 GeV/c and $E_{HAD}/E_{EM} \leq 0.125$. This trigger path has a total acceptance rate at Level 3 of 1 Hz.

- MUON_CMUP_18, which looks for a muon passing through the CMU and CMP muon detectors, with transverse momentum $\geq$ 18 GeV/c spatially matched to a stub in both detectors. The accept rate for this trigger is 0.15 Hz at Level 3.

- MUON_CMX_18, which looks for a muon passing through the CMX detector with transverse momentum $\geq$ 18 GeV/c and spatially matched to a stub in the CMX. The Level 3 accept rate is 0.1 Hz.

These three trigger paths are written to tape in two different data streams: the high-p_T electron and the high-p_T muon streams. We run over datasets made from these streams to do our analysis.

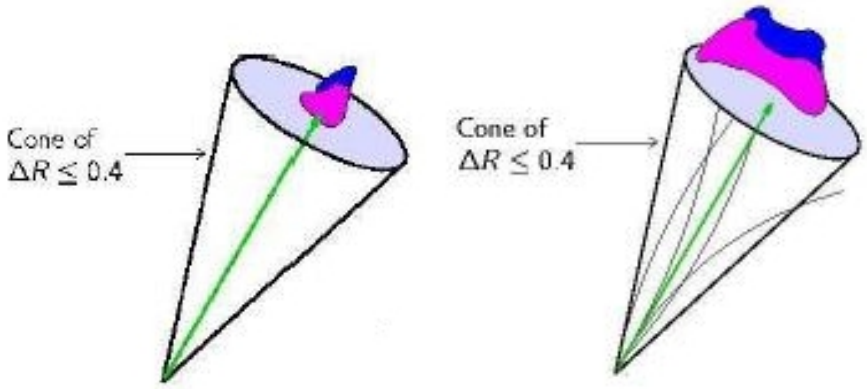
Chapter 3

General particle identification

3.1 Particle identification

While the particles that can be produced by a proton-antiproton collision include all those listed in the Standard Model (and possibly more!), the only particles that live long enough to be directly detectable fall into these five types: electrons, muons, photons, b-quark jets, and other jets. Jets are actually clusters of hadrons associated with quarks or gluons in the hard interaction. There is a sixth category of indirect detection for those particles that do not interact with anything at all, such as neutrinos and possible dark matter particles. We call combinations of these objects final states.

One very powerful means of separating leptons from jets is the isolation variable. This can be done in two ways: using track isolation and using calorimeter isolation.



(a) Isolated track with tight energy deposition

(b) Non-isolated track with wide energy deposition

To quantify the isolation of an object using track isolation we measure the momentum of all the tracks in a cone around the central track, excluding the central track. Because of the multiple charged particles that make up a jet, it will ordinarily have many tracks within a very narrow cone, which we define as

$$\Delta R_{cone} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

Typically we set this cone to be $\Delta R = 0.4$. If the ratio of the momentum in the isolation cone over the momentum of the central track is very small, the track is defined to be isolated.

To quantify the isolation of an object using the calorimeter we sum the energy in all the calorimeter towers in a cone of $\Delta R = 0.4$ and subtract the energy deposited in the cluster of towers, usually at least two and no more than three, that is associated with the lepton. If the ratio of the energy in the isolation cone to the cluster is a

small fraction, the object is defined to be isolated.

There are many other distinguishing variables we can use to identify each final state we are interested in. First I will give an overview of how we can use the detector to identify the different particles, and in the next chapter elaborate on the full lepton identification algorithm used to distinguish leptons from background by the previous CDF ZZ analysis on which we base our own analysis.

- Electrons

  We detect electrons by matching a track in the COT to a shower concentrated in the electromagnetic calorimeter. The electrons from weak boson decay are in general highly isolated. They continuously give off photon radiation due to bremsstrahlung in the detector material, but these photons will not leave tracks in the detector.

  In the EM calorimeter, electrons will deposit the vast majority of their energy in the same cluster as the energy of the collinear photons from bremsstrahlung. Therefore the energy associated with the electron tends to be larger than the momentum, which is quantified by the variable E/P. Furthermore, the shower shape of the deposited energy will have a narrower profile than that of a jet. The electron's track should point to this shower in the CES or PES.

  The electron should barely reach the hadronic calorimeter. It should never reach the muon chambers. There should be no stub in the muon chambers.

- Muons

Muons are primarily detected by matching an isolated track in the COT to corresponding hits in the muon chambers, and observing only a small amount of energy deposition in the calorimeters along the way. We call this lack of large energy in the calorimeter, along the path of a muon, minimum ionization. The minimum ionization is a characteristic of charged particles traveling through some material, in which particles with the mass of a muon and the typical momentum of a muon that is able to reach the calorimeter will experience the least stopping power from that material. This can be seen in Figure 3.1. Minimum ionization is a useful and powerful discriminant especially when the muon chambers fail to record the stub or the particle passes through an area not covered by any muon detector.

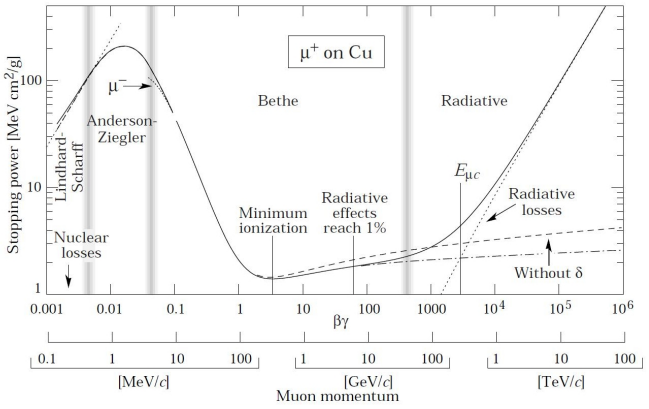


Figure 3.1: Ionization profile of a muon traveling through matter. The minimum ionization state occurs around the typical transverse momenta of muons emanting from Z decay and traveling through the detector.

The CMU muon chamber lies beyond the calorimeter, following its geometry,

so muon fiduciality to the detector can be checked by looking for muon fiduciality to the CES or PES. But because the muon tends to leave very little energy in the calorimeter and undergo very little bremsstrahlung, the E/P for the muon tends to be less than unity. This is the opposite behavior as the electron.

- Jets

  Jets are produced whenever there is QCD interaction at high energy. Since the protons and antiprotons are strongly coupled particles they are the most commonly produced objects in the detector. When a quark or gluon gets knocked out of its containing hadron, the QCD color constraint means that the particle will simply pull quark-antiquark pairs out of the vacuum until color-neutral hadrons can be produced to contain all the color charge of the original particle. All these hadrons will travel in the same direction as the initial particle and will therefore leave a series of tracks within the same ΔR_{cone}. The energy of these particles will also be mostly deposited in the hadronic calorimeter within a radius of a few calorimeter towers of the central tower hit.

  Jets may have some real leptons among the multitude of tracks that typically comprise them, for example in b- or c-quark decay to muons. Furthermore, since the hadronic calorimeters have an absorption length of 4.5-7 λ, although it is rare, it is possible for a jet to "punch-through" the calorimeter and be

recorded in the tracking chambers for the muons. Therefore we can see muon stubs fiducial to the jet direction, but they are usually an indicator of fake muons.

In general, once all other particles have been identified, the remaining object will be a jet.

- Photons

  Photons travel through the silicon and COT without leaving any track, but deposit an electromagnetic shower in the EM calorimeter similar to that of an electron. The CES or PES can normally distinguish between electron and photons. However, if a photon is so far forward that it does not travel through the COT but it does leave energy in the forward calorimeter, it is easier to confuse such a photon for a forward electron.

- Missing transverse energy ($\not\!\!E_T$)

  Neutrinos and other chargeless weakly-interacting particles can be indirectly detected by taking advantage of momentum conservation, i.e., the total initial momentum before the collision must equal the total final momentum. We measure the momentum of muons using the track curvature. We measure the momentum of electrons and photons by measuring the transverse energy deposited. Since electrons have negligible mass compared to their energy if they manage to escape the solenoid's magnetic field and reach all the way to

the calorimeter, and photons are massless, by the equation

$$\frac{E^2}{c^2} - p^2 = m^2c^2$$

and setting $m = 0$, in units where $c = 1$, their transverse energy *is* their transverse momentum.

Since there is no initial momentum transverse to the z-direction, and because of conservation of momentum, we can sum the transverse momenta of all the detected particles in the final state of the event and measure an absence of expected momentum pointing in a particular direction. If only one neutrino in an event is expected, the resulting vector of missing transverse energy, or missing E_T ($\not\!\!E_T$), will point in the direction of the neutrino. If more than one is present, then only their collective transverse momentum can be measured.

$\not\!\!E_T$has a broad resolution, however, because of its dependence on the accuracy of multiple energy measurements. We therefore expect to see some in our analysis despite not having any physics reason to be there.

Our analysis does not directly include $\not\!\!E_T$, b-quarks, photons, or jets. We only observe final state electrons and muons. However, there is a background to our analysis that is comprised of detector mismeasurement, and this background happens to include photons and jets that can fake a lepton in the detector.

3.2 Analysis backgrounds

$ZZ \rightarrow 4\ell$ is a process that has no irreducible background in the Standard Model. There are no other expected processes that have the same signature as 4 isolated leptons, with 2 opposite charge, same lepton flavor pairs that reconstruct to a Z mass.

However, there are real contributions to the background from detector mismeasurement; the dominant such process is multiple jet mismeasurement as leptons, in the channel $Z + 2$ jets. The rate of such mismeasurement is quite low, ranging from 1% to 0.01% depending on the definition of the lepton, and in the ZZ analysis must be squared to find the full mismeasurement rate. However, as was seen in Figure 1.8, the Z cross-section is 5 orders of magnitude higher than the ZZ cross-section, so at the level of our analysis this process becomes non-negligible. The only other significantly contributing background is $Z + \gamma+$ jet, where the jet is misidentified as a lepton and the photon is misidentified as an electron. $Z + \gamma+$jet is a particular problem for far forward electrons where no track is found. We consider the contribution from top pair production decaying in the fully leptonic channel to be part of this category, but it is not a significant contribution.

As such, the only background we expect is detector-based misidentification of jets and photons as leptons.

3.2.1 Fake electrons

Electrons in CDF can be faked in three distinct ways.

- A photon from FSR or from $\pi^0 \rightarrow \gamma\gamma$ decay, common in jets, may convert within the detector to an e^+e^- pair, where one electron is subsequently not found. Furthermore a photon may hit the far forward region outside the COT and leave an energy deposition in the PES shaped like an electron's.

- A charged pion can leave a track in the detector and convert to a neutral pion in the calorimeter, which then decays to two photons and leaves the EM energy in the calorimeter in the process $\pi^+ + N \rightarrow \pi^0 + N' \rightarrow \gamma\gamma N'$.

- A b-quark may decay to a real electron + other particles within the b-jet via $b \rightarrow e\nu_e X$. It is possible for this electron to deviate from the path of the jet or for the rest of the particles to leave little trace in the calorimeter.

3.2.2 Fake muons

Muons in CDF can also be faked in three ways.

- An object such as a charged pion may very rarely make it through the entire hadronic calorimeter and leave a stub in the muon chambers.

- Certain hadrons, such as K mesons or pions, may decay very early in their flight into a real muon, which then propagates through the tracking and

calorimetry.

- A b-quark again may decay to a real muon + other particles within the jet via $b \rightarrow \mu\nu_\mu X$. Again, it is possible for this muon to be detected separately from the jet.

3.2.3 Initial State and Final State Radiation (ISR/FSR)

There are two significant sources of background to $ZZ \rightarrow 4\ell$: $Z + 2$jets and $Z + \gamma$+jet.

$p\overline{p} \rightarrow Z \rightarrow \ell\ell$ is a common process known as Drell-Yan. Figure 3.2 shows the Feynman diagram of its production.

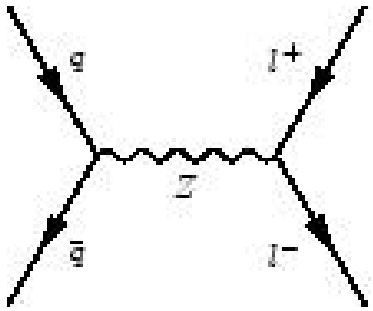
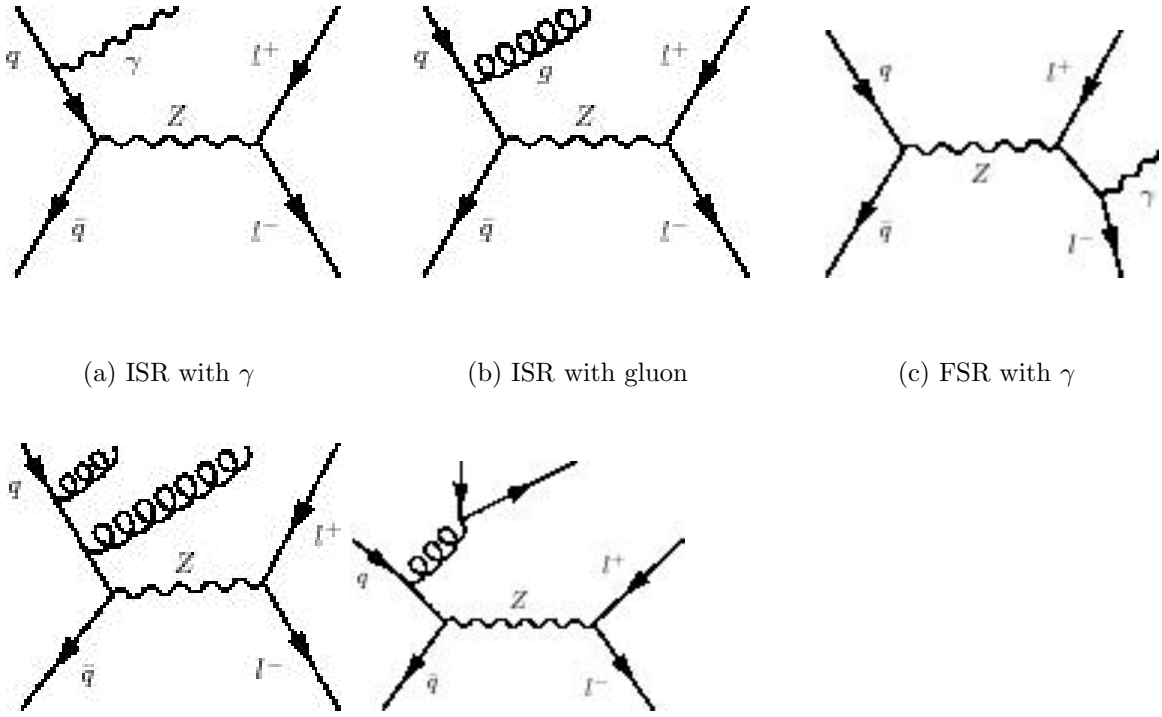


Figure 3.2: Feynman diagram of Drell-Yan process.

The incoming quark has both a strong and an electromagnetic charge, and can therefore release either a gluon or a photon as it de-excites. The gluon is emitted far more often due to its strong coupling with the quark. Figure 3.3 gives examples of both processes, and how they can produce backgrounds to this analysis.



(a) ISR with γ

(b) ISR with gluon

(c) FSR with γ

(d) One scenario for background

(e) Another scenario for background

Figure 3.3: Examples of ISR and FSR, and possible scenarios for backgrounds to the ZZ.

The outgoing lepton does not have a strong charge, so it can only release a photon as it propagates.

These extra particles released at the interaction point tend to have much greater transverse momenta than those radiated off within the detector after nuclear interaction, and therefore have the capacity to reconstruct the mass of a Z if they fake a lepton. Despite the confluence of events of two particles in the event radiating off the Drell-Yan process, both faking leptons, and reconstructing the Z mass, compared to the cross-section of $ZZ \to 4\ell$ it is non-negligible. However $Z + \gamma\gamma$ is too rare a process to be significant in our analysis.

Chapter 4

Lepton identification algorithms

4.1 Reasons for lepton ID improvement

We base our analysis off the work done by the group at CDF that combined their evidence of ZZ with the DØ result for the first observation of ZZ at a hadron collider. We also use the code framework developed by the $H \rightarrow WW$ group also based on the previous ZZ analysis. This simplified much of our work in calculating the efficiencies and fake rates of the lepton types we developed specifically for this analysis, to be described later, as well as the overall analysis structure, and we thank them heartily for it. The great contribution we made in return is redoing the lepton identification to improve acceptance.

The previous ZZ analyses shared the lepton identification criteria of the $H \rightarrow WW$ group for the common goal of maximizing signal over background for discovery. For

this analysis we instead chose to optimize our selection to measure as precise as possible a ZZ cross-section, and thus prioritized increasing acceptance over keeping the background low. We improved upon existing lepton types. We added three new lepton types entirely by modifying the types they were based upon in a smarter way. We also absorbed three extended types created by the the $H \rightarrow WW$ group.

The previous ZZ analysis found 3 candidate events with an expected background of 0.01 events in the central region and 0.2 events in the far forward region. We believe this tiny amount of background gives us some leeway in accepting more background along our increased signal acceptance, so we resolved to focus mostly on the latter instead of the former.

Since the Tevatron has since announced the 5 σ discovery of ZZ in the 4 lepton channel, and CDF and DØ have also individually discovered this signal, we felt free to optimize our analysis to most precisely measure the ZZ cross-section.

The cross-section and uncertainty of a process is

$$\sigma_{ZZ\times BR(ZZ\rightarrow 4\ell)} = \frac{N - B}{\int \mathcal{L}dt \times A \times \epsilon} \pm err(\delta_{ZZ\times BR(ZZ\rightarrow 4\ell)})$$

where N is the total number of events found at the end of the analysis, B is the expected number of background events in the end sample, $\int \mathcal{L}dt$ is the integrated luminosity over the running time of CDF, and $A \times \epsilon$ is the geometric and trigger acceptance times our lepton ID efficiency, which owing to the difficulty of separation we fold into one value.

Now we concentrate on the uncertainty, expanding and discarding quantities

contributing negligible uncertainties:

$$\frac{\delta\sigma}{\sigma} = \sqrt{\left(\frac{\delta(N-B)}{N-B}\right)^2 + \left(\frac{\delta\int\mathcal{L}dt}{\int\mathcal{L}dt}\right)^2 + \left(\frac{\delta(A\times\epsilon)}{A\times\epsilon}\right)^2}$$

The uncertainty on the integrated luminosity is 5.9%, which is constant for all analyses at CDF. Therefore the square of this uncertainty is negligible compared to the expected statistical uncertainty on the signal. Likewise, based on the previous analysis we do not expect the systematic uncertainty on our acceptance to be greater than 15%, which also gives a negligible amount when squared. Finally, we decided to not allow our background to reach any higher than around 15% of our signal in order to keep our analysis fairly clean. We also use a Gaussian approximation to estimate our cross-section uncertainty, ignoring the effect of Poisson statistics, because given the time elapsed since the last analysis and our expanded acceptance, we expect to see 3 to 4 times more events than previously. We therefore expect the absolute uncertainty on the background to be small compared with the uncertainty on the number of expected events.

When we eliminate all these terms we are left with

$$\frac{\delta\sigma}{\sigma} = \left(\frac{1}{N-B}\right)(\delta N)$$

By definition $N-B$ can be written as S, or the number of signal events in the final observed sample. Using Gaussian statistics we set the the uncertainty on N to be the square root.

$$\frac{\delta\sigma}{\sigma} = \frac{1}{S}\sqrt{N}$$

yielding, finally, the statistical uncertainty on our measured cross-section of

$$err(\delta_{ZZ}) = \frac{1}{S}\sqrt{S+B}$$

In 6.1 fb^{-1} of data we expect approximately 38 $ZZ \rightarrow 4\ell$ events to be produced. Using the efficiency of the previous analysis, which was 16%, then approximately 6 signal events and 0 background events should be observed. This would give a fractional uncertainty on the cross-section uncertainty of 41%. With an optimistic hypothetical increase of 15% acceptance per lepton and even 1000% increase in background, we can predict observing 10 events of signal and 1 event of background. This would shrink the uncertainty on our measured cross-section to 33%, a 24% gain in precision over the previous analysis if it were extended to the present amount of data.

4.2 Using probes to identify areas for improvement

We started our search for ways to increase the lepton acceptance by looking at probe leptons in the detector. These are objects that have been identified by the detector as being possible leptons but have not yet been filtered out from jets. They are the whole superset of lepton-like objects before lepton identification criteria have been applied. They pass some basic detector cuts which are outlined below, and have been discussed previously in sections “Muons” and “Electrons”. They

essentially have 100% detection efficiency.

Central Electron Probe

- A track is fiducial to the CES.
- The p_T of the track is 5 GeV/c^2 or greater.
- The closest approach of the track to the beamline is within 60 cm of the center of the detector.

Forward Electron Probe

- A shower is fiducial to the PES.
- The proportion of energy deposited in the hadronic calorimeter to the energy deposited in the electromagnetic calorimeter is 12.5%.
- The η at which the shower is recorded is between 1.2 and 2.8 in the detector.

Central Muon Probe

- A track is not fiducial to the PES.
- There are at least 20 hits out of a possible 96 in the COT.
- The closest approach of the track to the beamline is within 60 cm of the center of the detector.

Forward Muon Probe

- A track is fiducial to the PES.
- At least 60% of the possible wires on the COT have been hit by the track.
- The closest approach of the track to the beamline is within 60 cm of the center of the detector in z.

We reasoned that while the $ZZ \to \ell\ell\ell\ell$ acceptance of the previous analysis was only 16%, leaving us with a large phase space of recoverability to work with, the only truly recoverable leptons were those that (a) fell within the geometric acceptance of the CDF detector and (b) were reconstructed by the CDF software. Many leptons are not even detectable by CDF, for example because their η in the detector is too far forward and they escape down the beamline. The set of probes gives us objects that at least are detectable by CDF, allowing us to refine the identification of these objects from there.

The probes are divided into 4 different nonoverlapping categories which cover the basic divisions between the leptons and within the CDF detector. However, the previous analysis used more precise criteria for the fully identified leptons and divided them into 9 different lepton categories. We therefore measured the number of probe leptons fiducial to the specific geometric area to which each identified lepton category was fiducial, using a PYTHIA Monte Carlo simulation of ZZ for the study. We decided that the most useful category to study was where we found

Potential gain in ZZ from lepton recovery in 3l + p								
	TCE +LCE	PHX	PEM	CMUP	CMX	CMIO-CES	CMIO-PES	Crk-Trk
%ZZ gain	2.1	2.3	0.4	1.0	0.4	0.7	1.9	2.4

Table 4.1: Potential gain in ZZ from lepton recovery in 3ℓ+probe, in percent. Terms will be explained later in this chapter.

three identified leptons and a fourth probe that did not pass the identification cuts.

When we counted the number of ZZs that were recoverable in these categories, we found that some yielded a greater boost to the analysis than others. The results are in Table 4.1.

As is shown in Table 4.1, these all add up to 15% acceptance on the ZZ level, so the maximum possible ZZ acceptance with recovered leptons that we found in our studies was 31%, a possible 100% improvement over the previous analysis.

To further assess the potential gain in acceptance, we calculated how many events in our ZZ Monte Carlo with four probes ended up passing all the identification cuts. We divided the events into all four identified leptons, as well as three fully identified leptons with one probe lepton, two fully identified leptons with two probe leptons, and so on. We were surprised to find that the majority of events, about 55%, were in fact already being fully identified. Additionally about 30% were 3-lepton 1-probe, so recovery of a single probe lepton in each event would

immediately increase our total ZZ acceptance. Additionally, recovery of one probe would upgrade the 2-lepton 2-probe category to 3-lepton 1-probe, and bring that section of events closer to full recoverability as well.

In order to recover these probes we have to understand how leptons become identified. The standard method of lepton identification is to set cut threshholds on various quantities, and then eliminate all objects, however otherwise lepton-like, that do not pass those cuts. These discriminating variables are different for each category, and as set for the previous analysis follow the standard cuts at CDF for high-quality central electrons, central muons, etc. Since again we are more concerned with quantity over quality for this analysis, then to make the best measurement on the ZZ cross-section possible, we can and must open up those lepton identification cuts.

4.3 The likelihood method of lepton identification

As was mentioned above the standard method of lepton identification is to eliminate all objects that do not pass the cut threshholds on various quantities. However, if a lepton is very lepton-like but happens to fail just one cut, or fails a few cuts by just a small amount, then this might be one direction to recover lost acceptance.

The likelihood method enables us to make this recovery. We multiply the normalized distributions of all discriminating variables for a dataset of known signal

events, and then do the same for known background events. The sample of true electrons used to build the signal probability distributions was derived from a sample of Z decays in which one electron is fully identified and one is a probe. For the background sample jets opposite to a trigger jet were used. The normalized plots of the discriminating variable for signal and background events are useful because for each value of the variable along its axis, we can read the probability of an event with that variable to be either signal or background.

$$L = \frac{\prod_i P(signal)}{\prod_i P(signal) + \prod_i P(backg)}$$

For this method we assume the distributions for all these discriminating variables are uncorrelated, which is not completely true but is a good assumption. This gives us two distributions, of which one is highly signal-like and the other is highly background-like. In the resulting distribution the likelihood is the ratio of the signal-like distributions to the sum of the signal-like and background-like distributions.

This means that the points along the x-axis give the likelihood of an event falling on that line for being signal-like or background-like. Since the likelihood method carries all the information that would otherwise have been contained within our individual discriminating variables, we can therefore condense all our cuts on various lepton quantities to only one cut, on the likelihood of being signal. As is clear from Figure 4.1, this method gives a very clear and convenient separation between the two.

We select only those variables which are best able to differentiate between signal

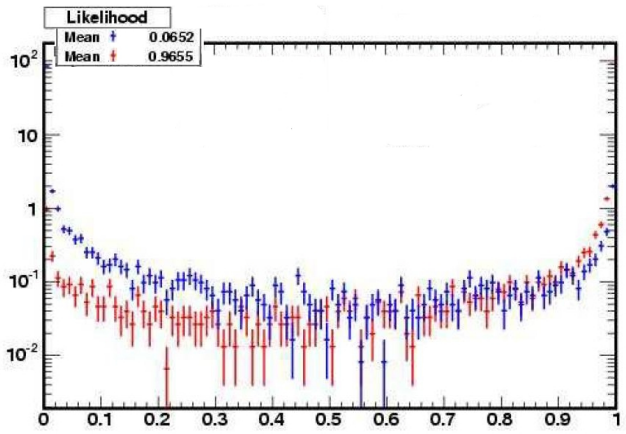


Figure 4.1: Signal and background likelihoods for LBE. The signal is in red, and dwarfs the blue points on the log scale as both approach a likelihood value of 1. Similarly, the background in blue dwarfs the red points as the likelihood approaches 0.

and background events to use in likelihood method, to optimize our separation in the likelihood.

The $H \rightarrow WW$ group at CDF has made use of this likelihood method to improve on the standard algorithm for searching for central electrons. We use their new central electron category, and we decided to apply the likelihood method to improve the forward electron categories as well. Another advantage about this method is that despite being overall more accepting, or looser, than the standard central electron, the new selection loses a negligible amount of trigger efficiency (to be explained in section 5.3). This means that it is looser than the standard cut-based algorithm, but not as loose as the selection of the trigger that selects the dataset.

4.4 Lepton identification that this analysis is based on

Since different parts of CDF detect particles differently, instead of searching for generic electrons or muons, they are divided into different nonoverlapping categories depending on to which part of the detector they are fiducial. The previous analysis had 9 of these categories: 4 electron types, 4 muon types, and one type that cannot be identified as either but must be a lepton. We refer to each lepton in the analysis by the category that it passes.

4.5 Electrons

TCE→LBE

The original tight central electron (TCE) category was defined by twelve cuts on various quantities that the electron candidate had to pass. These quantities are related broadly to the parts of the detector that deal with electron identification. These quantities are listed in table 4.2.

- Region - There are two regions of the detector, as has been pointed out elsewhere in this thesis: Central and Plug.
- Fiducial - The subdetector to which the object's track is pointing. In this case

	TCE
Region	Central
Fiducial	Track Fiducial to CES
Track p_T	>= 10(5 if $E_t < 20$)
Track $\|z_0\|$	<= 60 cm
#Ax SL (5hits)	>= 3
#St SL (5hits)	>= 2
Conversion	! = 1
Had/Em	$<= 0.055 + 0.00045E$
Iso/E_T	<= 0.1
Lshr	<= 0.2
E/P	$< 2.5 + 0.015 * E_T$
Signed CES ΔX	$-3 <= q\Delta X <= 1.5$ cm
$CES\|\Delta Z\|$	< 3 cm
Tracks	BcTrk (Larry's Correction if Data)

Table 4.2: Old cut-based selection for tight central electrons.

it is sufficiently inside the CES to get an accurate measurement.

- Track p_T - The transverse momentum of the track. For it to pass the trigger at all p_T must be ≥ 9 GeV, so this cut merely eliminates some of the trigger turn-on curve objects.

- Track $|z_0|$ - The point in the z-direction where the track makes the closest approach to the beamline. If the track is smooth it will indicate the position of the collision, i.e., the primary vertex, that produced the charged particle. This will on average be at $z = 0$ but there is a large spread due to the proton and antiproton bunch envelopes.

- Ax SL (5 hits) - This stands for Axial Superlayers, which are the half of the 8 layers in the COT that have wires in the axial direction, i.e., strictly along the z-axis. Each one is capable of recording 12 hits, so here we specify the minimum number of hits to be counted at 5.

- St SL (5 hits) - This stands for Stereo Superlayers, which are the half of the 8 layers in the COT that have wires in the stereo direction, i.e. at some small angle to the z-direction that enables the detector along with the axial information to measure the three-dimensional position of the hits.

- Conversion - Electrons can be faked by $\gamma \rightarrow e^+e^-$ conversions in the material before the calorimeter if one of the electrons has a higher p_T than the other, and the low-p_T electron is lost. The conversion-finding algorithm looks for two

tracks close together, consistent with originating from a common vertex displaced from the beam.

- Had/Em - This is the ratio of energy deposited in the hadronic calorimeter to energy deposited in the EM calorimeter. For electrons, this should be a small fraction.

- Iso/E_T - This is the isolation of the energy in the calorimeter tower in proportion to the transverse energy of the electron. The higher the transverse energy the more likely for other tracks to be collinear to the track in question, so we weight the isolation by the energy. We also use energy as a measure instead of momentum because of the bremsstrahlung difficulty with electrons.

- Lshr - the Lateral Shower Profile for electromagnetic objects. This is a measure of the width of the shower as measured in the CES, and is a distinguishing factor between electron and jet shower shapes.

- E/P - The ratio of the object's energy over its momentum. For a particle with such a negligble mass as the electron this should be equivalent to 1; however, again because of bremsstrahlung, the momentum tends to be carried off by the radiated photons that are recorded in the electron's energy cluster and therefore this ratio is usually slightly greater than 1. However a ratio far greater than 1 is usually indicative of a jet which may have many neutral particles contributing to the shower.

- Signed CES ΔX - The difference in the x-direction between the CES

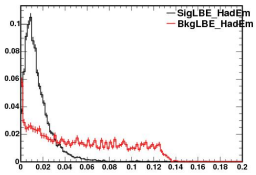
shower-maximum position and the track extrapolated to that point, multiplied by the charge of the electron. This quantity identifies electrons whose tracks deviate from their path due to bremsstrahlung. The asymmetry in the cut is used to reject jets. The x-direction is used and not η or ϕ in this case because it is the local coodinate in the CES plane.

- CES $|\Delta Z|$ - The same as above, but unsigned, and the difference is in the $z-$direction.

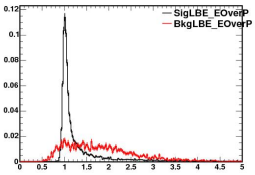
- χ^2 COT - A cut measuring the quality of the track in the COT.

For our analysis we decided to employ the likelihood method as described above and as implemented by the $H \rightarrow WW$ group to be able to accept electrons that might barely fail one or more cuts, but that overall are electron-like objects. We changed the name of this electron from TCE to LBE, or Likelihood-Based Electron. (The "central" part was left out due to its implementation for central electrons by $H \rightarrow WW$ before its usefulness was extended to forward electrons.)

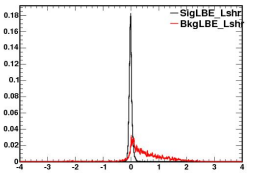
We set large cuts to weed out the least electron-like objects, shown in Table 4.3, and then incorporated 11 cuts into the likelihood, shown in Figure 4.5. We set the single likelihood cut very loosely at 0.7.



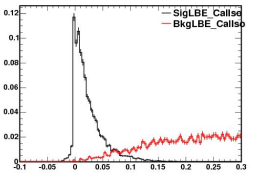
(a) LBE signal:bckgnd for Had/EM



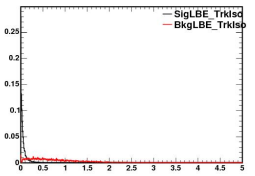
(b) LBE signal:bckgnd for E/P



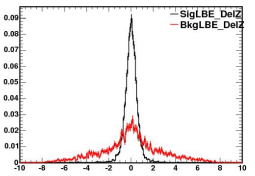
(c) LBE signal:bckgnd for Lshr



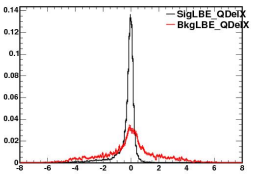
(d) LBE signal:bckgnd for calo isolation



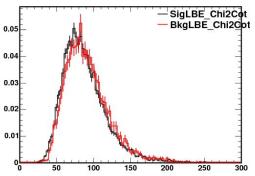
(e) LBE signal:bckgnd for track isolation



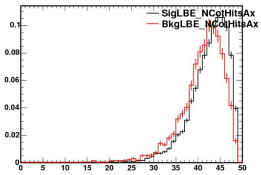
(f) LBE signal:bckgnd for CES ΔZ



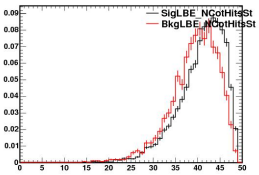
(g) LBE signal:bckgnd for CES $Q|DeltaX$



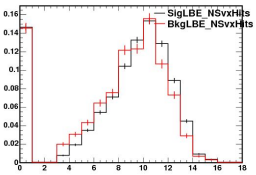
(h) LBE signal:bckgnd for COT χ^2



(i) LBE signal:bckgnd for num axial COT hits



(j) LBE signal:bckgnd for num stereo COT hits



(k) LBE signal:bckgnd for num Si hits

Figure 4.2: Signal:background distributions for all cuts used in making LBE category.

	LBE
Region	Central
Fiducial	Track Fiducial to CES
Track p_T	> 10(5 if $E_t < 20$)
Track $\|z_0\|$	< 60 cm
Had/Em	$<= 0.125$
Iso/E_T	< 0.3
Likelihood	> 0.70

Table 4.3: Preselection for central LBE electrons.

LCE→destroyed

The previous analysis used a loose central electron category to catch electrons that barely failed the cuts for TCE, as shown in Table 4.4. However this category was found to have very little yield, and therefore we chose to eliminate it and use the LBE category to recover these events along with its expanded acceptance.

PHX→PLBE

The PHX electron category covers electrons that are forward enough to strike the plug calorimeter, but not so much that they do not leave a track in the detector, mostly in the silicon strips. They are named for the Phoenix tracking algorithm, which starts outside with the shower position for these forward electrons and works

	LCE
Region	Central
Fiducial	Track Fiducial to CES
Track p_T	>= 10(5 if $E_t < 20$)
Track $\|z_0\|$	<= 60 cm
#Ax SL (5hits)	>= 3
#St SL (5hits)	>= 2
Conversion	! = 1
Had/Em	$<= 0.055 + 0.00045E$
Iso/E_T	<= 0.1
Tracks	BcTrk (Larry's Correction if Data)
Passes TCE cuts	FALSE

Table 4.4: Old cut-based selection for loose central electrons.

inward, as opposed to the standard tracking algorithm for central charged particles that starts at the outer edge of the COT and works inward.

We took the idea of the central likelihood electron from the $H \rightarrow WW$ group and extended it to the two forward electron categories. As before, we used the variables that would best differentiate between signal and background events, which mostly overlapped with the ones used for the cut-based selection. Some of these variables are the same as in the LBE; the ratio of energy in the hadronic calorimeter to the EM calorimeter, for example, will always be a good determining characteristic between electrons and jets. However, some variables are specific to the forward calorimeter. The definition is in Table 4.5 and the explanation of the variables is below.

- Region - As above, this is not a cut. We are just specifying that this electron is in the plug region of $|\eta| \geq 1.2$, not the central region.
- Pes2DEta - The detector η of the particle as measured by the PES, which is a 2D set of scintillator strips in what is locally the U and V perpendicular directions.
- PEM 3x3 Fit Tower - Whether there exists a 3x3 tower area in the PEM in which the shower shape can be fitted to the expected shape using χ^2.
- PEM $\chi^2 \times 3$ - The χ^2 of this fit
- Pes5x9U - The ratio of energy in the central five strips of the cluster to the

	PHX
Region	Plug
Pes2DEta	$1.2 < \|\eta\| < 2$
Had/Em	<= 0.05
PEM3x3FitTower	true
$PEM3x3\chi^2$	<= 10
Pes5x9U	>= 0.65
Pes5x9V	>= 0.65
Iso/Et	<= 0.1
ΔR(Pes,PEM)	<= 3.0
Track Match	True
NSiHits	>= 3
Track $\|Z0\|$	<= $60cm$

Table 4.5: Old cut-based selection for plug electrons.

total 9 strips, in the U-direction of the PES

- Pes5x9V - The ratio of energy in the central five strips of the cluster to the total 9 strips, in the V-direction of the PES

- ΔR(PES,PEM) - The difference in R (where $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$) between the shower recorded in the PEM and the position recorded in the PES.

- Corrected d0 - The closest distance between the collision point (the primary vertex) and the curved track extrapolated back to this point. If the track originated at the primary vertex, this should be very small.

- N Si Hits - The number of hits from the track recorded by the silicon strips. We would like this number to be large but we understand that it drops off with increasing η.

Again, we decided to use the likelihood method to extend the acceptance of our electrons in the forward direction. This time we used 10 variables to make the likelihood, shown in Figure 4.5, with a preselection described in Table 4.6, and set the final cut on the likelihood to be 0.7. We changed the name of our new electron type as well, from PHX to PLBE which stands for Plug Likelihood-Based Electron.

One additional update we made in this category has to do with the η of the electron. For an electron that is produced with some forward boost at $z = 0$ in the detector, the last tracking it will encounter is the silicon at $|\eta| = 2.0$. However, while studying the probe leptons in our Monte Carlo sample we noticed an additional feature in the

	PLBE/PNTLBE
Region	Plug
Pes2DEta	$1.2 < \|\eta\| < 2.8$
Had/Em	< 0.05
Iso/Et	< 0.3
PLBE only	
Track Match	True
NSiHits	$>= 3$
Track $\|Z0\|$	$< 60cm$
Likelihood	> 0.70

Table 4.6: Preselection for plug PLBE electrons.

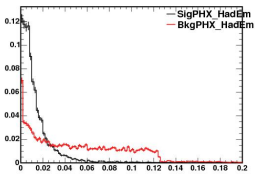
forward electron definition that we could take advantage of. If the primary vertex is significantly displaced from 0, and we allow its z-position to vary up to 60 cm in either direction along the z-axis, then the detector η of a electron with a track reconstructed by the Phoenix algorithm hitting the edge of the silicon or even COT could vary up to the edge of the forward electron calorimeter. Simply by extending the η range of the phoenix electron we were able to recover significant efficiency. See Figures 4.4,4.5 for a visual description of the problem.

PEM→PNTLBE

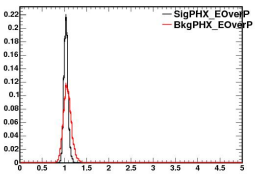
The PEM is a category that extends out further in η than the PHX because it takes advantage of the far forward calorimetry to detect electrons, without having to match to a track in the silicon or in the COT, which do not go out so far in η. It simply refers to the calorimeter in which it deposits its energy, since it is undetectable using other detector components of CDF.

Naturally, since it does not require the presence of a track, this is a category that tends to accept a lot of background from forward photons. However due to the overall 4-lepton requirement in this analysis, the total background is suppressed. The variables in the PEM are identical to those in the PHX, except all cuts having to do with track quality are of course left out. These variables are listed in Table 4.7.

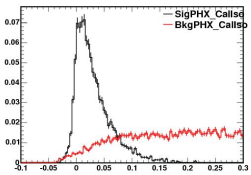
Once again we used the likelihood method to form our new category. We decided to call it PNTLBE, for Plug (No-Track) Likelihood-Based Electron. This time we did



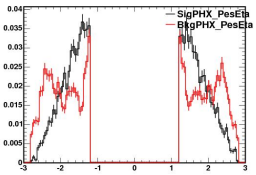
(a) PLBE signal:bckgnd for Had/EM



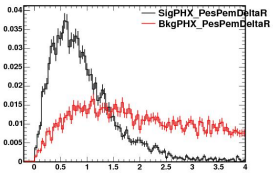
(b) PLBE signal:bckgnd for E/P



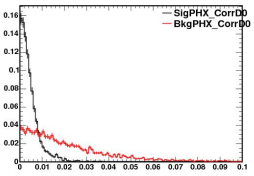
(c) PLBE signal:bckgnd for calorimeter isolation



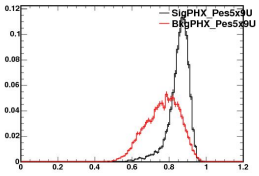
(d) PLBE signal:bckgnd for PES-measured η



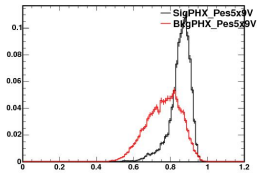
(e) PLBE signal:bckgnd for ΔR(PES,PEM)



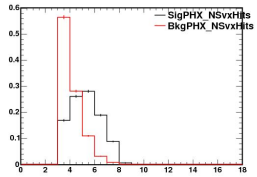
(f) PLBE signal:bckgnd for corrected d0



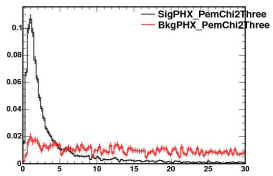
(g) PLBE signal:bckgnd for PES 5x9U



(h) PLBE signal:bckgnd for PES 5x9V



(i) PLBE signal:bckgnd for num Si hits



(j) PLBE signal:bckgnd for PEMχ^2

Figure 4.3: Signal:background distributions for all cuts used in making PLBE category

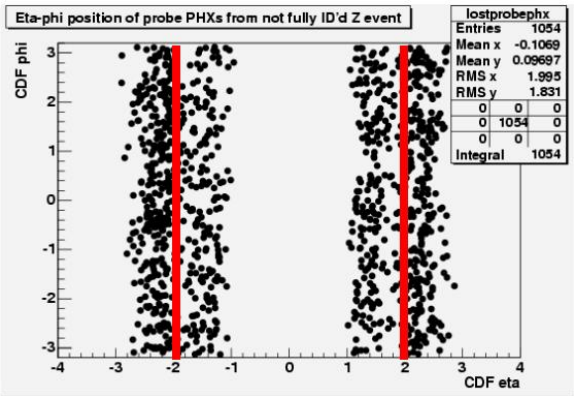


Figure 4.4: Detector position of PHX probes for almost-identified Zs in initial studies. The red line indicates the previous eta cut on PHX.

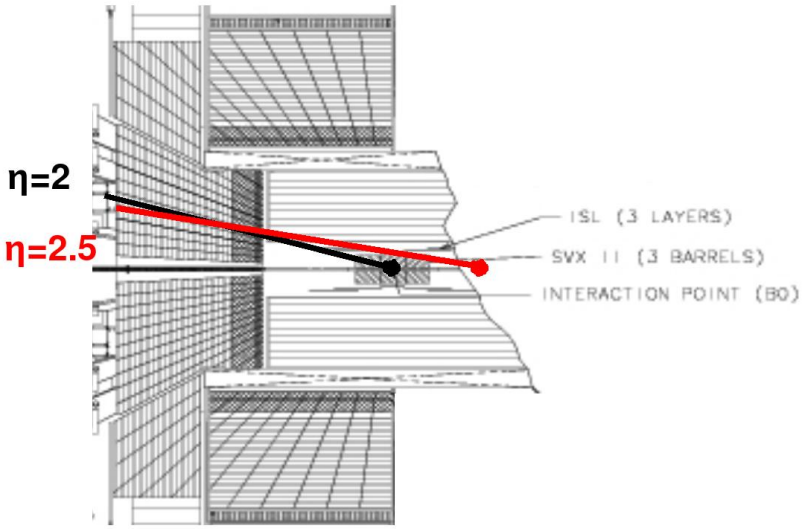


Figure 4.5: Trajectory of PHX electrons originating from z=0 and from elsewhere in the silicon barrel.

	PEM
Region	Plug
Pes2DEta	$1.2 < \|\eta\| < 2.8$
Had/Em	$<= 0.05$
PEM3x3FitTower	true
PEM3 $\times$ 3χ^2	$<= 10$
Pes5x9U	$>= 0.65$
Pes5x9V	$>= 0.65$
Iso/Et	$<= 0.1$
ΔR(Pes,PEM)	$<= 3.0$

Table 4.7: Old cut-based selection for far forward electrons without a track.

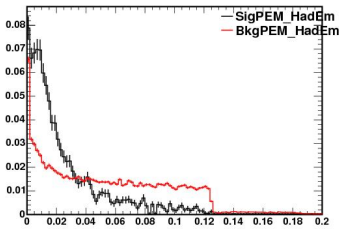
	PLBE/PNTLBE
Region	Plug
Pes2DEta	$1.2 < \|\eta\| < 2.8$
Had/Em	< 0.05
Iso/Et	< 0.3
PLBE only	
Track Match	True
NSiHits	$>= 3$
Track $\|Z0\|$	$< 60cm$
Likelihood	> 0.70

Table 4.8: New preselection for far forward PNTLBE electrons without a track.

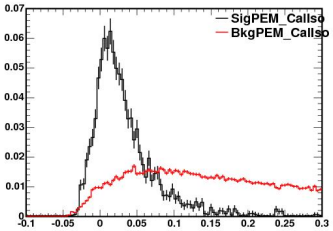
not touch the η cut for the category because we did not need to. We used only 7 variables for the PNTLBE, shown in Figure 4.5, and the same preselection as the PLBE, and again set the cut on the likelihood to 0.7.

4.6 Muons

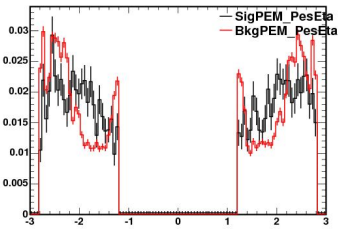
Muons are, in general, a much easier lepton to detect. They have the advantage of having their own detectors, which other charged particles can almost never reach. Furthermore, they tend to leave a very clean signal in the detector, leaving hardly any energy in the calorimeter. They are not typically produced at the interaction



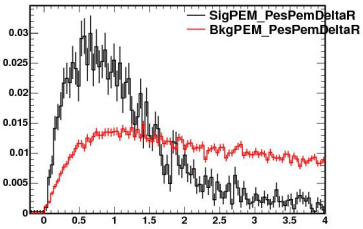
(a) PNTLBE signal:bckgnd for Had/EM



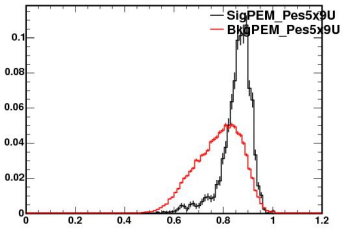
(b) PNTLBE signal:bckgnd for calorimeter isolation



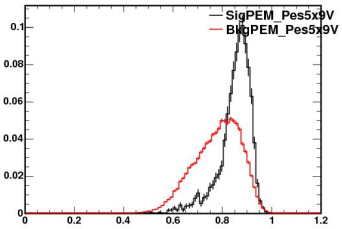
(c) PNTLBE signal:bckgnd for PES-measured η



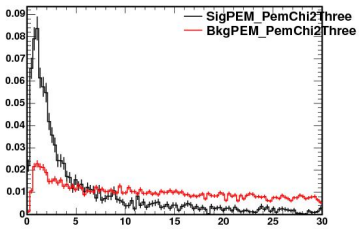
(d) PNTLBE signal:bckgnd for ΔR(PES,PEM)



(e) PNTLBE signal:bckgnd for PES 5x9U



(f) PNTLBE signal:bckgnd for PES 5x9V



(g) PNTLBE signal:bckgnd for PEMχ^2

Figure 4.6: Signal:background distributions for all cuts used in making PNTLBE category

point with jets but due to semileptonic decay are occasionally found within b- and c-quark jets.

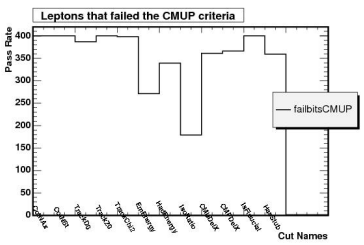
The most important cut variable for muons is the track isolation. As mentioned before, this is the amount of additional momentum carried by other tracks within a certain cone around the central muon candidate track.

Significantly loosening our lepton identification naturally means increasing our background. However, the total background on the previous analysis was only 0.01 events in the central lepton category and 0.09 events in the forward electron category. It became clear that signal improvements such as a 20% increase in acceptance even when accompanied by a ten-fold increase in background would still improve the precision in our ZZ cross-section measurement by a few percent.

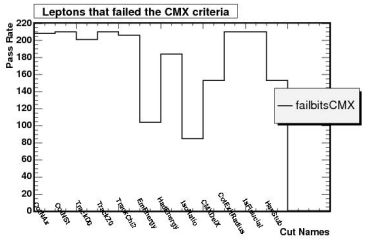
To find the areas of greatest possible increase in acceptance we began studying, for each muon probe category, which individual identification cuts cut out the most probes. The plots for each are in Figure 4.6 below.

We found that the muon categories were the most likely to benefit from loosening individual cuts, since each one only has at most three dominant lepton reducers. The electron categories tend to have more across-the-board reduction. We therefore chose to simply loosen the cuts on the dominant variable reducers in the muon categories and not to use the likelihood method to increase our muon acceptance.

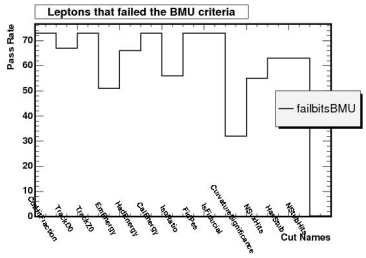
The $H \rightarrow WW$ group also recently added new muon categories to fine-tune geometric acceptance. We decided to incorporate the BMU muon category into our



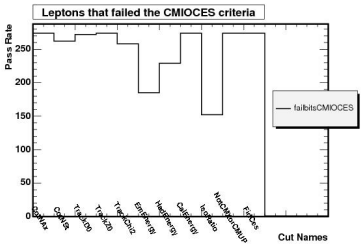
(a) Largest individual cut losses for CMUP



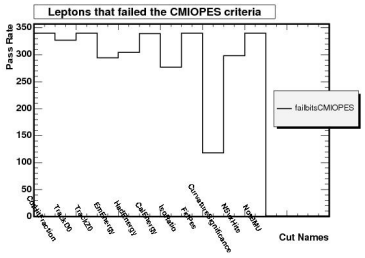
(b) Largest individual cut losses for CMX



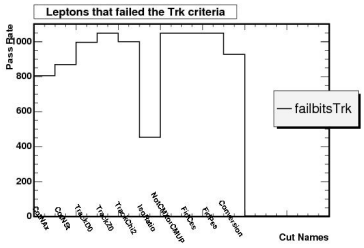
(c) Largest individual cut losses for BMU



(d) Largest individual cut losses for CMIOCES



(e) Largest individual cut losses for CMIOPES



(f) Largest individual cut losses for IsoCrkTrk

Figure 4.7: Large dips in each plot indicate a high rate of failure by fiducial probe leptons to pass the associated cuts

analysis, most of which was formerly CMIOPES. We loosened this category according to our prescription. We also accepted two more muon types from the $H \rightarrow WW$ group: CMP and CMXMsKs. We decided not to loosen these types because they are small and were really added to help the trigger for $H \rightarrow WW$. So of the now seven muon categories, four are based on previous types but with two or three loosened cuts in each. We will show in the next chapter how they increase our lepton acceptance.

CMUP→LooseCMUP,CMP

The CMUP, named after the two muon chambers that it must be fiducial to and leave stubs in, is the cleanest type we have available. Just like the TCE, it is defined by several cuts that broadly display how the components of the CDF detector relate to the identification of a muon. Most of these quantities are measurements of the quality of the track, and they are listed in Table 4.9.

- E_{em} - The amount of energy deposited in the EM calorimeter. Because the muon is a minimum ionizing particle this is not expected to be a large amount.
- E_{had} - The amount of energy deposited in the hadronic calorimeter, also typically small compared to other particles of similar transverse momentum.
- Iso/p_T - The isolation of the muon, using the ΔR cone in the calorimeter corrected for the small energy left by the muon.

	CMUP	Loose CMUP
CMU Fid	yes	
CMP Fid	yes	
E_{em}	$< 2+$	$< 3+$
	$max(0, (p-100)*0.0115)$	$max(0, (p-100)*0.0115)$
E_{had}	$< 6+$	$< 8+$
	$max(0, (p-100)*0.028)$	$max(0, (p-100)*0.028)$
Iso/Pt	< 0.1	< 0.2
NAxL(5 hits)	$>= 3$	
NStL(5 hits)	$>= 2$	
Track $\|Z_0\|$	$< 60cm$	
Track $\|D_0\|$	$< 0.2cm$ (0.02cm if NSiHit> 0)	
χ^2/dof	< 3 (4 if run $<=$ 186598)	
$\|\Delta X_{CMU}\|$	$< 7cm$	
$\|\Delta X_{CMP}\|$	$< 5cm$	
CMP veto	No Bluebeam in CMP	
	for run $<$ 154449	
Tracks	BcTrk (Larry's Correction if Data)	

Table 4.9: Selection for CMUP muons, old→new.

- NAxL(5 hits) - The number of axial superlayers of the COT in which at least 5 hits were recorded. This is a track quality measure and is identical to that used for the TCE and LBE electrons.

- NStL(5 hits) - The number of stereo superlayers of the COT in which at least 5 hits were recorded. Again, this is a track quality measure identical to that used for the TCE and LBE electrons.

- Track $|z_0|$ - The location along the z-axis where the collision point happened. It is extrapolated to the beamline from the curved track.

- Track $|d_0|$ - Just like in the electron, this is the distance between the beam axis and the point of closest approach of the track back to the beam.

- χ^2/dof - This is another track quality measurement

- $|\Delta X_{CMU}|$ - The difference in the x-direction between the track extapolated to the CMU and the stub recorded there.

- $|\Delta X_{CMP}|$ - The difference in the x-direction between the track extapolated to the CMP and the stub recorded there.

- BcTrk, Larry's correction - Beam-corrected track using the algorithm of the current convenor of the Electroweak working group at CDF.

The CMP-only category was added by the $H \rightarrow WW$ group as a means of recuperating those muons that did not leave a signal in the CMU, but otherwise is a

good muon. $H \rightarrow WW$ was able to create a trigger for this category, which we do not use. The trigger for the CMU-only category was not implemented till very recently so we do not include that in our analysis. The CMP is not a very well-populated category, and if eliminated these muons would be recovered by the CMIOCES (to be explained shortly). However the addition of the data from the CMP detector enables us to reduce the background on muons in this category. The selection for this is listed in Table 4.10.

CMX→LooseCMX,CMXMsKs

The CMX muons are, again, named for the part of CDF through which they pass and are recorded. The additional variables needed for this category are below, and the rest of the criteria are in Table 4.11.

- $|\Delta X_{CMX}|$ - The difference in the x-direction between the track extapolated to the CMX and the stub recorded there.
- ρ_{exit} - The radius at which the track exits the COT, in effect requiring that the track passes through all the superlayers of the COT.

The CMX Miniskirt and Keystone detectors were actually added to CDF part of the way through Run II, and the software to recognize muons that pass through these chambers was not implemented until about 3 fb^{-1} of data was taken. Previously, these lepton types were included in the CMIOCES category, and including them in

	CMP
CMU Fid	no
CMP Fid	yes
CMPBluebeam Fid	yes
E_{em}	$< 2 + max(0, (p - 100) * 0.0115)$
E_{had}	$< 6 + max(0, (p - 100) * 0.028)$
Iso/Pt	< 0.1
NAxL(5 hits)	$>= 3$
NStL(5 hits)	$>= 2$
Track $\|Z_0\|$	$< 60cm$
Track $\|D_0\|$	$< 0.2cm$ (0.02cm if NSiHit> 0)
χ^2/dof	< 3 (4 if run <= 186598)
$\|\Delta X_{CMP}\|$	$< 6cm$
Phi gap	within 2 degrees of gap
Active	run number >= 229764

Table 4.10: Selection for CMP muons.

	CMX	Loose CMX
CMX Fid	yes	
E_{em}	$< 2+$	$< 3+$
	$max(0, (p-100)*0.0115)$	$max(0, (p-100)*0.0115)$
E_{had}	$< 6 + max(0, (p-100)*0.028)$	
Iso/Pt	< 0.1	< 0.2
NAxL(5 hits)	$>= 3$	
NStL(5 hits)	$>= 2$	
Track $\|Z_0\|$	$< 60cm$	
Track $\|D_0\|$	$< 0.2cm$ (0.02cm if NSiHit> 0)	
χ^2/dof	< 3 (4 if run <= 186598)	
$\|\Delta X_{CMX}\|$	Max$(6.0, 125/\mu p_T)$	$< 10cm$
ρ_{exit}	> 140 cm if CMX	
CMX veto	No CMX for run < 150144,	
	No Miniskirt, No Keystone	
Arches	Arches only for all run range	
	Arches removing wedge 14 on	
	West Side for run > 190697	
Tracks	BcTrk (Larry's Correction if Data)	

Table 4.11: Selection for CMX muons, old→new.

their own categories mostly jsut reduces the background from insufficient detection. Therefore, although they do not cover an insignificant surface area of CDF, they do not provide a huge additional yield in our analysis. See Table 4.12 for cuts.

CMIOCES→LooseCMIOCES

Central Minimum IOnizing muons fiducial to the CES. This category encompasses all central muons that are not fiducial to any muon detector, but have an isolated track pointing at a lack of energy deposition in the calorimeter.

All the differentiating variables used in this selection (Table 4.13) have been described in earlier muon categories.

CMIOPES→LooseCMIOPES,BMU

Minimum ionizing muons fiducial to the PES. This category, similarly to the previous, encompasses all forward muons that are not fiducial to any muon detector, which is basically just the BMU in this region, but have an isolated track pointing at a lack of energy deposition in the calorimeter. Variables below and in Table 4.14.

- NSvxHits - The number of hits recorded in the silicon strips.
- Curvature significance - The radius of the track curvature divided by its uncertainty.
- Cot Hit Fraction - The fraction of the wires in the COT that show a hit on

	CMXMsKs
CMX Fid	yes
CMX Arches	no
CMX Miniskirt	yes
	OR
CMX Keystone	yes
E_{em}	$< 2 + max(0, (p - 100) * 0.0115)$
E_{had}	$< 6 + max(0, (p - 100) * 0.028)$
Iso/Pt	< 0.1
NAxL(5 hits)	$>= 3$
NStL(5 hits)	$>= 2$
Track $\|Z_0\|$	$< 60cm$
Track $\|D_0\|$	$< 0.2cm$ (0.02cm if NSiHit> 0)
χ^2/dof	< 3 (4 if run <= 186598)
$\|\Delta X_{CMX}\|$	$< 6cm$
ρ_{exit}	> 140 cm if CMX
Active	run >= 227704

Table 4.12: Selection for CMXMsKs muons.

	CMIOCES	Loose CMIOCES
E_{em}	$< 2+$	$< 3+$
	$max(0, (p-100) * 0.0115)$	$max(0, (p-100) * 0.0115)$
E_{had}	$< 6+$	$< 8+$
	$max(0, (p-100) * 0.028)$	$max(0, (p-100) * 0.028)$
Iso/Pt	< 0.1	< 0.2
Uniqueness	Not a CMUP or CMX muon	
Track $\|Z_0\|$	$< 60cm$	
Track $\|D_0\|$	$< 0.2cm$ (0.02cm if NSiHit> 0)	
$E_{em} + E_{had}$	$> 0.1GeV$	
Central	Track CES Fiducial	
NAxL(5 hits)	$>= 3$	
NStL(5 hits)	$>= 3$	
χ^2/dof	< 3	
	BcTrk (Larry Correction if Data)	

Table 4.13: Selection for central minimum ionizing muons, old → new.

	CMIOPES	Loose CMIOPES
Forward	Track PES Fiducial	
E_{em}	$< 1 + max(0, (p - 100) * 0.0115)$	
E_{had}	$< 6 + max(0, (p - 100) * 0.028)$	
Iso/Pt	< 0.1	< 0.2
Uniqueness	Not a BMU muon	
Track $\|Z_0\|$	$< 60cm$	
Track $\|D_0\|$	$< 0.2cm$ (0.02cm if NSiHit> 0)	
NSvxHits	$>= 3$	
$E_{em} + E_{had}$	$> 0.1GeV$	
Curvature significance	> 12.0	> 2.0
	BcTrk (Larry Correction if Data)	
Cot Hit Fraction	> 0.6	
	No beam constraint on IO tracks	

Table 4.14: Selection for forward minimum ionizing muons, old → new.

the track.

- $E_{em} + E_{had}$ - The total amount of energy found in both calorimeters.

BMU muons are fiducial to the BMU detector. See Table 4.15.

4.7 Either lepton

Finally we turn our attention to the last category, called CrkTrk in the previous analysis. This category consists of all leptons that leave a track in the COT, but those tracks point to the cracks in electromagnetic and hadronic calorimeters through which the phototubes are routed. They are also defined to not be fiducial to the muon detectors. Therefore, it is not possible to distinguish whether any particular CrkTrk lepton is a muon or an electron.

CrkTrk→LooseIsoCrkTrk

Since the only information we have to go on for the CrkTrk category is the track, it is necessary to make sure the track passes high-quality tracking cuts, such as isolation, listed in Table 4.16.

However there is an issue that has to do with the clustering algorithm at CDF. Normally, when an EM particle goes through the calorimeter it deposits most of its energy in one cone and a small amount in surrounding cones. But because between each tower in the ϕ direction there is a gap for the phototube routing, the clustering

	BMU	Loose BMU
Forward	Track PES Fiducial	
Stub	Has BMU Stub	
	(not before run 162312)	
E_{em}	$<= 2+$	$<= 3+$
	$max(0, (p - 100)$	$max(0, (p - 100)$
	$*0.0115)$	$*0.0115)$
E_{had}	$<= 6 + max(0, (p - 100) * 0.028)$	
Iso/Pt	< 0.1	< 0.2
Track $\|Z_0\|$	$<= 60cm$	
Track $\|D_0\|$	$<= 0.2cm$ (0.02cm if NSiHit> 0)	
NSvxHits	$>= 3$	
$E_{em} + E_{had}$	$< 0.1GeV$	
Curvature significance	> 12.0	> 2.0
	BcTrk (Larry Correction if Data)	
Cot Hit Fraction	> 0.6	
	No beam constraint on IO tracks	

Table 4.15: Selection for BMU muons.

algorithm does not consider two towers across a ϕ gap to be in the same cluster, or in other words originate from the same lepton. Normally this is a very safe assumption.

We realized, however, that the strict isolation criterion on this category using a cone of $\Delta R = 0.1$ could be self-defeating if the CrkTrk lepton were an electron that happened to laterally deposit sufficient electromagnetic energy in the calorimeter to form a cluster on either side of the detector crack in phi.

Since the CDF energy cluster definition always counts energy deposited on opposite sides of a crack as separate objects, the isolation criterion will cause this single electron depositing two very proximate clusters to self-veto. For a visual explanation of this see Figure 4.8.

Instead we wrote a new isolation definition to apply to the CrkTrk category only that would accept leptons with energy deposition spanning four calorimeter towers across a ϕ crack in a square nearest to the tower fiducial to the associated track. We applied this new isolation definition to the CrkTrk leptons to recover some of the electrons that would otherwise have been self-vetoed, and moreover loosened other selection criteria according to the probe plot in Figure 4.6 to gain as much acceptance as possible. The efficiency gained from the new isolation definition alone was not as great as we expected, hovering about 3%. The new lepton type we renamed IsoCrktrk.

	CrkTrk	IsoCrkTrk
Iso/Pt	< 0.1 *or*	< 0.125 *or*
	trkIso < 0.1	< 0.125 using new cal iso
		and trkIso < 0.125
Track $\|z_0\|$	$<= 60cm$	
Track $\|d_0\|$	$<= 0.2cm$ (0.02cm if NSiHit> 0)	
χ^2/dof	< 3	
NAxL(5 hits)	≥ 3	≥ 2
NStL(5 hits)	≥ 3	≥ 2
Uniqueness	Not a CMUP or CMX muon	
Is in Crack	Not Track CES or PES Fiducial	
Conversion	$! = 1$	
	BcTrk (Larry Correction if Data)	

Table 4.16: Selection for crack track leptons → IsoCrkTrk.

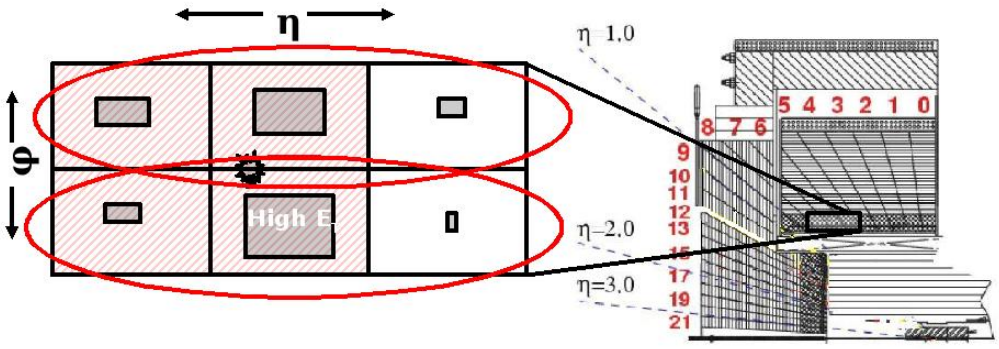


Figure 4.8: A cartoon of an electron striking in between two calorimeter towers in ϕ. Each black box is one tower. The ellipses denote the three-tower clusters created by the electron clustering algorithm. The size of the boxes represent the amount of E_T deposition within each tower. The pale 4-tower square is the new cluster that we have created in IsoCrkTrk to accept such events which would otherwise fail isolation.

4.8 Summary

We therefore have 11 lepton types in total: four which are loosened versions of the lepton types of the previous analysis (muons), four which are likelihood-altered versions of corresponding previous lepton types (electrons and IsoCrkTrk), one which is removed entirely due to negligible contribution (LCE), and three (CMP,CMXMsKs,BMU) which are new types added with lower backgrounds.

4.9 Data sets and Monte Carlo sets

The datasets we used for our final analysis cover approximately 6.1 fb^{-1} of luminosity. They use the following three trigger paths:

- ELECTRON CENTRAL 18
- MUON CMUP18
- MUON CMX18

where the 18 refers to the minimum p_T of the track to pass the trigger. These are the high-p_T central electron and high-p_T central muon trigger paths. Our analysis, like the previous ZZ search, has divided up lepton types into categories mirroring the CDF geometry. This means for our analysis that we have three triggerable lepton types–the LBE, the CMUP, and the CMX.

We decided to perform a blind analysis. This means that we develop and refine our selection entirely by using simulation and non-signal data. It also means that we are free to tweak our lepton identification criteria without biasing our result towards finding particular ZZ events in the data, which could throw off the final cross-section measurement.

The non-signal data samples we use are the 4 jet datasets that correspond to the 4 trigger paths that have been established at CDF. These have a minimum p_T for the triggering jet of 20, 50, 70, 100 GeV/c^2. There is some discrepancy in the

measurement of all four datasets, so we must use all four and add the uncertainty in the measurement between them as a systematic to our analysis.

In order to compare our measured results against our expected results, to see if there is any deviation from the theory, we must simulate those results. All the simulation samples we use are based on PYTHIA Monte Carlo, which is a Monte Carlo developed by theorists that includes many leading-order processes in the Standard Model that are producible by proton-antiproton hard scatter. A Monte Carlo process is essentially a random generator that performs numerical integrals of some theory that are too difficult to calculate analytically. PYTHIA is also capable of modeling the QCD fragmentation and showering that produces jets through quark hard scatter. However a PYTHIA-simulated sample of $Z+$ jets, for example, is not quite trustworthy enough to use on its own as part of a background measurement.

The samples that are processed by PYTHIA are then run through the GEANT CDF simulator to create simulated events that can be compared with real data events.

For estimating the expected signal yield, we use PYTHIA-created ZZ Monte Carlo filtered for the presence of at least two leptons with $p_T \geq 1$ GeV/c^2 in order to only process the data we need. For measuring efficiencies, we use the $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ sets. For measuring the contribution of real electroweak signal to the fake background, we use luminosity-scaled $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ sets. And for the final analysis cross-checks we use $W \rightarrow \ell\nu + \gamma$.

One final technical note: since the fraction of events in the datasets that pass our four-lepton requirement is quite small, and the number of variables we are considering for our analysis relatively limited, we skim the information in the large datasets into a format we call Dbntuples for greater processing efficiency. We separate them into different streams based on how many identified or probe leptons they contain. For this analysis we use the stream with two well-identified leptons and one loose lepton for the final result, the stream with one well-identified lepton and one loose lepton for the efficiency measurement, and the stream with a single loose lepton for the fake rate calculation.

Chapter 5

Efficiencies, Scale factors, and fake rates

5.1 Lepton efficiency

When we perform an analysis we make a prediction of the number of events we expect from simulation and compare it to the observed data. However, we need to make sure that the number of events from simulation is normalized to the integrated luminosity we have, so we can make sensible comparisons. In order to do so we use quantities called scale factors that adjust for the differences between simulation and data for each lepton type and period of data taking, and to measure scale factors, we need to measure the efficiencies for each lepton type.

The efficiency is essentially the ratio of the number of identified leptons to the total

number of known probe leptons. To measure it we use a procedure called "tag and probe". We select samples of known $Z \to e^+e^-$ or $Z \to \mu^+\mu^-$ in Monte Carlo or high-p_T leptons in the data where one lepton is well-identified and the other is a probe lepton, using the same definition of probe as was used in the section on probes to determine CDF maximum geometrical acceptance. Zs are very clean objects in all data samples; so if both legs are opposite sign, same flavor leptons that reconstruct to a Z mass, we have our base sample. Now the ratio becomes the number of probe legs that are identified as lepton types over the total number of probe legs.

Only three of our eleven lepton types are able to pass the high-p_T lepton trigger (LBE, CMUP, CMX), so we will only ever be able to identify single Zs that contain at least one of these triggerable types. The overall formula to calculate efficiency is therefore

$$\frac{2N_{GG} + N_{GT}}{2N_{GG} + N_{GT} + N_{GP}}$$

We use the following notation for these equations:

- G = triggerable lepton
- T = tight ("tagged") lepton
- P = probe lepton, failing tight cuts
- N = number of events

Six of the eleven lepton types, the non-triggerable ones, have no N_{GG} contribution, and therefore the formula to calculate those efficiencies reduces to

$$\frac{N_{GT}}{N_{GT} + N_{GP}}$$

We have to take into account a complication when calculating the final efficiency, which is the background contribution. Since the N_{GP} distribution has one leg failing the tight selection, the background shows up most prominently in that set of data, and must be subtracted out. Previously, the background was modeled as a straight line, using the integral of sidebands around the central Z peak.

Since the previous ZZ analysis also measured the efficiency of its lepton categories, measuring the efficiency of our new ones and comparing the two should also give us some idea of the potential gain of our analysis.

However, we discovered that for some of our lepton categories the background distribution overwhelmed the signal peak to such an extent that this subtraction method ceased to yield any reasonable efficiency results. We therefore developed a new method of background subtraction, modeling it as an exponential curve with three free parameters instead of a straight line. We fit the $M_{\ell\ell}$ plots for each lepton type and probe pass/fail category with the exponential in the sideband regions, then extrapolated that to the signal region, in order to subtract the background fit from the signal region. We then plugged this result into the other quantities in the efficiency equation and propagated the resulting uncertainty to yield our own efficiency results. We only apply this new subtraction method to data and not

Eff for 3.6 fb^{-1}

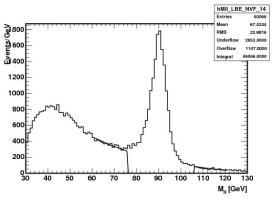
	data	MC	Scale Fac
LBE	0.905 ± 0.001	0.893 ± 0.001	1.013 ± 0.001
PLBE	0.859 ± 0.002	0.896 ± 0.001	0.958 ± 0.002
PNTLBE	0.724 ± 0.006	0.817 ± 0.001	0.887 ± 0.007
IsoCrkTrk e	0.933 ± 0.005	0.896 ± 0.001	1.042 ± 0.006

Table 5.1: Efficiency tables for new electron types in 3.6 fb^{-1}.

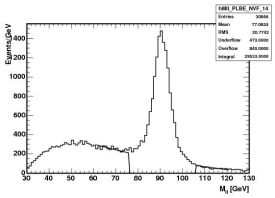
Monte Carlo sets, because there is no background contribution in the Monte Carlo. Table 5.5 compares both old and new efficiencies over approximately 3.6 fb^{-1} of data, more than half our final sample. It makes clear that for certain lepton categories the previous method was oversubtracting the background because the straight-line model was overestimating the background content of the N_{GP} quantity in the efficiency formula. The N_{GP} plots with exponential fit applied to the sideband regions can be found in Figure 5.1.

The reason we are showing the efficiencies for a representative portion of the data instead of the entire data set is that we divided the data into different periods based on the timescale of the history of the detector. In principle there should be no change in the efficiency over these different periods of data taking, but in fact we discovered some small change which we elaborate on further in Section 7.1.

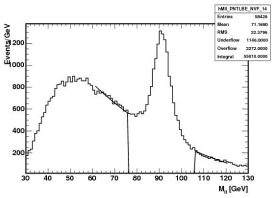
As we can see from Tables 5.1-5.4, we have gains in every single lepton category except the smaller types (CMP, CMXMsKs), which we did not change. These gains



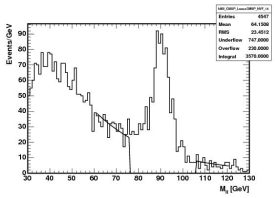
(a) N_{GP} for LBE



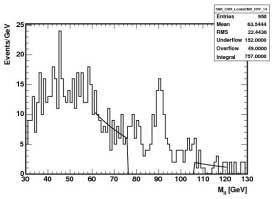
(b) N_{GP} for PLBE



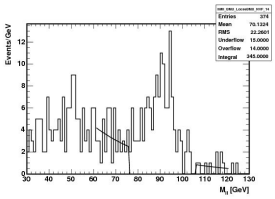
(c) N_{GP} for PNTLBE



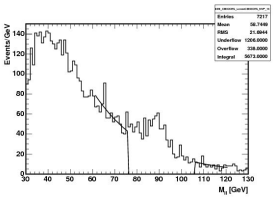
(d) N_{GP} for CMUP



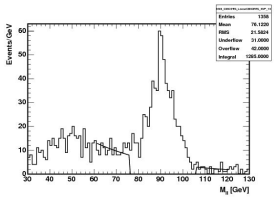
(e) N_{GP} for CMX



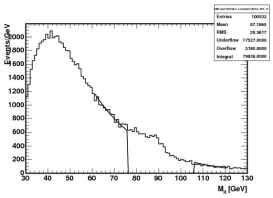
(f) N_{GP} for BMU



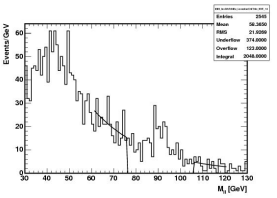
(g) N_{GP} for CMIOCES



(h) N_{GP} for CMIOPES



(i) N_{GP} for IsoCrkTrk



(j) N_{GP} for IsoCrkTrk

Figure 5.1: The background and its subtraction for the tag + unindentified probe Z mass reconstruction

Eff for 3.6 fb^{-1}

	data	MC	Scale Fac
TCE	0.854 ± 0.001	0.840 ± 0.001	1.016 ± 0.002
PHX	0.777 ± 0.001	0.842 ± 0.001	0.923 ± 0.002
PEM	0.617 ± 0.005	0.703 ± 0.001	0.878 ± 0.008
CrkTrkElec	0.779 ± 0.004	0.759 ± 0.001	1.027 ± 0.006

Table 5.2: Efficiency tables for old electron types in 3.6 fb^{-1}.

Eff for 3.6 fb^{-1}

Loose types	data	MC	Scale Fac
CMUP	0.844 ± 0.008	0.913 ± 0.001	0.924 ± 0.009
CMP	0.823 ± 0.015	0.896 ± 0.002	0.919 ± 0.017
CMX	0.927 ± 0.016	0.927 ± 0.001	1.000 ± 0.017
CMXMsKs	0.806 ± 0.019	0.888 ± 0.002	0.907 ± 0.021
BMU	0.835 ± 0.017	0.728 ± 0.002	1.148 ± 0.023
CMIOCES	0.897 ± 0.014	0.824 ± 0.001	1.089 ± 0.017
CMIOPES	0.661 ± 0.009	0.613 ± 0.001	1.079 ± 0.015
IsoCrkTrk μ	0.962 ± 0.013	0.943 ± 0.001	1.020 ± 0.014

Table 5.3: Efficiency tables for new muons in 3.6 fb^{-1}.

Eff for 3.6 fb^{-1}

	data	MC	Scale Fac
CMUP	0.803 ± 0.008	0.871 ± 0.001	0.922 ± 0.009
CMP	0.823 ± 0.015	0.896 ± 0.002	0.919 ± 0.017
CMX	0.886 ± 0.015	0.888 ± 0.002	0.998 ± 0.017
CMXMsKs	0.806 ± 0.019	0.888 ± 0.002	0.907 ± 0.021
BMU	0.777 ± 0.018	0.691 ± 0.002	1.125 ± 0.026
CMIOCES	0.768 ± 0.012	0.702 ± 0.001	1.093 ± 0.017
CMIOPES	0.604 ± 0.010	0.569 ± 0.001	1.061 ± 0.018
CrkTrkMu	0.591 ± 0.007	0.581 ± 0.001	1.018 ± 0.012

Table 5.4: Efficiency tables for old muons in 3.6 fb^{-1}.

range from 5% to 63% for each type. However to gauge the overall increase in acceptance generated by this analysis it is better to compare the change in the overall ZZ yield between this and the previous version. We can find this comparison in Chapter 8.

5.2 Scale factors

The difficulty with using just Monte Carlo datasets to measure the efficiency of our lepton detection for different lepton types is that the Monte Carlo is never a 100% trustworthy tool for describing the CDF detector and associated detector issues. We

Comparison for 3.6 fb^{-1}

	data (new eff)	data (old eff)
LBE	0.905 ± 0.001	0.910 ± 0.001
PLBE	0.859 ± 0.002	0.852 ± 0.002
PNTLBE	0.724 ± 0.006	0.747 ± 0.005
IsoCrkTrk e	0.933 ± 0.005	1.233 ± 0.014
CMUP	0.844 ± 0.008	0.888 ± 0.007
CMX	0.927 ± 0.016	0.984 ± 0.014
BMU	0.835 ± 0.017	0.890 ± 0.018
CMIOCES	0.897 ± 0.014	1.053 ± 0.014
CMIOPES	0.661 ± 0.009	0.657 ± 0.012
IsoCrkTrk μ	0.962 ± 0.013	1.048 ± 0.012

Table 5.5: Comparison of efficiencies using the old and new methods for lepton types in 3.6 fb^{-1} The new method tends to yield a smaller result because the old method was overestimating the background and removing signal. A tip-off for this was that the old efficiency for some types is greater than 1.

therefore measure the efficiency of each lepton in both the data and the Monte Carlo. We perform the same subtraction methods as described above on these efficiencies. We then take the ratio of the efficiency results in data over the Monte Carlo to arrive at the scale factor, which we measure for each data period. Finally we apply this scale factor to the $ZZ \to 4\ell$ level of the analysis, run over using ZZ Monte Carlo, and scale the expected yield of ZZ using the scale factors.

5.3 Trigger Efficiencies

Trigger efficiencies only play a small role in this analysis. We use three trigger paths, which have been combined in two lepton streams; these are the central electron and the two types of central muon triggers. We use the same set of standard CDF trigger efficiency measurements as the $H \to WW$ group. Our analysis is particularly insensitive to the trigger efficiency because we search for four leptons and the likelihood that one of those leptons will trigger goes as

$$1 - (1 - \epsilon)^4$$

where ϵ is the trigger efficiency for each trigger path. We are therefore extremely insensitive to changes in the trigger efficiency, arising from the changes we have made to loosen the central electron and muon types. Moreover, a study done by the $H \to WW$ group shows that the change in trigger efficiency between the old type TCE and their new type LBE is effectively negligible. The trigger efficiencies that we apply are taken from earlier CDF calculations.

5.4 Fake rates and background estimation

As we have stated earlier, $ZZ \rightarrow 4\ell$ is a remarkably clean signal. No other identified Standard Model process has the same signature. In principle any four isolated leptons that reconstruct to two Z bosons should be part of our analysis; however, because of the prevalence of objects in the detector faking leptons, we must balance identification and acceptance to cut out this background.

Most of our categories, despite our loosening of their cuts, are still quite low in background. However two of our categories have a relatively high fake rate, PNTLBE and IsoCrkTrk. The reason for this is that PNTLBE has no track requirement, so it is less difficult for a forward photon to fake it in the detector. Similarly, the IsoCrkTrk category has no calorimetric requirement, so the background may be higher there. In addition, the fact that an electron propagating through a crack in the calorimeter can travel farther than if it were going through the lead-scintillator accordion of the calorimeter proper, means that in order to measure the isolation of the category properly, we need to create a cluster using the energy deposited in the hadronic calorimeter. If we remove this cluster from the isolation energy, then we let in more background from energy in the hadronic calorimeter that may be due to jets.

To determine the background estimate in our analysis, we employ the method of fake rates. This is a procedure that has become standard at CDF in which we take as pure a sample as possible of jet-triggered events and search for roughly isolated

tracks with a few additional very loose cuts within the sample. We take the subset of events that contain these objects, which we use as denominators for our fake identification ratio. These denominator objects are useful because with them we can more reliably extrapolate the fake rate than from the entire jet sample. Furthermore, the denominator objects will cancel out when we measure the total background, so we need not concern ourselves with their exact selection. The denominator object definitions are similar to the probe definitions, listed in the section on probes.

We have to use this data-driven method, and not a Monte Carlo sample of pure jets, since it is hard to simulate rare QCD processes reliably. Even measuring jets in the data is difficult to manage–there are four jet triggers at different p_T: 20, 50, 70, and 100 GeV/c^2, all of which have some slight inconsistency in the jet events they accept. We must use all of these to calculate our fake rate and take the difference between them as a systematic.

Once we have this subset of events we then need to find the yield of identified leptons within it. This will give us the ratio of denominators faking leptons to first order. However we do need to take into account the fact that there will be some electroweak contribution in the jets sample, if for example an event is triggered on a jet but also contains leptons in a separate process. This contribution dominantly comes from $W+$ jets.

What we must then do is use a Monte Carlo sample of $W \to \ell\nu$ scaled to the

luminosity of the jet sample in order to subtract off the true leptons in the jet sample. We can apply scale factors to these quantities and plug them into the following equation:

$$f_i = \frac{N_{ID} - N_{EWK}}{N_{den} - N_{den\_EWK}}$$

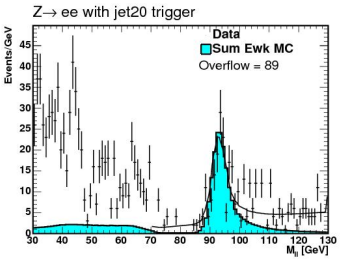
where N_{ID} is the number of identified leptons from the sample, N_{EWK} is the number of actual electroweak leptons in the sample, N_{den} is the number of denominator objects, and N_{den_EWK} is the number of denominator objects that are actually electroweak.

Before subtracting the electroweak contribution we do a cross-check to reconstruct the Z peak in the jet data sample and compare the results between data and scaled Monte Carlo. We find that before subtraction the peak is well modeled and we can be confident we are subtracting the right amount of events. These plots are summarized in Figure 5.4.

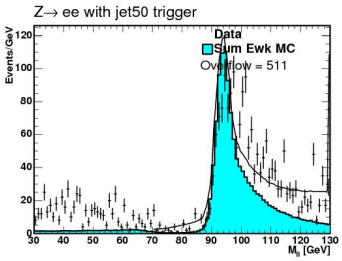
Finally we can calculate the fake rate for each lepton type. This quantity varies according to p_T, and is influenced by various competing detector effects.

So the fake rate for each lepton is taken from the plots shown in Figure 5.4. The four points for each p_T correspond to the four jet trigger samples, and the gray band measuring the spread between them is our systematic.

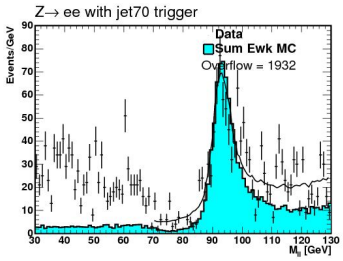
In order to apply the fake rates to the measurement, we must calculate the total background. We consider the two physics backgrounds relevant to this analysis, $Z\gamma$ + jet (faking a lepton) and Z + 2 jets(both faking leptons). We need to add both



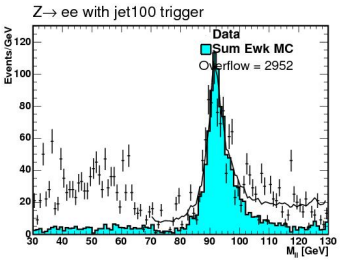
(a) M_{ee} in jet 20 modeled



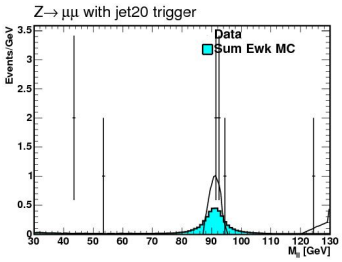
(b) M_{ee} in jet 50 modeled



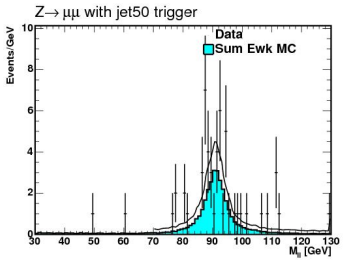
(c) M_{ee} in jet 70 modeled



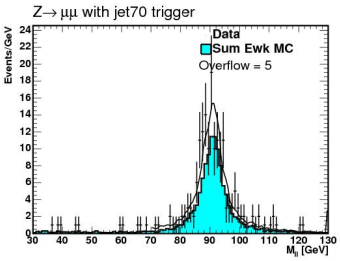
(d) M_{ee} in jet 100 modeled



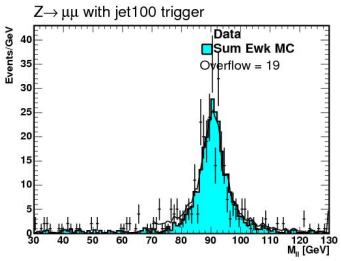
(e) M_{ee} in jet 20 modeled



(f) M_{ee} in jet 50 modeled

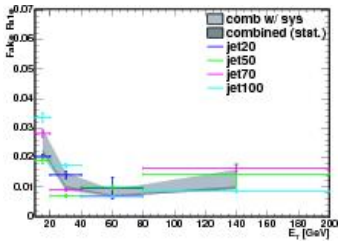


(g) M_{ee} in jet 70 modeled

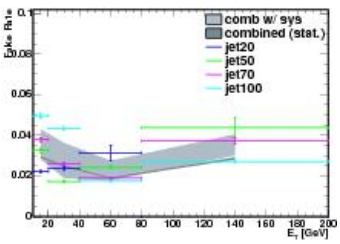


(h) M_{ee} in jet 100 modeled

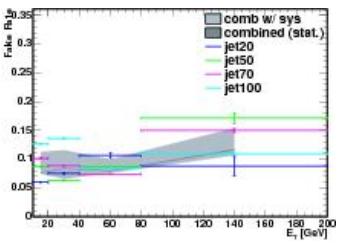
Figure 5.2: Z peak in jets modeled by MC fit by data



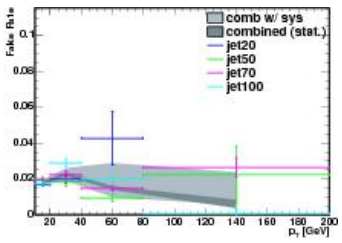
(a) LBE Fake Rate



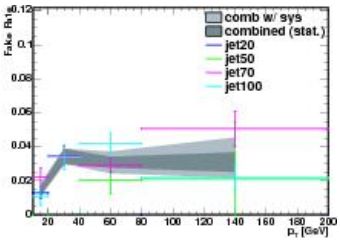
(b) PLBE Fake Rate



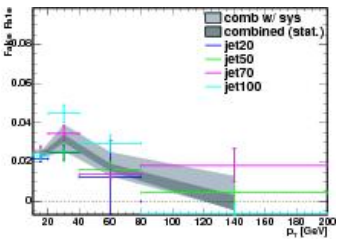
(c) PNLBE Fake Rate



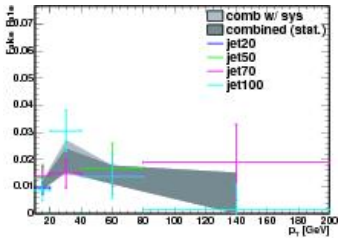
(d) CMUP Fake Rate



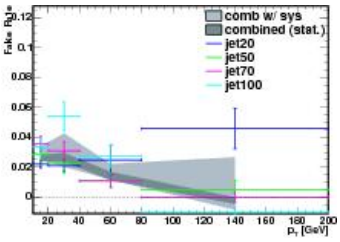
(e) CMP Fake Rate



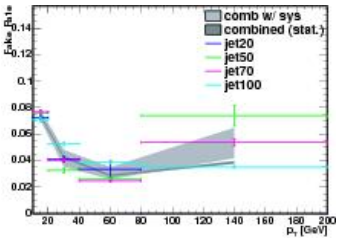
(f) CMX Fake Rate



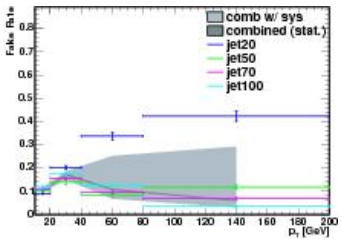
(g) CMXMsKs Fake Rate



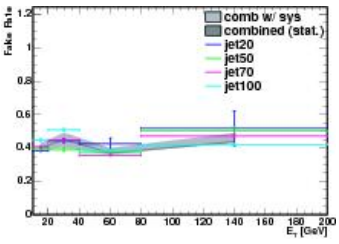
(h) BMU Fake Rate



(i) CMIOCES Fake Rate



(j) CMIOPES Fake Rate



(k) IsoCrkTrk Fake Rate

Figure 5.3: Fake rates for the various lepton types

and apply the fake rates to them according to the proportion of objects potentially faking leptons within them. However in our analysis we cannot count the number of events coming from each background, but we can count the number of events with identified leptons and denominators. So we measure the total background by counting these objects in the following fashion.

$$\text{Total background} = N_{Z\gamma+1fake} + N_{Z+2fakes} = N_{Z\gamma+f} + N_{2l+2f} = fN_{Z\gamma+d} + f^2N_{2l+2d}$$

where f is the fake rate for each lepton and d is the corresponding jet denominator with the potential to fake a lepton. We can measure N_{3l+d} and also N_{2l+2d} in our analysis, so must convert the background into these quantities.

If we break down the components of N_{3l+d}, we get

$$N_{3l+d} = N_{Z\gamma+d} + N_{2l+1f+1d} = N_{Z\gamma+d} + 2fN_{2l+2d}$$

(The 2 enters from the combinatorics of which denominator out of the two faked the lepton.) Solving for $N_{Z\gamma+d}$, and plugging back into the top equation we get

$$\text{Total background} = fN_{3l+d} - f^2N_{2l+1d}$$

We can now list our first preliminary result for the analysis, the measured background using $3l + d$ and $2l + d$.

$$N_{Z\gamma+1fake} + N_{Z+2fakes} = 0.59 - 0.05 = 0.54 \text{ events.}$$

However we are not yet done! We must consider one more aspect of the background; namely, that there is some contamination by real ZZ events that failed one lepton's full identification.

This quantity must be subtracted from the fN_{3l+1d} background. We do not calculate the real ZZ contribution to the f^2N_{2l+2d} because of the negligible effect on such a small quantity. When we calculate this quantity using the ZZ Monte Carlo we find that the expected contribution for fN_{3l+1d} is 0.29 events. After subtracting this leaves us with a total background of 0.25 events.

Chapter 6

Event selection

6.1 Event Selection

There are only three criteria for each Z reconstruction once lepton selection has been performed. They are:

1. the primary vertex of the Z has to be within the detector, for a range of $\pm$ 60 cm around the center point
2. the leptons that reconstruct the Z must be opposite sign, same flavor
3. the Z mass must be within a certain mass range.

To gain further acceptance, and taking advantage of the cleanliness of the four-lepton event as well as the broadness of the Breit-Wigner Z mass resonance, we specify that the primary Z must be within a 30 GeV mass region, but the second Z

can be much less constrained–it can fall within a 100 GeV region around the Z resonance. For $ee\mu\mu$ events, there is no ambiguity in assigning ZZ reconstruction; the leptons of the same flavor go together. However, this is not the case for events with $4e$, 4μ, or 2ℓ2trk. We must instead minimize the χ^2 for the ZZ mass equation

$$\chi^2 = \frac{(M_{12} - M_Z)^2}{\sigma(M_{12})^2} + \frac{(M_{34} - M_Z)^2}{\sigma(M_{34})^2}$$

Once this is done then the event has fit the requirements necessary for a fully identified ZZ.

Chapter 7

Analysis cross-checks and systematics

7.1 Analysis Cross-Checks

At CDF it is standard procedure, before unblinding the results of an analysis, to test the analysis algorithm by attempting to predict the signal and background in a region that is orthogonal to the signal region. For instance, a WW analysis may use as such a control the region of low $\not\!\!E_T$. However in our analysis there is no such obvious region. We therefore perform two cross-checks in the Z region to justify our analysis algorithms.

As our first cross check we compare the Drell-Yan yield in the Z signal region between data and Monte Carlo over all periods of data taking. When we first

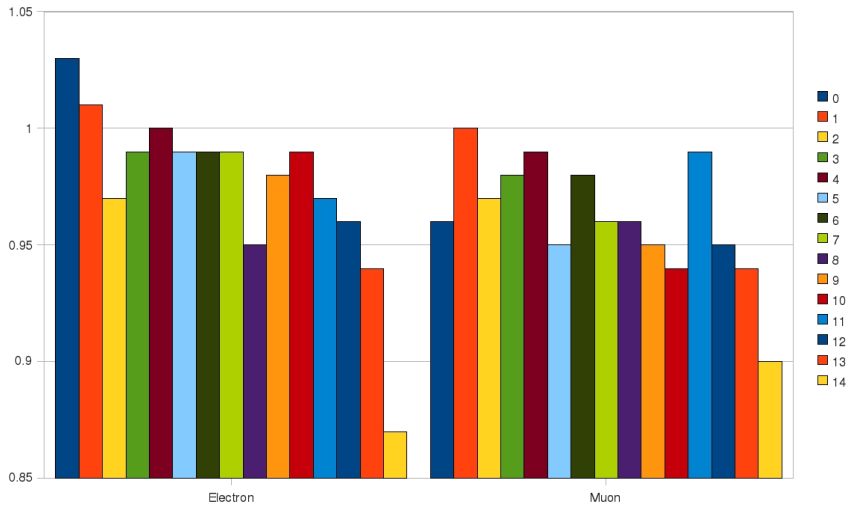


Figure 7.1: Ratio of yields in the data divided by yields in the Monte Carlo, normalized to integrated luminosity over all data periods in Drell-Yan cross-check in the Z signal region.

measured these ratios we decided upon a 7% per-Z systematic for lepton identification efficiency to cover the uncertainty between periods for each Z, or about 15% for the full analysis. We can see the results in Figure 7.1.

However it should not be the case that the efficiency varies significantly with time, and yet we see a clear decrease in the data in the yield of Drell-Yan events over both electrons and muons. We tried several routes to finding a possible reason for this decrease and came up short. Moreover, we discovered that two other groups at CDF also saw a mysterious decrease in the Drell-Yan yield over time. We therefore decided to simply adjust the scale factor for each period such that the Drell-Yan yield was the same in data and in Monte Carlo. The result of this adjustment can be seen in Figure 7.2 and is described in the section on the systematics.

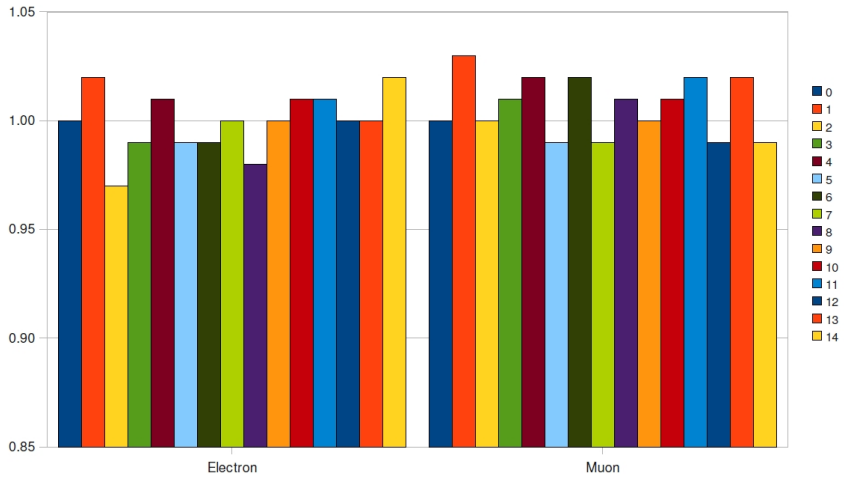


Figure 7.2: Ratio of data/MC yields over all data periods in Drell-Yan cross-check in the Z signal region, using scale factors for each lepton corrected for the decrease in Drell-Yan yield over the different periods.

As our second cross-check we investigate the same-sign region in the Z mass signal area, and sum up the background contributions to this region and match to the data. The same-sign region is that where a Z mass is reconstructed using only same-sign lepton objects. This is clearly a background region, so we wanted to be sure that we could model this region with our new lepton types as well.

There is an inherent systematic of approximately 30% in the $W + \gamma$ Monte Carlo, one of our predicted backgrounds, and an approximately 25% uncertainty on the W+jets background; keeping this in mind we see that even after adjusting the scale factors for the first cross-check the data does appear to follow our predicted yield for each lepton type in this region. Figure 7.3 shows the result.

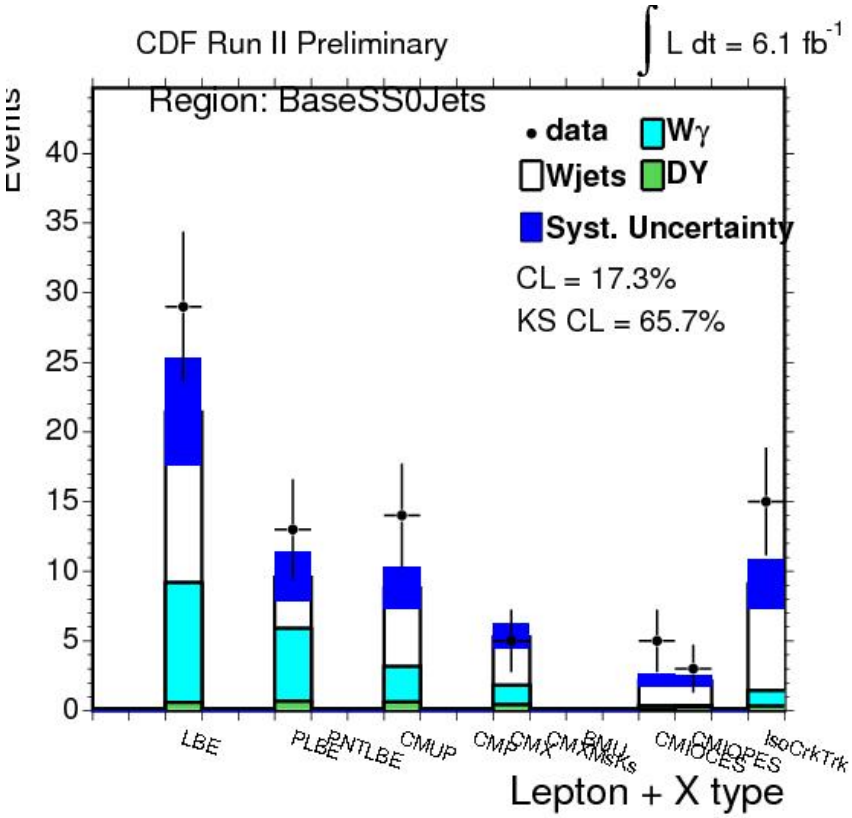


Figure 7.3: Comparison of data and MC prediction for same-sign region.

7.2 Systematics

We consider multiple sources for our systematics.

- Lepton ID efficiency

  The dominant contribution comes from the uncertainty in our lepton efficiency measurement. When we first calculated this uncertainty, we took the uncertainty on the ratios of data/MC Drell-Yan yield over all periods by eye and estimated the uncertainty to be 7% for each Z, or 15% for the full ZZ analysis. However, as stated in the section on cross-checks, we decided to adjust the scale factors of each period to conform to the decrease in Drell-Yan yield over time. Since we corrected the Monte Carlo yield and our background prediction is still well-modeled, we decided that we had accounted for the discrepancy and could then lower this systematic. We calculated the uncertainty on the corrected Drell-Yan yield, which was 2% per Z due to rounding uncertainties during the calculation, and 4% for the ZZ analysis. We then took the average of the the two uncertainties for a final systematic of 10%.

- NLO Acceptance

  This quantity comes from the fact that the PYTHIA Monte Carlo is a leading-order simulation. This means that it simulates tree-level diagrams and neglects loop diagrams. However in reality these loop diagrams, although having a much smaller amplitude than the leading-order diagrams, do

contribute some amount towards the ZZ production cross-section, and therefore we include it as a systematic. The systematic uncertainty cited in Table 7.1 was calculated by a collaborator at CDF by varying other Monte Carlo simulations.

- Trigger efficiency

  This systematic was calculated by other groups at CDF for the triggers that we use in our analysis. Since even with our looser lepton definitions than the norm, we find that they pass the trigger requirements, we are able to keep the calculated trigger efficiency. When we calculated the uncertainty on that efficiency for our analysis, we got the result of 0.3%.

- PDF Uncertainty

  Similarly, this systematic was calculated by a collaborator at CDF and deals with the uncertainty on the parton distribution function used by the Monte Carlo to model the interior of the proton at the moment of collision.

- Luminosity

  This systematic was calculated by the Luminosity group at CDF and is standard across all analyses at the detector.

- Background uncertainty

  The uncertainty in our fake background is obtained in a slightly complicated way. There are three pieces that need to be considered: the uncertainty that is shown in the gray band around the fake rate in Figure 5.4; the uncertainty on

the ZZ contribution to the fake background; and finally we must take into consideration the magnitude of the variance of our fake rates over all our lepton types. Our measurement of the uncertainty on the fake background is 24% of the total 0.54 calculated. The uncertainty on the ZZ contribution to the background is dominated by the measured uncertainty of the ZZ cross-section itself, resulting from the differences between the four jet triggers. Another component of this measurement is the uncertainty on the efficiency with which each lepton is identified, which is the lepton ID systematic. Finally the fact that the fake rate of our IsoCrkTrk category is an order of magnitude larger than almost any other fake rate gives it undue weight in the background estimation algorithm. However our background estimation happened not to include any fake IsoCrkTrk leptons. Therefore we set an upper limit on its occurrence at 68% confidence at 1.3 fake events and multiply by its fake rate. If we had seen one fake IsoCrkTrk we could have used the central value instead. This is the largest contribution to the background systematic. We add all in quadrature for a final systematic of 4.3%.

All that remains is to plug these numbers into the cross-section formula and propagate the uncertainties to a standard systematic.

Systematics	% error on σ_{ZZ}
Lepton ID efficiency	10
NLO Acceptance	2.5
Trigger efficiency	0.3
PDF Uncertainty	5.4
Luminosity	5.9
Background uncertainty	4.3
Total	13.7

Table 7.1: Systematics for $ZZ \to \ell\ell\ell\ell$.

Chapter 8

Results

We see 14 events in 6.1 fb^{-1} with an expected background of 0.25 events. No other collaboration has to date seen this number of ZZ candidate events in the four lepton channel.

The kinematic distribution of these events, including the $M_{4\ell}$ plot, can be found in below in Figures 8.

To measure the final cross-section we use the formula

$$\sigma_{ZZ} \times BR_{ZZ\to 4\ell} = \frac{N_{obs} - B}{\int \mathcal{L} \times \epsilon}$$

where N_{obs} is the observed yield of ZZ, B is the measured background, and ϵ is the efficiency of our analysis to find ZZ at CDF.

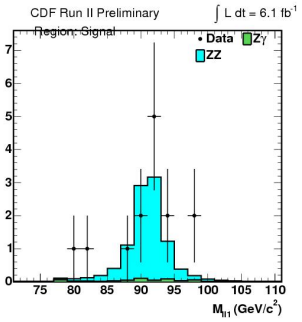
The overall efficiency of our analysis we can get by using the equation

$$\epsilon = fracN_{exp}\sigma_{ZZ} \times BR_{ZZ\to 4\ell} \times \int \mathcal{L}$$

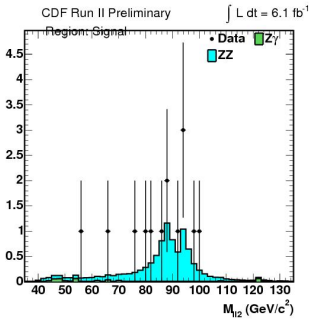
Here N_{exp} is the expected yield of ZZ running the Monte Carlo through this analysis. When we calculate this number we get a result of 23% for the efficiency, which is a 44% jump over the 16% efficiency of the previous analysis.

To measure the uncertainty on the cross-section we use the Feldman-Cousins procedure at the 68th percentile, which is equivalent to one sigma on a Gaussian. We add the systematics listed in the previous section in quadrature to measure its own contribution to the uncertainty on the central result.

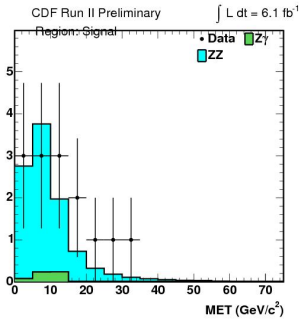
The final cross-section result is $2.18^{+0.64}_{-0.63}$ pb (stat) ± 0.30 (sys). This result is slightly greater than one sigma above the predicted theoretical cross-section of 1.4 pb.



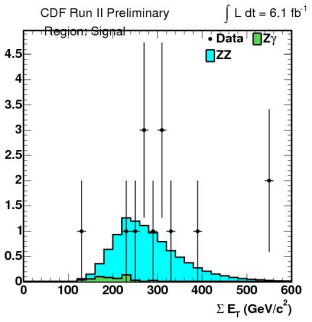
(a) Z_1 mass plot



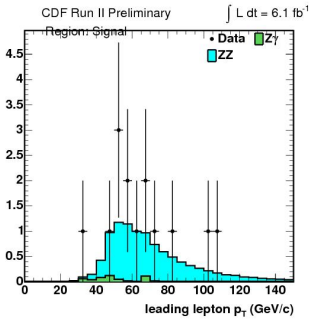
(b) Z_2 mass plot



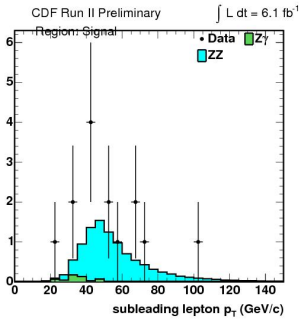
(c) ZZ MET



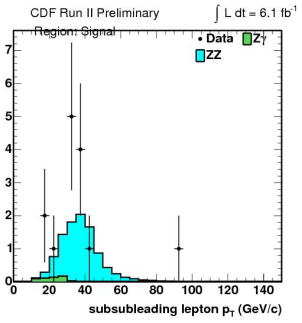
(d) ZZ Sum Et



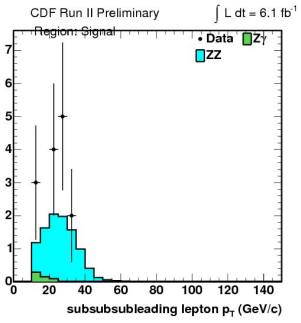
(e) Highest lepton p_T



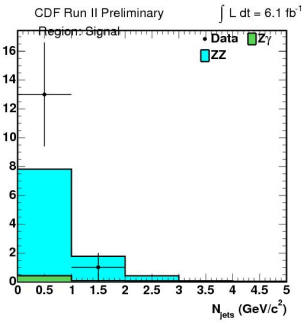
(f) Second-highest lepton p_T



(g) Second-lowest lepton p_T

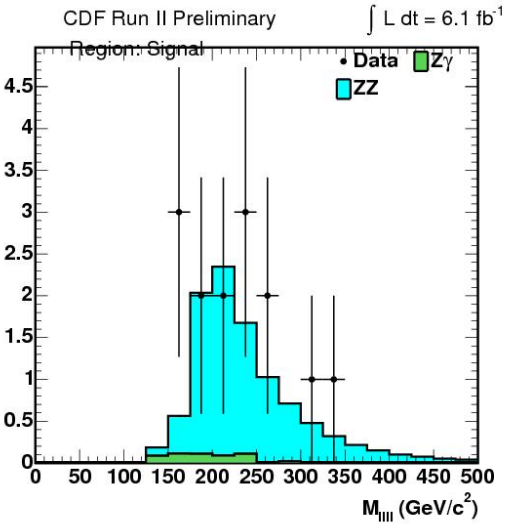


(h) Lowest lepton p_T

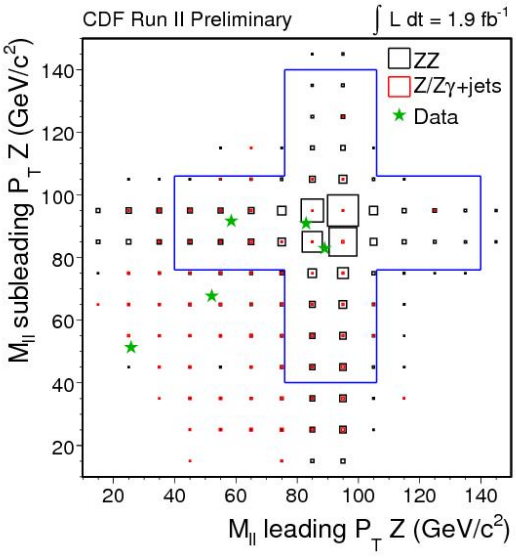


(i) Number of jets in each ZZ event

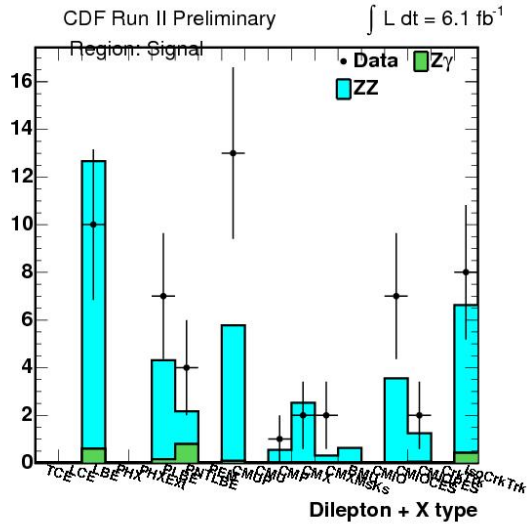
Figure 8.1: Kinematic plots for observed ZZ events



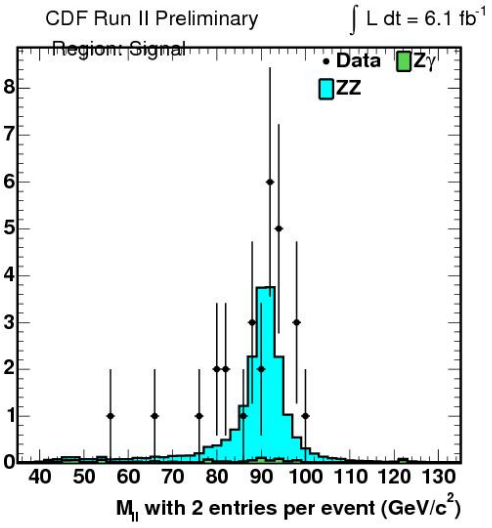
(a) ZZ 4 lepton mass plot



(b) Z_1 v. Z_2 mass plot



(c) Lepton types forming ZZ candidates



(d) Mass of all Z candidates

Bibliography

[1] CDF/ANAL/EXOTIC/CDFR/10407.

[2] CDF/PUB/ELECTROWEAK/PUBLIC/10344 (Contribution to proceedings of ICHEP 2010).

[3] P. Langacker, Rev. Mod. Phys. **81**, 1199-1228, (2008)

[4] Biswarup Mukhopadhyaya, Satyanarayan Mukhopadhyay, Phys. Rev. D **82**, 031501 (2010)

[5] CDF/THESIS/EXOTIC/PUBLIC/9619.

[6] CDF/ANAL/ELECTROWEAK/CDFR/8538.

[7] CDF/ANAL/ELECTROWEAK/CDFR/8539.

[8] CDF/ANAL/ELECTROWEAK/CDFR/8720.

[9] CDF/ANAL/ELECTROWEAK/CDFR/8924.

[10] CDF/ANAL/ELECTROWEAK/CDFR/9137.

[11] CDF/PHYS/ELECTROWEAK/CDFR/10477.

[12] www.fnal.gov/pub/about/faqs/index.html

[13] Jianming Qian, Eur. Phys. J. C. **34** 1,39-48 (2004)

[14] K. Gounder, Nuclear Physics B. **117**, 875-877 (2003)

[15] Accelerator Concepts,
www-bdnew.fnal.gov/operations/rookie_books/Concepts_v3.6.pdf

[16] CDF facebook page, https://www.facebook.com/pages/Collider-Detector-at-Fermilab-CDF/359360509462

[17] F.M.Bieniosek, Fermilab-TM-1857 (1993)

[18] K.Y. Ng, Fermilab-TM-2199 (2003)

[19] N.I.M. **556** 459-481 (2006)

[20] K. Nakamura et al., J. Phys. **G 37**, 075021 (2010)

[21] F. Abe et. al., FERMILAB-Pub-96/390-E(1996)

[22] D. Ambrose, N.I.M. **518**, 117-120 (2004)

[23] The CDF II Collaboration, FERMILAB-Pub-96/390-E (1996)

[24] Newman-Holmes, C., Freeman, J., Fermilab-Conf-87/231 (1987)

[25] The LEP collaboration, Physics Letters B **572**, 133-144 (2003)

[26] The L3 Collaboration, CERN-EP 2003-2034, (2003)

[27] G. Feldman and R. Cousins, Phys. Rev. D **57**, 3873-3889 (1998)

[28] K. Yasuoka, S. Mikamo, T. Kamon, and A. Yamashita, N.I.M. A, **267**, 315-329 (1988)

[29] CDF/DOC/TRIGGER/CDFR/2909.

[30] H.U. Bengtsson, T. Sjstrand., Comp. Phys. Comm. **46**, 43 (1987)