Search for the Standard Model Higgs Boson in the Diphoton Final State in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using the CDF II Detector

by

Karen R. Bland

Department of Physics, Baylor University

Advisor: Dr. Jay R. Dittmann

Defended in May 2012

Search for the Standard Model Higgs Boson in the Diphoton Final State
in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV
Using the CDF II Detector

by

Karen R. Bland, B.S.

A Dissertation

Approved by the Department of Physics

Gregory A. Benesh, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

Jay R. Dittmann, Ph.D., Chairperson

Gerald B. Cleaver, Ph.D.

Kenichi Hatakeyama, Ph.D.

Linda J. Olafsen, Ph.D.

Jack D. Tubbs, Ph.D.

Accepted by the Graduate School
August 2012

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

Abstract

We present a search for the Standard Model Higgs boson decaying into a pair of photons produced in $p\bar{p}$ collisions with a center of mass energy of 1.96 TeV. The results are based on data corresponding to an integrated luminosity of 10 fb^{-1} collected by the CDF II detector. Higgs boson candidate events are identified by reconstructing two photons in either the central or plug regions of the detector. The acceptance for identifying photons is significantly increased by using a new algorithm designed to reconstruct photons in the central region that have converted to an electron-positron pair. In addition, a new neural network discriminant is employed to improve the identification of non-converting central photons. No evidence for the Higgs boson is observed in the data, and we set an upper limit on the cross section for Higgs boson production multiplied by the $H \to \gamma\gamma$ branching ratio. For a Higgs boson mass of 125 GeV/c^2, we obtain an observed (expected) limit of 12.2 (10.8) times the Standard Model prediction at the 95% credibility level.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to first thank my advisor, Dr. Jay Dittmann, for helping lead me to the path of experimental high energy physics (HEP) and encouraging me to stick with it when I wanted to quit. Throughout my time as a graduate student, he has been a constant source of advice and knowledge on the tools needed to perform a successful analysis in this field. These tools include the physics of particle interactions, programming, statistics, Linux, LaTeX, and even writing and grammar. Dr. Dittmann has furthermore taught me to pay more attention to detail and to persevere in solving difficult problems. He has extended consistent encouragement and patience in my struggles as a graduate student, both academic and personal. It has been a pleasure to work with him and I cannot imagine having had a better advisor.

I next express my gratitude to Dr. Ray Culbertson and Dr. Craig Group. There are not words to describe my appreciation of their patience with me and support over the course of this analysis, answering my numerous (and repetitive) questions and meeting with me on a weekly basis. By their guidance and expertise, I have learned how to perform an analysis and am able to confidently present the results of this dissertation. The original code framework for this analysis was based on a previous beyond-the-Standard-Model (BSM) Higgs boson search using photons that was performed by Dr. Culbertson, Dr. Group, and a high school student they mentored, Jamie Ray. They also later worked with Jamie (who became a college student) to develop the improved central photon identification using a neural network algorithm (see Chapter 4).

I thank my postdoc, Dr. Azeddine Kasmi, for the opportunity to work with and learn from him, and for exercising much patience with me. Dr. Nils Krumnack was a postdoc for Baylor before Dr. Kasmi and I am grateful for his guidance and criticism in my first year of research. I also thank Dr. Costas Vellidis for providing feedback and supporting this analysis.

It has been a wonderful privilege to perform my research with and under the guidance of the CDF collaboration (Appendix E). In addition to those already men-

tioned, I thank the leadership (spokespersons) of this collaboration, Dr. Rob Roser and Dr. Giovanni Punzi, for helping develop a fruitful training ground for young scientists. I next thank the leaders (conveners) of the Higgs group, Dr. Eric James, Dr. Ben Kilminster, and Dr. Group, for critiquing and approving the results of this analysis. During internal presentations, numerous other colleagues in the Higgs group also provided feedback to help improve upon our methods. For publishing, I thank the three members of the collaboration that volunteered to understand and critique our analysis at a deeper level (our "godparents"): Drs. Rick St. Denis, Max Goncharov, and Ben Auerbach. Other members of the collaboration also helped improve our presentation of the analysis by commenting and providing feedback on drafts of the publication. I thank Nils Holgersson, from Baylor's Classroom and Technology Services, for providing the ability to communicate with colleagues from the CDF collaboration via video conferencing.

During the one year that I lived at Fermilab, I learned a great deal about the detector, accelerator, and data-taking process from the CDF operations team. My three-month "Ace" shift time was a wonderful experience for me and I thank those who contributed to teaching me. I especially thank Dr. Eugene (JJ) Schmidt for guiding Ace meetings, and for many enjoyable conversations throughout my time at Fermilab. I additionally thank those who built my understanding of the XFT trigger system over the course of that year, in particular Dr. Dittmann (who helped build this subdetector) and Dr. Homer Wolfe. Dr. Wolfe spent much time training me and was extensively patient with my numerous anxious calls over this subdetector. Dr. Krumnack, Dr. Jonathan Lewis, and Baylor graduate students also taught me much about this system.

Dr. Martin Frank and Dr. Sam Hewamanage are recent graduates of Baylor that have acted like CDF postdocs to me, helping me with so much of the frustrating details that can be involved with an analysis. I am also grateful to the other members of Baylor's experimental HEP group not already mentioned: Zhenbin (Ben) Wu, Tara Scarborough, Dr. Hongxuan Liu, and Dr. Ken Hatakeyama. They have provided feedback on my analysis and I have also learned many HEP concepts through them.

I next wish to acknowledge the Baylor Physics Department chair, Dr. Greg Benesh, for helping provide an environment that allowed me to enjoy my time as a graduate student at Baylor. I also thank the professors of the Physics Department who instructed me in my coursework and well prepared me for the Baylor qualifying exams. The office staff, Mrs. Marian Nunn-Graves and Mrs. Chava Baker, keep this department running smoothly and I am exceedingly grateful for their friendly help and support over the years. I thank the professors that served on my dissertation committee, Dr. Dittmann, Dr. Gerald Cleaver, Dr. Hatakeyama, Dr. Linda Olafsen, and Dr. Jack Tubbs. Dr. Tubbs is the department chair of the Baylor Statistics Department and all other committee members are from the Physics Department.

For my first three years as a graduate student, I was funded to teach labs for the Baylor physics department. I thank Dr. Linda Kinslow and Mr. Edward Schaub for allowing me the opportunity to serve as a teaching assistant under them. I truly enjoyed my teaching experience and believe that no physics student should earn a Ph.D. without a similar experience. I also thank Mr. Schaub for the numerous friendly conversations over the years that have blessed my time here at Baylor.

Of the things I will walk away with from my time as a graduate student, one is a much greater confidence in public speaking. I thank Dr. Dittmann and Dr. Olafsen for leading the graduate colloquium in my first couple of years at Baylor. Since then, I thank the CDF collaboration for training me in this area and for providing dozens of opportunities to give presentations either internally or publicly. I especially learned from Dr. Dittmann, Dr. Culbertson, Dr. Group, and the Higgs conveners and spokespersons of CDF. I also thank Dr. Bennie Ward, Dr. Hatakeyama, and Dr. Group for providing me opportunities to give HEP colloquium talks.

I have spoken thus far of my time as a graduate student; however, I would not be here were it not for the numerous professors that taught me as an undergraduate. I loved my time at James Madison University and hold only pleasant memories of the Physics Department there. I thank Dr. Chris Hughes, my undergraduate advisor, and Dr. Bill Ingham and the late Dr. Don Chodrow. Each of these men greatly supported me as an undergraduate and helped me apply to graduate schools. I also thank the

many professors that allowed me to perform summer research projects with them. Of these, I particularly thank Dr. Paolo Desiati for guiding my first experimental particle physics project, which sparked my initial interest in the HEP field.

There are many other people over the years that have helped build me into who I am today. I cannot name them all, but I thank all of you for being a part of my life and for caring for me. I thank my high school friends for helping give me a love of learning and for loving my crazy family: Jocelyn, Ingrid, Carina, Susan, and Amy. Since then, I thank those that have preached the word of God into my heart and have consistently fed my spirit. This includes the preachers and members of my church families in VA, TX, and IL. I particularly thank Dr. Qianyu (Theresa) Ma, a fellow Baylor student that began the same year as me and who brought me many prayers and a constant friend at the office. From Illinois, I also thank Christina, Renee, Bethany, Ben E., and Richard for your friendship. While at Baylor, I have had much fellowship and encouragement from Cristina, Yiliu, Holly, Amy, Joy, Alessandra, Anna, Autumn, Chava, William Q., the Ellis family, the Martin family, and the Hunts. I love each of you and thank you for supporting me through this.

So much of who we are is influenced by those who raised us. My family has always been consistently proud of me no matter what I do or pursue. It is from this foundation of love and encouragement that I have been able to succeed as a graduate student. I especially thank my mom, whose sweet demeanor with others I can only wish to better model; my dad, who has taught me to be a hard worker and who has loved me through this graduate work; and my sister, who has been so faithful to pray for and call me.

I wish to end by thanking and honoring my best and dearest friend, William Low. I do not know if I would be writing this dissertation were it not for you walking with me through it. Thank you for being with me in both my deepest struggles and my greatest joys. Other than my parents, there is no other person in my life that I have learned more from. Thank you for helping feed my spirit and keeping me grounded over the past few years, reminding me of the bigger and grander things that I pursue in this short life.

To the Lord Jesus Christ

He made the earth by his power;
he founded the world by his wisdom,
and stretched out the heavens by his understanding.
Jeremiah 10:14, John 1:1-18

Chapter 1

Introduction and Theoretical Motivation

The concept that all matter is composed of fundamental particles is thought to have originated from the Greeks in the fifth century BC. In particular, Democritus called these particles atoms, which in Greek means "indivisible" or "uncuttable." This concept remained philosophical, however, until advancements in technology over the past century have finally allowed it to be tested.

The Standard Model (SM) of particle physics is a modern theory that very accurately describes the known fundamental particles of nature and their interactions. The Higgs boson is the only particle predicted by the theory that has not yet been observed in nature. Observation of this particle would be a validation of the Higgs mechanism, which was introduced into the theory in order to explain why different particles have different masses.

A summary of the SM theory is provided in this chapter, including a description of the Higgs mechanism and current searches for the Higgs boson. A motivation for the search for the Higgs boson using photons is also provided.

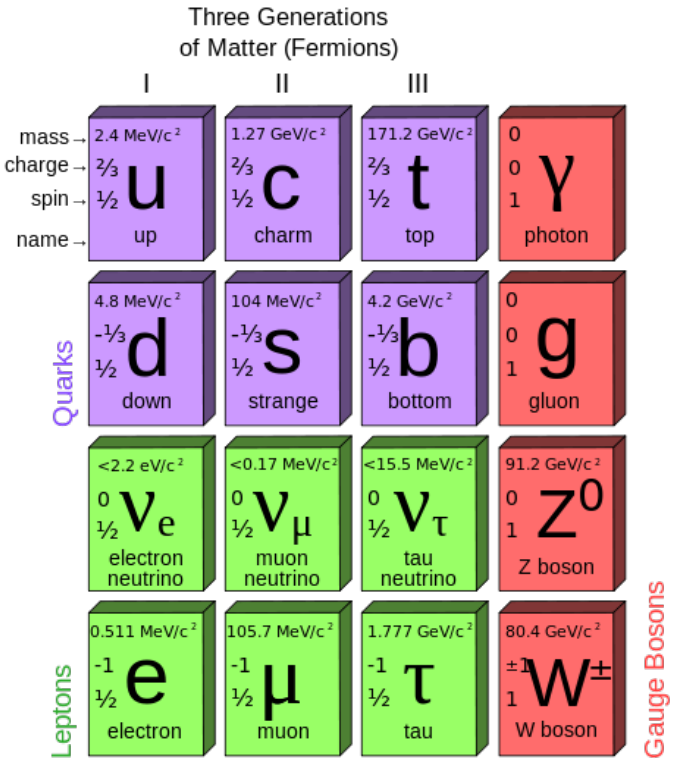


Figure 1.1: The elementary particles that have been predicted by the SM and verified by experiment [1, 2, 3, 4]. The matter particles are the quarks and leptons and the force carrier particles are the gauge bosons. The quarks and leptons also have antimatter counterparts, which have the same properties but with opposite electric charge.

1.1 The Standard Model of Particle Physics

Figure 1.1 provides a summary of the elementary particles described by the SM of particle physics. The name of each particle is given with its symbol, mass, electric charge, and intrinsic spin. These particles are divided into the matter particles and the gauge bosons. The matter particles consist of the quarks (q) and leptons (l), which are grouped into three generations, generally based on their mass and stability. All quarks and leptons have spin 1/2 and are therefore fermions, a term used to describe particles with half-integer spin and that obey Fermi-Dirac statistics (i.e. the Pauli exclusion principle). The gauge bosons are elementary particles that mediate the fundamental forces and are thus responsible for particle interactions (see Figure 1.2).

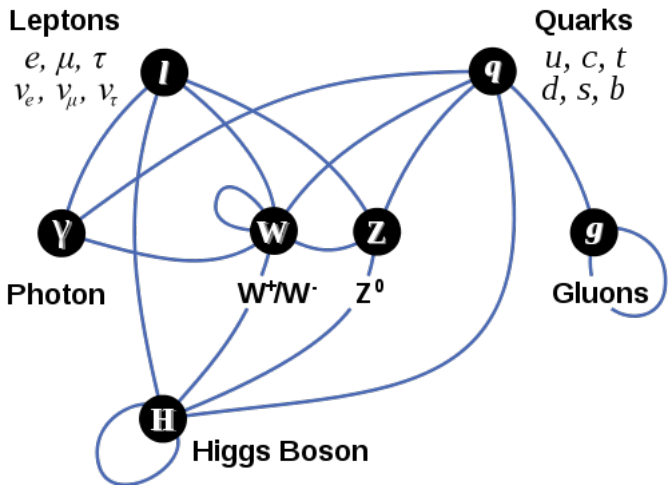


Figure 1.2: A diagram showing the ways in which the elementary particles interact with one another, including interactions with the Higgs field [5]. The Higgs boson is the only particle predicted by the SM that has not yet been observed in nature.

They each are spin-1 particles and are therefore classified as bosons, a term used to describe particles with integer spin and which obey Bose-Einstein statistics.

The elementary matter particles are the SM theory's answer to "What is everything made of?" We now know that the protons and neutrons of atomic nuclei are each made up of three quarks; the proton is composed of two up quarks and a down quark (uud) and the neutron is composed of one up quark and two down quarks (udd). The electrons that we are familiar with are a type of lepton and they surround the nucleus of the atom. It is the up quark, down quark, and electron that make up stable (ordinary) matter. From Figure 1.1, we can see that these particles are among what is called the first generation of matter particles — they are the lightest and the most stable. The other matter particles are produced at higher energies, such as in particle collisions, and they decay to more stable particles that can be observed and reconstructed in particle detectors (as described in the next two chapters).

There are a host of other composite particles[1] that are made of combinations of either two or three quarks, similar to the proton and neutron. These are called

[1] So many that Enrico Fermi once told a student, Leon Lederman, "Young man, if I could remember the names of these particles, I would have been a botanist!" [6].

hadrons; those containing two quarks are called mesons and those with three quarks are called baryons. This analysis collects data from a particle accelerator that collides a proton (p) and an antiproton ($\bar{p}$) together and it is, therefore, called a hadron collider. Later, we'll see two examples of mesons, the neutral pion (π^0) and eta (η) particles, that can mimic the likeness of a photon in the detector, creating a background to a search for true photons produced in a collision. The quark composition for the π^0 meson is a mixture of up and down quarks, $\frac{u\bar{u}-d\bar{d}}{\sqrt{2}}$, and that of the η meson is a mixture of up, down, and strange quarks, $\frac{u\bar{u}+d\bar{d}-s\bar{s}}{\sqrt{6}}$.

All matter particles also have antimatter counterparts: particles with the same properties, but with opposite electric charge. Here we denote antimatter particles either by a placing a bar over the particle, such as $\bar{p}$ for an antiproton, or by putting the actual charge with the particle's symbol, such as e^- for a normal electron and e^+ for the antielectron, which is more commonly called a positron.

The fundamental interactions between matter particles in nature are gravity, electromagnetism, the weak force, and the strong force. Gravity is thought to be mediated by the graviton, however, this force is the only fundamental interaction that is not described by the SM — a weakness of the theory. The dynamics of SM particles and their non-gravitational interactions (Figure 1.2) are described by relativistic quantum field theories (QFT), where it is through a perturbative version of the theory that interactions between particles are described by the exchange of field quanta, the gauge bosons. Though they are not categorized with the matter particles, they can have mass (the W and Z bosons in particular) and can be produced in high energy collisions and reconstructed in particle detectors from their stable final-state products.

The strong force is the force responsible for holding the protons and neutrons of the atom together. It acts on particles with a fundamental property called color charge, held by both quarks and gluons. The quarks and gluons can each have a color charge of three different values, called red, green, and blue. Quantum chromodynamics (QCD) is the field theory describing the exchange of gluons between these colored particles (thus gluons interact with other gluons). The strong force

has the interesting property that it increases as colored particles move apart. The potential energy between two such particles increases sufficiently that as they move apart, there is enough energy to produce a $q\bar{q}$ pair. This process, called fragmentation or hadronization, can continue as long as their is sufficient potential energy between pairs of colored particles. As we will see in Chapter 3, when a quark or gluon is produced in a collision or interaction, the result is a spray of particles in the detector, which is called a jet.

Quantum electrodynamics (QED) is the field theory describing the exchange of a neutral boson, the photon, between particles with electric charge. It, therefore, describes the electromagnetic force that we are familiar with. The weak force is mediated by three bosons: the W^+ which has an electric charge of $+1$, the W^- which has an electric charge of -1, and the Z which is neutral. Due to the fact that the W boson is charged, the weak force is the only force that can change a quark from one type to another (i.e. change its flavor), such as a u quark changing to a d quark in beta decay. The weak force is, therefore, responsible for radioactivity. At higher energy, the electromagnetic and weak forces are unified to form the electroweak theory.

SM field theories are called gauge theories because they are obtained by requiring the Lagrangian describing their dynamics to be invariant under a generalized (local) gauge transformation. The gauge transformation produces a new massless spin-1 vector field or fields, which is the photon in QED and the 8 gluons in QCD. The W and Z bosons are known to have mass, however, and therefore this process alone fails to correctly describe the weak interactions. In the SM, their masses are instead obtained by applying both gauge invariance and spontaneous symmetry breaking of a new spin-0 scalar field that is introduced, called the Higgs field. The next section describes local gauge invariance using QED as an example. The subsequent section introduces the Higgs field and the Higgs mechanism, from which the electroweak field theory is derived.

1.2 Local Gauge Invariance

The Dirac Lagrangian[2] for a single free spin-1/2 particle of mass m in the absence of any interactions is

$$\mathscr{L} = i\psi^\dagger\gamma^\mu\partial_\mu\psi - m\psi^\dagger\psi, \tag{1.1}$$

where $\psi = \psi(x)$ is a complex spinor, $\psi^\dagger = \psi^\dagger(x)$ is its complex conjugate, x is the set of space-time indices, μ is an index running over these indices such that $\frac{\partial}{\partial x_\mu} \equiv \partial_\mu = \partial_0 - \partial_1 - \partial_2 - \partial_3$, and γ^μ are the gamma matrices. A global U(1) phase transformation of the form $\psi \to e^{-iq\lambda}\psi$ (λ a constant) leaves the Lagrangian invariant; however, this is not true if we instead apply a *local* phase transformation [8, 9] of the form $\psi \to e^{-iq\lambda(x)}\psi$. The derivative, ∂_μ, now acts on the $\lambda(x)$ to add an extra term to the Lagrangian:

$$\mathscr{L} \to \mathscr{L} + (q\psi^\dagger\gamma^\mu\psi)\partial_\mu\lambda. \tag{1.2}$$

In order to remove this, something has to be added to the original Lagrangian to cancel this term. In particular, we add a new vector field A_μ such that Equation (1.1) is modified to

$$\mathscr{L} = \left[i\psi^\dagger\gamma^\mu\partial_\mu\psi - m\psi^\dagger\psi\right] - (q\psi^\dagger\gamma^\mu\psi)A_\mu \tag{1.3}$$

where A_μ transforms as $A_\mu \to A_\mu + \partial_\mu\lambda$ under the local phase transformation $\psi \to e^{-iq\lambda(x)}\psi$. With this choice, Equation (1.1) becomes invariant both globally and locally. The Lagrangian is complete with the addition of a term[3] describing the free vector field for A_μ:

$$\mathscr{L}_{\text{QED}} = \left[i\psi^\dagger\gamma^\mu\partial_\mu\psi - m\psi^\dagger\psi\right] - \left[\frac{1}{4}F^{\mu\nu}F_{\mu\nu}\right] - (q\psi^\dagger\gamma^\mu\psi)A_\mu. \tag{1.4}$$

The first and second brackets contain the free terms for a Dirac field (spin-1/2 particle) and the new massless vector field, respectively, while the last term describes their interaction.

[2] This section follows the explanation provided in David Griffiths' *Introduction to Elementary Particles* [7].

[3] Since A_μ is a vector field, this comes from the Proca Lagrangian for spin-1 particles, however the mass term is left out because the mass term is not invariant under a local phase transformation.

We have just demonstrated that when we start with a Lagrangian, such as that for Dirac fields, and require invariance under a local phase transformation, a new massless field arises along with an interaction between the fields. Equation (1.4) is in fact the Lagrangian for QED [10] and describes the dynamics between Dirac and Maxwell fields, whose quanta are electrons and positrons for the Dirac fields and photons for the Maxwell fields. Notice that the new term in Equation (1.2) came about because of the partial derivative, $\partial_\mu \psi \rightarrow e^{-iq\lambda}\left[\partial_\mu - iq(\partial_\mu \lambda)\right]\psi$. We can obtain the same results if we introduce a covariant derivative to replace the ∂_μ in Equation (1.1):

$$D_\mu = \partial_\mu + iqA_\mu \tag{1.5}$$

This is the covariant derivative for a U(1) local phase transformation, a mechanism which produced the photon and its interaction with electrons or positrons.

In 1954, Yang and Mills applied this technique [11] to *two* spin-1/2 particles of equal mass with an SU(2) transformation, leading to a Lagrangian describing the interaction between two Dirac fields of the same mass and three massless vector fields. There are no such massless vector fields observed, however, nor are there two spin-1/2 particles of the same mass that interact as Yang and Mills predicted. Nevertheless, their work provided the framework for the general application of this gauge theory to non-Abelian groups,[4] allowing the later development of QCD [12, 13] and electroweak interactions [14, 15, 16]. The former was obtained by requiring invariance under an SU(3) local phase transformation on three Dirac fields from which came a set of eight massless vector fields. In other words, the theory described three colors of a single quark flavor and their interaction with the gluons. The electroweak field gauge theory was obtained by requiring invariance of both a U(1) and an SU(2) local phase transformation, denoted U(1) $\otimes$ SU(2), in addition to spontaneous symmetry breaking via the Higgs mechanism [17, 18, 19]. This is described in the next section.

[4] The example provided for QED is based on a U(1) transformation, which is Abelian (i.e. commutative). This is not the case, however, for SU(2), so their theory led to a technique for handling what are called non-Abelian transformations.

1.3 The Higgs Mechanism

In order to explain[5] the origin of the heavy masses for the W and Z gauge bosons, while leaving the photon massless, a new field (or fields) is introduced which is called the Higgs field. The simplest approach, which is used by the SM, is to have two complex scalar fields,

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \tag{1.6}$$

with a Lagrangian of the form

$$\mathscr{L}_{\text{Higgs}} = (\partial_\mu \phi^\dagger)(\partial^\mu \phi) - V(\phi^\dagger, \phi) \tag{1.7}$$

where the first term is the kinetic term and the second is the potential term. The Higgs field potential (a "Mexican hat" potential) is selected such that its minimum is non-zero, as in

$$V(\phi^\dagger, \phi) = \frac{1}{2}\lambda^2(\phi^\dagger\phi - \frac{1}{2}\nu^2)^2 \tag{1.8}$$

where λ and ν are constants. This gives a minimum field configuration of $|\phi| = \nu/\sqrt{2}$ and the Lagrangian is symmetric in this form (see Figure 1.3).

The dynamics of an interaction in the SM are calculated from Feynman calculus (calculation of Feynman diagrams), which is based on deviations from *some* ground state. We must, therefore, choose one of the continuous states, and thus break the symmetry.[6] The choice is to let one of them have a minimum of zero, say ϕ_1, such that $\phi_1^{\min} = 0$. The other is real with $\phi_2^{\min} = \nu/\sqrt{2}$. The Higgs field ϕ is then expanded about this ground state such that

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix}. \tag{1.9}$$

We next use the lessons learned from Section 1.2 and apply both a U(1) and SU(2) local gauge transformation and demand that the Lagrangian be invariant. Similar to

[5] The primary references used for the description provided here are [7], [20], and [21].

[6] This is called spontaneous symmetry breaking because there is no external agency that causes it [7]. The choice shown here removes massless Goldstone scalar boson fields, which do not exist in nature. The degrees of freedom from these Goldstone fields will instead be transferred to vector boson fields (that were created by requiring local gauge invariance) in order to give them mass.

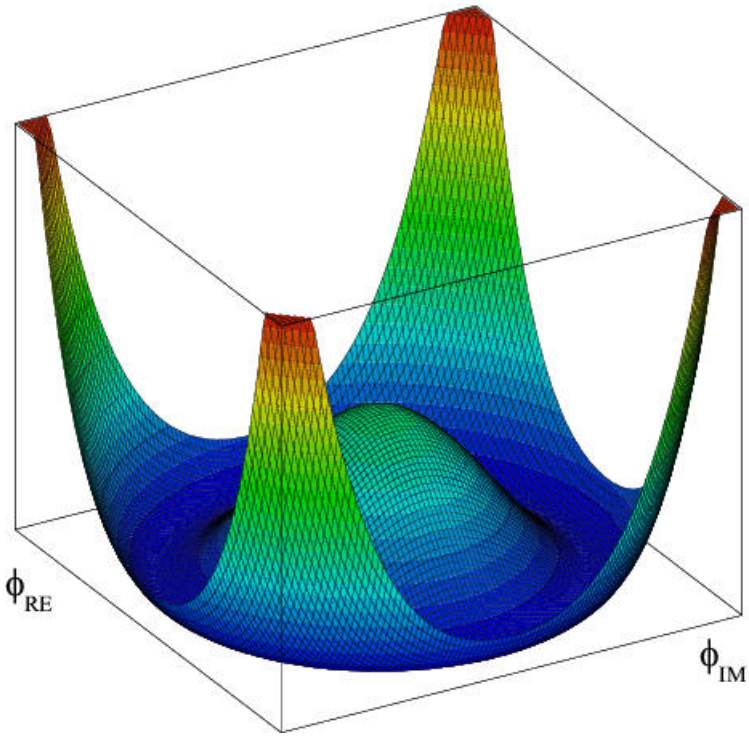


Figure 1.3: The shape of the Higgs potential when one field in Equation (1.6) is set to zero and the value of λ in Equation (1.8) is fixed. The figure [22] then shows the potential as a function of different values of the real and imaginary parts of the remaining scalar field.

Equation (1.5), and based on the work of Yang and Mills, the covariant derivative for such a transformation is

$$D_\mu = \partial_\mu + \frac{1}{2} i g_1 B_\mu + \frac{1}{2} i g_2 \mathbf{a} \cdot \mathbf{W}_\mu \tag{1.10}$$

where g_1 and g_2 are coupling constants similar to charge in Equation (1.5), $\mathbf{a}$ are the three Pauli matrices, B_μ is a new scalar field, and $\mathbf{W}_\mu$ are three new vector fields such that the term $\mathbf{a} \cdot \mathbf{W}_\mu$ indicates the sum $a_1 W_\mu^1 + a_2 W_\mu^2 + a_3 W_\mu^3$. From this we obtain

$$\begin{aligned} D_\mu \phi &= (\partial_\mu + \frac{1}{2} i g_1 B_\mu + \frac{1}{2} i g_2 \mathbf{a} \cdot \mathbf{W}_\mu) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} \qquad (1.11) \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \partial_\mu h \end{pmatrix} + \frac{1}{2\sqrt{2}} i g_1 (\nu + h) \begin{pmatrix} 0 \\ B_\mu \end{pmatrix} + \frac{1}{2\sqrt{2}} i g_2 (\nu + h) \begin{pmatrix} W_\mu^1 - i W_\mu^2 \\ -W_\mu^3 \end{pmatrix}. \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \partial_\mu h \end{pmatrix} + \frac{1}{2\sqrt{2}} i (\nu + h) \begin{pmatrix} g_2 W_\mu^1 - i g_2 W_\mu^2 \\ g_1 B_\mu - g_2 W_\mu^3 \end{pmatrix} \end{aligned}$$

Since we are most interested in demonstrating the mechanism used to explain the masses of the electroweak gauge bosons, in the following, the interaction terms between the Higgs field and the new fields created by the local gauge transformation are excluded. We therefore obtain

$$\begin{aligned}(D_\mu\phi)^\dagger(D_\mu\phi) &= \frac{1}{2}(\partial^\mu h)(\partial_\mu h) + \frac{1}{2}\left(\frac{g_2\nu}{2}\right)^2 \left|W_\mu^1 - iW_\mu^2\right|^2 \\ &\quad + \frac{1}{2}\left(\frac{\nu}{2}\right)^2 \left|g_1 B_\mu - g_2 W_\mu^3\right|^2 + \text{other terms.}\end{aligned} \tag{1.12}$$

This is simplified with the introduction of gauge fields that are linear combinations of the B_μ, W_μ^1, W_μ^2, and W_μ^3 fields. The first two gauge fields are the charged vector boson W_μ^- and its complex conjugate W_μ^+:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \tag{1.13}$$

Two neutral gauge bosons are also introduced with the following definitions:

$$Z_\mu = \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_2 W_\mu^3 - g_1 B_\mu) = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W \tag{1.14}$$

$$A_\mu = \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_1 W_\mu^3 + g_2 B_\mu) = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W \tag{1.15}$$

where $\sin\theta_W = g_1/\sqrt{g_1^2 + g_2^2}$ and $\cos\theta_W = g_2/\sqrt{g_1^2 + g_2^2}$ (and θ_W is called the weak mixing angle). Equation (1.12) then simplifies to

$$\begin{aligned}(D_\mu\phi)^\dagger(D_\mu\phi) &= \frac{1}{2}(\partial^\mu h)(\partial_\mu h) + \left(\frac{g_2\nu}{2}\right)^2 (W_\mu^+ W_\mu^-) \\ &\quad + \frac{1}{2}\left(\frac{g_2\nu}{2\cos\theta_W}\right)^2 |Z_\mu|^2 + \text{other terms}\end{aligned} \tag{1.16}$$

which can be inserted into the Higgs field Lagrangian from Equation (1.7).

With the inclusion of free terms for the vector fields themselves,[7] the Higgs and electroweak Lagrangian is

$$\begin{aligned}\mathscr{L}_{\text{Higgs+Ewk}} &= \left[\frac{1}{2}(\partial^\mu h)(\partial_\mu h) - \frac{1}{2}(\lambda\nu)^2 h^2\right] \\ &\quad -\frac{1}{4}\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu} + \left(\frac{g_2\nu}{2}\right)^2 (W_\mu^+ W_\mu^-) \\ &\quad +\frac{1}{2}\left(\frac{g_2\nu}{2\cos\theta_W}\right)^2 |Z_\mu|^2 - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\ &\quad + \text{other terms}\end{aligned} \tag{1.17}$$

(the other terms are self-interaction terms for the Higgs field and interaction terms between the Higgs and vector fields). Equation (1.17) is formatted such that we can read off the mass terms for the gauge bosons, which are expected to be of the form $M_W^2 W_\mu^+ W_\mu^-$ for the charged bosons and of the form $\frac{1}{2}M_Z^2 Z_\mu^2$ and $\frac{1}{2}M_A^2 A_\mu^2$ for the neutral gauge bosons [21]. Their masses are then

$$M_{W^+} = M_{W^-} = \frac{g_2\nu}{2} \tag{1.18}$$

$$M_Z = \frac{g_2\nu}{2\cos\theta_W} \tag{1.19}$$

$$M_\gamma = 0 \tag{1.20}$$

where the last line associates A_μ with the photon and acknowledges the lack of a mass term as indicative that it has zero mass, which is in fact what is observed.

Applying the local gauge invariance of Section 1.2 alone gives rise to the addition of massless vector bosons. While the photon is indeed found to be massless, the W and Z boson of the weak force are very heavy and the gauge theory alone does not provide an adequate model of experimental observation. The technique provided in this section introduced a manner in which the W and Z bosons acquire mass, yet leave the photon massless. In addition to requiring local gauge invariance of the Lagrangian, the Higgs field is included into the theory and the symmetry of its potential is broken with the choice of a particular ground state.

[7] These are terms like that introduced for Equation (1.4) in the QED Lagrangian. Here, they take the form $-\frac{1}{4}\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}$ and are the kinetic and self-interaction terms for the $W^\pm, Z$, and γ gauge bosons.

This mechanism — called the Higgs mechanism — provides an explanation for how elementary particles of the SM obtain mass. Here, it was demonstrated for the gauge bosons of the electroweak theory. However, it is also applied to fermions to explain how leptons and quarks obtain different masses from one another.

1.4 The Higgs Boson

In the Lagrangian of Equation (1.17), we also find a term for the neutral Higgs boson mass which is

$$M_{\text{Higgs}} = \lambda\nu \tag{1.21}$$

where λ and ν are shown in Equation (1.8) and are a Higgs field self-coupling parameter and the vacuum expectation value, respectively. Based on measurements of muon decays, the latter is constrained to $\sim$246 GeV [3], but λ is unknown. Therefore, though the SM predicts a manner in which fundamental particles acquire mass through the Higgs mechanism, the mass of the Higgs boson itself remains a free parameter of the theory, to be determined experimentally if it does in fact exist. An observation of the Higgs boson would be a direct validation of the SM theory and it is, therefore, one of the most sought after discoveries in science.

We next discuss the properties of the Higgs boson that are predicted by the SM theory. Current searches for the Higgs boson are also described, along with a motivation for searching for this particle using photons.

1.4.1 Coupling to Gauge Bosons and Fermions

Going from Equation (1.11) to (1.12), the focus was on the manner in which the gauge bosons obtain mass from the Higgs field vacuum expectation value, where the masses were then determined from the third and fourth terms of Equation (1.17). Writing Equation (1.12) more completely, however, we can obtain the interaction terms between the Higgs field and the gauge bosons:

$$\mathscr{L} = \left(\frac{g_2(\nu+h)}{2}\right)^2 (W_\mu^+ W_\mu^-) + \frac{1}{2}\left(\frac{g_2(\nu+h)}{2\cos\theta_W}\right)^2 |Z_\mu|^2 . \tag{1.22}$$

Then expanding $(\nu + h)^2$ and applying Equations (1.18) and (1.19) gives

$$\begin{aligned}\mathscr{L} &= M_W^2(W_\mu^+ W_\mu^-) + \frac{2M_W^2}{\nu}(W_\mu^+ W_\mu^-)h + \frac{M_W^2}{\nu^2}(W_\mu^+ W_\mu^-)h^2 \\ &+\frac{1}{2}M_Z^2 \left|Z_\mu\right|^2 + \frac{1}{2}\frac{2M_Z^2}{\nu}\left|Z_\mu\right|^2 h + \frac{1}{2}\frac{M_Z^2}{\nu^2}\left|Z_\mu\right|^2 h^2\end{aligned} \tag{1.23}$$

The first and fourth terms are just the mass terms we saw in Equation (1.17), but we now find two interaction vertex terms both for the W fields with the Higgs field and for the Z fields with the Higgs field. The strength of a gauge boson interaction with a single Higgs boson (the second and fifth terms) is then given by

$$g_{HVV} = \frac{2M_V^2}{\nu}. \tag{1.24}$$

where V refers to the W or Z vector boson, with $\nu \sim 246$ GeV as mentioned above and with $M_W = 80.4$ GeV/c^2 and $M_Z = 91.2$ GeV/c^2. A similar process can be performed to determine the strength of the interaction of the Higgs boson with fermions, and it is found that

$$g_{Hf\bar{f}} = \frac{M_f}{\nu} \tag{1.25}$$

with M_f equal to the mass of the fermion (given in Figure 1.1).

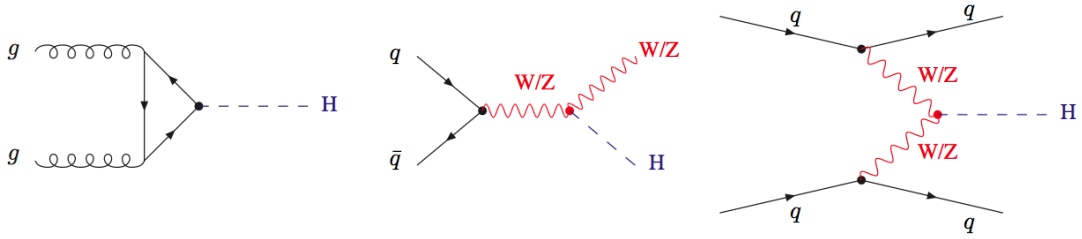
The dominant production and decay processes for the Higgs boson are then predicted by the SM to involve a coupling to one of the massive gauge bosons or with a coupling to one of the heavy quarks, in particular the top quark, which has a mass of $\sim$173 GeV/c^2. There is no direct coupling to the massless gluons or photons; however, the Hgg and $H\gamma\gamma$ couplings can be induced though a loop involving the heavy gauge bosons or quarks.

1.4.2 Higgs Boson Production

The data used in this analysis are obtained from the Fermilab Tevatron accelerator, located in Batavia, IL, near Chicago. Proton and antiproton beams are accelerated in the Tevatron ring to a center of mass energy ($\sqrt{s}$) of 1.96 TeV and collided together. These collisions are energetic enough to allow the production of massive particles, including the Higgs boson. In addition to three valence quarks, protons and antiprotons are also composed of gluons, which bind the valence quarks together, and a sea

of transient quark pairs. A collision is really a scattering interaction between partons (subparticles) of the proton and antiproton.

The most likely mechanisms for the production of the SM Higgs boson at the Tevatron are shown in Figure 1.4. The first diagram shows the gluon fusion (GF) process, which is the interaction of two gluons with a virtual (intermediate) $t\bar{t}$ pair that fuses to produce the Higgs boson.[8] The associated production (VH) mechanism is shown in the second diagram. In this process, a $q\bar{q}$ pair annihilate and produce either a W or Z boson which then radiates the Higgs boson such that the Higgs boson is produced in association with a vector boson. The third diagram shows the vector boson fusion mechanism (VBF), where a pair of vector bosons is produced and couple to form the Higgs boson. In this process, the Higgs boson is present alongside the original $q\bar{q}$ pair.



(a) Gluon fusion (b) Associated production (c) Vector boson fusion

Figure 1.4: The dominant production mechanisms at the Tevatron for the SM Higgs boson [3, 23, 24]. We search for a Higgs boson originating from any of these three processes.

A measure of the likelihood of one of these interactions is called the cross section (σ). The cross section depends on such quantities as the type of colliding particles ($p\bar{p}$), the $\sqrt{s}$ of the collision (1.96 TeV), the coupling strengths of the interacting partons, and the mass of the Higgs boson. The cross section for a particular scattering process has units of area, where a unit called the barn (b) is most often used in high

[8] The loop can contain other $q\bar{q}$ pairs, but since the top quark is the most massive, it has the largest coupling to the Higgs boson.

energy physics such that

$$\begin{aligned} 1 \text{ b} &= 10^{-28} \text{ m}^2 \\ 1 \text{ fb} &= 10^{-15} \text{ b} \\ 1 \text{ pb} &= 10^{-12} \text{ b}. \end{aligned} \tag{1.26}$$

The production cross sections for the Higgs boson at the Tevatron are shown in Figure 1.5 for the different mechanisms. From this plot, it is seen that the GF process ($gg \to H$) is the most dominant production mode, followed by VH ($q\bar{q} \to WH$ and $q\bar{q} \to ZH$) and VBF ($q\bar{q} \to q\bar{q}H$). In this analysis, we search for a Higgs boson produced from any of these three mechanisms; their predicted cross sections are provided in Table 1.1 as a function of the Higgs boson mass.

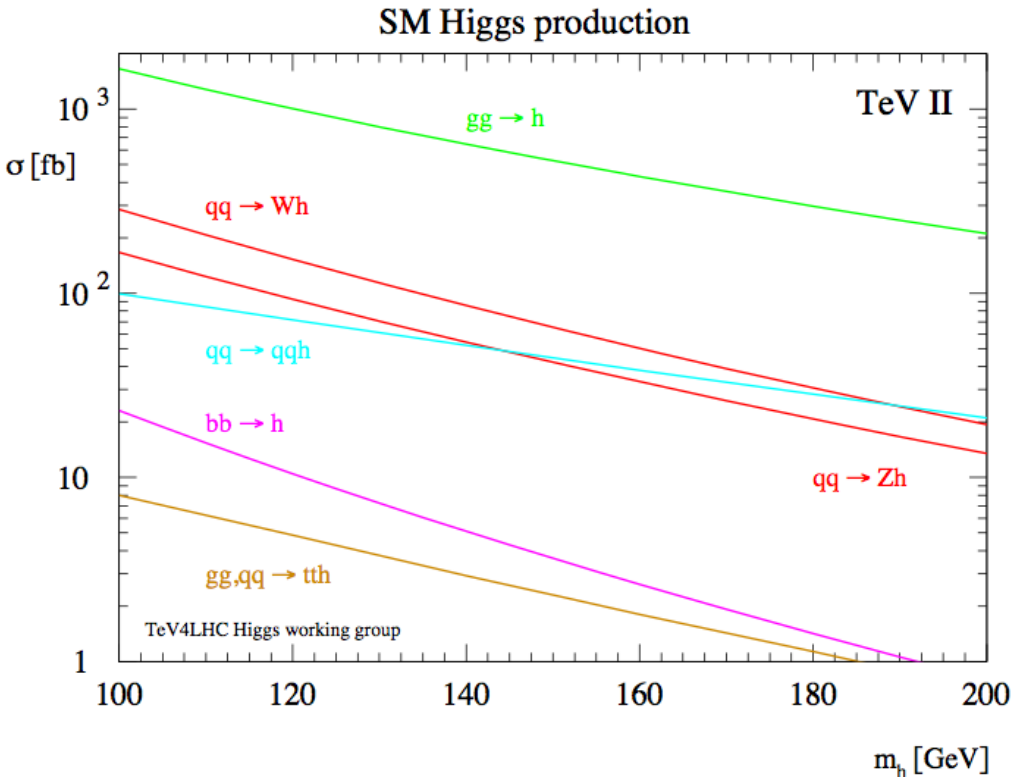


Figure 1.5: The dominant production mechanisms at the Tevatron for the SM Higgs boson [3, 25].

Table 1.1: For each Higgs boson mass hypothesis (m_H) tested, the SM prediction of the Higgs boson production cross section (σ) is provided for the GF, VH (WH and ZH), and VBF mechanisms, obtained from Reference [26] and references therein. The total SM Higgs production cross section for these mechanisms is the sum of the individual cross sections: $\sigma = \sigma_{GF} + \sigma_{WH} + \sigma_{ZH} + \sigma_{VBF}$. The diphoton branching ratio, $B(H \to \gamma\gamma)$, is also provided [27]. For the mechanisms that we consider in the analysis, the values in the last column provided the total SM $H \to \gamma\gamma$ cross sections, i.e., the SM prediction for the number of Higgs bosons decaying to photons to be produced in $p\bar{p}$ collisions at the Tevatron, per fb^{-1} of data.

m_H (GeV/c^2)	σ_{GF} (fb)	σ_{WH} (fb)	σ_{ZH} (fb)	σ_{VBF} (fb)	$B(H \to \gamma\gamma)$ (%)	$\sigma \times B(H \to \gamma\gamma)$ (fb)
100	1821.8	291.9	169.8	97.2	0.159	3.79
105	1584.7	248.4	145.9	89.7	0.178	3.68
110	1385.0	212.0	125.7	82.7	0.197	3.56
115	1215.9	174.5	103.9	76.4	0.213	3.35
120	1072.3	150.1	90.2	70.7	0.225	3.11
125	949.3	129.5	78.5	65.3	0.230	2.81
130	842.9	112.0	68.5	60.4	0.226	2.45
135	750.8	97.2	60.0	55.9	0.214	2.06
140	670.6	84.6	52.7	51.8	0.194	1.67
145	600.6	73.7	46.3	48.1	0.168	1.29
150	539.1	64.4	40.8	44.6	0.137	0.944

1.4.3 Higgs Boson Decay

The SM theory predicts that once the Higgs boson is produced, it decays very rapidly into other particles. Observation of the Higgs boson can only then be obtained through reconstruction of these decay products. A branching ratio (or branching fraction) is the probability of decay to a set of particles, given as a fraction between 0 and 1. This value is dependent on the coupling strengths, which are proportional to M_V^2 or M_f (see Equations (1.24) and (1.25)). Branching ratios are then largest for the heavier daughter products that are kinematically allowed, as shown in Figure 1.6 for different SM Higgs boson masses and decay modes.

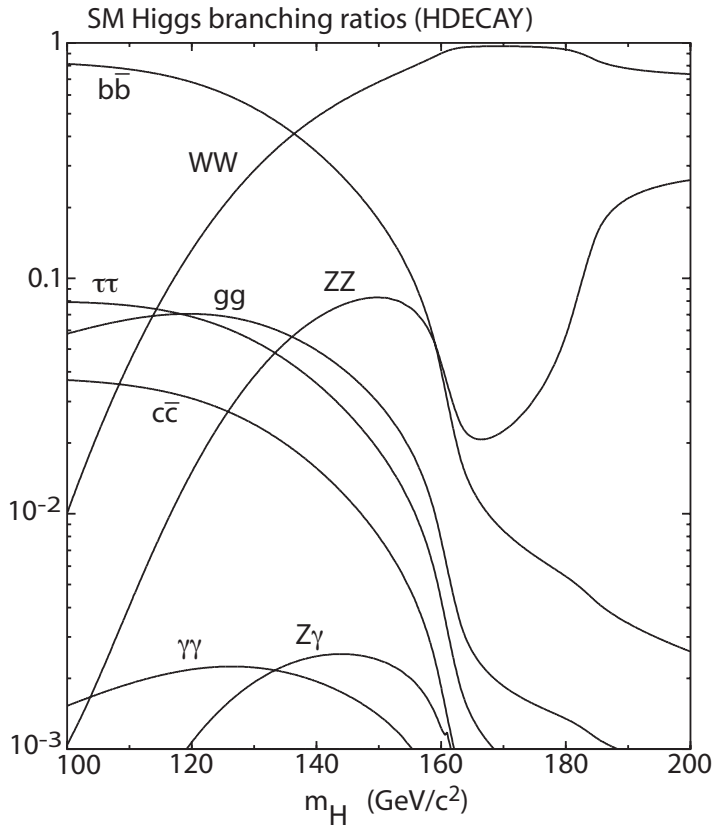


Figure 1.6: Possible decay modes for the SM Higgs boson for masses between 100 and 200 GeV/c^2 [27].

1.4.4 Higgs Boson Searches

The SM Higgs boson search strategy at the Tevatron is driven by the most dominant decay modes (the decay mechanisms with the largest branching ratios). These search channels provide the greatest predicted sensitivity for observing a Higgs boson signal in the data. Formally, the sensitivity of a particular search channel is determined from the calculation of an expected limit. The interpretation of expected limits is discussed in Chapter 11. At lower mass ($m_H < 135$ GeV/c^2), the $H \to b\bar{b}$ mechanism provides the greatest sensitivity; this is true even though the search channel generally includes only the WH or ZH production methods, which help better distinguish the desired $b\bar{b}$ pair from other processes. At higher mass ($m_H > 135$ GeV/c^2), the $H \to WW$ mode provides the greatest sensitivity. Further sensitivity to a SM Higgs

boson observation is obtained by the inclusion of more challenging search modes such as $H \rightarrow \tau^+\tau^-$, gg or $q\bar{q} \rightarrow t\bar{t}H$, and $H \rightarrow \gamma\gamma$. The combination of these more challenging channels together provides a sensitivity like that of a dominant search mode and each, therefore, is considered an important contribution to the overall Higgs boson search at the Tevatron.

Because the Higgs boson does not couple directly to photons (since they are massless), the $H \rightarrow \gamma\gamma$ decay mode is obtained most often through a W loop, as shown in Figure 1.7. As Figure 1.6 illustrates, the branching ratio for the $H \rightarrow \gamma\gamma$ channel has its primary sensitivity for Higgs boson masses between 110 and 140 GeV/c^2 and is, therefore, considered a "low mass" search channel. Its branching fraction peaks at $\sim$125 GeV/c^2 with a value of 0.23% (see Table 1.1). This is a very small branching fraction; however, the clean profile ("signature") a photon leaves in particle detectors makes it an appealing and competitive search mode. As described in Section 1.1, b quarks will fragment to produce jets in the detector, which are difficult to fully reconstruct. Photons, on the other hand, deposit an isolated cluster of energy in the detector, which allows for a larger fraction of $H \rightarrow \gamma\gamma$ events to be reconstructed compared to the $H \rightarrow b\bar{b}$ process. The clean photon signature also leads to a narrow diphoton mass ($m_{\gamma\gamma}$) peak in the data, which is a powerful discriminant against smoothly falling diphoton backgrounds from other SM processes with two photons (or photon-like objects). It is these experimental features that help make the diphoton final state one of the most promising search modes for a Higgs boson discovery at the ATLAS and CMS experiments at CERN for $m_H < 135$ GeV/c^2.

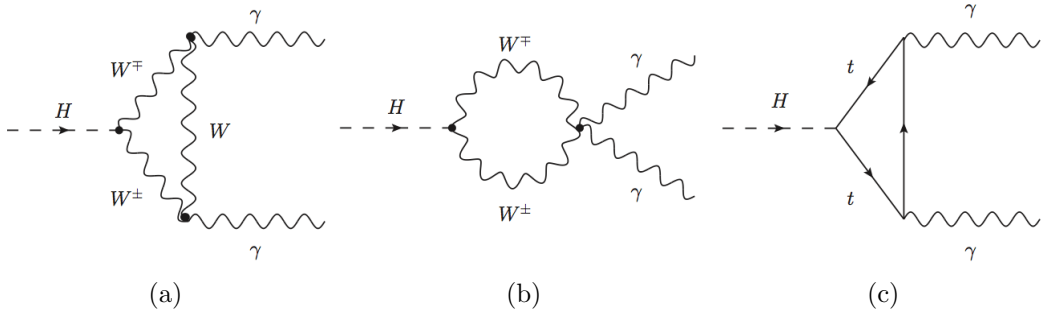


Figure 1.7: Diagrams of the Higgs boson decay to a pair of photons, obtained most often through a W or top loop [28].

Figure 1.8 shows the result of an indirect measurement of the Higgs boson mass based on electroweak constraints (as of July 2011). This plot indicates that the SM theory prefers a Higgs boson with lower mass (where the $\Delta\chi^2$ approaches a minimum), in a region where the diphoton channel provides sensitivity. In particular, if the Higgs boson exists, the SM theory prefers a Higgs boson mass below about $m_H < 169$ GeV/c^2 at 95% confidence level (C.L.) [29]. This figure also indicates the mass region below 114.4 GeV/c^2 that has been excluded at 95% C.L. by direct searches at the Large Electron-Positron Collider (LEP) at CERN.[9] Furthermore, Figure 1.8 shows a region excluded by direct searches from combined D0 and CDF data at the Tevatron as of July 2011.

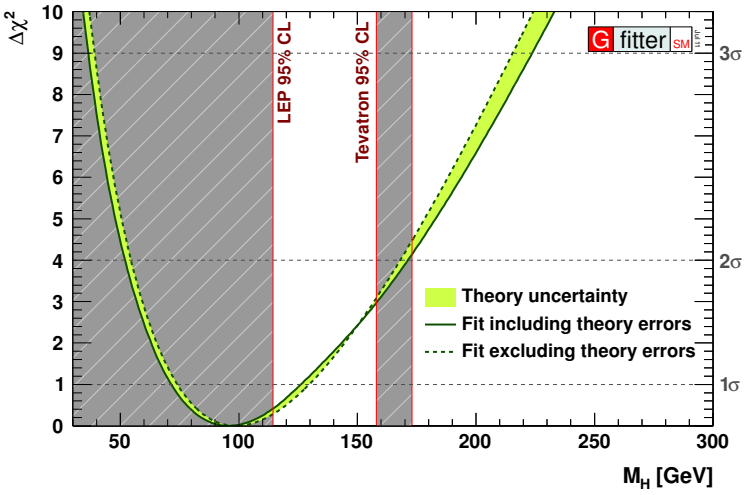


Figure 1.8: An indirect constraint on the Higgs boson mass is obtained from a global electroweak fit performed by the Gfitter collaboration. From July of 2011, the Gfitter collaboration uses a $\Delta\chi^2$ statistic to fit electroweak parameters of the SM theory to the corresponding measured values [29]. The preferred value of the Higgs boson mass from this indirect electroweak measurement is the mass for which the $\Delta\chi^2$ is minimized (96 GeV/c^2). With 95% confidence level (see the 2σ horizontal line), this plot demonstrates that if the Higgs boson exists, the SM theory prefers its mass to be below about $m_H < 169$ GeV/c^2. See Reference [30] for a preliminary updated plot using results from more recent direct searches and electroweak measurements.

[9] The exclusion regions quoted in this section are obtained from either a frequentist or Bayesian technique. The former are technically obtained with a 95% confidence level and the latter are technically obtained with a 95% credibility level. The acronym for both is C.L.

In March 2012, the Tevatron results were updated to exclude the region between $147 < m_H < 179$ GeV/c^2 at 95% C.L. [31]. Furthermore, the ranges 110.0–117.5 GeV/c^2, 118.5–122.5 GeV/c^2, and 127–600 GeV/c^2 were excluded at 95% C.L. by direct searches from the ATLAS and/or CMS experiments [32, 33], both of which are located at the Large Hadron Collider (LHC) at CERN. The range of possible mass values for the Higgs boson is rapidly decreasing and hints of an excess in the low mass range ($115 < m_H < 135$ GeV/c^2) appear both in the Tevatron and CERN data as of this writing [31, 32, 33].

In this thesis, we present a search for the SM Higgs boson decaying to a photon pair in data obtained by the Collider Detector at Fermilab (CDF). Though the sensitivity for observing $H \to \gamma\gamma$ signal is smaller at the Tevatron than at the LHC, the results provide a statement of the sensitivity of the CDF data to the $H \to \gamma\gamma$ process, demonstrating the unique experimental properties of a search in the diphoton decay mode relative to other search modes. The results presented were combined with other Higgs boson searches at CDF and D0 in order to gain as much sensitivity as possible to a Higgs boson observation at the Tevatron [31]. The techniques described have been published using a smaller set of data [34] and the results shown here are an updated version using the full (and final) data available at CDF.

Chapter 2

The Tevatron Collider

2.1 History

The Tevatron[1] was the world's most powerful collider from a time period between July 1983 [35] and November 2009 [36], and was the second most powerful collider[2] from 2009 until it shut down on September 30, 2011. The Tevatron began accelerating protons to an energy of 512 GeV in 1983, reaching a world record at that time. It was not until 1992, however, that the first significant collider physics program began at Fermilab ("Run I") in which data were collected from both the CDF and D0 experiments from $p\bar{p}$ beams with $\sqrt{s} = 1.8$ TeV. The Run I physics program lasted from 1992–1996, during which the top quark was discovered by CDF and D0. After many upgrades, the second significant collider physics program ("Run II") began at Fermilab with $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. This run started in 2001 and ended on September 30, 2011. In addition to particle discoveries and precision measurements during this run, the combined data from the CDF and D0 experiments began excluding mass regions for the Higgs boson, suggesting a light Higgs boson

[1] The dates provided here for the Tevatron $p\bar{p}$ collider physics program are obtained from Reference [35], which is a Fermilab website providing an interactive timeline for Tevatron physics.

[2] The first pp collisions at the LHC occurred on November 30, 2009 with $\sqrt{s} = 2.36$ TeV [36].

with a mass between 114 GeV and 157 GeV. Physicists at CDF and D0 continue to analyze Run II data delivered from Tevatron $p\bar{p}$ collisions between 2001–2011.

2.2 Acceleration Chain

The acceleration chain for the $p\bar{p}$ collisions at the Tevatron is described next.[3] Several parts of this chain are still in operation at Fermilab for other accelerator physics programs. Use of the past tense in this description then refers to the acceleration chain for $p\bar{p}$ collisions.

The proton and antiproton beams were obtained from a chain of accelerators shown in Figure 2.1, starting with a standard tank of hydrogen gas located at the Cockcroft Walton preaccelerator. The preaccelerator first ionized the hydrogen to make H^- ions and then passed them to a short linear transfer line that accelerated them to an energy of 750 keV using a constant electric field. They were then transferred to a long linear accelerator called the Linac. The Linac used alternating electric fields to further accelerate the ions to 400 MeV ($\sim$70% the speed of light) along a 150 m linear radio frequency (RF) cavity. At the end of the cavity was a carbon foil that stripped the H^- ions of their electrons, leaving only protons.

In order to accelerate these protons to near the speed of light, a linear cavity would have to be extremely long;[4] the remaining accelerators were therefore circular. The first of the chain of circular accelerators was the Booster, with a circumference of 1,500 ft, where protons traveled about 20,000 times to increase their energy from 400 MeV to about 8 GeV. The protons were then transferred to the Main Injector, with a circumference of about 2 mi. The Main Injector ring had four purposes: (i) it accelerated 8 GeV protons from the booster to 150 GeV, (ii) it accelerated 8 GeV protons from the booster to 120 GeV, which were then used for antiproton production, (iii) it accelerated 8 GeV antiprotons from the antiproton source or Recycler to

[3] References [2] and [37] are recommended for more details on the acceleration chain.

[4] The longest linear accelerator (the Stanford Linear Collider) is 2.0 mi long and accelerates electrons and positrons to 50 GeV.

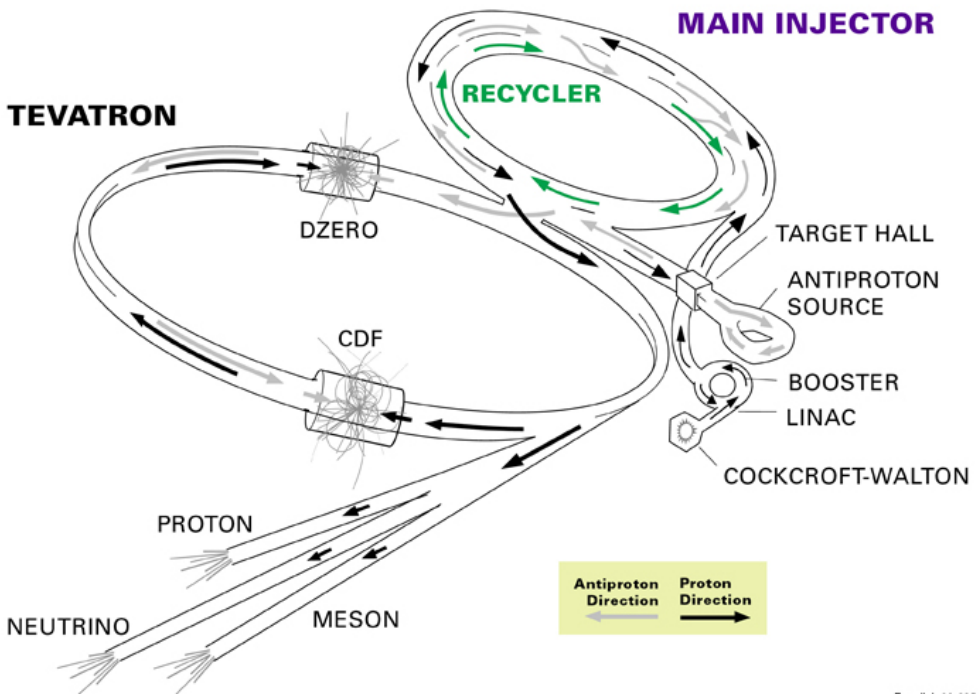


Figure 2.1: Fermilab's $p\bar{p}$ accelerator chain [2, 37]. See the text for more detail.

150 GeV, and (iv) it transferred 150 GeV protons and antiprotons to the Tevatron ring. In the final acceleration step, the Tevatron increased the energy of the proton and antiproton beams (moving in opposite directions) to an energy of 980 GeV, 99.999954% of the speed of light. This provided a center of mass energy of 1.96 TeV for collisions between the beams.

The Tevatron is a circular tunnel ∼4 mi in circumference, located about 25 ft below ground. It contains a vacuum pipe for the proton and antiproton beams to travel through. Superconducting electromagnets surround most regions of this beam pipe. There were over 1,000 such magnets used, kept at a temperature of about 4 K. The beams maintained their (slightly) circular path due to dipole magnets. Quadrupole magnets were used for focusing the beams to a transverse[5] cross section small enough for collisions. Fine positioning of the beams in the transverse plane was achieved with correcting magnets.

[5] In the cylindrical tunnel geometry, the z axis is along the circular tunnel path and the transverse plane (the $r - \phi$ plane) is perpendicular to the z axis.

When this positioning was complete, the beams were ready for collider physics. These proton and antiproton beams remained in the Tevatron ring, cycling over and over until there were not enough particles per beam remaining for collisions to be produced at a sufficient rate. The beams were then carefully dumped.[6] The time period for a pair of beams to be used for collisions was called a store, and was dependent on the initial number of protons and antiprotons per beam. Stores with a higher number of initial particles may last roughly 18 to 20 hours, while those with much smaller initial number of particles may last roughly 8 to 10 hours.

2.3 Antiproton Production and Storage

While a store was in the Tevatron, a set of antiprotons would be accumulating to be used in the next store. The larger the number accumulated, the more intense the beam would be (i.e. the higher its instantaneous luminosity, as will be discussed in Section 2.5). A large number of antiprotons may take roughly 18 hours to accumulate, though smaller amounts were often used. This section discusses how they were obtained.

Protons with 120 GeV of energy from the Main Injector were directed to a fixed nickel target, which produced many secondary particles (including antiprotons) with a large angular spread (see Figure 2.2). A lithium lens focused these particles, and dipole magnets separated 8 GeV antiprotons from the other particles. Only about 1 antiproton was obtained for every 50,000 protons sent to the target. The unused particles were sent to the beam dump and the 8 GeV antiprotons were sent to the Debuncher, which resides on the triangular storage ring shown in Figure 2.1. The momentum spread of the antiprotons was large and they were subject to a cooling in the Debuncher, which helped reduce this spread. Antiprotons produced in this manner were transferred to the Accumulator, a storage area which also resides on this ring. The cooling process continued in the Accumulator as more antiprotons

[6] Sometimes a mechanical or electrical malfunction would actually cause the beam to be dumped by accident.

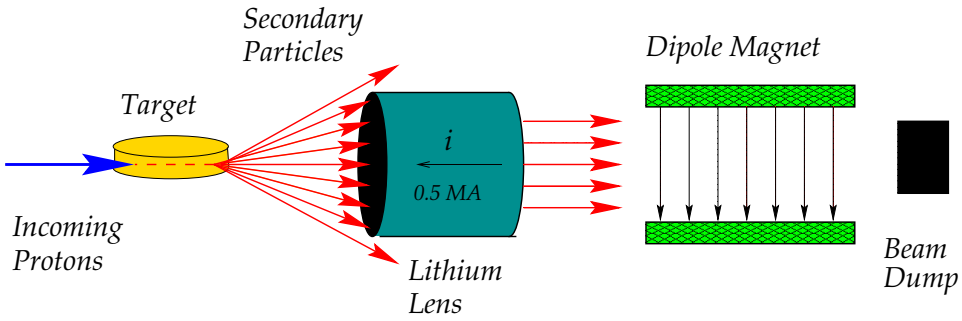


Figure 2.2: Mechanism for antiproton extraction [37]. Protons were sent to a fixed target, a lithium lens was used to reduce the spread of secondary particles, and a dipole magnet was used to separate antiprotons with 8 GeV of energy. These antiprotons were sent to the Debuncher (see text) and all other secondary particles were sent to an area for disposal (a beam dump).

were transferred from the target, maintaining and further reducing the momentum spread.

When a sufficient number had been accumulated, they were then either transferred to the Main Injector for acceleration or sent to a larger storage ring called the Recycler, which resides in the same ring as the Main Injector. The antiprotons stored in this ring were kept at an energy of 8 GeV.

2.4 Beam Structure

Both proton and antiproton beams moved in a closed helical path inside the Tevatron beam pipe, but in opposite directions. Their paths intersected at two locations along the Tevatron ring, B0 and D0, which are shown in Figure 2.3. The CDF experiment has a detector located at B0. The name of the D0 experiment comes from the location of its detector along the ring. This figure also shows how each beam was actually divided into three trains of 12 bunches, giving 36 proton bunches and 36 antiproton bunches. The bunches were separated such that there was 396 ns between each bunch crossing; this gave a bunch crossing rate of 2.5 MHz.

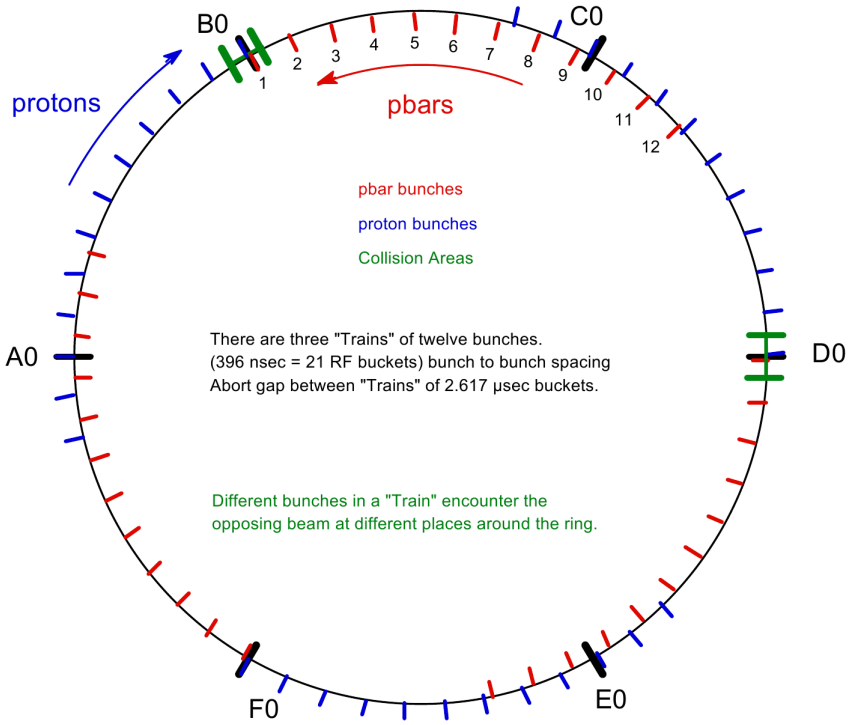


Figure 2.3: Proton and antiproton beams are shown [37] divided into three trains of 12 bunches each. This gave 36 proton bunches and 36 antiproton bunches. A gap between trains allowed for beams to be cleanly dumped (aborted).

2.5 Luminosity

One of the goals of the Accelerator Division at Fermilab was to improve the density and stability of the proton and antiproton beams in order to provide more collisions for the CDF and D0 experiments. Over the course of Run II, for example, a major improvement was obtained by increasing the initial number of antiprotons available for a store (on the order of 10^9). Other improvements included decreasing the size and spread of the beams, using momentum scraping of the two beams, and reducing proton and antiproton losses throughout a store [38]. This increased a quantity called the instantaneous luminosity, a plot of which is shown in Figure 2.4 for Run II. The Tevatron continually broke world records for reaching the highest instantaneous luminosity of any collider, the last of which was in April 2010.

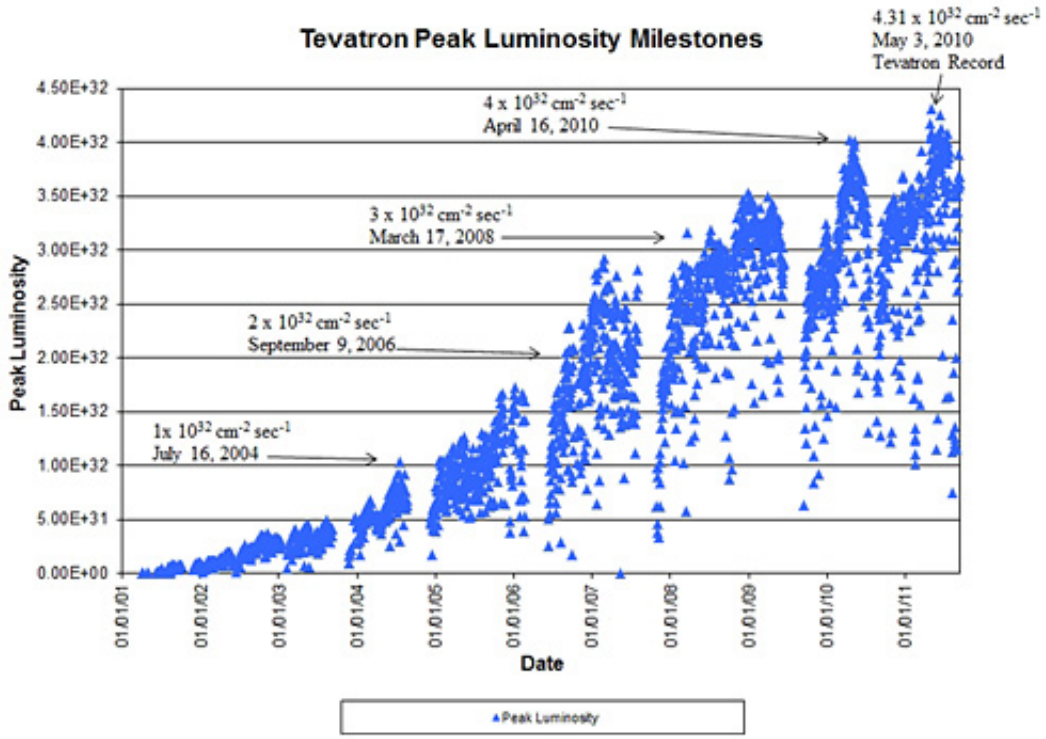


Figure 2.4: Instantaneous luminosity for each store over the course of Run II [35].

Instantaneous luminosity can be thought of as follows. If there were a single proton bunch with N_p particles and a single antiproton bunch with $N_{\bar{p}}$ particles, and they were allowed to travel around the ring (in opposite directions) and overlap just once, then the number of particles passing each other per unit area (assuming no crossing angle) would be given by

$$\frac{N_p N_{\bar{p}}}{A} \tag{2.1}$$

where A is the transverse area of overlap between the two beams. The $p\bar{p}$ beams, however, each consist of n bunches and are guided to overlap more than once, at a bunch crossing rate of f_{BC} times per second. The number of particles that pass each other per unit area per second is called the instantaneous luminosity (L) and is given by

$$L = f_{\mathrm{BC}}\, n\, \frac{N_p N_{\bar{p}}}{A}. \tag{2.2}$$

The collision rate (the number of interactions per second) is derived from the product of the instantaneous luminosity and the likelihood of interaction between

particles in the $p\bar{p}$ beams. The latter is the total $p\bar{p}$ cross section ($\sigma_{\text{tot}} = 81.90 \pm 2.30$ mb for $\sqrt{s} = 1.96$ TeV [39]), and a subset of these collisions will produce an inelastic scattering process of interest, such as the production of the Z (or Higgs) boson. The number of such interactions per second ($\dot{N}$) for a physics process with cross section σ is given by

$$\dot{N} = \sigma L. \tag{2.3}$$

The total number of interactions (N) for a physics process is then obtained from

$$N = \sigma \mathcal{L} \tag{2.4}$$

where

$$\mathcal{L} = \int L \, dt \tag{2.5}$$

is referred to as the integrated luminosity and describes the quantity of data available. Over the course of Run II, the Tevatron delivered an integrated luminosity of about 12 fb^{-1} to both the CDF and D0 experiments.[7]

As an example, the SM predicts a Higgs boson produced from the gluon fusion process ($p\bar{p} \to gg \to H$) to have a cross section of 1,072.3 fb^{-1} for a mass of 120 GeV/c^2. With an amount of data corresponding to $\mathcal{L} = 10$ fb^{-1}, we would expect 10,723 Higgs bosons to be produced just from the gluon fusion process. Only a tiny fraction of these (0.225%) would decay to photons as predicted by the SM, which means that about 24 events are obtained from the $p\bar{p} \to gg \to H \to \gamma\gamma$ interaction for 10 fb^{-1} of data. As will be described later, only a portion of these are reconstructed by the CDF detector using diphoton selection requirements.

2.6 Particle Detectors

Collisions between partons of the proton and antiproton produce highly energetic particles that cannot be studied in lower energy particle experiments. The high center of mass energy of the $p\bar{p}$ beams allows for the production of massive objects

[7] The integrated luminosity is often truncated throughout this paper as simply luminosity.

such as the top quark, the W and Z bosons, and perhaps the Higgs boson. Such collisions might also produce new phenomena that is not described by the SM. Data obtained from Tevatron collisions has fueled a large amount of research at Fermilab surrounding precisions measurements of the properties of known particles and the search for new particles such as the Higgs boson.

Often, the products of Tevatron collisions produce unstable particles that quickly decay to more stable particles. Detectors (of about 5,000 tons) are built around the collision point (or points) in order to measure the properties of final-state particles such that the particles produced in the primary interaction can be reconstructed. Each detector is designed in a similar manner, and having two such detectors provides the ability to compare methods and results between the two collaborations. The results presented here are from data collected by members of the CDF collaboration (see Appendix E). This group is comprised of about 430 scientists from $\sim$60 institutions around the world.

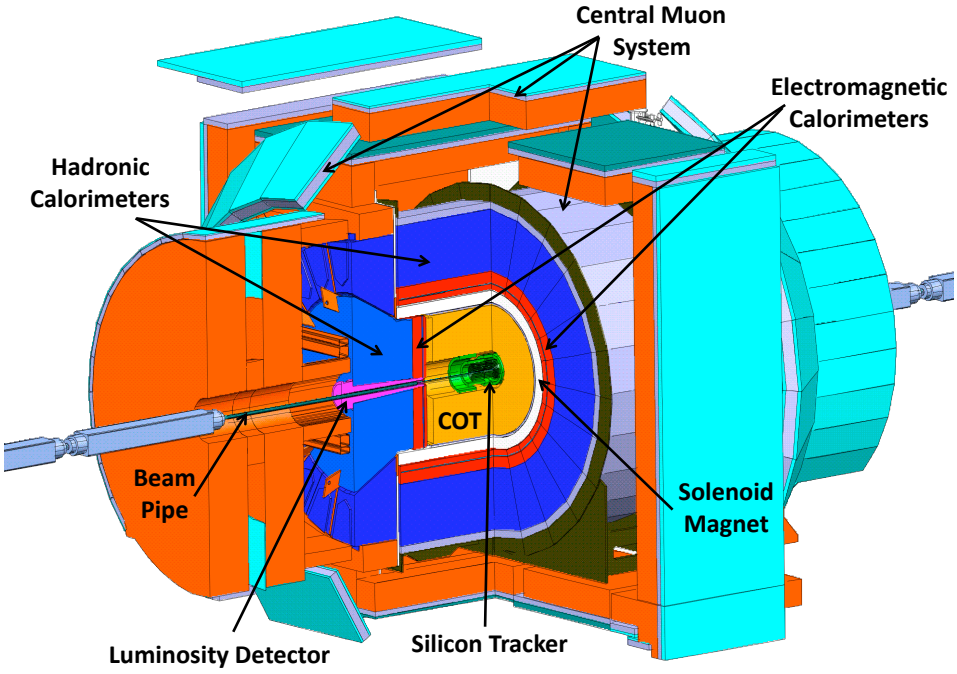
Chapter 3

The CDF Experiment

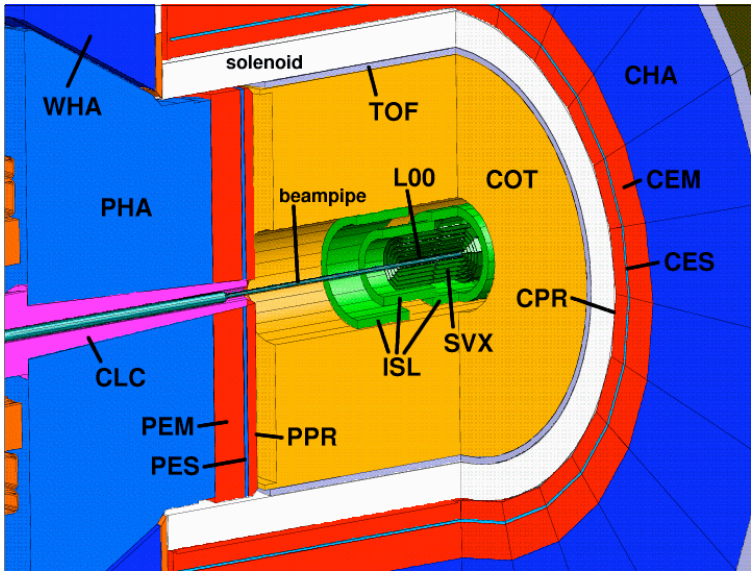
3.1 Detector Description

3.1.1 Overview

The CDF detector (Figure 3.1) is designed to measure the properties of final-state particles produced in $p\bar{p}$ collisions. It is a cylindrical detector, built in layers surrounding the collision region. The inner region contains two detectors designed to measure the path of a charged particle (called a track) and to identify the origin of the $p\bar{p}$ interaction (the primary vertex). These detectors are surrounded by a 1.4 tesla superconducting solenoid magnet. As a charged particle passes through the tracking detectors, its path bends to produce a curved, helical trajectory. Surrounding the tracking detectors and magnet are calorimeters designed to stop both neutral and charged particles and measure their energy. The calorimeters are divided into electromagnetic (EM) calorimeters and hadron (HAD) calorimeters. Charged particles that are not stopped by the calorimeters are most likely muons, and therefore special detectors (the muon chambers) are built in order to stop and measure the energy of these particles. Neutrinos do not interact with the detector, but their presence is inferred through an imbalance in the measured momentum of other particles.



(a)



(b)

Figure 3.1: Schematic of the CDF detector [40]. See the text for more detail.

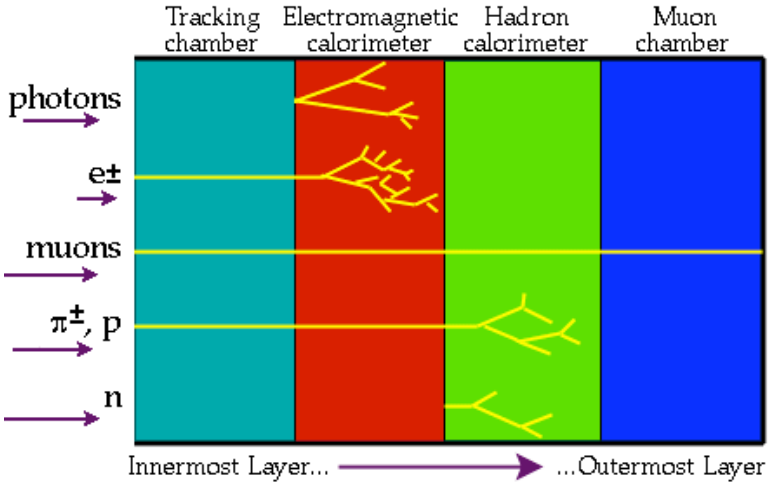


Figure 3.2: Diagram [40] showing which detector regions are sensitive to different types of particles. The left side of the figure corresponds to the innermost cylindrical layer of the detector, near the beam line, and the right side of the figure corresponds to the outermost layer of the detector.

Figure 3.2 shows an example of how different types of particles can be identified using different regions of the detector. Photons will only interact with material in the EM calorimeter. Electrons have an essentially identical signature as photons in the EM calorimeter, but because they are charged, they also produce a track that points to the energy cluster in the calorimeter. Charged hadrons have a track in the tracking chamber, interact only a small amount in the EM calorimeter, and deposit the majority of their energy in the HAD calorimeter. (Charged pions and a proton are shown as an example in the figure.) Neutral hadrons, such as a neutron, mostly interact in the HAD calorimeter.

The coordinate system used for the detector is shown in Figure 3.3. The z axis is along the beam line with $+z$ in the direction that protons travel, phi (ϕ) is the azimuthal angle, and θ is the polar angle. The physical center of the detector is at $x = y = z = 0$. For a particular type of interaction, stable particles have a predicted distribution in a kinematic quantity called rapidity which is given by

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{3.1}$$

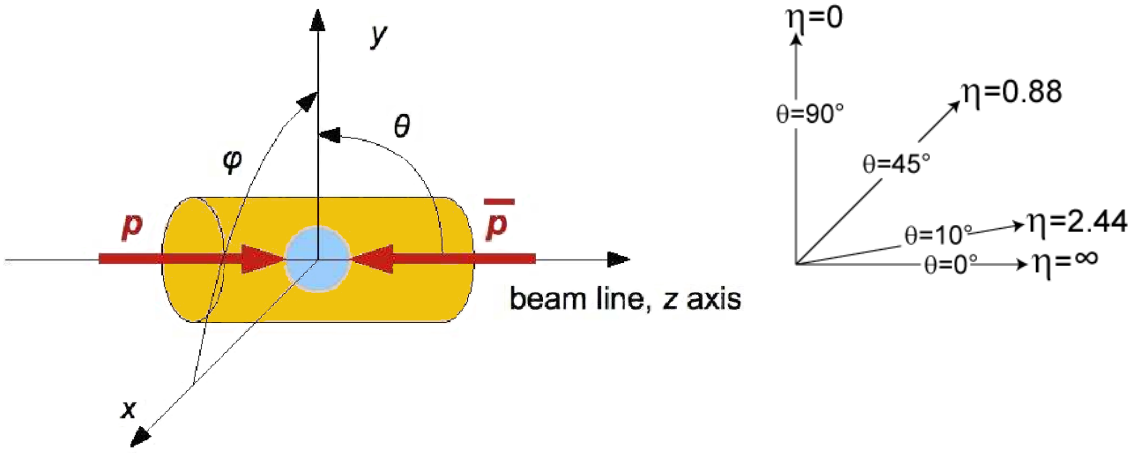


Figure 3.3: The CDF coordinate system [40].

where Δy is invariant under a Lorentz boost along the z axis. These particles generally have high enough momenta such that $E \approx p$ (the massless limit). The rapidity is then approximated by a spatial quantity called pseudorapidity,

$$\begin{aligned} \eta &= \frac{1}{2}\ln\left(\frac{p+p_z}{p-p_z}\right) = \frac{1}{2}\ln\left(\frac{1+\cos\theta}{1-\cos\theta}\right) \\ &= -\ln\left[\tan\frac{\theta}{2}\right] \end{aligned} \tag{3.2}$$

with $\cos\theta = p_z/p$. The polar angle is, therefore, directly mapped to pseudorapidity and we commonly replace references to θ with references to η. However, kinematic quantities in the transverse plane (the plane perpendicular to the z axis) are defined from θ as $p_T = p\sin\theta$ and $E_T = E\sin\theta$.

A more detailed description of detector components in Figure 3.1 is now provided, with an emphasis on those detectors most relevant for photon identification.

3.1.2 Tracking

The path of charged particles through the detector can be reconstructed from patterns of position measurements ("hits") made in the tracking detectors. Complex algorithms [41] are used to reconstruct a track from these hits to form a helical curve described by five parameters (see Section 3.4.1). Parametrized tracks provide the momentum of a particle, the position of its origin (vertex), and errors on these quantities.

Measurements from two different detectors at CDF are used to reconstruct charged particle tracks. The inner detector is made of silicon microstrips and the outer detector, called the Central Outer Tracker (COT), is a wire drift chamber. Figure 3.4 shows the η coverage and positions of these detectors.

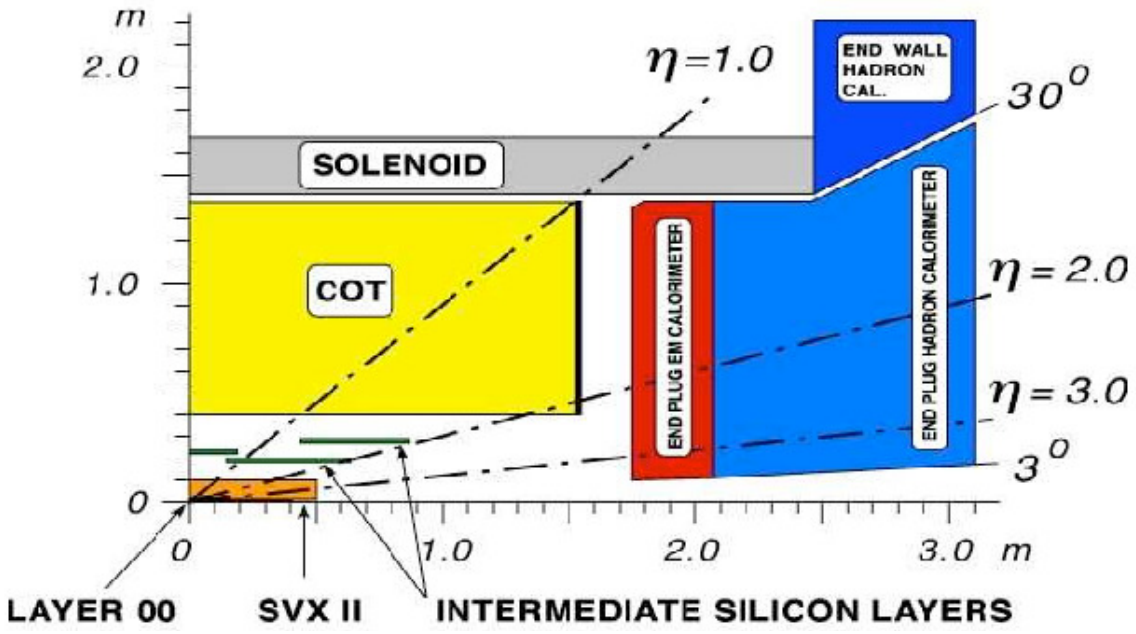


Figure 3.4: An $r-z$ view of one quadrant of the detector showing the CDF tracking coverage [40]. The COT detector provides a coverage of $|\eta| < 1.0$ and the silicon detector provides a coverage of $|\eta| < 2.0$.

Silicon tracking

The inner tracking region is composed of silicon microstrip sensors layered in a cylindrical (or barrel) geometry around the beam line. As a charged particle passes a silicon sensor, it ionizes molecules on the surface, creating electron/hole pairs in the silicon material. In the presence of an applied electric field, the charge carriers in the silicon move to produce a current which can be amplified, creating a silicon hit.

There are three parts to this detector: Layer 00 (L00), the Silicon Vertex Detector (SVXII), and the Intermediate Silicon Layers (ISL). The combined silicon system

provides coverage up to $|\eta| < 2$ and coverage along the z axis such that nearly all $p\bar{p}$ interactions fall into a well-instrumented region of the detector. Their positions are shown in Figures 3.1 and 3.4, with more detail given in Figure 3.5.

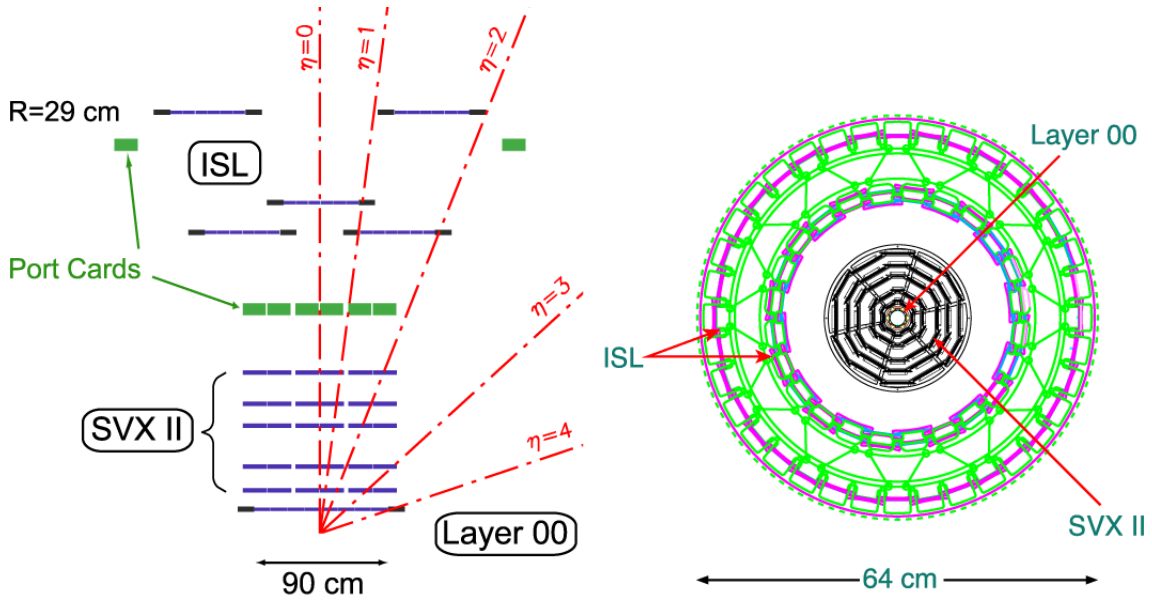


Figure 3.5: Schematic [40] of the silicon track detector in the $r-z$ view (left) and $r-\phi$ view (right).

The first layer, the L00 detector, touches the beam pipe and is closest to the beam line and the primary collision. It is designed to perform well under high radiation doses and is the only layer with single-sided microstrips. The SVXII detector consists of the next five layers of silicon and the ISL consists of the remaining three, all of which contain double-sided microstrips and allow three-dimensional reconstruction of particle paths. The ISL extends the coverage of the silicon tracking region up to $|\eta| = 2.0$ and improves the ability to link silicon tracks with those from the outer tracker.

The silicon layers together provide precise measurements of a track's impact parameter (d_0) and z_0 position (see Section 3.4.1). This ensures good vertex resolution and allows identification of displaced vertices from particles decaying in the inner tracking region of the detector. The d_0 resolution is 25 μm and the z_0 resolution is 50 μm. The silicon layers also allow for silicon-only tracking out to an $|\eta|$ of 2.0.

The Central Outer Tracker

The COT is a cylindrical wire chamber (Figure 3.6) that surrounds the silicon region and provides a coverage of $|\eta| < 1.0$. It is designed (i) for general purpose tracking of charged particles at large radii and (ii) to provide accurate momentum reconstruction of charged particles. The chamber is composed of 8 superlayers of wire cells, where each cell contains anode sense wire panels adjacent to cathode field wire panels, as shown in Figure 3.7. The COT detector is filled with a gas mixture. As charged particles traverse the controlled electric fields of the chamber, the gas is ionized and electrons from the gas molecules drift toward a sense wire. This produces a current in the sense wire which is amplified, creating a COT hit.

As charged particles travel through the COT detector, they produce hits on sense wires along the way. For each superlayer, a portion of the particle's path is constructed

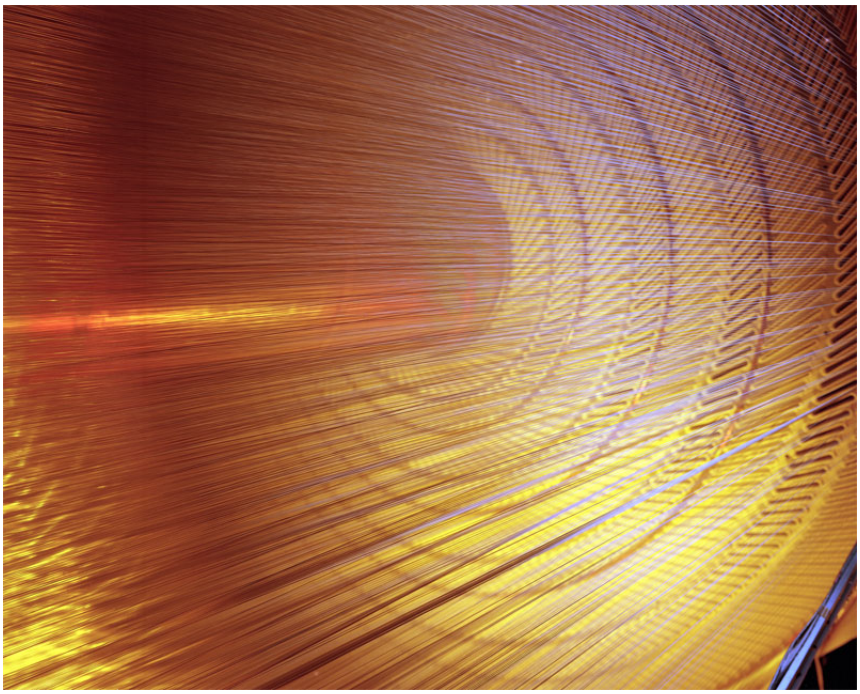
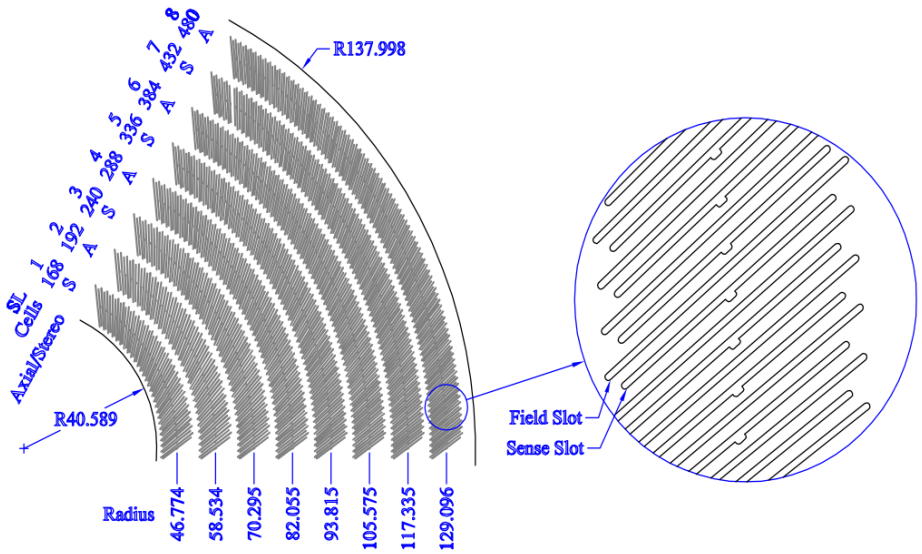
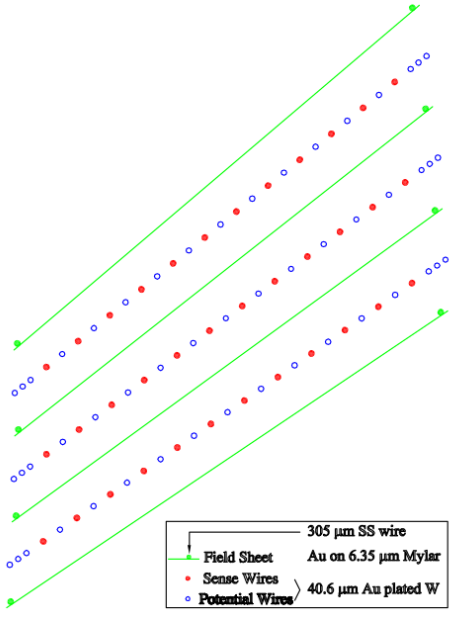


Figure 3.6: A picture of CDF's Run I Central Tracking Chamber (CTC), which sits in the museum area of the main CDF building (B0) [40]. The Run II COT design is very similar, but with a larger number of cells and a larger number of wires for stereo layers, allowing for a smaller maximum drift distance [42].



(a)



(b)

Figure 3.7: (a) The upper figure shows the a portion of the COT in the $r-\phi$ plane. The number of cells per superlayer is given, along with whether the superlayer has axial or stereo wires. The radius of each layer relative to the beam line is also shown (in cm). The blow-up subfigure shows the geometry of the sense and field planes. (b) The lower figure provides a diagram demonstrating more detail of the field/sense planes, where three cells are shown. Each figure shown here is from Reference [42].

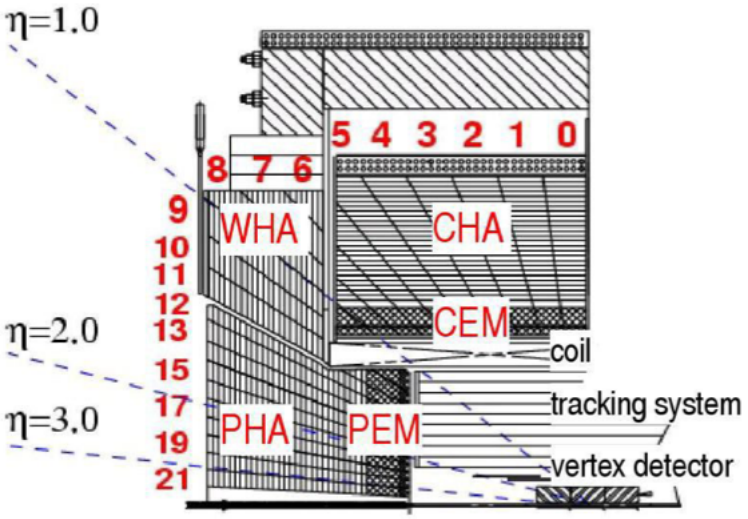
based on these hits, creating a segment. Linking segments together contributes to the track reconstruction. As Figure 3.7 illustrates, there are four even-numbered axial superlayers and four odd-numbered stereo superlayers. The former have wires that are positioned parallel to the beam line and, therefore, only provide tracking reconstruction for the $r - \phi$ plane. The stereo wires are angled relative to the z axis to allow for three-dimensional track reconstruction.

For $|\eta| < 1.0$, the COT provides 96 possible hit measurements between a radius of 44 and 132 cm from the beam line. The hit resolution is about 140 μm, which allows for good momentum resolution, $\Delta p_T/p_T^2 \sim 0.0015$ (GeV/c)$^{-1}$. The d_0 and z_0 resolutions are 250 μm and 5 mm, respectively.

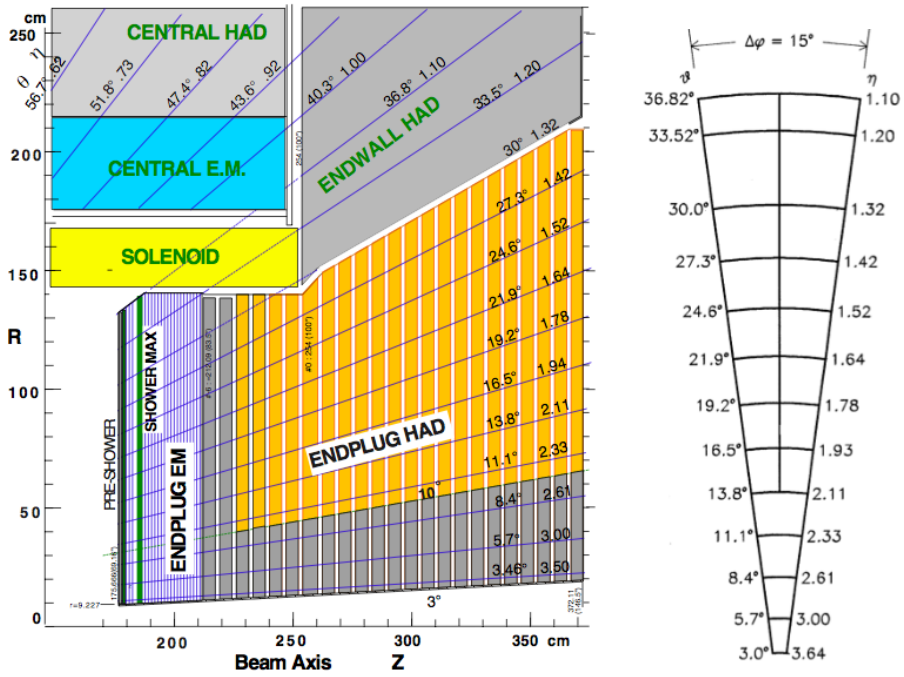
3.1.3 Calorimetry

The calorimeters are built outside of the solenoid with the purpose of measuring the energy of both neutral and charged particles. As incident particles enter the region of the calorimeters, they interact with dense slabs of material with large atomic Z, called an absorbing material because its purpose is to rapidly stop particles. Interactions with this material produces a shower of other particles with lower energy, a cascade that continues until an energy threshold is met (dependent on the material). At CDF, calorimeters contain alternating layers of the absorber and a scintillating material. As particles from the shower pass through the calorimeter, a portion of its energy is absorbed by the scintillating material which emits this energy in the form of light (luminescence). This light is collected by photomultiplier tubes (PMTs) and is converted to an electric analog signal that is amplified for electronic readouts. The total energy of an incident particle or incident particles can be determined from these readouts.

Figures 3.8 (a) and (b) show one quadrant of the CDF calorimeter system, divided into five parts, extending out to $|\eta| = 3.6$. The CEM and PEM are the central and (end)plug EM calorimeters, respectively, and the CHA, WHA, and PHA are the central, endwall, and (end)plug hadron calorimeters, respectively.



(a)



(b) (c)

Figure 3.8: (a) One quadrant of the CDF detector showing the different calorimeter subdetectors. (b) Similar to (a), but more detail is shown for the plug calorimeters [40]. (c) The plug calorimeter segmentation for a single ϕ wedge, also showing θ and η boundaries for each tower in the wedge [43].

Figure 3.9 shows one slice of the central calorimeter; this slice is called a wedge and covers an azimuthal angle ϕ of 15°. The full transverse plane is covered with 24 wedges, minus small cracks between wedges. The upper (lower) region of the wedge is a portion of the CHA (CEM), where it is seen that the amount of material needed to fully contain a hadronic shower is much higher. The wedge has subregions corresponding to values of η, which are each aligned to point back to the center of the detector near the primary collision. The η divisions can also be seen in Figure 3.8, which shows 22 total η regions with the numbering scheme shown in red in the upper subfigure. The divisions in η and ϕ are called towers and determine the sampling segmentation of the calorimeter detectors. Some portions of plug wedges are actually divided into two subregions to provide slightly finer segmentation, as shown in Figure 3.8 (c). The granularity of the calorimeters is such that an EM shower generally deposits energy in only a few towers while a hadronic shower deposits energy

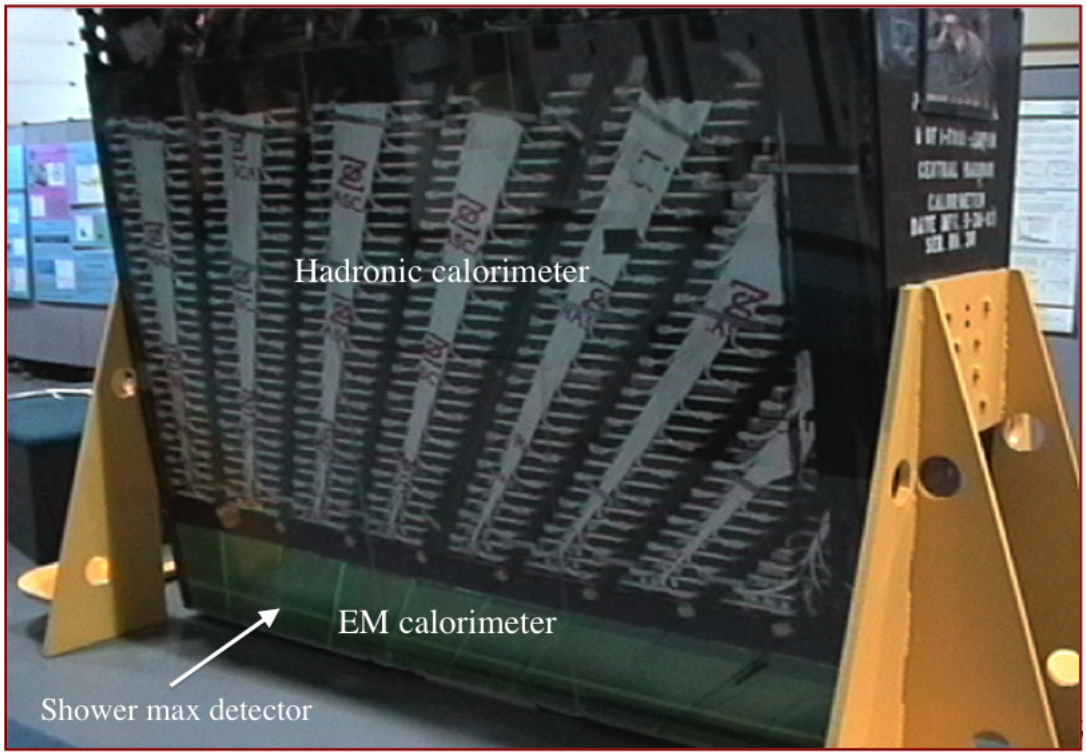


Figure 3.9: An example of a central calorimeter wedge [40]. The lower (upper) region is a portion of the EM (HAD) calorimeter. The shower maximum detector is located six radiation lengths into the EM calorimeter.

over many towers. This granularity also provides a coarse position measurement for particle showers.

A more detailed description is provided next for the EM and HAD regions, with an emphasis on the EM calorimeters, which measure photon energies. This is followed by an explanation of the shower maximum detector, located within the EM calorimeter as shown in Figures 3.1, 3.8 (b), and 3.9.

Electromagnetic calorimeters

Particles such as photons and electrons[1] interact with lead slabs in the EM calorimeter to produce a rapid cascade of lower-energy particles. Figure 3.10 shows an example of this. Electrons interact with lead nuclei via an electromagnetic process called bremsstrahlung. The first diagram of Figure 3.11 shows this process where an electron interacts with a lead nucleus to lose energy through photon radiation.[2] Photons interact electromagnetically with lead nuclei via the photoelectric effect, Compton scattering, and pair production. The second diagram of Figure 3.11 shows the pair production process where a photon interacts with a nucleus and converts to an e^+e^- pair. The parameter X_0 in Figure 3.10 is the radiation length for the material, defined as the average distance for a high-energy electron to lose all of its energy via bremsstrahlung except $1/e \sim 37\%$ of it. This length also corresponds to 7/9 of the mean free path for a high-energy photon undergoing pair production [3].

Table 3.1 summarizes the CEM and PEM segmentation shown in Figure 3.8. The CEM is segmented with 24 ϕ wedges (towers) of $\Delta\phi = 15^\circ$ and 10 η towers of $\Delta\eta = 0.11$. In the CEM, $\eta - \phi$ towers then have a size of $0.11 \times 15^\circ$. The majority of each wedge in the plug region is subdivided into two regions to make ϕ towers corresponding to 7.5°. The size of the η towers varies. EM showers are well contained, mostly falling within only a few neighboring $\eta - \phi$ towers.

[1] Throughout this thesis, the word electron most often refers to either an electron or positron.

[2] Bremsstrahlung radiation is discussed in other parts of this analysis and refers to an electron radiating a photon in the presence of a nucleus. Bremsstrahlung is a German word meaning braking or deceleration radiation.

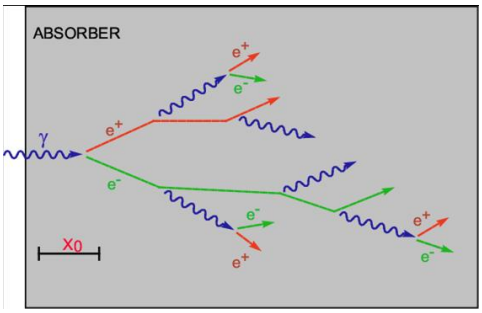
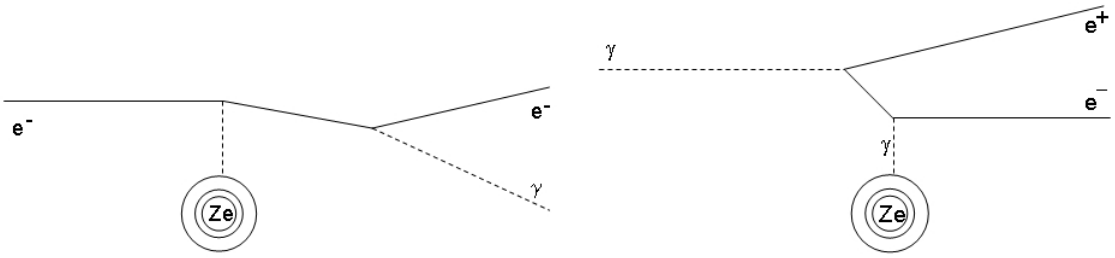


Figure 3.10: Example electromagnetic cascade in the EM calorimeters in the presence of absorber material (lead).



(a) Bremsstrahlung Radiation (b) Pair production

Figure 3.11: (a) The interaction process for an electron in the presence of absorber material. (b) An example interaction process for a photon in the presence of absorber material.

Table 3.1: Summary of the calorimeter detector segmentation and coverage. The PEM $\Delta\eta$ segmentation varies as shown in Figure 3.8 (c). It is ∼0.1 for $1.1 < |\eta| < 1.6$, ∼0.15 for $1.6 < |\eta| < 1.9$, and ∼0.2–0.6 for $1.9 < |\eta| < 3.6$.

Detector	Coverage	# of ϕ Towers	$\Delta\phi$	# of η Towers	$\Delta\eta$
CEM	$\|\eta\| < 1.1$	24	15°	10	∼0.1
PEM	$1.1 < \|\eta\| < 3.6$	24 or 48	15° or 7.5°	12	varies

The resolution of the energy measurement of the EM calorimeters is roughly proportional to the number of particles produced by the EM shower. The latter is Poisson distributed and, therefore, the resolution can be expected to be proportional to $\sigma_N/N = 1/\sqrt{N}$. The CEM and PEM σ_E/E resolution is $\frac{13.5\%}{\sqrt{E_T}} \oplus 2\%$ and $\frac{16.0\%}{\sqrt{E_T}} \oplus 1\%$, respectively, where both improve for more energetic EM particles. Ta-

ble 3.2 provides a summary of the parameters describing the central and plug EM calorimeters.

Table 3.2: Summary of the EM calorimeter features for the central and plug detectors.

	CEM	PEM
Coverage	$0 < \|\eta\| < 1.1$	$1.1 < \|\eta\| < 3.6$
Thickness	$19X_0$, 1λ	$21X_0$, 1λ
Absorber (Pb.)	$0.6X_0$	$0.8X_0$
Scintillator	5 mm	4.5 mm
Energy Resolution	$\frac{13.5\%}{\sqrt{E_T}} \oplus 2\%$	$\frac{16.0\%}{\sqrt{E_T}} \oplus 1\%$

Hadron calorimeters

Incident neutral or charged hadrons interact with iron slabs in the hadron calorimeters via inelastic nuclear processes to break up iron nuclei [3]. The energy of these particles is then rapidly decreased by the production of multiple less-energetic hadrons such as neutral and charged pions, protons, neutrons, and kaons. Hadronic showers in the calorimeters contain both an electromagnetic and hadronic component. Neutral pions or eta mesons decay almost 99% of the time to a pair of collinear photons, which interact electromagnetically to produce an EM cascade (some of which is produced and measured in the EM calorimeter). Charged secondaries (pions, protons, and kaons) will interact hadronically and transfer energy through ionization and excitation of iron nuclei; they may also produce more protons or neutrons through interactions with nuclei.

Table 3.3 summarizes the η coverage and $\eta - \phi$ segmentation for the CHA, WHA, and PHA. Their energy resolutions are limited by sampling and intrinsic fluctuations of the detector and readout material, producing a much poorer energy resolution compared to EM cascades [3]. The CHA, WHA, and PHA σ_E/E resolution are $\sim\frac{50\%}{\sqrt{E_T}}\oplus 3\%$, $\sim\frac{75\%}{\sqrt{E_T}}\oplus 4\%$, and $\sim\frac{80\%}{\sqrt{E}}\oplus 5\%$, respectively. Despite lower energy resolution

from hadron calorimeters, a complete reconstructed picture of the primary interaction is impossible without it.

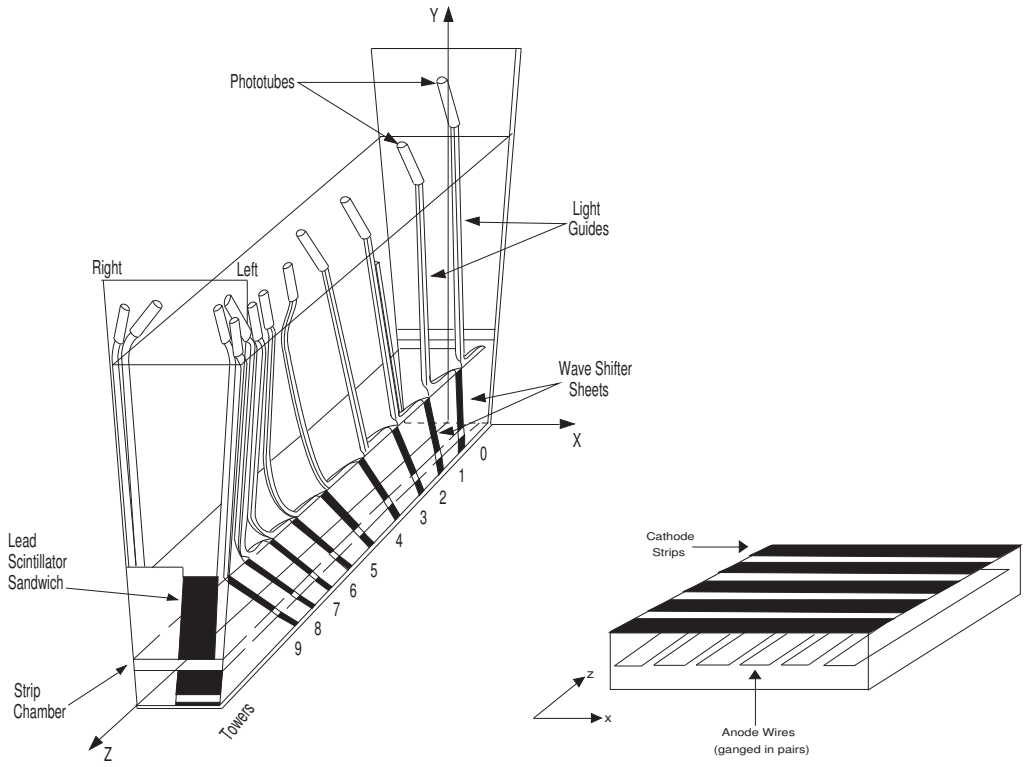
Table 3.3: Summary of the calorimeter detector segmentation and coverage. The PHA $\Delta\eta$ segmentation varies as shown in Figure 3.8 (c). It is $\sim$0.1 for $1.2 < |\eta| < 1.6$, $\sim$0.15 for $1.6 < |\eta| < 1.9$, and $\sim$0.2–0.6 for $1.9 < |\eta| < 3.6$. Some detectors overlap in η due to detector geometry (see Figure 3.8).

Detector	Coverage	# of ϕ Towers	$\Delta\phi$	# of η Towers	$\Delta\eta$
CHA	$\|\eta\| < 0.9$	24	15°	8	$\sim$0.1
WHA	$0.66 < \|\eta\| < 1.32$	24	15°	6	$\sim$0.1
PHA	$1.2 < \|\eta\| < 3.6$	24 or 48	15° or 7.5°	11	varies

3.1.4 Shower Maximum Detector

As discussed in the previous section, hadronic showers contain a substantial EM component from neutral pions and eta mesons decaying to pairs of collinear photons. The EM calorimeter segmentation is designed to capture and contain such showers and provides only coarse position measurements. They are, therefore, poor at distinguishing the difference between isolated direct photons (from the event vertex) and photons from neutral meson decays. A separate detector is designed for this purpose and to provide a precise position measurement for EM showers. This shower maximum detector (SMX) is built at the location of the EM shower maximum, six radiation lengths into the EM calorimeter as shown in Figures 3.8 and 3.9 for the plug and central regions, respectively.

The CES detector is a gas chamber composed of orthogonal anode strips and cathode wires. Figure 3.12 shows the local coordinate system for a CEM wedge and CES chamber, where there are 64 wires and 128 strips per wedge. The anode wires run in the direction of the z axis and measure the local CES x position, providing the $r - \phi$ location of the shower. The cathode strips run parallel to the local x axis and measure the local CES z position, providing the z position of the shower in the detector. The position resolution is about ± 1 mm for 50 GeV EM showers.



(a) Detail of a CEM wedge (b) Portion of the CES chamber

Figure 3.12: Diagrams showing the orientation of the cathode strips and anode wires of the CES detector, also indicating the local CES x and z axes [40]. The CES detector is located 6 radiation lengths into the CEM.

The PES detector is a chamber composed of two layers of scintillating strips. As shown in Figure 3.13, this detector is divided into 8 sectors, each covering an azimuthal angle of 45°. The two sets of strips are called U and V layers and are oriented 45° to one another. (Better $\pi^0 \to \gamma\gamma$ separation is obtained for larger U − V crossing angles, but it is mechanically difficult to incorporate sectors above 45° [44].) Each 45° sector contains 400 U and 400 V strips, providing a position resolution for high momentum EM showers of about ±1 mm.

3.1.5 Cherenkov Luminosity Counters

CDF obtains its own instantaneous and integrated luminosity measurements using the Cherenkov Luminosity Counter (CLC) detector. An instantaneous luminosity

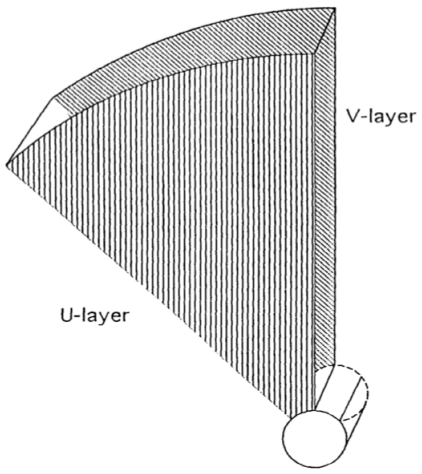


Figure 3.13: Diagram showing one 45° sector of the PES detector, located 6 radiation lengths into the PEM [43]. The two scintillator layers are oriented 45° to one another.

measurement is obtained from

$$L = \frac{f_{\mathrm{BC}} \cdot \mu}{\sigma^{\mathrm{CLC}}} \tag{3.3}$$

where f_{BC} is the bunch crossing frequency, μ is the estimated average number of $p\bar{p}$ interactions per bunch crossing, and σ^{CLC} is the measured cross section by the CLC for inelastic $p\bar{p}$ interactions. The σ^{CLC} measurement is based on the number of the hits in the CLC detector; these hits result in Cherenkov radiation which produces a conical wave of light that is collected by PMTs. Figure 3.1 shows the CLC located around the beam line in a conical hole within the plug calorimeters. References [45] and [46] provide more detail on this detector and the luminosity measurement.

3.2 Trigger and Data Acquisition

As particles from a collision pass through the detector, readouts from different detector components are converted to measurements such as energy and momentum. Not every collision is of interest, however, because particles produced in $p\bar{p}$ collisions usually come from common low-energy interactions that are well understood. Instead, a filtering scheme is incorporated which is designed to very rapidly distinguish a po-

tentially interesting event from the common events and make a decision on whether it is deemed worthy to be stored on disk for later analysis. With $p\bar{p}$ bunches crossing every 396 ns, the potential rate of data to be stored is about 2.5 MHz (2.5 million events per second). To make such a rapid decision, a three-tier trigger system is used which reduces the rate, one level at a time, to less than about 250 Hz ($< 0.01\%$ of all collisions). This is a rate sufficiently low to allow time for writing to disk [43]. Figure 3.14 shows a block diagram for the first two levels of the trigger system.

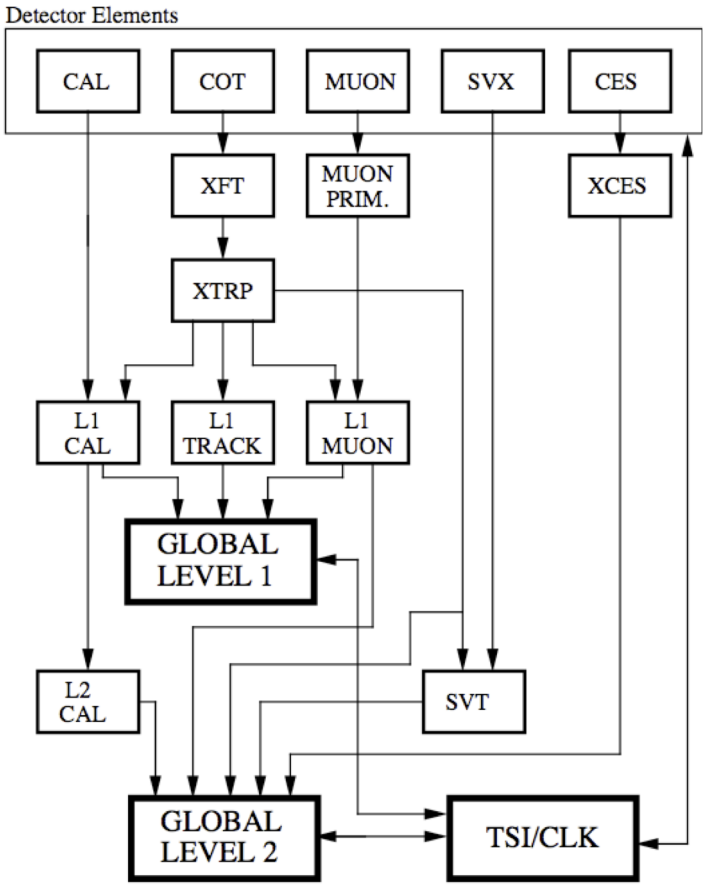


Figure 3.14: A block diagram showing detector inputs for the first two levels of the trigger system [43]. The third level of the trigger system uses all detector components.

The first level (L1) begins with a simple reconstruction of the event, such as counting the number of calorimeter trigger towers with an E_T of at least 12 GeV. The detector inputs for the L1 decision are from the calorimeter, COT, and muon readouts. Calorimeter trigger towers are formed by combining physical towers to

create a trigger segmentation of $\Delta\eta \times \Delta\phi \sim 0.2 \times 15^\circ$ (energies measured by physical towers are summed together to form the larger trigger tower). At L1, a count is then obtained on the number of trigger towers above a given E_T ($= E \sin\theta$) threshold, with θ being the angle between the z axis and the line formed from the calorimeter tower position and the center of the detector ($z = 0$). A COT track at L1 is obtained from both axial and stereo wires using the eXtremely Fast Tracker (XFT),[3] a processor designed for rapid reconstruction of high p_T tracks. These tracks are then extrapolated (using the XTRP) to the radius of the calorimeter and muon detectors to identify electron and muon objects.

If an event is accepted at L1, it is sent to the second level (L2), which forms a more refined picture of the event. The L2 decision contains all information available at L1, in addition to CES information and output from a processor designed to look for secondary vertices in the silicon, the Silicon Vertex Tracker (SVT). It additionally clusters nearby calorimeter trigger towers to form a L2 EM trigger object from which the L2 E_T is obtained.

If an event is accepted at L2, the third level (L3) fully reconstructs the event to make a decision on whether to store it on disk. The L3 system has access to all detector elements and uses a computing farm to form the L3 decision.

The trigger and data acquisition systems are synchronized using the Trigger Supervisor (TS) and CDF clock. Whether a $p\bar{p}$ collision event passes a given level is based on preselected trigger paths, which provides the ability to store and study a wide variety of physics processes.

[3] The Baylor high energy physics group was primarily responsible for the maintenance of the XFT. In particular, during data-taking, we shared the responsibility of monitoring its efficiency for track finding and of answering calls at any time of day or night when a problem occurred. During my one year residence in Batavia, I spent much time gaining expertise with this system (from other students, postdocs, and senior scientists), monitoring its efficiency, sharing the responsibility of carrying a pager, and documenting examples of both good and bad performance to better understand how to identify future problems.

3.3 Good Run Bits

During a Tevatron store, the period of time during which the CDF detector is on and is actively collecting data is called a run and each of these time ranges is given a run number. Each run can contain anywhere from a few dozen collision events to millions of events depending on the duration of a store and how long CDF detector components are working properly during this store. At the end of each run, good run bits for each detector component are set to true if they were functioning properly during the run; portions of runs can also be marked as good. Whether a run is good for physics analysis depends on which components were working and the type of analysis being performed.

3.4 Offline Reconstruction

Event reconstruction in real data-taking time is performed with CDF's online computing systems. After data are stored on disk, offline computing and software systems are used. The basic principles of track parametrization, vertex reconstruction, and calorimeter clustering are described here for offline reconstruction. Online algorithms for the trigger response are similar, but slightly different due to the need for quick decision processing.

3.4.1 Track Reconstruction

CDF tracks are classified into different (orthogonal) groups based on their silicon hit information and the track reconstruction algorithm [41, 47, 48]. Because the density of hits is much higher for smaller radii near the collision point, track reconstruction begins with an outside-in (OI) algorithm, starting with hits in outer COT superlayers or layers 6 and 7 of the silicon. For the central region, COT tracks are first formed by linking segments between superlayers and/or using a histogram linking algorithm based on hit patterns. The COT track is then used as a seed to search for a track fit including silicon hits. Outside the acceptance of the COT, silicon standalone (SiSA)

tracks are also obtained using an OI approach with only silicon hits, starting with two three-dimensional measurements from the outer two silicon layers and a third point obtained from the beam line. This provides an initial prediction of the particle track, which is modified with the addition of more hits from other silicon layers. The pattern recognition for SiSA tracks is much harder due to having a maximum of only 8 measurements, while COT tracks can be reconstructed from up to 96 measurements. Some lower p_T tracks traverse inner COT layers and fail to be reconstructed by OI algorithms. Inside-out (IO) algorithms are also then employed to improve track reconstruction, particularly in the plug region [49]. The IO algorithm starts with SiSA tracks and searches outward to add COT hits. Using multiple algorithms provides a combined track reconstruction efficiency of near 100% for the central region and up to about 75% in the plug region [49].

Charged particle tracks at CDF are described using five parameters defined at the helix's point of closest approach to the origin of the detector [50] (at perigee as shown in Figure 3.15):

- c_0: the signed half-curvature of the track. The curvature is defined as $1/r$, with r equal to the xy radius of the helix. The half-curvature c_0 is $q/(2r)$ where the track charge $q = \pm 1$. The charge is determined from the direction the track

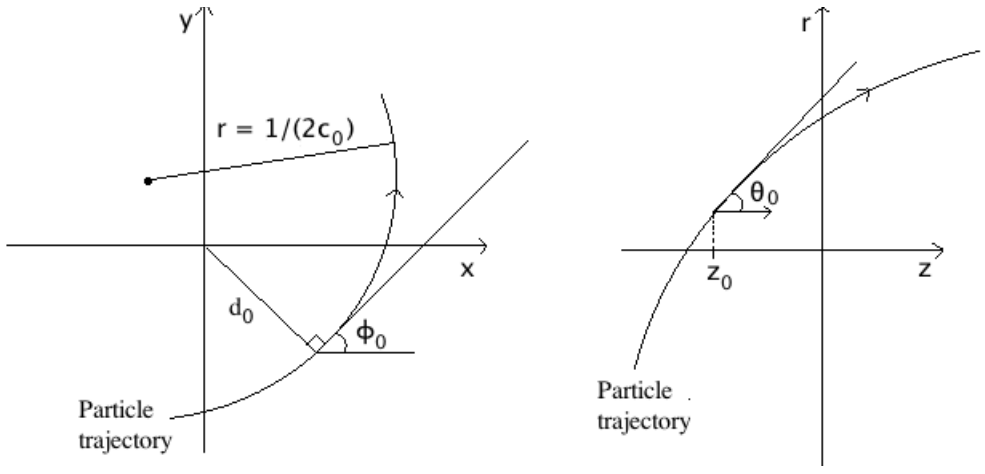


Figure 3.15: Diagram showing $r - \phi$ (left) and $r - z$ (right) track parameters.

bends since positively charged particles bend one direction in the solenoid's magnetic field and negatively charged particles bend the other way.

- ϕ_0: the azimuthal angle at the point of closest approach to the origin of the detector. This angle gives the direction of the track in the $r - \phi$ plane at minimum approach.

- d_0: the signed impact parameter. This quantity is the radial distance between the detector origin and the helix curve at the point of closest approach. Figure 3.16 shows the sign given to the impact parameter for the four different scenarios obtained when q is either $+1$ or -1 and when the helix curve either contains or does not contain the z axis. If the sign of d_0 and q are not the same, then the helix curve contains the z axis and the distance to the helix center can be calculated as $\rho = r - |d_0|$ (as in Figure 3.15). If the sign of d_0 and q are the same, then the helix curve does not contain the z axis and $\rho = r + |d_0|$.

- $\cot\theta_0$: the cotangent of the polar angle at the point of closest approach.

- z_0: the z position of the helix at the point of closest approach.

These parameters are stored for each track along with a 5×5 covariance matrix describing their errors.

Given these parameters, we know the trajectory of the charged particle and its vector momentum. With knowledge of the region over which the magnetic field acts and does not, the track can be extrapolated to any position in the detector. The d_0 and z_0 positions determine the vertex of the track and the other parameters are used for momentum reconstruction. The particle's transverse momentum is obtained from

$$p_T = |q|Br. \tag{3.4}$$

In units of eV/c, this is can be written as

$$p_T = \frac{Br}{c} = \frac{B}{2c}\frac{1}{|c_0|} \tag{3.5}$$

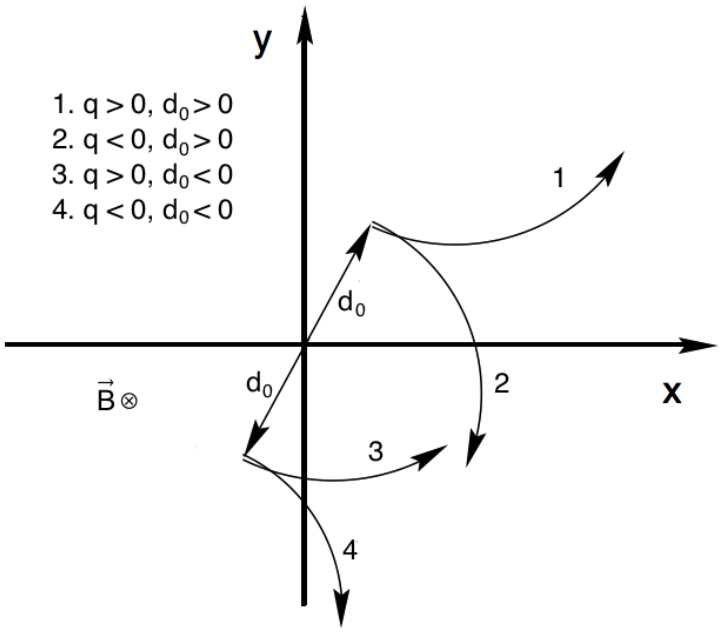


Figure 3.16: The sign of the impact parameter, d_0, is shown for the four scenarios obtained when $q = \pm 1$ and the track curves towards or away from the z axis [50]. The CDF magnetic field points along the $-z$ axis (into the page).

where c is the speed of light. The p_T of a charged particle is then just a constant multiplied by $1/c_0$. The total momentum is obtained from

$$p = \frac{p_T}{\sin\theta_0} \tag{3.6}$$

and the momentum components are

$$\begin{aligned} p_x &= p_T \cos\phi_0 \\ p_y &= p_T \sin\phi_0 \\ p_z &= p \cos\theta_0. \end{aligned} \tag{3.7}$$

With these quantities, the fourth parameter is also sometimes written as

$$\cot\theta_0 = \frac{p_z}{p_T}. \tag{3.8}$$

3.4.2 Vertex Reconstruction

Within a single bunch crossing, multiple $p\bar{p}$ interactions can occur which are recorded as a single event. Even if there is only one high p_T interaction, remnants of the beam can also contribute lower p_T interactions, which means that more than one vertex is reconstructed per event. (Section 8.1 describes these concepts in more detail.)

We use a z vertex finding algorithm [51] that has the purpose of determining the z position of primary vertices and counting the number of reconstructed vertices. The z position of the primary vertex is important for measuring the E_T of EM and HAD calorimeter towers, as shown in Figure 3.17.

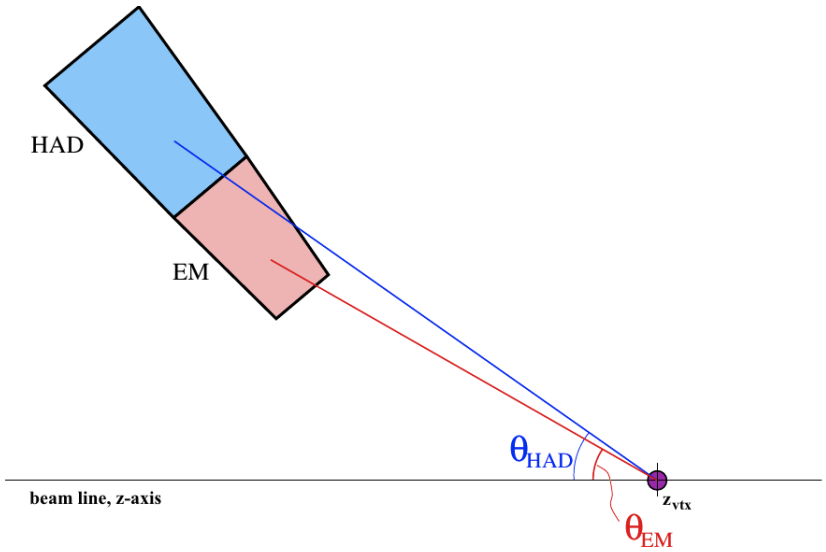


Figure 3.17: The E_T measurement for an EM cluster is calculated from $E_T = E \sin\theta$ with the EM polar angle obtained relative to the primary vertex location [40]. For trigger towers, the primary vertex is assumed to have $z_{\rm vtx} = 0$. During offline EM calorimeter clustering, this also true, but once an EM cluster is obtained, its E_T is calculated using the reconstructed primary vertex.

Preliminary vertices are formed from COT and/or silicon three-dimensional (stereo) hits using a histogram algorithm: the z position of preliminary tracks from these hits are used to fill a histogram and resonances in the histogram provide the position of preliminary vertices. Tracks are associated to these seed vertices if their z_0 position is within a minimum distance (of a few cm) of the preliminary z vertex and if they have

a minimum p_T. The vertex z position is then recalculated from an error weighted average of the z_0 positions of associated tracks clustered to the vertex:

$$z_{\rm vtx} = \frac{\sum_i z_0^i/(\Delta z_0^i)^2}{\sum_i 1/(\Delta z_0^i)^2}. \tag{3.9}$$

The vertices are classified according to the number of tracks with silicon and/or COT hits. For example, Quality 0 refers to all vertices and Quality 7 refers to vertices having ≥ 6 tracks with silicon hits and ≥ 1 track with COT hits. As with many CDF analyses, we use Quality 12 vertices which have ≥ 2 tracks with COT hits (tracks with COT hits reduce fake rates). This class of vertices is chosen because it provides a balance between true vertex reconstruction efficiency and the number of fake vertices reconstructed. A linear relationship is also observed between the number of Quality 12 reconstructed vertices in an event ($N_{\rm vtx}$) and instantaneous luminosity [51].

Each reconstructed vertex is assigned a vertex p_T defined as the sum of the transverse momentum of its associated tracks ($\sum_{\rm tracks} p_T$).

3.4.3 Electromagnetic Calorimeter Clustering

The primary particles of interest in this analysis are photons, some of which convert to an electron-positron pair before reaching the calorimeters. Both photons and electrons are dependent on energy measurements of the EM calorimeters. This section first describes how such EM objects are obtained by clustering energy in calorimeter towers. It ends with a description of the resulting detector coordinates of the cluster. Later chapters describe how particles are identified from these clusters.

Figure 3.18 provides an example of an EM shower in the central calorimeters. The shower profile is shown in an $\eta-\phi$ grid with the amount of energy from the EM (HAD) calorimeter indicated in pink (blue). Clusters are formed from energy measurements made in such towers. The $\eta-\phi$ towers available for calorimeter clustering are given in Figure 3.19 for the CEM (blue) and PEM (green). For clustering, an E_T measurement for each tower is obtained as shown in Figure 3.17, with $z_{\rm vtx} = 0$. For a given event, clusterable towers are those which have an E_T greater than 100 MeV. Clustering in both the CEM and PEM begins by searching for all towers with $E_T > 3$ GeV, called

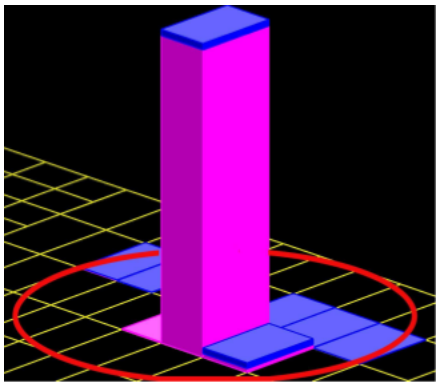


Figure 3.18: "Lego" plot of an EM shower in the calorimeter. The yellow grid shows the $\eta - \phi$ segmentation, so each rectangular region represents a calorimeter tower [40]. Pink represents the amount of EM E_T and blue represents the amount of HAD E_T. The tower blocks provide an example of an EM shower in the central calorimeters. The energy of an EM shower is generally contained within a few towers. Only a small amount of this energy leaks into the hadron towers. The red circle indicates a region in $\eta - \phi$ space defined by $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

seed towers. The seed towers are then sorted by decreasing E_T such that clustering begins with the largest E_T tower. Neighboring towers, called daughter towers, are grouped with the seed tower if they are in the same detector (CEM or PEM) and not already included with another cluster.

In the CEM, daughter towers are defined as border towers with the same ϕ value as the seed tower such that the difference is η and ϕ index numbers from the seed is $\Delta n_\phi = 0$ and $\Delta n_\eta = 1$. Central EM clusters, therefore, contain a minimum of one tower (the seed) and a maximum of three towers (when both neighbors have $E_T > 100$ MeV), all in the same ϕ wedge.

The PEM detector geometry is different from the CEM, however, and a clustering algorithm is desired that allows daughter towers with $\Delta n_\phi = 1$ and $\Delta n_\eta = 1$, which includes towers that either border or share a corner with the seed. With all daughter towers being equal, this would lead to a 3×3 structure, with the seed at the center. This cluster size was considered too large, however, and the PEM clustering algorithm most often constrains the size to a square 2×2 structure. The 2×2 structure starts by obtaining the bordering daughter tower with the largest E_T among the seed's

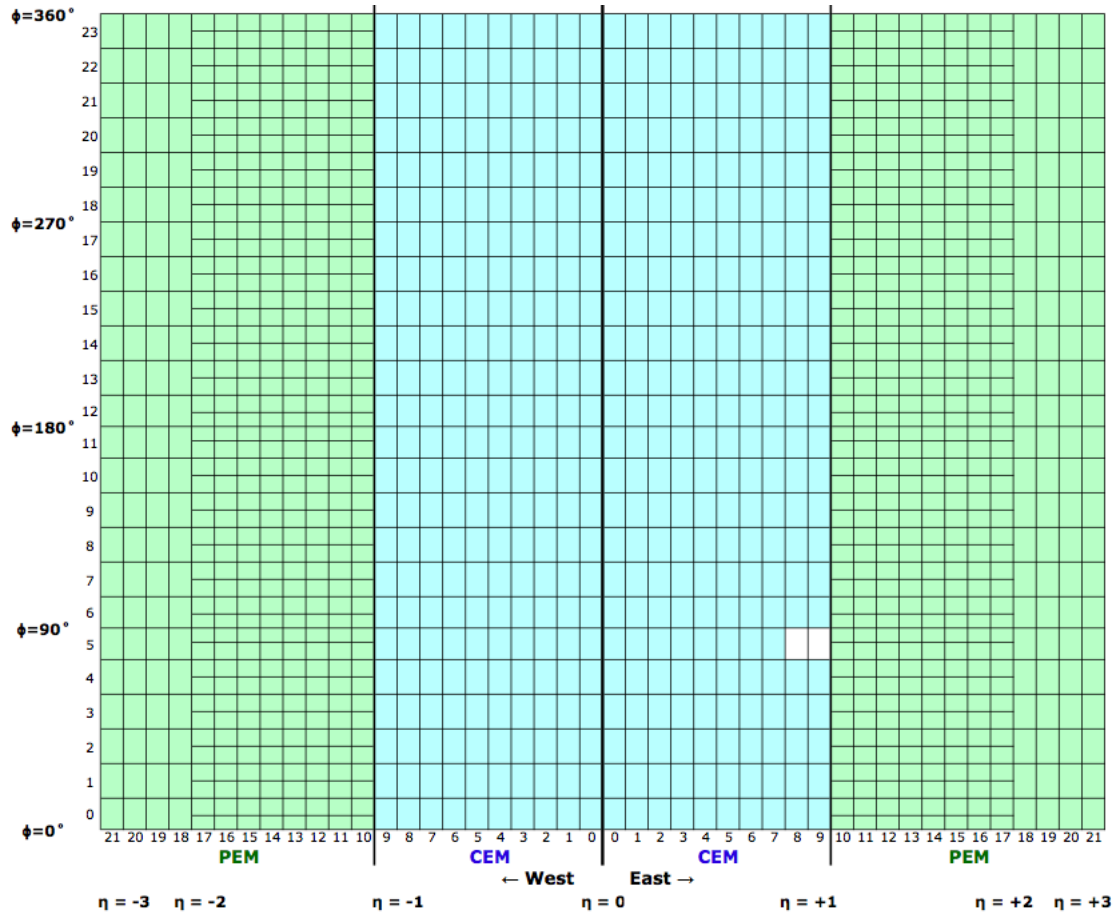


Figure 3.19: Calorimeter segmentation in $\eta - \phi$ space for the central (blue) and plug (green) electromagnetic calorimeters, CEM and PEM, respectively. Calorimeter wedge numbers (0–23) are shown on the left vertical axis labels, each covering 15° in ϕ. For a portion of the plug region, each wedge contains two towers such that the tower ϕ segmentation is 7.5°. The x axis shows the west (left) and east (right) η tower numbers, from 0–21 as shown in Figure 3.8, with a coverage of $0 < |\eta| < 3.6$. The white box, called the chimney, is a region excluded from use in particle identification. It is a gap in the detector used for cables and cryogenic utilities needed for the solenoid.

neighbors. The cluster is completed by searching for a pair of towers that neighbor both the seed and this daughter and have total $E_T > 100$ MeV. If there are two such sets, the pair with higher E_T is selected. The 3×3 structure is allowed in the PEM if there are no daughter towers with these requirements or if the seed is on the boundary in the PEM where the ϕ segmentation goes from 7.5° to 15° (Figure 3.19).

Central and plug energy clusters obtained in this way [52] are called EM clusters, or EM objects throughout this paper. Their total EM energy ($E_{\rm EM}$) is the sum of the energies of each EM calorimeter tower included in the cluster. The total amount of hadronic energy ($E_{\rm HAD}$) of the cluster is obtained from the sum of the associated HAD calorimeter towers. The total cluster energy is $E = E_{\rm EM} + E_{\rm HAD}$. Since a real EM object deposits essentially all of its energy in the EM calorimeters, we take the total energy of such a particle to be the measurement from the EM calorimeters alone and often refer to it as just E rather than $E_{\rm EM}$ (or E_T rather than $E_T^{\rm EM}$).

The η and ϕ detector coordinates of the cluster are obtained from an energy-weighted method. The EM cluster positions are from the EM energy-weighted sum of the η and ϕ detector positions of each tower in the cluster,

$$\eta_{\rm EM} = \frac{\sum_i E^i_{\rm EM} \times \eta^i}{\sum_i E^i_{\rm EM}} \tag{3.10}$$

$$\phi_{\rm EM} = \frac{\sum_i E^i_{\rm EM} \times \phi^i}{\sum_i E^i_{\rm EM}}. \tag{3.11}$$

A similar calculation is performed for the hadronic cluster positions:

$$\eta_{\rm HAD} = \frac{\sum_i E^i_{\rm HAD} \times \eta^i}{\sum_i E^i_{\rm HAD}} \tag{3.12}$$

$$\phi_{\rm HAD} = \frac{\sum_i E^i_{\rm HAD} \times \phi^i}{\sum_i E^i_{\rm HAD}}. \tag{3.13}$$

The detector coordinates for the cluster are then obtained from

$$\eta = \frac{E_{\rm EM} \times \eta_{\rm EM} + E_{\rm HAD} \times \eta_{\rm HAD}}{E} \tag{3.14}$$

$$\phi = \frac{E_{\rm EM} \times \phi_{\rm EM} + E_{\rm HAD} \times \phi_{\rm HAD}}{E} \tag{3.15}$$

which are called detector η and detector ϕ.

3.4.4 Clustering in Shower Maximum Detectors

Energy clusters are also formed in the CES and PES from hits in their strips and wires. These SMX clusters are later matched to their corresponding CEM or PEM cluster based on their proximity to these clusters [52].

For each wedge, CES clusters are obtained from either a strip-based algorithm or a track-based algorithm [53]. The former is considered an unbiased approach and the latter is used for charged EM objects, particularly electrons. The strip-based algorithm sorts the energy of strips from largest to smallest and seed strips are obtained from those which have an energy above a certain threshold. Strips surrounding the seed are grouped to form a 1D cluster, and this is repeated for the wire layer to form its own 1D shower. For both layers, 11 wire and 11 strips are used to form a cluster. The two energy showers are then matched to form a 2D CES cluster with the following energy-weighted local CES x and z coordinates:

$$x_{\mathrm{CES}} = \frac{\sum_i^{11} x_i E_i^w}{\sum_i^{11} E_i^w} \quad z_{\mathrm{CES}} = \frac{\sum_i^{11} z_i E_i^s}{\sum_i^{11} E_i^s} \tag{3.16}$$

where E_i^w (E_i^s) is the energy measured by the i^{th} wire (strip). This is repeated for other seeds, excluding any already-used strips or wires. The track-based algorithm extrapolates tracks to the location of the CES, and the strip or wire near this location is used as a seed. Strips and wires near the seed are then grouped as they are with the seed-based algorithm to form a 2D cluster, and coordinates are obtained from Equation (3.16).

PES clustering uses a strip-based algorithm, forming both a U and V layer cluster from 9 strips with a seed at its center. The energy for each layer is obtained by summing the energies of the corresponding strips of the cluster. As with the CES, the two 1D clusters of the U and V layers are matched to form a 2D PES cluster.

Chapter 4

Central and Plug Photon Identification

We refer to photons with $|\eta| < 1.1$ as central photons and those with $1.1 < |\eta| < 3.6$ as plug photons, where η refers to the detector EM cluster position. This chapter discusses the reconstruction of photons that do not convert to an e^+e^- pair in the detector, and the next chapter discusses the reconstruction of central photon conversions. A description of photon kinematic variables and photon identification (ID) is first provided in this chapter, followed by a description of the ID efficiency.

4.1 Kinematic Variables

The total energy of an EM shower is the sum of energies of all towers in the cluster. Before applying the offline photon selection, we make corrections to the energy of each EM shower. The first of the corrections shown here is actually made online [54] and is applied only to data. The remaining two are applied offline for EM clusters both in the collision data and in any Monte Carlo (MC) simulated data.

- *Time-dependent gain variations*: The gain variations of phototubes for each tower are monitored over time and corrected during run-time [54].

- *Face (or map) correction*: A correction of a few percent is made to the energy offline based on the location of the shower in the tower. If a shower is near a phototube, there is a higher measurement of energy than if the same shower energy were in the center of a tower. Since towers vary in size, corrections are applied per tower and are based on 50 GeV electron test beam measurements and cosmic ray studies [55, 56].

- *Z-based energy correction*: After other corrections are made, calibrations are performed by plotting the reconstructed Z boson mass from $Z \to e^+e^-$ decays. For different data-taking periods, the Z mass is measured from the mean of a fit made to the data, and this measurement is compared to the accepted PDG value of 91.2 GeV/c^2 [3]. The CEM (PEM) energy is tested if an electron falls into the central (plug) region. A separate correction is applied to each data-taking period and is obtained from the ratio of the PDG mass to the reconstructed mean obtained from the data. A single correction is also applied to MC events. This correction is obtained from the ratio of the PDG mass to the reconstructed mass obtained from Z MC samples.

As previously described, the E_T of the shower is obtained from $E \sin\theta$, where the calorimeter energy includes the energy corrections just discussed. For central and plug offline reconstruction, the polar angle θ is formed relative to the z position of the primary vertex ($z_{\rm vtx}$) and the position of the cluster in the shower maximum detector (see Figure 3.17). In particular,

$$\sin\theta = \frac{R_{\rm SMX}}{\sqrt{R^2_{\rm SMX} + z^2_{\rm vtx}}} \tag{4.1}$$

where $R_{\rm SMX}$ is the radial distance from the z axis to the shower position in the CES or PES detector. The selection of the primary vertex is described in Section 6.4.

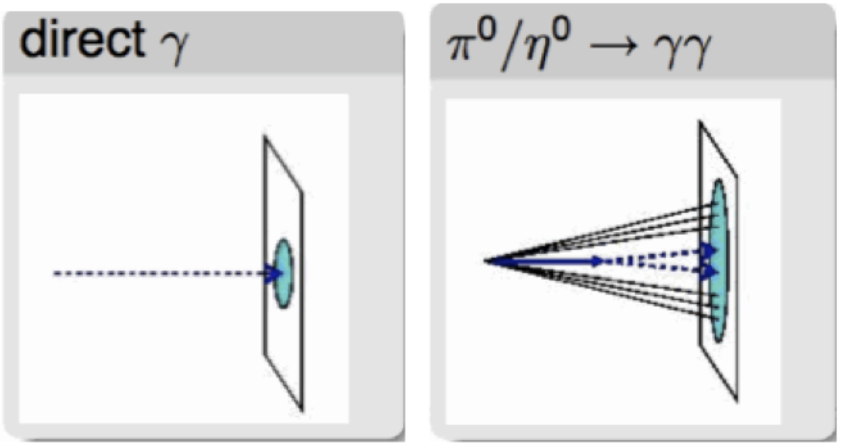
A real photon is massless and, therefore, the total vector momentum of photon candidates is set to the energy obtained from the EM calorimeter energy measurement: $p = E$ and $p_T = E_T$ (with $c = 1$). The photon four-momentum is then

$$P \;=\; (E, p_x, p_y, p_z) = (E, E_T\cos\phi, E_T\sin\phi, E\cos\theta) \tag{4.2}$$

where ϕ is determined from the azimuthal position of the EM cluster (Equation (3.11)).

4.2 Photon Identification

Photon reconstruction is based on a set of detector variables that are designed to distinguish true photons from other objects (called backgrounds). In general, photons are differentiated from these other particles because they deposit an isolated energy cluster in the EM calorimeter alone and have no tracks pointing to them. Photon identification then consists of requirements based on how isolated the photon is in the calorimeter, the amount of energy in the EM vs. HAD calorimeters, and the presence of tracks. The selection is also designed to distinguish photons originating from the primary collision and collinear photon pairs from the decay of a neutral meson in the core of a jet. The former are called direct, or prompt, photons. An illustration of each is given in Figure 4.1. A set of variables is also included that require the profile of the EM shower to be consistent with that of a prompt photon.



(a) Isolated Direct Photon (b) Collinear Photon Pair in a Jet

Figure 4.1: (a) Illustration of an isolated energy deposition from a direct (prompt) photon from the primary collision. (b) Illustration of an energy deposition from a collinear pair of photons from the decay of a neutral meson in the core of a jet [40]

The typical approach for photon identification at CDF is to make requirements, called cuts, on each individual detector variable. The selected cut for a particular variable is based on the different spread of values measured for photons versus backgrounds. Looser cuts on these variables provide a high efficiency for true photon reconstruction, but also allow for a higher fraction of fake photons to be reconstructed.

Tighter cuts (a more stringent selection) on these variables result in a lower efficiency for selecting pure photons, but better reject the fake photon background. This cut-based approach is very effective and we apply the standard CDF tight selection for reconstructing plug photons. For central photons, we obtain an even better efficiency for identifying photons and rejecting backgrounds by using a multivariate (MV) analysis technique.[1] Both central and plug photon ID for this analysis are now described in more detail.

4.2.1 Central Photon ID

Multivariate algorithms in particle physics are typically given an input set of variables sensitive to distinguishing a signal from backgrounds, and they output a single continuous variable with a value that describes how signal-like or background-like a candidate object or event is [58]. These algorithms "learn" the statistical relationships (such as correlations) between input variables in order to extract as much information as possible to distinguish signal and background candidates. For particle identification, a single cut can be applied on the MV output. The MV central photon ID developed for CDF is summarized here and described in more detail in References [59] and [60].

Neural net training

Both electrons and jets can be backgrounds to the identification of prompt photons. Electrons, however, have an almost identical signature in the EM calorimeters as photons and are effectively rejected with the standard central CDF tight cuts on track-based variables (about 99% rejected). We furthermore want a photon ID that can be slightly modified to identify electrons so that we can later use electrons from Z boson decays to determine the efficiency of the photon ID selection. In particular, input variables for the MV algorithm were selected that have an almost identical distribution for electrons as for photons. Standard CDF tight cuts are maintained

[1] Reference [57] describes a MV ID for plug photons that was recently studied.

for removing electrons and the MV tool is instead used to better reject the largest background to photon reconstruction: jet backgrounds mostly due to π^0 and η meson decays. In order to let the MV algorithm concentrate on distinguishing harder cases where a jet looks like a real photon, loose cuts are first applied on central photon variables not already used for electron rejection.

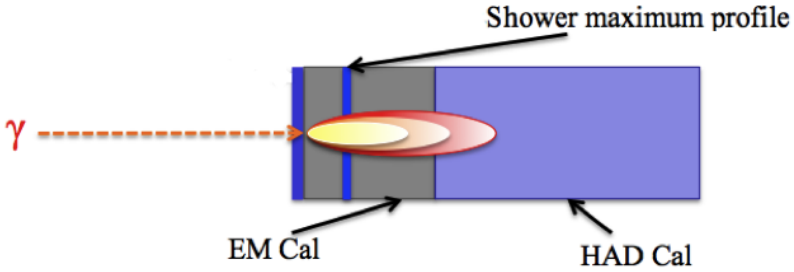
After this initial selection, several MV algorithms were "trained" using simulated photon and jet MC samples; six variables sensitive to differences between photons and jets were given as inputs. Based on this training, an artificial neural network (NN) was found to provide both the best efficiency for identifying photons (signal efficiency) and efficiency for rejecting jet backgrounds (background rejection) [59, 60]. The variables used as inputs for the NN are described next. For these descriptions, the corrected energy and E_T are obtained as described in Section 4.1.

- *Had/Em*: Jets have a large fraction of energy in the HAD calorimeter while photons do not. The Had/Em variable is the ratio of hadronic to EM energy, $E_{\rm HAD}/E_{\rm EM}$. These energy values are defined in Section 3.4.3.

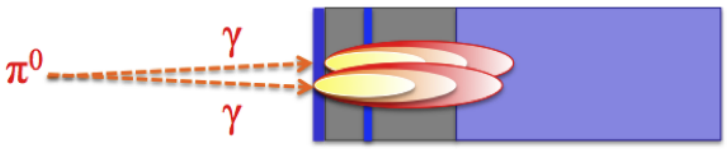
- *Calorimeter Isolation E_T*: Jets generally deposit energy over a wide region of the calorimeters while photons have compact energy clusters. The calorimeter isolation variable quantifies the amount of energy near an EM shower. After the EM cluster's E_T has been subtracted, calorimeter isolation E_T is defined as the sum of E_T in towers surrounding the cluster. The towers included in this definition are those within a cone of radius 0.4 around the (η, ϕ) position of the EM cluster's seed tower: $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.4$. This quantity is corrected for the amount of energy from a shower that is expected to have leaked into adjacent wedges (and is, therefore, not included in the cluster's E_T calculation). A correction to the isolation E_T is also applied based on the number of reconstructed vertices in the event; this is to account for the contribution of energy arising from underlying event and multiple $p\bar{p}$ interactions in a single bunch crossing (see Section 8.1).

- *Track Isolation* p_T: Jets have many tracks associated with a calorimeter cluster while photons do not.[2] Tracks with a z_0 position within 5 cm of the z position of the primary vertex are extrapolated to the EM calorimeter. The track isolation p_T is obtained from the sum of p_T from tracks that are within a cone of $\Delta R < 0.4$ around the EM cluster centroid.

- *Chi2Strip* and *Chi2Wire* (CES Strip and Wire χ^2): These variables help better distinguish prompt photons from photon pairs produced in neutral meson decays. Figure 4.2 illustrates how the shower profile of collinear photons in the CES is generally different from that of a direct photon because it has two energy clusters. The expected energy profile for a single EM shower was determined early in Run II using test beam data and the profile was measured separately for the strip and wire CES layers. A χ^2 is obtained by comparing the lateral shape of a CES energy cluster to the profile measured from the test beam data, separately for the strip and wire views.



(a) Isolated Direct Photon



(b) Collinear Photon Pair in a Jet

Figure 4.2: Illustration of the shower profile in the shower maximum detector for (a) a direct (prompt) photon and (b) collinear pairs of photons from the decay of a neutral meson [40].

[2] This track isolation variable is also typically used to reject electrons; however, it can be modified for electron ID studies using the NN by subtracting the electron's track p_T.

- *Lshr*: The purpose of this variable is to distinguish the photon and electron shower development in the EM calorimeter to that of jets, based on the lateral sharing of energy in towers adjacent to the EM cluster's seed tower. The Lshr variable (pronounced *el share*) is a measure of how well the EM shower profile agrees with the expected profile for a single EM shower. Lshr is defined as

$$\text{Lshr} = \frac{0.14 \sum_i (E_i^{\text{meas}} - E_i^{\text{exp}})}{\sqrt{(0.14\sqrt{E_{\text{EM}}^{\text{meas}}})^2 + \sum_i (\Delta E_i^{\text{exp}})^2}} \tag{4.3}$$

  where the sum is over towers in an EM cluster adjacent to the seed tower and in the same wedge as the seed tower (i.e. over either a one or two tower sum). The value E_i^{meas} is the measured energy in an adjacent tower, E_i^{exp} is the expected energy in the adjacent tower obtained from test beam data, $E_{\text{EM}}^{\text{meas}}$ is the total measured EM energy of the 1–3 tower cluster, and ΔE_i^{exp} is an estimate of the uncertainty in E_i^{exp}.

- *CES/CEM*: This is the ratio of energy measured by the CES to the energy measured by the CEM. This variable helps distinguish prompt photons and photons from neutral meson decays.

For photon and jet MC samples, Figure 4.3 shows the input variable shapes and Figure 4.4 shows the corresponding NN output. The background generally has a NN value near zero while the signal peaks near one.

Central photon ID cuts

For central photon candidates in both data and MC events, we apply the same initial selection as was used in training: standard CDF tight cuts to remove electrons and loose cuts to remove photon candidates that are more easily identifiable as jets. Each of the variables to which cuts are applied is next described.

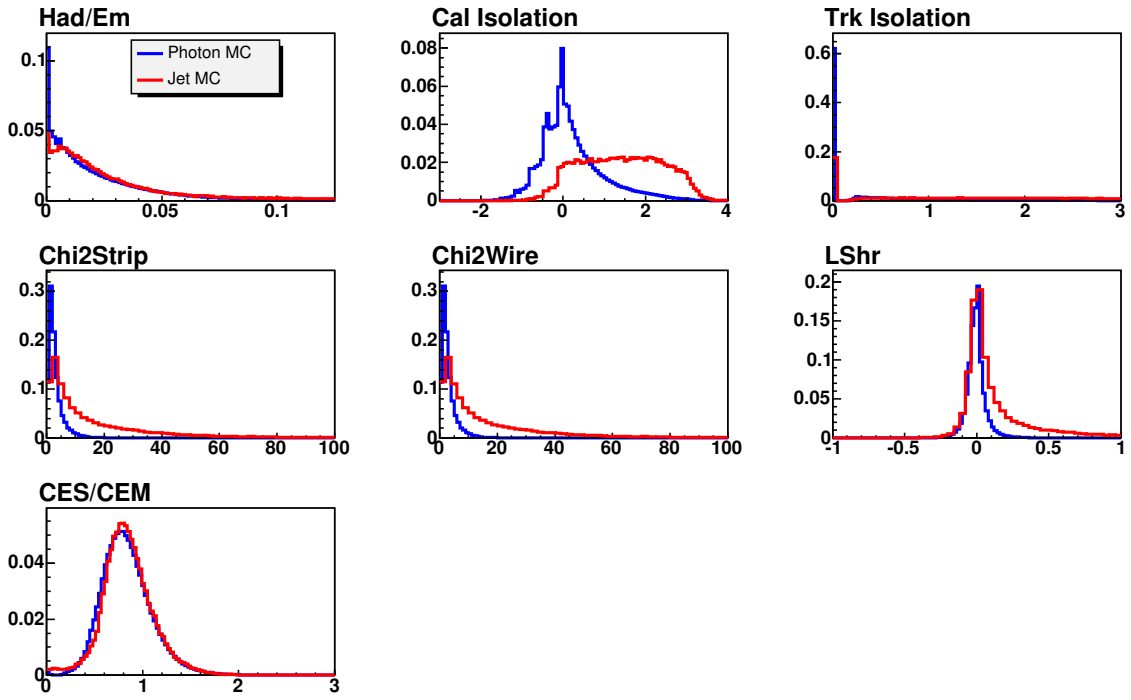


Figure 4.3: Distributions of variables used as inputs to the NN, shown for photon and jet MC samples. Each distribution is normalized to unit area.

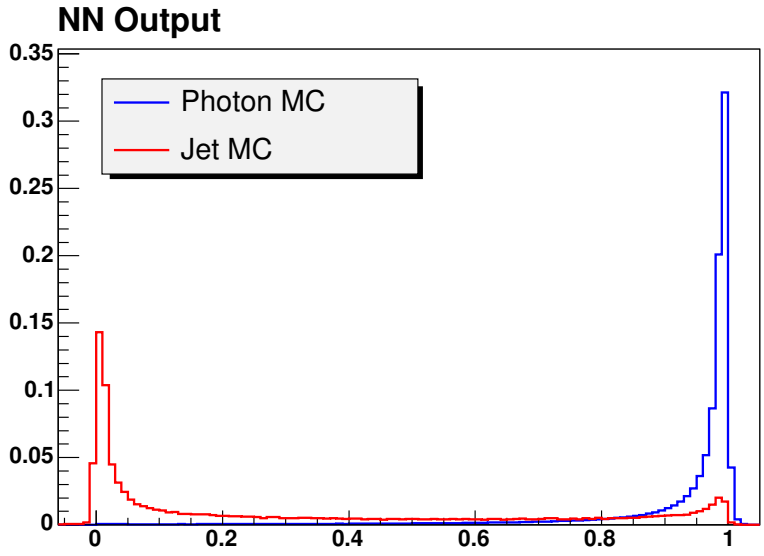


Figure 4.4: Distribution of the NN output for photon and jet MC samples, normalized to unit area.

- E and E_T: The corrected energy and E_T are obtained as described in Section 4.1. We select energetic photons by requiring their E_T to be at least 15 GeV.[3]

- *CES Fiducial*: The EM shower reconstructed in the CES detector is required to have a cluster position that is in the fiducial (well-instrumented) region of the CES detector. This is obtained by making cuts on the local x and z positions of the CES cluster.

- *Had/Em*: Since this is one of the variables used as an input for the NN, we apply the standard CDF loose cut on this variable.

- *Calorimeter Isolation* E_T: A sliding cut is made on this variable because the isolation profile changes for higher E_T EM showers. We apply the standard CDF loose cut since this is one of the variables used as an input for the NN.

- *Track Isolation* p_T: Since this is one of the variables used as an input for the NN, we apply the standard CDF loose cut on this variable.

- *N Tracks (N3D) and Track* p_T: Three-dimensional tracks (i.e. tracks with at least one stereo segment) are extrapolated to the plane of the CES, and N3D is the number of tracks that have a position within 5 cm of the CEM's seed tower boundary in the z direction. For a true photon, there should be no tracks pointing to the EM cluster. However, one track is allowed if it has small transverse momentum. Tight cuts are made on these variables to remove electrons.

- *2nd CES cluster* E_T: We have made a tight cut on this variable to remove photons from meson decays that tend to produce two CES energy clusters. This cut is made by restricting the E_T of the second highest CES strip or wire

[3] For the diphoton selection, we use a trigger that selects events with at least one EM shower of $E_T > 25$ GeV. At least one photon, then, will have $E_T > 25$ GeV and the other must have an E_T of at least 15 GeV. See Section 8.3.3 for how the trigger efficiency is applied to the $H \to \gamma\gamma$ MC prediction.

cluster that can be matched to the EM calorimeter cluster. The CES E_T is obtained from the energy of a CES cluster multiplied by $\sin\theta$ (Equation (4.1)).

- *NN Output*: As previously discussed, we place a cut on the NN output to better remove jet backgrounds. A figure of merit, called significance, was calculated from $s/\sqrt{b}$ where s is the expected number of $H \to \gamma\gamma$ signal events (for $m_H = 120$ GeV/c^2) and b is the approximate number of diphoton candidate events in the data in which one or both reconstructed photons are from a jet. This figure of merit was plotted as a function of the NN cut and the maximal value of the significance was obtained using a cut of 0.74.

Table 4.1 summarizes the central photon ID selection.

Table 4.1: Central photon selection cuts listed in the order that they are applied.

Central Photon Variable	Cut
E_T	> 15 GeV
CES Fiducial	$\lvert x_{\rm CES}\rvert < 21$ cm, $9 < \lvert z_{\rm CES}\rvert < 230$ cm
Had/Em	< 0.125
Calorimeter Isolation	$< 0.15E_T$ GeV for $E_T \leq 20$ GeV
	$< 3.0 + 0.02(E_T - 20.0)$ GeV for $E_T > 20$ GeV
Track Isolation	< 5 GeV/c
N track (N3D)	≤ 1
Track p_T (if N3D = 1)	$< 1.0 + 0.005E_T$ GeV/c
2nd CES Cluster E_T	$< 0.14E_T$ GeV for $E_T < 18$ GeV
	$< 2.4 + 0.01E_T$ GeV for $E_T > 18$ GeV
NN Output	> 0.74

4.2.2 Plug Photon ID

We apply a cut-based approach for plug photon ID using the standard CDF plug photon selection. Some variables were already described for central photons, and a description of those that were not is given here.

- *PES Fiducial*: EM showers in the plug region are required to pass through the fiducial region of the PES: the detector η value of the shower as measured by the PES is to be in the range $1.2 < |\eta| < 2.8$.

- *PEM* $\chi^2_{3\times3}$: A cut on the lateral shower profile in the PEM helps reduce $\pi^0/\eta \to \gamma\gamma$ backgrounds. The energy distribution of the 3×3 block of towers surrounding the seed tower of an EM cluster is fit to profiles from electron test beam data to construct a χ^2.

- *PES 5 by 9 U and V*: A PES energy cluster is obtained from 9 strips surrounding a seed strip, separately for the U and V layers. For each layer, the ratio of the energy of the middle 5 strips to the energy of all 9 strips is used to distinguish prompt photons from neutral meson decays.

Table 4.2 summarizes the cuts on plug photons.

Table 4.2: Plug photon selection cuts listed in the order that they are applied.

Plug Photon Variable	Cut
E_T	> 15 GeV
PES Fiducial	$1.2 < \|\eta\| < 2.8$
Had/Em	< 0.05 for $E \leq 100$ GeV
	$< 0.05 + 0.026 \times \ln(E/100)$ for $E > 100$ GeV
Calorimeter Isolation	$< 0.1E_T$ GeV for $E_T < 20$ GeV
	$< 2.0 + 0.02(E_T - 20.0)$ GeV for $E_T > 20$ GeV
PEM $\chi^2_{3\times3}$	< 10
PES 5 by 9 U and V	> 0.65
Track Isolation	$< 2.0 + 0.005E_T$ GeV/c

4.3 Photon ID Efficiency

We next discuss the efficiency of this identification for true photons. Based on the results of this study, we apply a correction factor to $H \to \gamma\gamma$ simulated events based

on differences between the measured efficiency in MC samples relative to the data. For the high E_T range that we wish to study, there is not a large sample of pure photons available in the data. We instead take advantage of similarities between electron and photon showers in the EM calorimeter and use a photon-like ID to reconstruct electrons.

4.3.1 Method

To measure the photon ID efficiency, a high-purity sample of $Z \to e^+e^-$ events are obtained using a "tag and probe" method. We first reconstruct a "tag" electron leg: one that is identified using tight cuts and in a region of the detector where the identification efficiency is high (i.e. the central region). The second electron leg is considered the "probe" leg and we measure the photon ID efficiency from this electron. The invariant mass of the e^+e^- pair is used to select events around the Z pole, which reduces backgrounds for this study. This method has been used by many previous photon analyses at CDF and we have repeated it for this analysis.

The efficiency calculation is performed in both data and simulated events where the data are obtained from a high p_T electron trigger[4] and the simulated events are from $Z \to e^+e^-$ MC samples. For this study, we use data from the same time periods used in the $H \to \gamma\gamma$ analysis (data-taking period numbers 0 through 38). Essentially the same good run list is used (see Section 6.5) and the z position of the event vertex must be within 60 cm of zero, as required for the $H \to \gamma\gamma$ analysis (Section 6.4). The following modifications are made to the photon ID for electrons.

- *Track Isolation*: The only modification made to this variable is to subtract the p_T of the highest p_T track associated with the EM cluster. This is done for both the central and plug photon ID selection. For the central selection, the modified track isolation p_T is also used as an input to the NN.

[4] The `ELECTRON_CENTRAL_18` trigger is used to collect data from CDF b-stream datasets in Stntuple format: `bhelbd`, `bhelbh`, `bhelbi`, `bhelbj`, `bhelbk`, `bhelbm`, `bhelap`. This comprises CDF data-taking periods 0–38, which were collected from February 2002 through September 2011.

- *N tracks (N3D) and Track p_T*: The requirements on these variables are identical to those applied for the central photon selection except rather than cutting on the highest p_T track, cuts are made on the second highest p_T track.

- E/p: For high momentum electrons, $p \gg m_e$ and, therefore, $E/p \to 1$. A cut on the ratio of the energy of the EM cluster to the momentum of the highest p_T track is added to the central electron selection only. In the presence of detector material, an electron can radiate a photon via bremsstrahlung such that its momentum is reduced. The momentum measurement of the electron's track is lower as a result, however the energy measurement from the calorimeter may remain unchanged because the photon's energy is included in the same calorimeter cluster as the electron's. This skews the measurement of the photon ID efficiency and a cut on E/p reduces the number of these events (a decreased p results in a higher ratio.)

Loose and tight photon-like identification for central and plug electrons is given in Tables 4.3 and 4.4.

Figure 4.5 shows the NN response for electrons from Z events in data and MC samples and photons from a $H \to \gamma\gamma$ MC sample. This plot provides an example of the similarity of the electron response to the photon-like cuts compared to photons. This figure also gives an indication of how the MC simulation compares to the data for the central ID selection.

$Z \to e^+e^-$ events are divided into two categories, one in which two central electrons are found (CC events) and one in which a central and plug electron are found (CP events). In both cases, the central tag leg must pass tight central photon-like ID. In CC (CP) events, central (plug) photon-like ID cuts are then applied to the probe leg to measure efficiencies.

In all events, the probe electron leg must pass at least the loose cuts. For those that pass loose cuts and each remaining cut, we fill histograms of the invariant mass distributions of the e^+e^- pair (M_{ee}). Figure 4.6 shows examples of such histograms. Each histogram is fit to the sum of a Gaussian function (for the Z pole) and a

Table 4.3: The photon-like central electron selection cuts, listed in the order that they are applied. The loose selection is defined by the first three cuts and the tight selection consists of these loose cuts in addition to the remaining cuts shown.

Central Electron Variable	Cut
E_T	> 15 GeV
CES Fiducial	$\lvert x_{\rm CES}\rvert < 21$ cm, $9 < \lvert z_{\rm CES}\rvert < 230$ cm
E/p	> 0.9 and < 1.1
Had/Em	< 0.125
Calorimeter Isolation	$< 0.15E_T$ GeV for $E_T \leq 20$ GeV
	$< 3.0 + 0.02(E_T - 20.0)$ GeV for $E_T > 20$ GeV
Track Isolation – Highest Track p_T	< 5 GeV/c
N track (N3D)	≤ 2
2nd Track p_T (if N3D = 2)	$< 1.0 + 0.005E_T$ GeV/c
2nd CES Cluster E_T	$< 0.14E_T$ GeV for $E_T < 18$ GeV
	$< 2.4 + 0.01E_T$ GeV for $E_T > 18$ GeV
NN Output	> 0.74

Table 4.4: The photon-like plug electron selection cuts, listed in the order that they are applied. The loose selection is defined by the first two cuts and the tight selection consists of these loose cuts in addition to the remaining cuts shown.

Plug Electron Variable	Cut
E_T	> 15 GeV
PES Fiducial	$1.2 < \lvert\eta\rvert < 2.8$
Had/Em	< 0.05 for $E \leq 100$ GeV
	$< 0.05 + 0.026 \times \ln(E/100)$ for $E > 100$ GeV
Calorimeter Isolation	$< 0.1E_T$ GeV for $E_T < 20$ GeV
	$< 2.0 + 0.02(E_T - 20.0)$ GeV for $E_T > 20$ GeV
PEM $\chi^2_{3\times3}$	< 10
PES 5 by 9 U and V	> 0.65
Track Isolation – Highest Track p_T	$< 2.0 + 0.005E_T$ GeV/c

linear function (to describe backgrounds). The number of signal events that pass a particular cut is first estimated from the integral of the Gaussian over the range

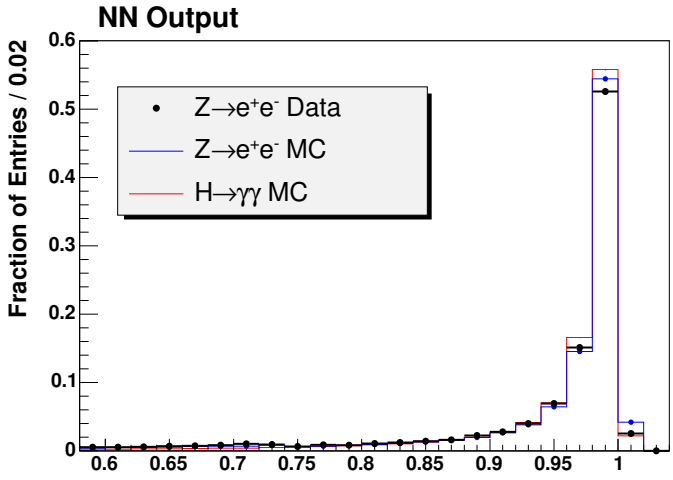


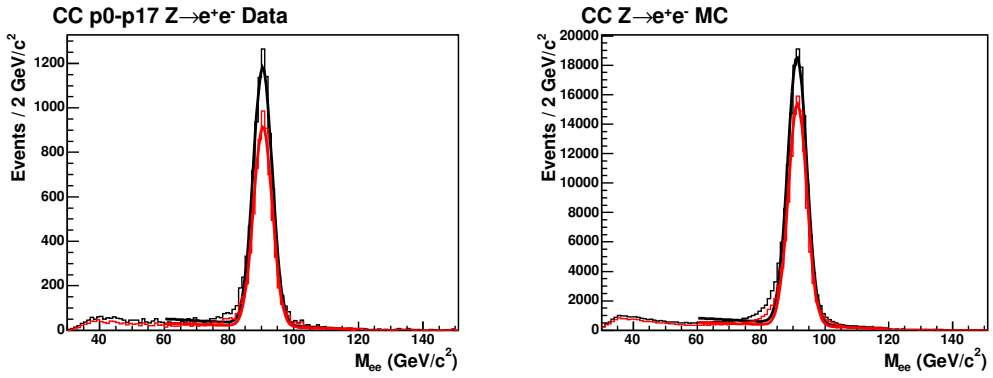
Figure 4.5: After all other cuts have been applied, the NN output for central electrons from Z decays in data and MC simulation is shown, demonstrating the similarity of the MC prediction to the data. The NN output for central photons from a $H \to \gamma\gamma$ MC sample is also shown to demonstrate the similarity of the electron and photon response to the central ID selection. All histograms are normalized to unit area.

$86 < M_{ee} < 98$ GeV/c^2. The background is then subtracted from this value by averaging the number of events away from the central Z mass, on either side of its mean value: $66 < M_{ee} < 72$ GeV/c^2 and $112 < M_{ee} < 118$ GeV/c^2. The background-subtracted number of signal events is labeled N_{Ti} where the T indicates that at least one central leg has passed tight cuts and i indicates that these are events for which the probe leg has passed the photon-like selection at least up through the i^{th} cut. If the probe leg has passed the photon-like selection at least up through the loose (tight) cuts, then the corresponding number of background-subtracted events is labeled N_{TL} (N_{TT}).

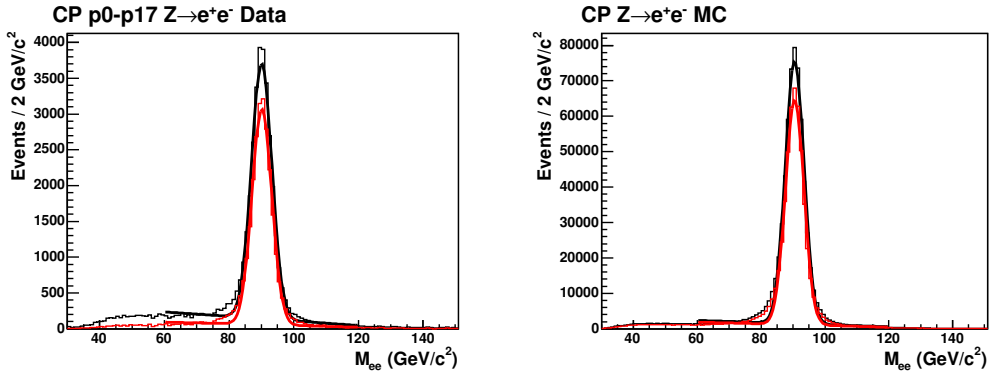
We begin with the efficiency equation for plug photons which is

$$\epsilon_i^{\text{PEM}} = \frac{N_{Ti}}{N_{TL}} \tag{4.4}$$

where ϵ_i^{PEM} is the cumulative plug photon ID efficiency obtained for the photon-like selection applied to the probe leg up through the i$^{\text{th}}$ cut. For the central ID efficiency, the equation is slightly more complicated to account for the fact that both electrons are in the central region and that the tag leg is already required to pass the tight



(a)



(b)

Figure 4.6: Example $Z \rightarrow e^+e^-$ invariant mass (M_{ee}) plots are shown for the (a) central and (b) plug photon ID efficiency study. The black (red) histograms of each plot are from cases in which the probe electron leg passes loose (tight) cuts. The fit discussed in the text is also shown for each histogram. Plots obtained from the data (MC simulation) are shown on the left (right).

central photon-like ID. This bias is avoided with the use of a corrected efficiency formula (derived in Appendix A):

$$\epsilon_i^{\mathrm{CEM}} = \frac{N_{Ti} + N_{TT}}{N_{TL} + N_{TT}} \tag{4.5}$$

where $\epsilon_i^{\mathrm{CEM}}$ is the cumulative central photon ID efficiency obtained for the photon-like selection applied to the probe leg up through the i$^{\mathrm{th}}$ cut.

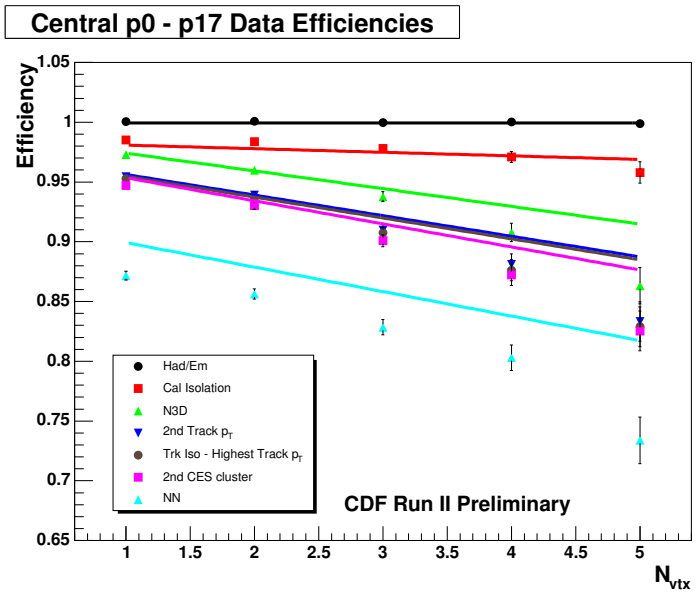
4.3.2 Efficiency Results

We have calculated the data efficiency separately for two different time periods. This is because there was a modification in track reconstruction after CDF's data-taking period 17 and we, therefore, show efficiencies before and after this change. Periods 0 through 17 comprise data taken from February 2002 through April 2008 and periods 18 through 38 comprise data taken from April 2008 through September 2011. Table 4.5 gives the central and plug photon ID efficiencies for the data, after all cuts have been applied to the probe electron leg. Table 4.6 shows the corresponding efficiencies obtained from the MC prediction. Efficiencies are calculated as a function of the number of reconstructed vertices in the event (N_{vtx}) because overlapping collisions affect the ability to distinguish an isolated photon. The photon ID efficiency is lower for higher N_{vtx} because of this overlapping activity in the detector.

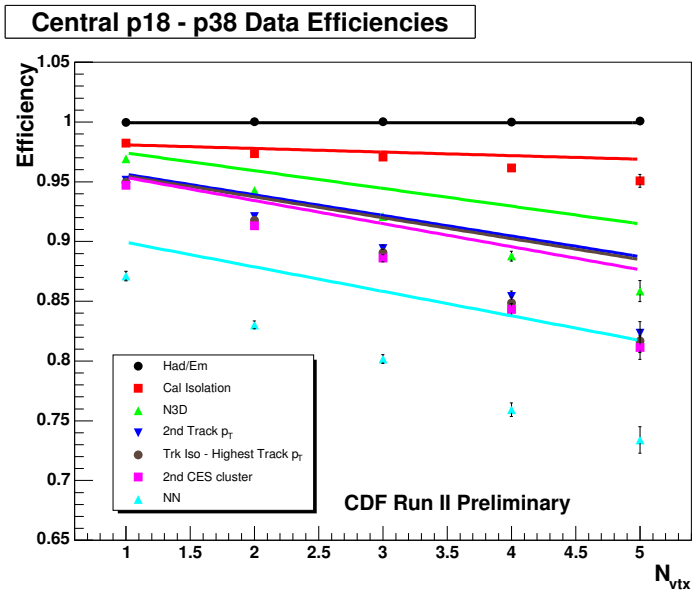
Table 4.5: Efficiencies obtained from $Z \to e^+e^-$ data as a function of N_{vtx}, for period ranges before and after the tracking reconstruction change. Binomial statistical errors are provided.

	Central Data Efficiency (%)		Plug Data Efficiency (%)	
N_{vtx}	p0–p17	p18–p38	p0–p17	p18–p38
1	87.2 ± 0.4	87.1 ± 0.4	82.8 ± 0.2	79.8 ± 0.3
2	85.6 ± 0.4	83.0 ± 0.3	75.6 ± 0.3	71.2 ± 0.2
3	82.8 ± 0.6	80.2 ± 0.4	69.0 ± 0.4	63.3 ± 0.3
4	80.3 ± 1.1	75.9 ± 0.6	61.8 ± 0.7	56.3 ± 0.4
5	73.4 ± 2.0	73.4 ± 1.1	52.8 ± 1.2	51.5 ± 0.7

The central and plug photon ID efficiencies are plotted as a function of N_{vtx} in Figures 4.7 and 4.8, respectively. For each figure, the upper subfigure shows efficiencies for periods 0 through 17 and the lower subfigure shows efficiencies for periods 18 through 38. In each subfigure, the cumulative efficiency for cuts applied up through the photon ID variable indicated are plotted for the data as dots (markers). The corresponding efficiency results from the MC prediction are represented by lines on each plot. The efficiency values for the last cut applied (NN for central photons and track

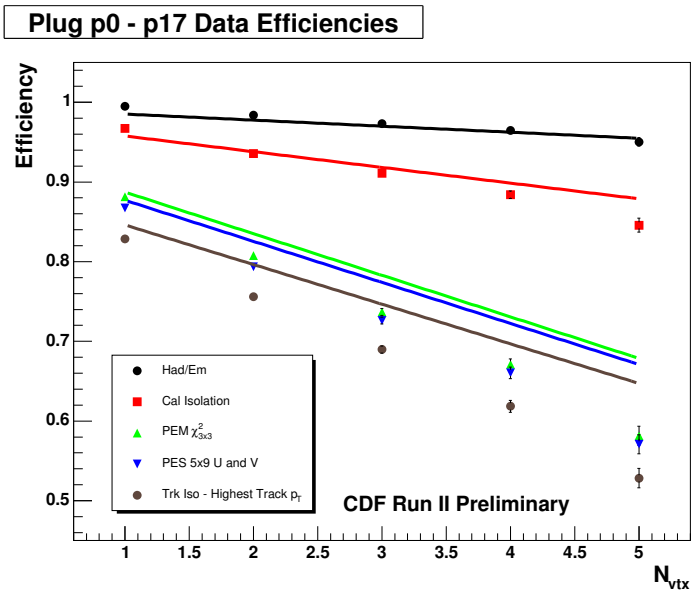


(a)

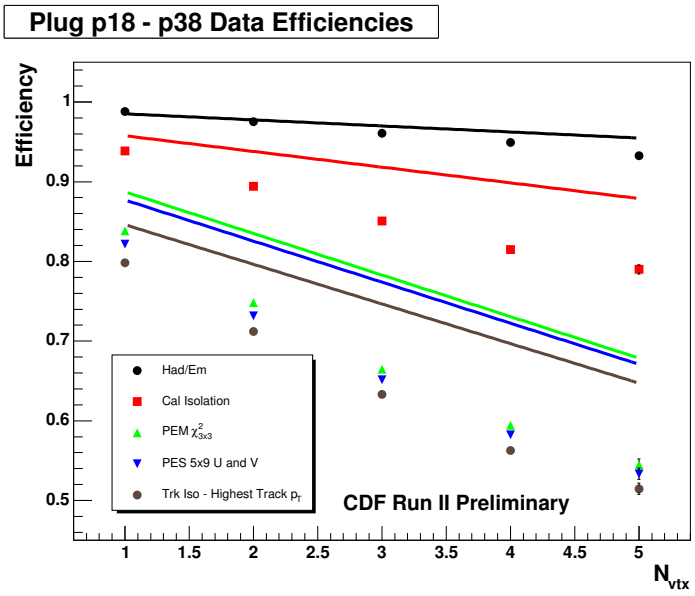


(b)

Figure 4.7: As a function of the number of vertices reconstructed in the event, cumulative central photon ID efficiencies are plotted for each cut made in the order they are shown. Efficiencies from the data are represented by markers (solid shapes) and efficiencies from the MC prediction are represented by lines. The same set of MC efficiency lines are drawn for both data periods. See the text for details on the upper and lower plots.



(a)



(b)

Figure 4.8: As a function of the number of vertices reconstructed in the event, cumulative plug photon ID efficiencies are plotted for each cut made in the order they are shown. Efficiencies from the data are represented by markers (solid shapes) and efficiencies from the MC prediction are represented by lines. The same set of MC efficiency lines are drawn for both data periods. See the text for details on the upper and lower plots.

isolation for plug photons) is the same as the net efficiency values given in Tables 4.5 and 4.6.

Based on the calculated difference in results for the $Z \to e^+e^-$ data relative to the $Z \to e^+e^-$ MC prediction, we apply a photon ID efficiency correction to the $H \to \gamma\gamma$ MC samples that is weighted by $N_{\rm vtx}$. Section 8.3.2 describes this scale factor in more detail.

Table 4.6: Efficiencies obtained from $Z \to e^+e^-$ MC simulation as a function of $N_{\rm vtx}$. Binomial statistical errors are provided.

$N_{\rm vtx}$	Central MC Efficiency (%)	Plug MC Efficiency (%)
1	89.7 ± 0.1	84.5 ± 0.1
2	88.1 ± 0.1	79.8 ± 0.1
3	86.0 ± 0.2	74.6 ± 0.1
4	83.9 ± 0.3	69.6 ± 0.2
5	81.5 ± 0.7	64.8 ± 0.5

4.3.3 Uncertainty in Efficiency

Several sources of systematic uncertainty in the efficiency measurement were considered.

- The efficiency measurements are based on fits made to the Z boson mass distribution (Figure 4.6). Variations in these fits result in an uncertainty of 0.2% (0.8%) applied to central (plug) photons.

- Based on the results of the previous section, a single scale factor is applied to the photon ID efficiency predicted by $H \to \gamma\gamma$ MC samples. Variations in the scale factor for different individual data-taking periods lead to an applied uncertainty of 1.5% (2.0%) for central (plug) photons.

- Efficiency measurements are based on electrons from Z boson decays; however, there are slight differences between electron and photon shower profiles in the

calorimeters and shower maximum detectors. In order to account for this, a 1.0% (2.6%) systematic uncertainty is applied for central (plug) photons, based on the difference in efficiency measured from photon and electron MC samples.

- In the comparison made between electrons and photons, $\gamma \to e^+e^-$ conversions were removed from the photon MC sample. An uncertainty on the efficiency of this removal is derived from the uncertainty in material included in the MC simulation of the detector. An uncertainty of 0.2% (3%) is obtained for the removal of central (plug) conversions in the photon MC sample.

Though this chapter has focused on central and plug non-converting photons, we also reconstruct conversion photons in this analysis. The identification of these photons is described in the next chapter.

Chapter 5

Central Conversion Photon Identification

5.1 Introduction

In the presence of detector material, a photon can convert into an electron-positron pair that is moving in nearly the same direction as the original photon. As discussed in Chapter 3, this is one of the primary mechanisms for creating electromagnetic showers in the EM calorimeter. However, some prompt photons interact with detector material before even reaching the calorimeters and the resulting e^+e^- decay products are generally rejected by the photon ID requirements described in Chapter 4. A separate ID selection is necessary for the reconstruction of these photons, which we call conversion photons or conversions.

Figure 3.4 gave a diagram of the CDF tracking volume that a photon traverses before arriving at the calorimeters. The probability of a photon converting is dependent on the amount of detector material that it travels through, as shown in Figure 5.1. We observe that if a photon goes through the central region of the detector ($|\eta| < 1.1$), the probability that it converts is about 15%. For a photon going through higher $|\eta|$ regions (the plug region), the probability that it converts is roughly double that for

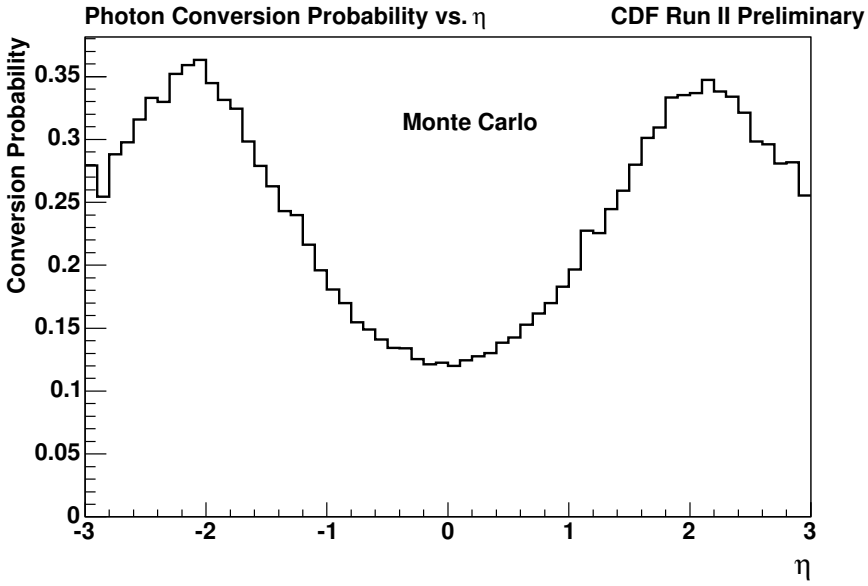


Figure 5.1: Photons from a $H \to \gamma\gamma$ MC sample are used to study the conversion probability as a function of η. We use generator-level ("truth") information from the simulated MC to select photons that convert. Using this truth information, the probability is obtained from the ratio of the η distribution for photons that convert to the η distribution for all generated photons.

the central region. We do not consider these plug conversions, however, due to the lower track reconstruction efficiency in this region.

In $H \to \gamma\gamma$ decays, one or both photons can convert. If we consider those decays that result in two photons in the central region of the detector, the probability that neither converts is roughly $(0.85)^2 \approx 72\%$, that only one converts is roughly $2(0.85)(0.15) \approx 26\%$, and that both convert is roughly $(0.15)^2 \approx 2\%$. We call the first case the CC category and it is these events that provide the greatest $H \to \gamma\gamma$ signal sensitivity for this analysis (the greatest sensitivity for observing Higgs boson signal in the diphoton data relative to the diphoton backgrounds). A considerable improvement to this sensitivity is expected, however, with the inclusion of the second case. This second case is called the C′C category and its contribution to the $H \to \gamma\gamma$ search in this analysis is discussed in Chapter 11. We do not attempt to reconstruct events for the third case because there is a negligible expected gain by doing so. We, therefore, consider only cases in which a single photon converts in the central region of

the detector. As will be described in Chapter 6, this contributes not only to the C′C category, but also to a category in which a plug (non-converting) photon is identified (C′P).

In this chapter, the conversion identification algorithm used for this analysis is described. An uncertainty on the ID efficiency in the MC prediction relative to the data is also derived.

5.2 Central Conversion ID

As previously discussed, the signature of a conversion in the detector is different from that of a non-converting photon. The diagrams in Figure 5.2 illustrate the type of objects that we search for. For a real conversion, the curved arrows represent the collinear charged tracks from an e^+e^- pair. The origin of the tracks indicates the location at which the photon converted. The radial distance to this location in the detector will be discussed more later. The two grey trapezoids represent calorimeter ϕ towers. A 1-tower conversion refers to cases in which two electrons have sufficient track p_T that they deposit their energy in the same ϕ tower. (As a reminder, electron refers to either e^- or e^+). A 2-tower conversion refers to cases in which one electron has lower p_T such that it deposits its energy in a different ϕ tower (not shown). We consider both types for this analysis and treat the reconstruction of the parent photon's energy differently for each case.

Because the e^+e^- tracks may not be separated enough to produce two EM clusters in the calorimeter, we begin the conversion selection by searching for a single EM cluster that has at least one associated track. If there is more than one associated track, then an electron candidate is defined by the EM cluster and the track with the highest p_T. Electron candidates are sorted by decreasing p_T and we next apply a set of track-based cuts to each candidate to determine if it is the result of a conversion. In particular, we search for a second, oppositely signed track originating from the same location as the electron candidate's track. If there is such a track, then a conversion candidate is considered found. If more than one conversion exists in the event, then

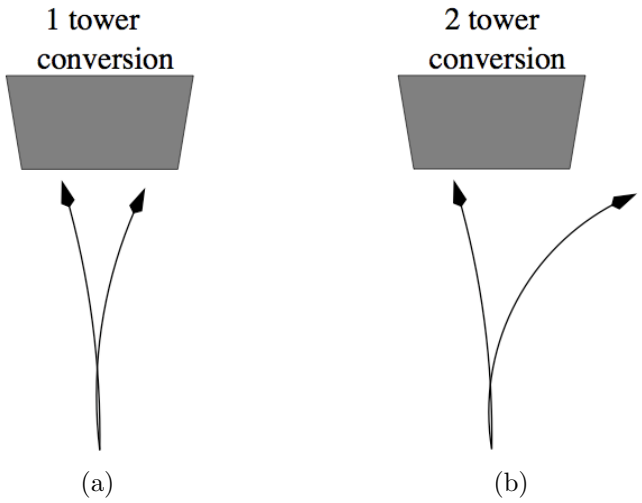


Figure 5.2: Diagram of two different types of conversion signatures in the detector [61]. The curved arrows represent the charged tracks from the e^+e^- pair. The grey trapezoids represent a calorimeter tower. (a) The two tracks point to the same calorimeter tower in 1-tower conversions. (b) The two tracks point to different towers in 2-tower conversions. The second tower is not shown in (b) because the second track does not necessarily point to an adjacent tower.

the one with the highest p_T electron candidate is chosen. A tighter selection based on both calorimeter and track variables is next applied to the conversion in order to better reject backgrounds. For each conversion photon that passes this selection, we reconstruct its kinematic variables and use it in the diphoton selection discussed in Chapter 6.

In addition to potential $H \to \gamma\gamma$ conversions, this selection will also reconstruct real SM photons that are produced through QCD interactions in the primary collision. Other objects can be misidentified as a prompt photon and are backgrounds that we desire to reject. These backgrounds consist of jets that contain $\pi^0/\eta \to \gamma\gamma$ decays with one or both photons converting, prompt electrons associated with a random track or other random electron not resulting from a conversion, tridents (described below), and Dalitz decays where a neutral meson is produced in the primary collision.

A more detailed description is now provided for the conversion reconstruction algorithm. The goal is to use quality measurements from the detector to construct real

conversions from the primary interaction; the identification selection is also designed to reduce the backgrounds just described.

Based on the track parameters and variables defined in Section 3.4, we apply an initial set of track-based cuts similar to that used in other analyses at CDF. We label the electron candidate discussed above as e_1 since it has higher p_T and has an EM cluster. It is also sometimes referred to as the leading electron and must have an EM cluster position of $|\eta| < 1.1$. A second track found near the origin of the leading electron's track indicates the presence of a second electron, labeled e_2. This secondary electron has smaller p_T and may or may not have its own EM cluster.

- *Track quality*: We define the quality of a COT track based on the number of stereo and axial segments reconstructed. For good $r - \phi$ tracking, we require all tracks to have at least two axial segments, each constructed from at least 5 hits. For good $r - z$ tracking, we require all tracks to have at least two stereo segments, one of which must have been constructed from at least 5 hits.

- *Oppositely signed*: The total charge of the two tracks must be zero: $q_1 + q_2 = 0$, where q_1 (q_2) is the track charge of e_1 (e_2).

- *sep*: The sep variable is the spatial separation in the $r - \phi$ plane between the two electron tracks produced from the conversion, at the radial location where they are parallel (see Figure 5.3). It is also known as xy separation and is defined as negative if the tracks cross each other. The distribution shown in Figure 5.4 shows this xy separation between the track from electron candidates (e_1) and any other track in the event. Real conversions originate from a single point and their sep value is, therefore, clustered around 0 cm. The red vertical lines indicate where a cut of |sep| < 0.2 cm is applied based on optimization studies from previous analyses [62]. Appendix B provides a calculation of this variable from track parameters.

- $\Delta \cot\theta$: The difference in the $\cot\theta$ between the two electron tracks from the conversion. Its distribution is shown in Figure 5.5 where true conversions are clustered around $\Delta \cot\theta = 0$. The red vertical lines indicate where a cut of

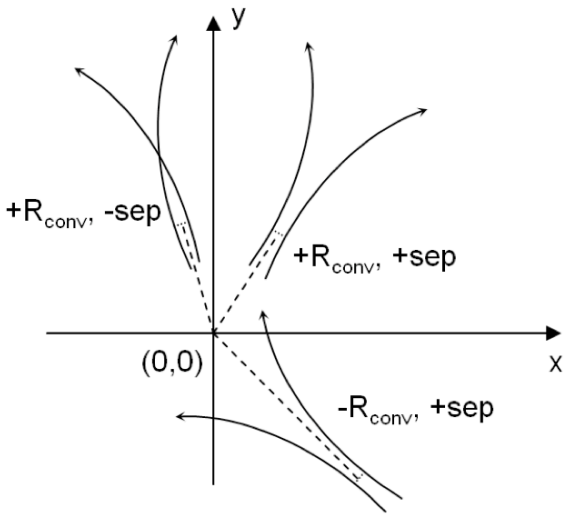


Figure 5.3: The radius of conversion, R_{conv}, is the radial distance from the center of the detector to the location where the two tracks are parallel. Sep is the spatial separation between the tracks in the $r-\phi$ plane at this location. Three different conversion examples are shown [63] illustrating the R_{conv} variable as a dashed line and the sep variable as a dotted line. The R_{conv} is positive (negative) if the conversion points away from (towards) the origin and sep is positive (negative) if the electron-positron tracks do not (do) cross one another.

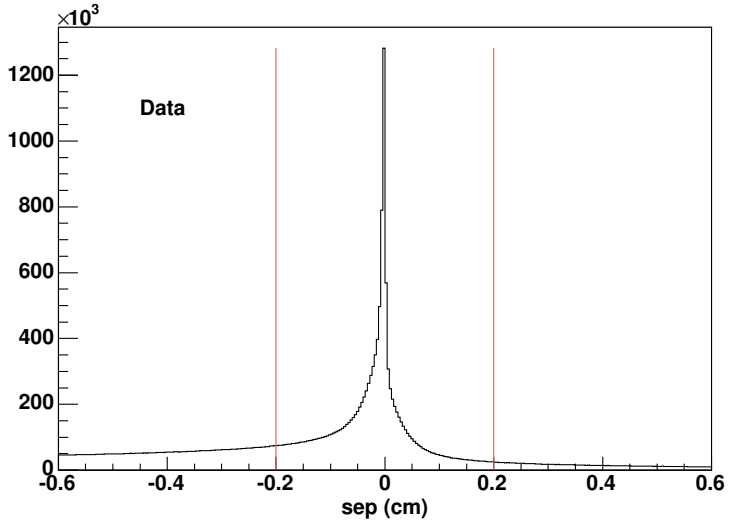


Figure 5.4: The sep distribution for electron candidate (e_1) tracks and any other track in the event. True conversions populate the sep = 0 cm region. Red lines show where cuts are made. (The data sample is described in Chapter 6.)

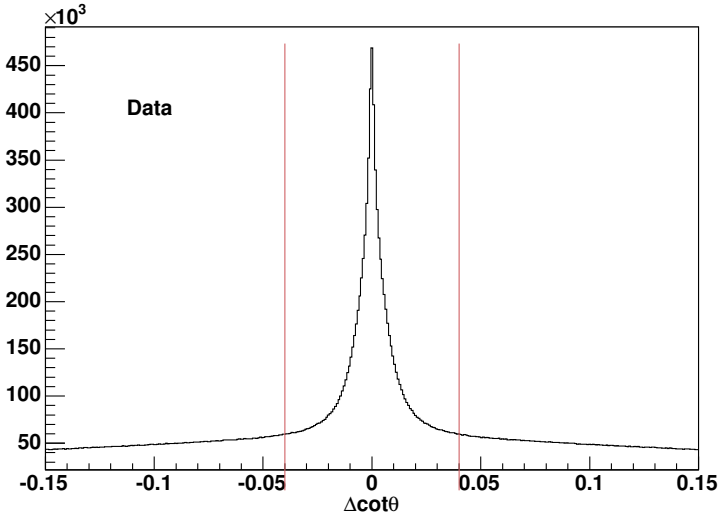


Figure 5.5: The $\Delta \cot\theta$ distribution for electron candidate (e_1) tracks and any other track in the event. True conversions populate the $\Delta \cot\theta = 0$ region. Red lines show where cuts are made. (The data sample is described in Chapter 6.)

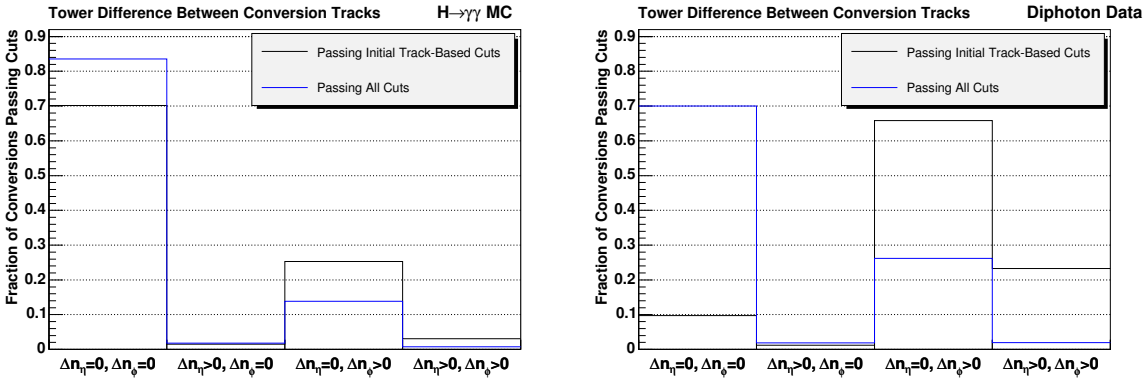
$|\Delta \cot\theta| < 0.04$ is applied based on optimization studies from previous analyses [62].

- *Trident veto*: A "trident" background occurs from an electron that radiates a photon via bremsstrahlung, which then converts to an electron-positron pair ($e^{\pm}\gamma \to e^{\pm}e^{+}e^{-}$). The three electron tracks are nearly parallel, and a selection similar to the above cuts is applied in order to remove cases where there is a third track (e_3) present. We search for a third track that passes the same track quality cuts as above and is of opposite sign with e_2. We next apply sep and $\Delta \cot\theta$ cuts between the second and third tracks. Tight cuts of $|\text{sep}| < 1.6$ cm and $|\Delta \cot\theta| < 0.2$ and are found to better reject a background from Z boson decays.

Cuts on these five variables define the initial track-based selection. Electron candidates (e_1) that pass this selection are considered conversion candidates. If there is more than one conversion candidate in an event, we select the one with the highest e_1 track p_T. A conversion candidate is then categorized as either a 1- or 2-tower

conversion, determined from the specific tower(s) with which the tracks are associated. The details on this categorization are described next.

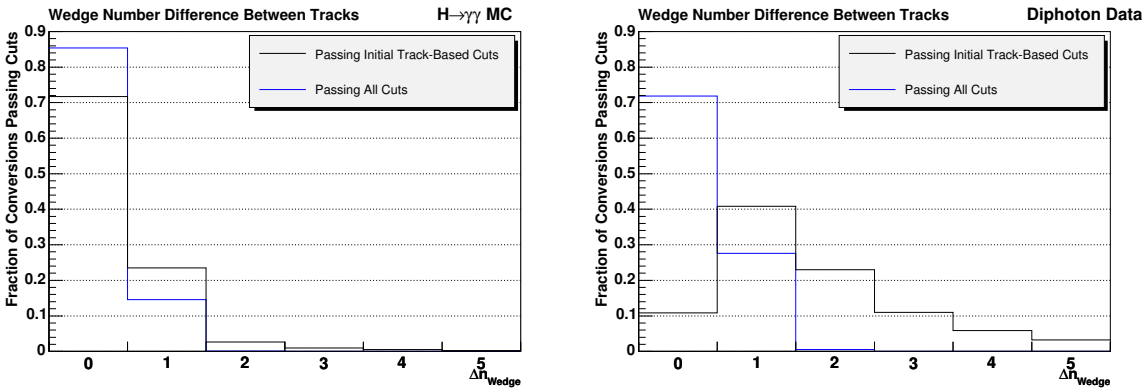
In order to determine the ϕ and η tower number (n_ϕ and n_η, respectively) that a track points to, the track trajectory is extrapolated to the plane of the CES. Figure 5.6 shows the fraction of conversions for four possible scenarios of Δn_ϕ and Δn_η between the two tracks. The majority of conversion candidates have both tracks pointing to the same η tower; these are the first and third scenarios of Figure 5.6, where $\Delta n_\eta = 0$. This is expected since the solenoid magnetic field causes tracks to curve in the azimuthal direction rather than in the polar direction. We then categorize conversions based on differences in Δn_ϕ. By this grouping, the first two scenarios are different types of 1-tower conversions and the second two scenarios are different types of 2-tower conversions.



(a) (b)

Figure 5.6: Different scenarios of Δn_η and Δn_ϕ for (a) $H \to \gamma\gamma$ conversions and (b) conversions reconstructed in the data described in Chapter 6. The $\Delta n_\eta = 0$, $\Delta n_\phi = 0$ scenario are 1-tower conversions in which the two tracks point to the same tower. Essentially all conversions in the $\Delta n_\eta > 0$, $\Delta n_\phi = 0$ scenario have $\Delta n_\eta = 1$, which means that most of the shower energy from e_2 is generally included in the 1–3 tower EM cluster for the leading electron. These are considered 1-tower conversions. The $\Delta n_\eta = 0$, $\Delta n_\phi > 0$ scenario are 2-tower conversions where the two tracks point to the same η tower, but different ϕ wedges. The last scenario shows another type of 2-tower conversion. The fraction of conversions with tracks in different ϕ towers reduces after the initial track-based selection due to additional cuts such as requiring a minimum p_T for the second track.

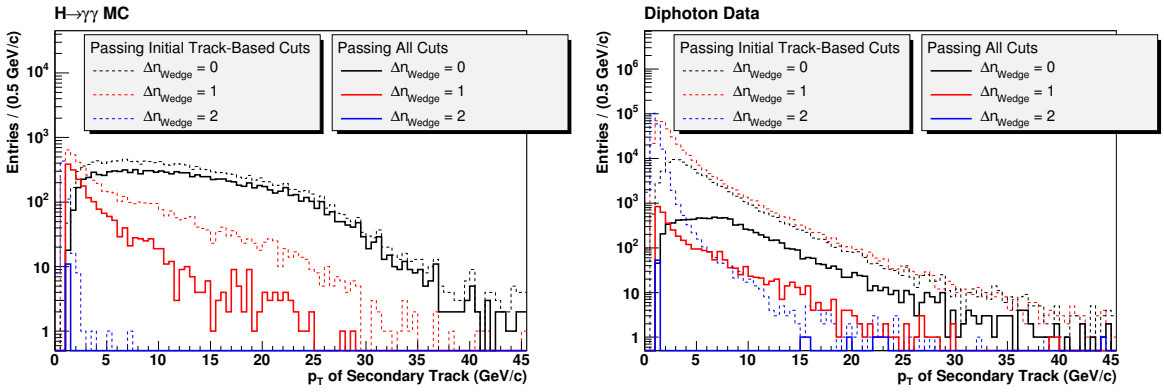
In the CEM, a ϕ tower corresponds to a single ϕ wedge. The information in Figure 5.6 is then more easily summarized by considering the wedge separation between the two tracks, Δn_{Wedge}. This is shown in Figure 5.7. The wedge separation is calculated by first obtaining the wedge number that the two tracks point to, n^1_{Wedge} and n^2_{Wedge}. We next define the variable $\Delta n_{\text{Wedge}} = |n^1_{\text{Wedge}} - n^2_{\text{Wedge}}|$. (The numbering scheme for CDF wedges, 0–23, is such that if $\Delta n_{\text{Wedge}} > 12$, then we redefine Δn_{Wedge} as $24 - \Delta n_{\text{Wedge}}$.) Whether or not the two tracks fall into the same wedge is correlated with the p_T of the second track (e_2). This is clearly seen in Figure 5.8 where the average p_T becomes much lower for conversions with $\Delta n_{\text{Wedge}} > 0$.



(a) (b)

Figure 5.7: The wedge number separation for (a) $H \to \gamma\gamma$ conversions and (b) conversions reconstructed in the data described in Chapter 6. Conversions from $H \to \gamma\gamma$ events are of higher p_T than those from backgrounds in the data. A larger fraction of $H \to \gamma\gamma$ conversions are, therefore, expected to be 1-tower types.

It is interesting to note that the majority of $H \to \gamma\gamma$ conversions in the MC simulation are found to be 1-tower types. This is shown in Figure 5.7 (a), where we find that about 72% of $H \to \gamma\gamma$ conversions that pass the initial track-based selection have a second track with sufficiently high p_T that the conversion is categorized as a 1-tower conversion. This increases to 86% after applying all cuts described in this chapter. The remaining conversions reconstructed are categorized as 2-tower conver-



(a) (b)

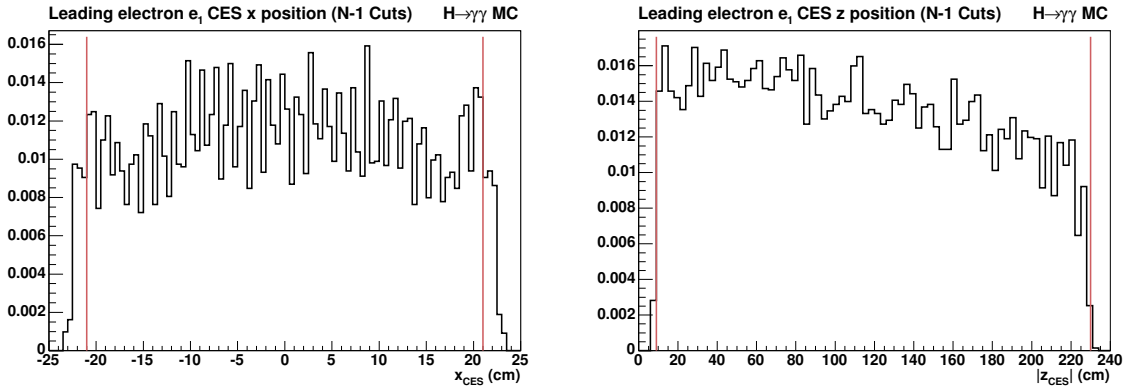
Figure 5.8: The p_T distribution of the second track of a conversion for different wedge separation numbers. This is shown for (a) $H \to \gamma\gamma$ conversions and (b) conversions reconstructed in the data described in Chapter 6. Conversions from $H \to \gamma\gamma$ events are of higher p_T than those from backgrounds in the data. The second track is, therefore, expected to have higher than average p_T than backgrounds in the data.

sions; from Figure 5.7 (a), we see that most of these conversions have a second track with sufficiently high p_T that the two tracks point to adjacent ϕ towers (wedges).[1]

After conversion candidates are categorized as 1- or 2-tower types, further requirements are applied in order to select higher quality conversions and to better reject backgrounds. The energy corrections of Section 4.1 applied to photon EM clusters are similarly applied to the electron EM clusters.

- *CES Fiducial (e_1)*: As with central photons, we require the electron candidate's EM shower to have a CES cluster position (Equation (3.16)) that is in the fiducial region of the CES detector. The fiducial requirement ensures that electron candidates are selected that have an EM cluster energy measurement from the well-instrumented region of the CEM. Figure 5.9 shows the CES x and z distribution for electron candidates in the $H \to \gamma\gamma$ MC prediction that have passed all of the conversion selection given in this chapter except for the cut on the distribution shown (called "$N-1$" cuts).

[1] Only about 0.2% of $H \to \gamma\gamma$ conversions that pass the initial track-based selection have a second track with sufficiently low p_T that it does not reach the calorimeters. This type of conversion is, therefore, not reconstructed.



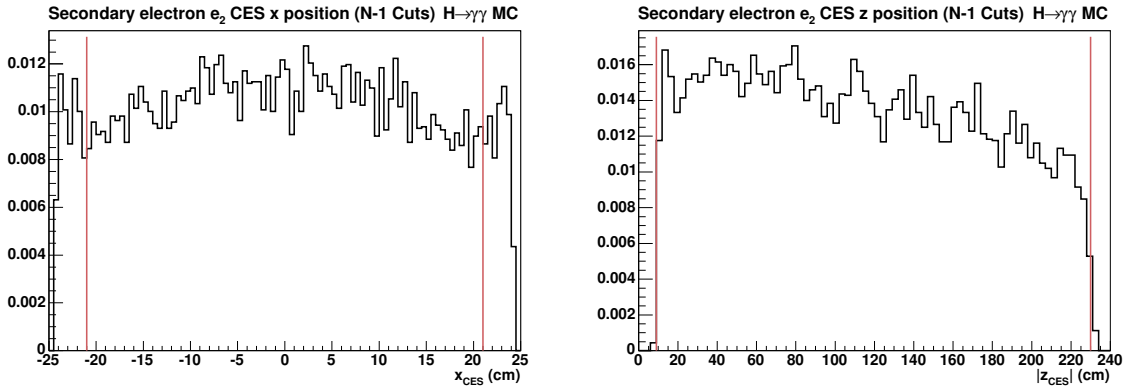
(a) (b)

Figure 5.9: The local x and z position of the electron candidate's EM cluster in the CES for conversions from a $H \to \gamma\gamma$ MC simulation. All cuts made in the conversion selection have been applied, except for the cut on the variable shown ("$N-1$ cuts"). Red lines show where cuts are made.

- *CES Fiducial (e_2)*: The second track is extrapolated to the plane of the CES, and the corresponding x and z positions in the CES (Figure 5.10) are required to be in a fiducial region. For 1-tower conversions, the second track points to the same cluster as e_1; requiring it point to a fiducial region of the CES is just an extension of the same cut for e_1 and ensures a quality EM energy measurement. For 2-tower conversions the fiducial requirement for e_2 improves the conversion calorimeter isolation variable that is described below.

- *Track p_T of e_2*: Figure 5.8 showed the track p_T distribution of the second track after the initial-track based selection and after all cuts have been applied. The pair finding efficiency of the sep and $\Delta \cot\theta$ requirements is much lower when the second electron has a p_T below about 1 GeV/c. We select conversions with a second track p_T above this value in order to improve the agreement in this efficiency for the MC prediction relative to the data [64].

- *Conversion p_T*: The conversion photon's momentum vector is reconstructed by first extrapolating the tracks to the radius of the conversion and then taking the vector sum of the momentum of the two tracks at this location. The conversion p_T is the transverse part of this vector and we select photons with



(a) (b)

Figure 5.10: The local x and z position in the CES that the second track of a $H \to \gamma\gamma$ conversion points to. All cuts made in the conversion selection have been applied, except for the cut on the variable shown ($N-1$ cuts). Red lines show where cuts are made.

$p_T > 15$ GeV/c (see Figure 5.11), the same requirement made for central and plug non-converting photons. The pair finding efficiency of the sep and $\Delta \cot\theta$ requirements is also found to be in better agreement between data and MC simulation for higher p_T values.

- *Had/Em for* e_1: As with isolated photons or electrons, jets that pass the track-based conversion selection can be better rejected by constraining the amount of energy in the hadron calorimeters. The Had/Em ratio for the electron candidate has a similar distribution as that for isolated electrons; the standard CDF tight cut is, therefore, applied to the leading electron. The Had/Em cut slides with energy because the fraction of energy from an EM shower that leaks into the hadron calorimeters increases for more energetic showers. Figure 5.12 (a) demonstrates that the Had/Em shape for electrons from a $H \to \gamma\gamma$ conversion is similar to that of isolated electrons. The Had/Em distribution for e_1 from $H \to \gamma\gamma$ conversions and jet conversion candidates is shown in Figure 5.12 (b).

- *Conversion* E/p: We construct an E/p variable similar to that used for isolated electrons; the ratio is obtained from the reconstructed conversion E_T to the conversion p_T (the latter is described above). For 1-tower conversions, the E_T

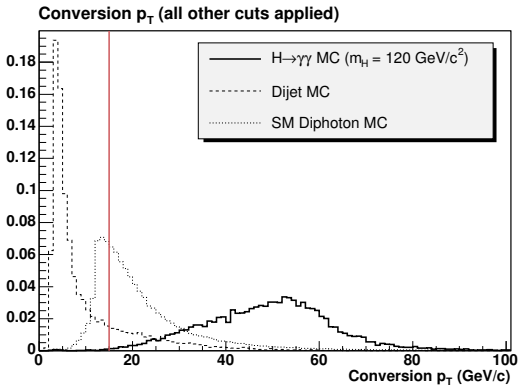
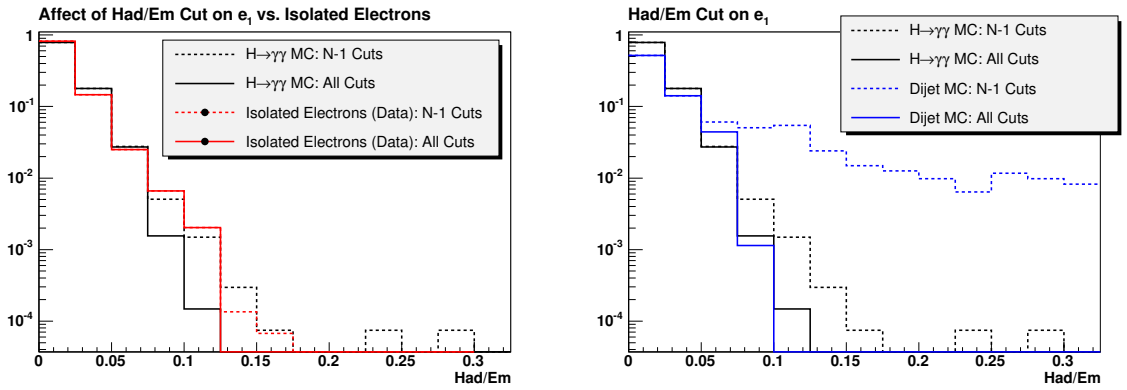


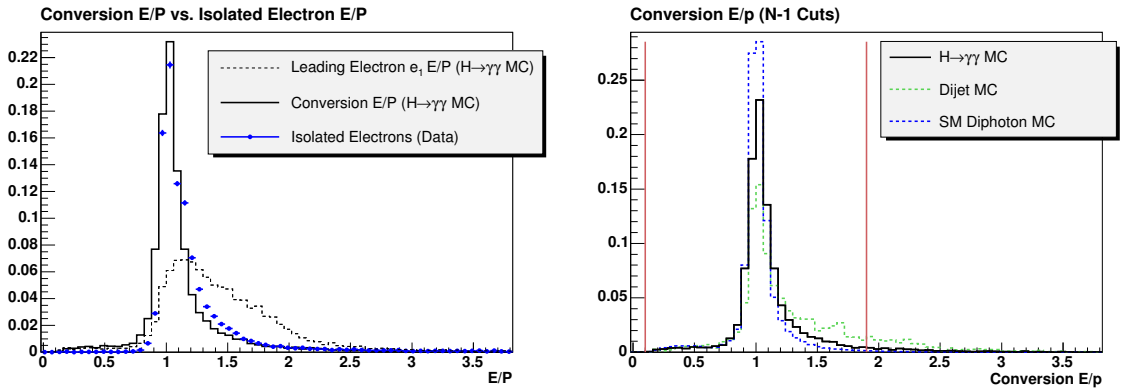
Figure 5.11: The p_T distribution for conversions reconstructed in $H \to \gamma\gamma$, dijet, and QCD diphoton MC samples; each histogram is scaled to unity. Conversion photons from Higgs boson decays have a higher than average p_T value compared to the backgrounds. Conversions from jets are from $\pi^0/\eta \to \gamma\gamma$ decays and populate the lower p_T region; much of these are removed with a cut on this variable. All cuts made in the conversion selection have been applied, except for the cut on the variable shown ($N-1$ cuts).



(a) (b)

Figure 5.12: The dashed lines in both figures show the distribution after all other cuts have been applied except for the Had/Em cut ($N-1$ cuts). The solid lines show the distribution after this cut has been applied. For the $H \to \gamma\gamma$ and dijet MC samples, the other cuts applied are those from this chapter. The isolated electrons come from $Z \to e^+e^-$ decays and the other cuts applied are those described in Section 4.3.

is calculated as the electron's candidate's EM cluster E_T. The p_T of the second track is added to this value for 2-tower conversions.[2] Figure 5.13 (a) shows the conversion E/p distribution for photons from $H \to \gamma\gamma$ MC simulation; the distribution from isolated electrons in data is provided for comparison, along with the electron candidate's E/p shape before the above corrections are applied. The E/p shape for two backgrounds in the $H \to \gamma\gamma$ search is given in Figure 5.13 (b). Figure 5.14 shows an optimization study performed on both this variable and conversion calorimeter isolation. For different E/p and isolation cuts applied to a conversion candidate, an estimation is made of the number of expected $H \to \gamma\gamma$ diphoton events (s) and the number of expected background events (b) in the diphoton data. We select a cut that maximizes the $H \to \gamma\gamma$ significance, which is defined as $s/\sqrt{b}$. For each trial, a symmetric cut is applied to the E/p shape, so that a cut of 1.4 on the x axis represents selecting conversions with $0.6 < E/p < 1.4$. This example cut would be a reasonable choice, however, a slight gain is obtained by requiring $0.1 < E/p < 1.9$.



(a) (b)

Figure 5.13: (a) The conversion E/p shape is compared to the leading electron's E/p shape and isolated electrons in the data. (b) The E/p shape for conversions reconstructed in $H \to \gamma\gamma$, dijet, and QCD diphoton MC samples. The red lines show where cuts are made on this variable.

[2] In 2-tower conversions, the resolution for the reconstructed second track p_T compared to the $H \to \gamma\gamma$ MC-generated p_T is found to be slightly better than the same comparison made for the second electron's cluster E_T.

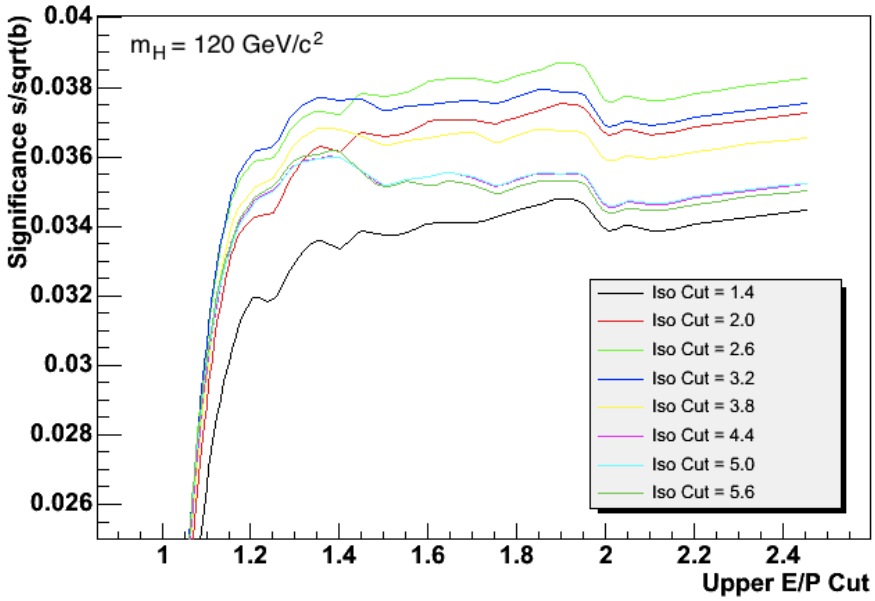


Figure 5.14: Optimization study on cuts applied to the conversion E/p and calorimeter isolation. For each trial, a symmetric cut is applied to the E/p shape, so that a cut of 1.9 on the x axis represents selecting conversions with $0.1 < E/p < 1.9$. The different colors represent different isolation cuts, such as requiring calorimeter isolation to be less than 2.6 GeV.

- *Conversion Calorimeter Isolation*: The shape of this variable is provided in Figure 5.15 for $H \to \gamma\gamma$ conversions and conversions from backgrounds. A cut on the conversion isolation is applied in order to better remove jet backgrounds. For 1-tower conversions, the isolation value is defined as the calorimeter isolation measured for the leading electron (e_1). The p_T of the secondary electron is subtracted from this value for 2-tower conversions. As with E/p, we use the results of Figure 5.14 to determine a cut on the isolation. For the E/p cut selected, the $H \to \gamma\gamma$ significance is maximized by requiring the isolation energy to be less than 2.6 GeV.

- $R_{\rm conv}$: The radius of the conversion is defined as the radial distance from the detector origin to the position where the two tracks are parallel (Figure 5.3). Appendix B provides a calculation of this variable from track parameters. The distribution of this variable is plotted in Figure 5.16 (a) for $H \to \gamma\gamma$, jet, and

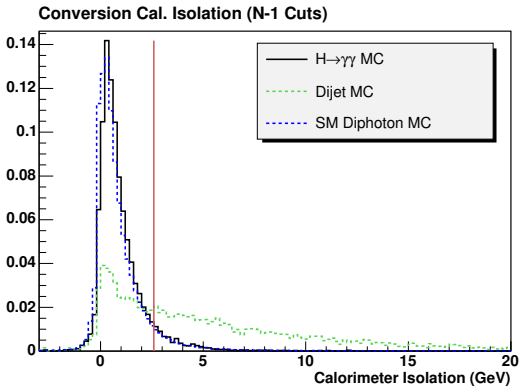
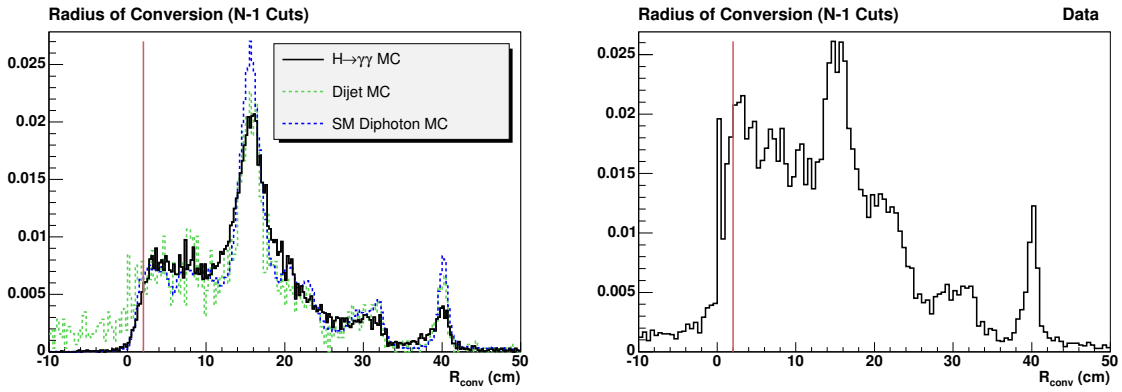


Figure 5.15: The calorimeter isolation distribution for conversions reconstructed in $H \to \gamma\gamma$, dijet, and QCD diphoton MC samples. The red line shows where a cut is made on this variable.

SM prompt conversion photons. The distribution for conversions reconstructed in the data using the conversion ID selection of this chapter is shown in Figure 5.16 (b). The negative tail in the R_{conv} distribution is mostly populated by fake conversions that tend to be symmetric about $R_{\text{conv}} = 0$. Additionally, prompt electron-positron pairs from the Dalitz decay of neutral pions ($\pi^0 \to e^+e^-\gamma$) tend to have $R_{\text{conv}} = 0$. A cut of $R_{\text{conv}} > 2.0$ cm removes much of these backgrounds.



(a) (b)

Figure 5.16: (a) The radius of conversion distribution for conversions reconstructed in $H \to \gamma\gamma$, dijet, and QCD diphoton MC samples. (b) The radius of conversion distribution for conversions reconstructed in the data described in Chapter 6. The red line shows where a cut is made on this variable.

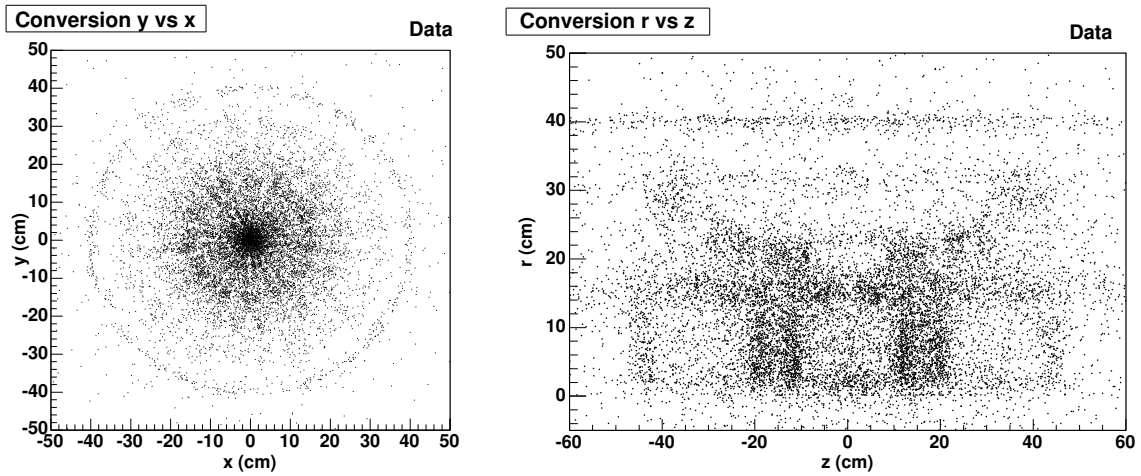
The full conversion selection is summarized in Table 5.1 and consists of the initial track-based selection and cuts made on the track- and calorimeter-based variables just described. After selecting conversions from this identification, we next define the kinematic variables used for the conversion photon.

Table 5.1: A summary of all cuts made for the conversion selection. The first cut selects central electron candidates; each of these candidates has an EM cluster with a track pointing to the cluster. The second level applies a set of track-based cuts to identify electron pairs originating from a conversion photon. The last set of cuts selects higher quality conversions and better rejects backgrounds.

	Variable	Cut
Central	$\|\eta\|$ of e_1	< 1.1
Initial Track-based Selection	N axial segments with ≥ 5 hits	≥ 2
	N stereo segments	≥ 2
	N stereo segments with ≥ 5 hits	≥ 1
	Oppositely signed tracks	
	$\|sep\|$	< 0.2 cm
	$\|\Delta \cot\theta\|$	< 0.04
	Trident removal	
Remaining Selection	CES Fiducial (e_1)	$\|x_{\rm CES}\| < 21$ cm, $9 < \|z_{\rm CES}\| < 230$ cm
	CES Fiducial (e_2)	$\|x_{\rm CES}\| < 21$ cm, $9 < \|z_{\rm CES}\| < 230$ cm
	Track p_T of e_2	> 1.0 GeV/c
	Conversion p_T	> 15.0 GeV/c
	Had/Em for e_1	$< 0.055 + 0.00045 \times E$
	Conversion E/p	> 0.1 and < 1.9
	Conversion Calorimeter Isolation	< 2.6 GeV
	$R_{\rm conv}$	> 2.0 cm

Before doing so, however, it is interesting to study the reconstructed radius of the conversion distribution, $R_{\rm conv}$, that was provided in Figure 5.16. As previously discussed, the probability of a photon converting depends on the amount of material it travels through. Resonances in the $R_{\rm conv}$ distributions of Figure 5.16 then indicate regions of the detector where there is a higher density of detector material: the L00 detector and first five layers of the SVX contribute to conversions with $R_{\rm conv} < 10$ cm;

port cards and cables in the silicon detector create the high peak at about 15 cm; the outer SVX and ISL layers contribute to conversions with $R_{\rm conv}$ between about 18 and 35 cm; the inner edge of the COT is at about 40 cm. We can also plot the conversion position in the $x-y$ and $r-z$ views, which is provided in Figure 5.17 for conversions reconstructed in the data. Though we do not do so in this analysis, conversions are often used to study the material distribution of the detector as shown in these plots.



(a) (b)

Figure 5.17: The reconstructed point of conversion in the (a) $x-y$ view and (b) the $r-z$ view for conversions reconstructed in data described in Chapter 6. All cuts described in this chapter have been applied except for the cut on R_{conv}.

5.3 Kinematic Variables

For all conversion candidates that pass the full selection of Table 5.1, we next construct the photon's kinematic variables for use in the diphoton analysis. In particular, the photon's four-momentum is constructed, which we define based on the smallest $H \to \gamma\gamma$ mass resolution obtained when reconstructing the diphoton mass from a central photon and central conversion. For this purpose, two variables are defined: (a) the track-based vector momentum, $\vec{p}_{\rm conv}$, and (b) the mostly calorimeter-based energy, $E_{\rm cal}$. The first is the conversion p_T that was already described, obtained from the vector sum of the two track momenta. For 1-tower conversions, the $E_{\rm cal}$ variable

is obtained from the electron candidate's EM cluster energy; the momentum of the second track, $p_{\rm conv}$, is added to this quantity for 2-tower conversions.

Figure 5.18 provides the diphoton mass shape obtained for $H \to \gamma\gamma$ events with various definitions for the conversion four-momentum (with $c = 1$). The best resolution (black histogram in the figure) is obtained by using $E_{\rm cal}$ to define the conversion's energy and then requiring that the conversion be massless; this requirement is satisfied by setting the magnitude of the vector momentum to the energy measurement:[3]

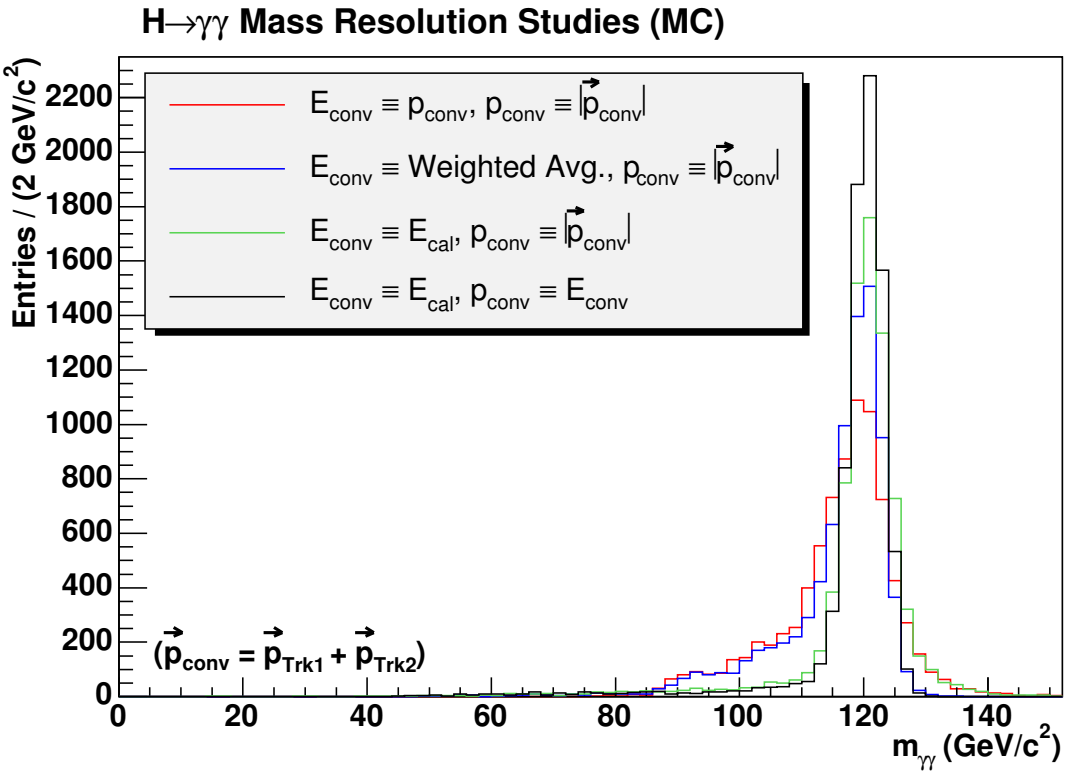


Figure 5.18: The Higgs boson mass for $m_H = 120$ GeV/c^2 is shown, reconstructed from a central photon and a central conversion. Each histogram shows a different manner in which the conversion four-momentum was defined based on calorimeter and/or tracking measurements. The "Weighted Avg." method refers to defining the conversion energy from an error-weighed average of the calorimeter energy and track momentum. The black histogram shows the method that provides the best resolution, which is discussed in the main text.

$p_{\text{conv}} = E_{\text{cal}}$. The direction of the photon (ϕ, θ) is still obtained from the track-based vector momentum. The conversion's four-momentum is then defined as

$$\begin{aligned} P &= (E, p_x, p_y, p_z) \\ &= (E_{\text{cal}}, E_{\text{cal}} \sin\theta \cos\phi, E_{\text{cal}} \sin\theta \sin\phi, E_{\text{cal}} \cos\theta). \end{aligned} \tag{5.1}$$

For conversion candidates that pass the selection of the previous section, this four-vector is constructed for use in the diphoton mass reconstruction discussed in the next chapter.

This now concludes the description of the conversion identification algorithm for the $H \to \gamma\gamma$ analysis. We next derive an uncertainty on the conversion identification.

5.4 Central Conversion ID Efficiency

The purpose of measuring the uncertainty in the conversion identification selection is to understand how well the MC simulation models the data response to our conversion ID. We later assign this uncertainty to the number of predicted $H \to \gamma\gamma$ events expected in the data for diphoton categories that contain a conversion photon. In this section, we first provide an introduction to the method, describing the data and MC samples used and the conversion variables that are studied. Two methods or approaches to studying the uncertainty are then described, which are labeled Method A and B. The results of these methods are used to constrain and validate the final uncertainty that is derived and applied to the $H \to \gamma\gamma$ MC samples.

[3] For the energetic conversion photons of interest to a $H \to \gamma\gamma$ analysis, we find that the reconstructed energy resolution is smaller than the reconstructed momentum resolution. This is suspected to be due to the fact that electrons from the conversion can radiate photons through bremsstrahlung, a process that reduces the track p_T measurement. On the other hand, the electron's energy before bremsstrahlung is often reconstructed because the radiated photon deposits its energy in the same cluster as the post-bremsstrahlung electron.

5.4.1 Introduction

As with the non-converting photon ID efficiency that was described in Section 4.3, we use an electron probe leg from $Z \to e^+e^-$ decays in order to understand how well the MC simulation models the data response to our conversion ID selection. For the tag leg, we identify a tight electron using the central selection of Section 4.3. For the probe leg, we search for a trident in which the other electron from the Z boson produces a photon via bremsstrahlung that converts to an electron-positron pair. An illustration of a trident is provided in Figure 5.19 and we use the radiated photon to test the efficiency of the conversion selection. For this study, the data corresponds to an integrated luminosity of 7.0 fb^{-1}, collected from the same trigger used for the non-converting photon ID studies; we use the same MC samples of Section 4.3.

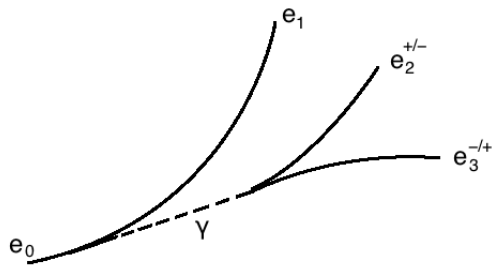


Figure 5.19: Illustration of a trident.

We focus in particular on the efficiency of the track-based selection for finding the second electron from a photon conversion. Calorimeter-based variables such as Had/Em, E/p, and calorimeter isolation have been studied in the past for EM objects like photons and electrons and are assumed here to be modeled well in the MC simulation. In each study, there are four electrons per event that are identified: e_{iso} refers to the central isolated electron identified using the selection in Table 4.3; the other three electrons are from the trident and are labeled e_1, e_2, and e_3 as shown in Figure 5.19. The two electrons from the conversion are then e_2 and e_3, a different numbering scheme from Section 5.2.

5.4.2 Method A

As Figure 5.20 illustrates, the photon from bremsstrahlung generally has lower p_T than the photons of interest from $H \rightarrow \gamma\gamma$ events. For the first study then, we aim to select higher p_T conversion candidates produced in the trident. In the next method, we will remove much of the initial selection applied to third track candidates in this study.

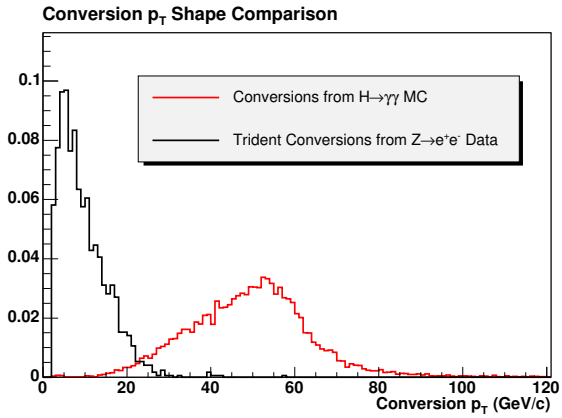


Figure 5.20: The conversion p_T distribution from $H \rightarrow \gamma\gamma$ photons versus photons from tridents.

In addition to already requiring a central tight isolated electron be reconstructed, Table 5.2 summarizes the loose set of requirements made on the trident selection. The first electron must be in the central region, have a fiducial CES cluster, and have only a small fraction of its energy in the hadron calorimeter. We also require that it have a high quality track (same definitions on track quality as discussed in Section 5.2) that is of opposite sign to that of the isolated electron. We then search for a second track that points to a fiducial CES cluster, has a track p_T smaller than that of the first electron, and is near the first electron's track at the radius where the two tracks are parallel. We next search for any third track that is opposite in sign to that of e_2, has at least 1.0 GeV/c transverse momentum, and has smaller p_T than the first electron. In order to select conversion candidates of higher momentum, the second and third track must construct a total vector momentum with $p_T > 10$ GeV/c. Similar to the construction of E/p and calorimeter isolation discussed in Section 5.2, the third track

Table 5.2: The loose selection made for the Method A efficiency study. The first set of cuts is applied to the leading electron in the trident, e_1. The second set of cuts search for a second track, e_2, that can be paired with the leading electron. The last set of requirements selects third track candidates that form a higher p_T conversion with e_2; no proximity cuts (i.e. sep or $\Delta \cot\theta$) are yet made on this third track.

Variable	Cut	Applied to
$\lvert\eta\rvert$	< 1.1	
CES Fiducial		
Track Quality		e_1
Oppositely signed track to that of $e_{\rm iso}$	$q_{\rm iso} + q_1 = 0$	
Had/Em	$< 0.055 + 0.00045 \times E$	
$\lvert{\rm sep}_{12}\rvert$	< 0.25 cm	
$\lvert\Delta \cot\theta_{12}\rvert$	< 0.1	e_2
CES Fiducial		
p_{T2}	$< p_{T1}$	
Oppositely signed track to that of e_2	$q_2 + q_3 = 0$	
p_{T3}	> 1.0 GeV/c	
p_{T3}	$< p_{T1}$	e_3
CES Fiducial		Candidates
Conversion p_T from e_2 and e_3	> 10.0 GeV/c	
Trident E/p	> 0.1 and < 1.9	
Trident Calorimeter Isolation	< 2.6 GeV	

must contribute to the trident E/p and isolation such that the trident passes the cuts of Section 5.2 on these variables.

The second track and any third track are considered a conversion candidate. We next fill a histogram of the $\Delta \cot\theta_{23}$ between these two tracks; if there is more than one potential third track, then it is filled more than once per event. The distribution for these entries is given in the upper left subplots of Figure 5.21 for the data and Figure 5.22 for the MC prediction. The number that are considered to be true conversions are the number in the peak of this distribution minus the estimated background; the background is estimated from the number outside the peak (the sidebands). This

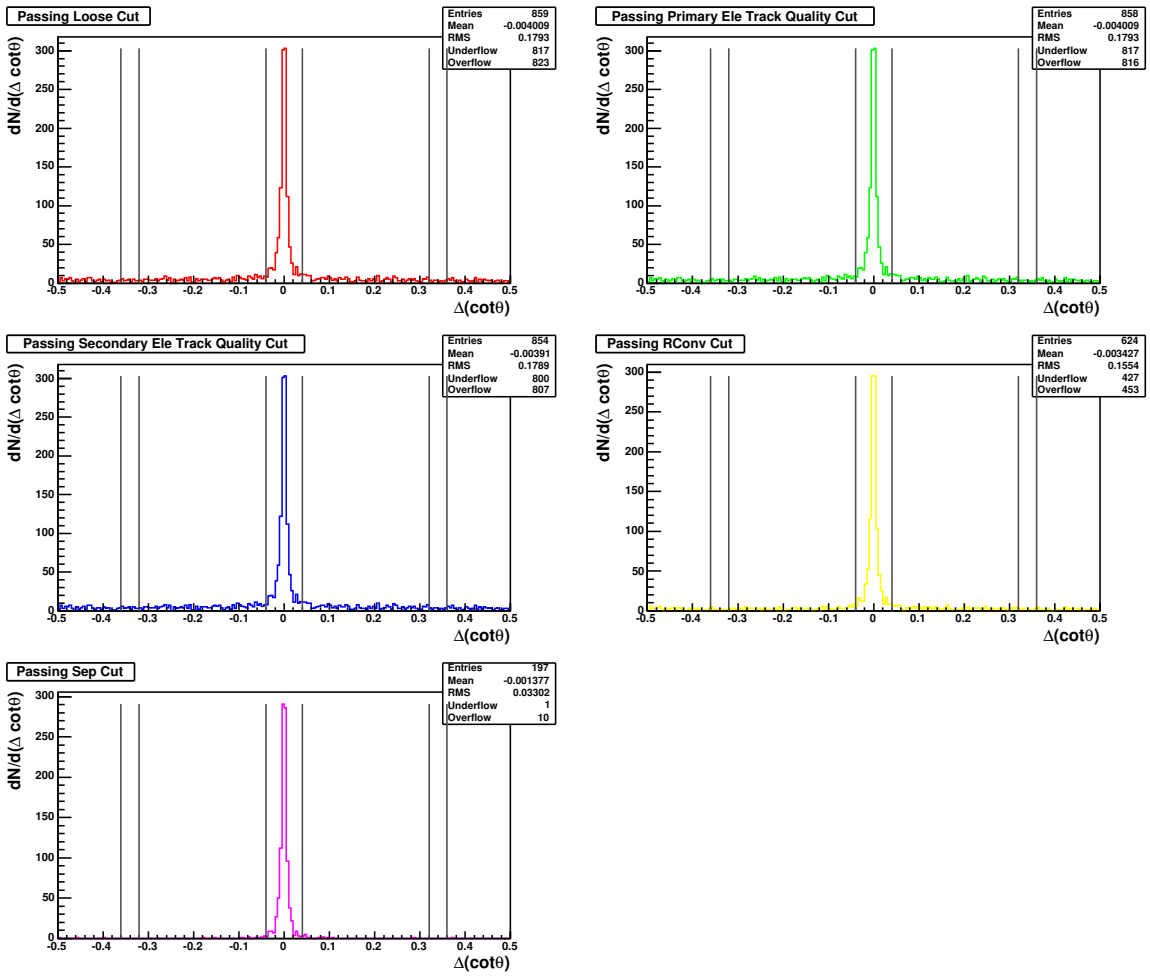


Figure 5.21: The $\Delta \cot\theta$ distribution for conversions in the $Z \to e+$trident data for each cut tested in Method A (see Table 5.3). Grey vertical lines show where the central and sideband windows are located. The central window is $-0.04 < \Delta \cot\theta < 0.04$; the left and right sideband windows are $-0.36 < \Delta \cot\theta < -0.32$ and $+0.32 < \Delta \cot\theta < +0.36$, respectively.

background-subtracted value is taken as the denominator of the efficiency calculation for additional cuts.

For each conversion candidate, the selection given in Table 5.3 is next applied and separate $\Delta \cot\theta_{23}$ distributions are constructed. These are the remaining subplots of Figures 5.21 and 5.22. The numerator of the efficiency calculation is the number of background-subtracted entries that pass each cut. The resulting efficiencies for both the data and MC prediction are provided in Table 5.4.

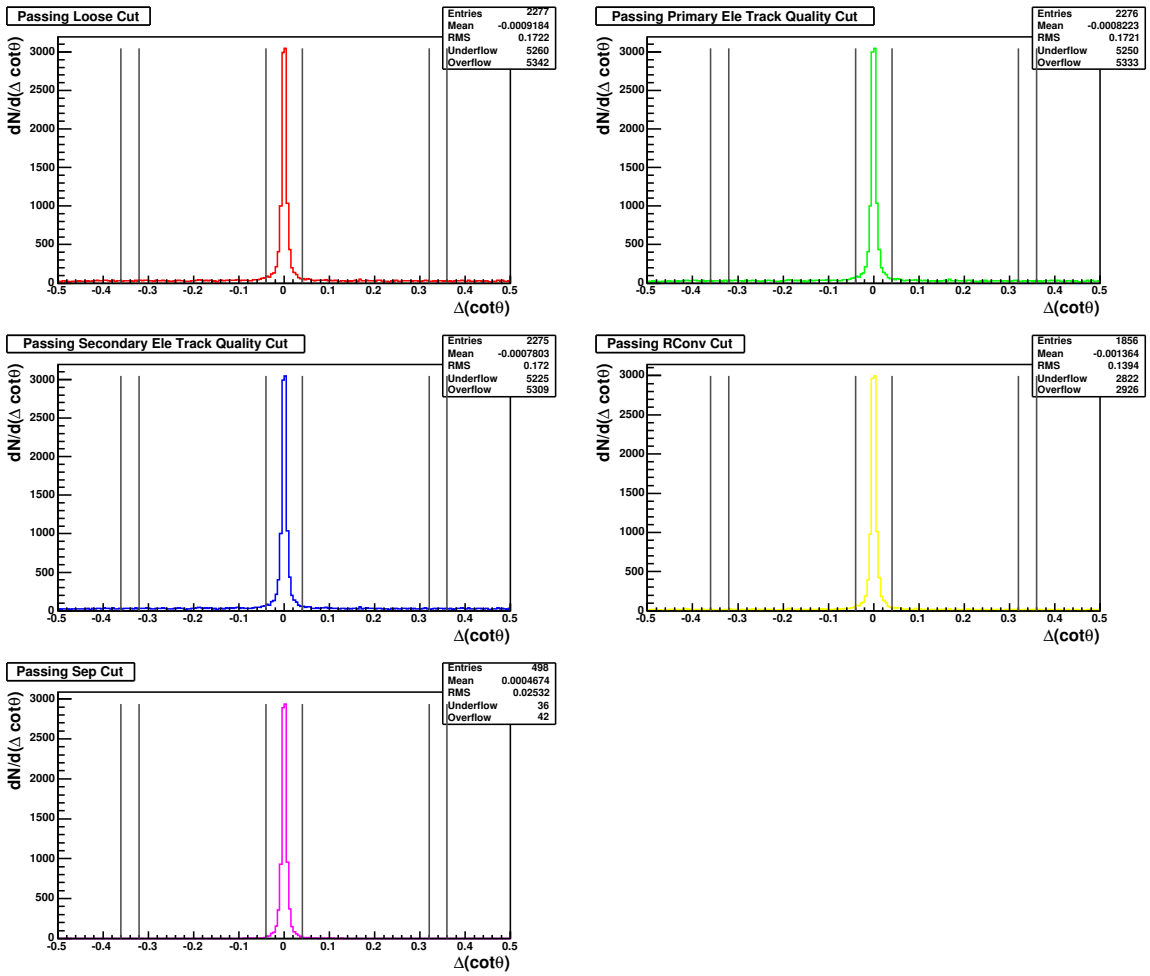


Figure 5.22: The $\Delta \cot\theta$ distribution for conversions in the $Z \to e +$ trident MC simulation for each cut tested in Method A (see Table 5.3). Grey vertical lines show where the central and sideband windows are located. The central window is $-0.04 < \Delta \cot\theta < 0.04$; the left and right sideband windows are $-0.36 < \Delta \cot\theta < -0.32$ and $+0.32 < \Delta \cot\theta < +0.36$, respectively.

The loose cuts we apply for this method select higher momentum conversion candidates. Considering this initial selection, the last column of Table 5.4 provides a suggested correction factor needed for the MC simulation for the tighter track-based cuts of Table 5.3; this correction factor is calculated as the efficiency measured from the data divided by the efficiency obtained from the MC prediction. We repeat the study in two ways: (a) using the sep distribution for background subtraction and applying instead a cut of 0.04 on $\Delta \cot\theta_{23}$ and (b) dividing the entries based on the

Table 5.3: The tight selection made for the Method A efficiency study.

Variable	Cut	Applied to
Track quality		e_2
Track quality		e_3
$\lvert\text{sep}_{23}\rvert$	< 0.2 cm	Conversion
R^{23}_{conv}	> 2.0 cm	Conversion

Table 5.4: Efficiency (Eff.) results obtained from the data and MC prediction for the cuts given in Table 5.3. These values were obtained using the $\Delta \cot\theta$ distributions of Figures 5.21 and 5.22 for background subtraction. The corresponding scale factors are also provided, calculated as the efficiency obtained from the data divided by the efficiency obtained from the MC simulation.

Cut	Data Eff. (%)	Z MC Eff. (%)	Scale Factor (%)
Loose	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
e_2 Track quality	100.0 ± 0.0	99.95 ± 0.02	100.1 ± 0.0
e_3 Track quality	99.91 ± 0.09	99.95 ± 0.02	99.96 ± 0.10
R^{23}_{conv}	95.24 ± 0.65	99.14 ± 0.09	96.07 ± 0.66
sep_{23}	90.21 ± 0.91	94.49 ± 0.23	95.46 ± 0.99

conversion candidate p_T. The resulting scale factors from these studies are provided in Table 5.5. These results suggest that the MC simulation better models the response to these cuts in the data when a conversion is of higher p_T.

5.4.3 Method B

In the previous section, we selected higher p_T conversion candidates and measured the efficiency of the track-based pair finding selection using either the sep or $\Delta \cot\theta$ distribution for background subtraction. The limitation of this method arises from the denominator for the efficiency calculation, which is based on a special set of conversion candidates. A second method is next described in which we measure the efficiency of the conversion selection for a wider sample of conversion candidates. In particular,

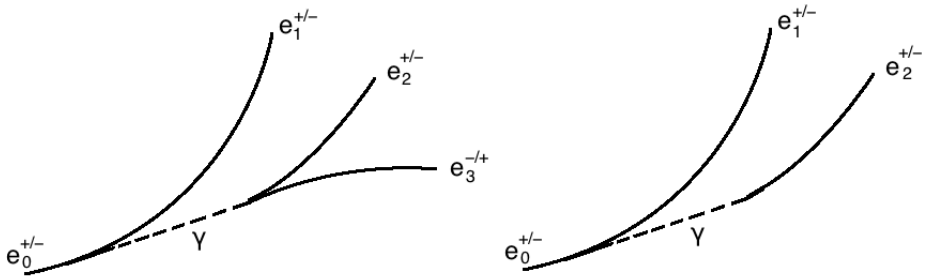
Table 5.5: After all cuts in Table 5.3 have been applied, the second column provides the results of Table 5.4 divided into conversion p_T ranges. The third column is similar to that of the second column, except (i) the sep distribution is used for background subtraction rather than $\Delta \cot\theta$ and (ii) the sep cut in Table 5.4 is replaced with a $|\Delta \cot\theta| < 0.04$ cut.

Conversion p_T	Scale Factor (%)	
Range (GeV/c)	Using $\Delta \cot\theta$	Using sep
10–12	92.4 ± 2.1	92.3 ± 1.9
12–15	94.9 ± 1.8	96.2 ± 1.6
15–200	96.4 ± 1.3	97.6 ± 1.3
10–200	95.5 ± 1.0	95.9 ± 0.9

we make no requirements on the third track in the loose selection; this removes the assumption of the third track in the denominator of the efficiency calculation. A limitation arises in this method from the inclusion of low p_T conversions not of interest to our study; however, the results are used to constrain the uncertainty for reconstructing conversions in the $H \to \gamma\gamma$ MC simulation.

Trident candidates in this study (Figure 5.23 (a)) are first selected from same-sign (SS) electrons that pass the selection given in Table 5.6. The aim of the loose selection in this table is to reconstruct the SS pair given in Figure 5.23 (b). Per event, the invariant mass distribution constructed from the SS pair and isolated electron is then filled. This is shown in the upper left subplot of Figure 5.24 for the data and Figure 5.25 for the MC simulation. Events that contain a true conversion from a trident are expected to populate the Z boson peak around 91 GeV/c^2. The number of events in this region is calculated from a Gaussian fit. The expected background is then subtracted from this value based on the number of events outside the mass peak. This background-subtracted value is taken as the denominator of the efficiency calculation.

After identifying events with a SS pair, the third, oppositely signed track of Figure 5.23 (a) is considered in order to complete the trident and to study the conversion ID efficiency. The selection that we will apply is given in Table 5.7 and is



(a) Trident (b) SS Electrons

Figure 5.23: (a) Illustration of a trident, which has two electron tracks of the same sign and a third track with opposite sign. (b) An illustration of the same-sign (SS) electron tracks that we reconstruct for Method B.

Table 5.6: The loose selection of Method B; the goal of these cuts is to select events for the denominator of the Method B efficiency calculation. The primary differences in these cuts relative to Table 5.2 are (a) there is no selection on a third track, so no cuts identifying a conversion photon are in the denominator, and (b) the first electron and second track are always of the same sign (SS). The first set of cuts are on the leading electron only (e_1). All second track candidates must be of the same sign as the leading electron, be a high quality track, and point to a fiducial CES cluster. A SS pair is formed from two same-sign tracks that are near one another (using sep and $\Delta \cot\theta$ cuts).

Variable	Cut	Applied to
$\lvert\eta\rvert$	< 1.1	
CES Fiducial		
Track Quality		e_1
Oppositely signed track to that of $e_{\rm iso}$	$q_{\rm iso} + q_1 = 0$	
Had/Em	$< 0.055 + 0.00045 \times E$	
Same signed track to that of e_1	$\lvert q_1 + q_2\rvert = 2$	
CES Fiducial		e_2
Track Quality		
p_{T2}	> 1.0 GeV/c	
$\lvert{\rm sep}_{12}\rvert$	< 0.2 cm	
$\lvert\Delta \cot\theta_{12}\rvert$	< 0.1	SS Pair
Number of SS candidates in event	$= 1$	

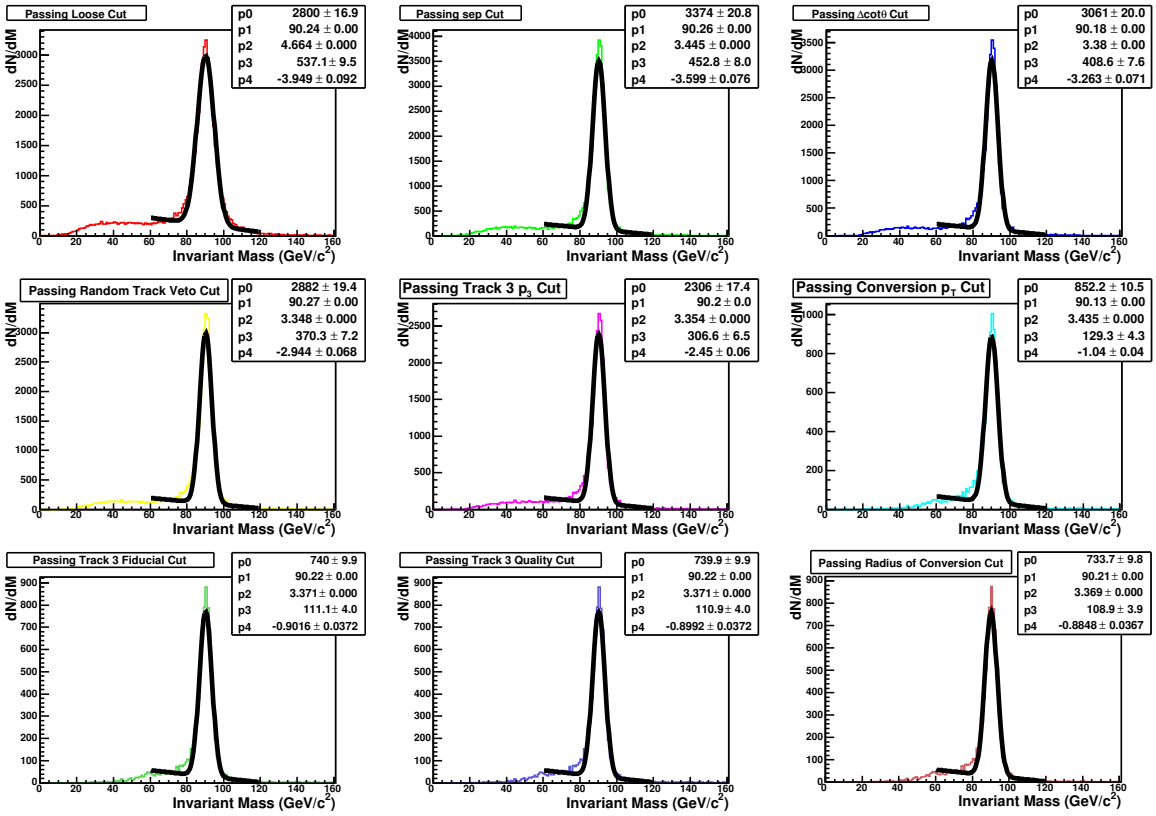


Figure 5.24: The invariant mass distribution of $Z \rightarrow e +$ trident events in the data. The upper left distribution includes only the loose cuts of Table 5.6, which are cuts applied to the same-sign pair of the trident. For the remaining distributions, consecutively tighter cuts on the conversion have been applied. These are the cuts given in Table 5.7. Each histogram is fit to the sum of a Gaussian function and a polynomial function. The number of signal events is obtained from the integral in the range $83.0 < m_{ee} < 99.0$ GeV/c^2, with the background subtracted from sideband regions $61.0 < m_{ee} < 69$ GeV/c^2 and $113.0 < m_{ee} < 121$ GeV/c^2.

based primarily on the track-based cuts of Section 5.2. This selection is applied to each oppositely signed track in the event, and the most likely third track candidate is taken to be the one that passes the highest selection, in the order given in this table. For each consecutive cut passed by this track, the invariant mass distribution of the isolated electron and trident is filled. The background-subtracted number of events that pass each cut is taken as the numerator of the efficiency calculation for this study.

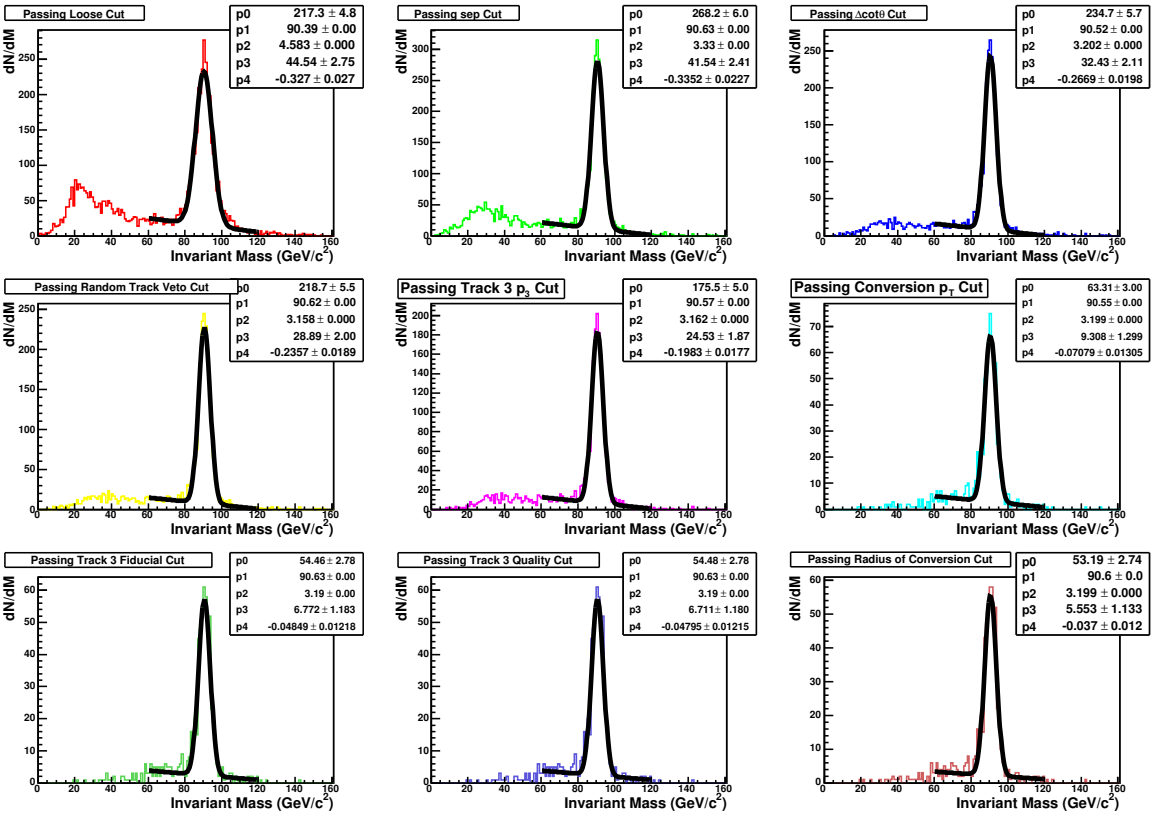


Figure 5.25: The invariant mass distribution of $Z \to e +$ trident events in the MC simulation. The upper left distribution includes only the loose cuts of Table 5.6, which are cuts applied to the same-sign pair of the trident. For the remaining distributions, consecutively tighter cuts on the conversion have been applied. These are the cuts given in Table 5.7. Each histogram is fit to the sum of a Gaussian function and a polynomial function. The number of signal events is obtained from the integral in the range $83.0 < m_{ee} < 99.0$ GeV/c^2, with the background subtracted from sideband regions $61.0 < m_{ee} < 69$ GeV/c^2 and $113.0 < m_{ee} < 121$ GeV/c^2.

Table 5.8 provides the results of the efficiency measurement, where it is found that up until the p_T cuts, the efficiency of the pair-finding selection is above 74–80%. Cuts on lower p_T conversions have the effect of reducing this efficiency significantly. Regardless, the results demonstrate that the scale factor after all applied cuts is above about 90%. For the higher p_T conversions from $H \to \gamma\gamma$, we expect the scale factor to be higher than this (based on the results of Table 5.5).

Table 5.7: All tighter cuts are made on either the conversion photon or on the third, oppositely signed track. Cuts are made in the order shown. In a true $H \to \gamma\gamma$ conversion event, the trident cut in Section 5.2 has the effect of removing random tracks near the conversion; for this study, we apply the same trident veto selection, but on a fourth track rather than a third track. For the efficiency study, this cut is labeled as the "random track veto."

Variable	Cut	Applied to
$\lvert\text{sep}_{23}\rvert$	< 0.2 cm	Conversion
$\lvert\Delta \cot\theta_{23}\rvert$	< 0.04	Conversion
Random Track Veto		Conversion
p_{T3}	> 1.0 GeV/c	e_3
p_T	> 10.0 GeV/c	Conversion
CES Fiducial		e_3
Track Quality		e_3
R_{conv}	> 2.0 cm	Conversion

Table 5.8: The conversion ID efficiency (Eff.) measured from the data and MC prediction using the Method B selection. In this method, no loose cuts are applied to the lower momentum track of the conversion photon. The corresponding scale factor for each cut applied is also provided, calculated from the ratio of the efficiency measured in the data to the efficiency obtained from the MC simulation.

Cut	Data Eff. (%)	Z MC Eff. (%)	Scale Factor (%)
Loose	100 ± 0	100 ± 0	100 ± 0
sep	96.2 ± 0.4	95.6 ± 0.1	100.7 ± 0.4
$\Delta \cot\theta$	81.2 ± 0.8	85.2 ± 0.2	95.3 ± 1.0
Random Track Veto	74.8 ± 0.9	79.6 ± 0.2	93.9 ± 1.1
Track 3 p_T	60.0 ± 1.0	63.8 ± 0.3	94.1 ± 1.7
Conversion p_T	21.9 ± 0.9	24.1 ± 0.2	91.0 ± 3.7
Track 3 Fiducial	18.8 ± 0.8	20.6 ± 0.2	91.4 ± 4.1
Track 3 Quality	18.8 ± 0.8	20.6 ± 0.2	91.4 ± 4.1
R_{conv}	18.4 ± 0.8	20.4 ± 0.2	90.3 ± 4.1

5.4.4 Conversion ID Uncertainty

The results of Methods A and B are now used to constrain and validate the final uncertainty derived for the $H \to \gamma\gamma$ conversion ID identification. For non-converting photons, the photon ID efficiency from the data was compared to the efficiency from the MC simulation in order to apply a correction (scale factor) to the number of $H \to \gamma\gamma$ signal events predicted to pass the photon selection. The efficiency study for conversions, however, is based on conversions with much lower p_T than those expected from $H \to \gamma\gamma$ decays. Therefore, we apply an uncertainty rather than a scale factor on the number of predicted $H \to \gamma\gamma$ events in the data.

We constrain this uncertainty based on the results of the two methods just described, where the uncertainty is taken from the difference in the calculated scale factor from 100%. The first study of Section 5.4.2 suggests a scale factor of at worst 94% for the cuts studied, giving a corresponding uncertainty of 6%. For the second study, however, the results of Table 5.8 constrain the uncertainty to at most 10%.

The final uncertainty is obtained using both the loose and tight trident selection described in Section 5.4.3, with the addition of a trident E/p and calorimeter isolation requirement. The goal is to compare the number of observed tridents that pass this full selection in the data to the number predicted to be in this data sample using the MC simulation. However, in order to remove any dependence on the uncertainty in the trigger efficiency, luminosity measurement, or Z boson cross section, we instead form a ratio from the number of reconstructed $Z \to e^+e^-$ events to the number of $Z \to e +$ trident events. This ratio is calculated from both the data and the MC simulation and the difference between these two values is found to be 7.0%. This value is consistent with the results of Methods A and B, and we apply this uncertainty to the conversion ID selection.

5.5 Summary

In this chapter, the method for identifying central conversion photons was discussed, along with the efficiency measurement of this identification. The previous chapter

discussed the reconstruction of central and plug photons and their corresponding ID efficiency. We are now prepared to describe how central, plug, and conversion photons are used in the diphoton analysis.

Chapter 6

Data Sample and Event Selection

6.1 Introduction

In this chapter, we begin the description of our search for $H \to \gamma\gamma$ events in the CDF data. The approach for this analysis is to select events with two photon candidates, construct the diphoton mass distribution for these events, and search for signs of a resonance (peak) in the data over the predicted background. The background is composed of SM diphoton candidates obtained from processes other than $H \to \gamma\gamma$ decays. Since the shape of the $m_{\gamma\gamma}$ distribution for $H \to \gamma\gamma$ decays is narrow, a resonance above this background would be evidence of the presence of the Higgs boson (or some other unknown particle).

The data sample used for the diphoton analysis is described in this chapter, along with how events in this data sample are selected. Chapter 7 describes the background events and the prediction of their number and shape in the data. Chapters 8 and 9 describe how the SM $H \to \gamma\gamma$ events (the signal) are modeled using simulation to predict the number and shape of diphoton events in the data. Chapters 10–12 provide results of the search and conclusions.

6.2 Diphoton Category Definitions

In the past two chapters, three types of photons have been described: central ($|\eta| < 1.1$), plug ($1.1 < |\eta| < 3.6$), and central conversion photons. These photon definitions provide four independent diphoton subsamples, based on their position in the detector (central or plug) and type (conversion or non-conversion):

- CC category: events with two central photons
- CP category: events with a central and plug photon
- C′C category: events with a central conversion and a central photon
- C′P category: events with a central conversion and a plug photon

The remainder of this chapter describes the selection of these diphoton events in the data.

6.3 Isolated Photon Trigger and Data Sample

As discussed in Chapter 3, the selection of collision data to be stored on disk is based on a three-level trigger system. We use data that has been collected from an isolated photon trigger, `PHOTON_25_ISO`, which has a minimum E_T threshold of 25 GeV on a single EM shower. Each level of this trigger is next described in more detail.

- Level 1 (L1): As discussed in Sections 3.2 and 3.4.3, the trigger tower E_T is obtained assuming $z_{\rm vtx} = 0$. At L1, $p\bar{p}$ collision events are selected that have a single EM trigger tower with $E_T > 8$ GeV.
- Level 2 (L2): At this level, the E_T threshold is raised from 8 to 21 GeV where the L2 E_T is determined from a trigger cluster rather than a trigger tower. The majority of the shower energy is also required to be in the EM calorimeter, obtained by selecting events with Had/Em < 0.125. The EM cluster must furthermore be isolated in the calorimeter with calorimeter isolation < 3 GeV

or $< 0.15E_T$. See Section 4.2.1 for a description of the Had/Em and calorimeter isolation variables.

- Level 3 (L3): The requirements at this level are similar to those at L2, but with more stringent thresholds on E_T, Had/Em, and calorimeter isolation. If the EM shower is in the central region of the detector, this level furthermore requires that the lateral shower profile in the CES be consistent with that of a single direct photon from the primary collision. For this purpose, the CES χ^2 variable is constructed, which is the average of the CES strip and wire χ^2 values. At L3, the CES χ^2 must be less than 20 for central EM showers. See Section 4.2.1 for a description of the CES strip and wire χ^2 definitions.

At L3, online data are divided into data streams based on similar triggers. We use CDF's c-stream and in particular the following data samples: `cph1ad`, `cph1ah`, `cph1ai`, `cph1aj`, `cph1ak`, `cph1am`, and `cph1ap`. These samples are structured so that they can be processed using CDF's Stntuple software. They contain a trigger bit variable that is set to true for each event that was collected by the `PHOTON_25_ISO` trigger. For this analysis, all events selected from these data samples are required to have this trigger bit set to true. Table 6.1 summarizes the `PHOTON_25_ISO` trigger selection.

6.4 Quality Vertex Requirement

Section 3.4.2 provided a description of vertex reconstruction, classification of vertices, and vertex p_T. For events selected by the `PHOTON_25_ISO` trigger, we choose the primary vertex to be the Class 12 vertex with the highest vertex p_T. This vertex is required to have $|z_{\rm vtx}| < 60$ cm, which retains about 97% of all events and ensures that tracks coming from the vertex are in a well-instrumented region of the detector.

Table 6.1: The PHOTON_25_ISO trigger selection that is required for at least one EM shower.

Trigger Level	Variable	Cut
L1	E_T ($z = 0$)	> 8.0 GeV
L2	E_T ($z = 0$)	> 21.0 GeV
	Had/Em	< 0.125
	Calorimeter Isolation	< 3.0 GeV or $< 0.15E_T$ GeV
L3	E_T ($z = 0$)	> 25.0 GeV
	CES χ^2 (if CEM)	< 20
	Had/Em	$< 0.055 + 0.00045 \times E$ for $E < 200$ GeV
		$< 0.2 + 0.001 \times E$ for $E > 200$ GeV
	Calorimeter Isolation	$< 0.10E_T$ GeV

6.5 Good Run Requirement

Not all data stored in the cph1 samples were actually taken during good detector conditions. Of the data stored, the Photon Group at CDF determines a set of runs that are good for photon physics analyses based on good run bits (see Section 3.3). These runs are organized into a text file, which is read in and stored in an array. The array is referenced on an event-by-event basis to determine if a particular event should be included in the analysis. Which runs are included determines the integrated luminosity of the data sample.

The specific detector components needed for an event in this analysis are dependent on what types of photons are included (central, plug, or conversion). We therefore use four different lists for the four different diphoton categories. Each list provides all of the available runs for that category and the corresponding integrated luminosity.

For CC and CP diphoton events, we start with the CDF Photon Group's goodrun_v44_pho_00.txt and goodrun_v44_phx_pho_00.txt lists, respectively. Each of these good run lists (GRLs) requires the following detector components to have

been functioning properly: trigger, CLC, central and plug calorimeter, and central and plug SMX. The COT is also required to have been marked as good, though COT compromised runs are allowed as long as the silicon was functioning properly; this ensures good vertex reconstruction efficiency. The list used for the CP category requires that the silicon always be marked as good to ensure quality track reconstruction in the plug region. For the the CC and CP categories, respectively, these version 44 GRLs initially provide data corresponding to an integrated luminosity of 9.48 fb^{-1} and 9.09 fb^{-1}. For diphoton events that have a central conversion, these files are modified to exclude the COT compromised runs; this ensures good track reconstruction efficiency for the e^+e^- pair.

We next expand the standard GRLs by studying more carefully runs not included in these lists [65]. A negligible loss in photon ID efficiency is observed for runs where a single central shower max card was not functioning during run time (studied with $Z \to e^+e^-$ events). These CES runs are added to the GRL for all diphoton categories, providing an additional ∼240 pb^{-1} of data. Another ∼289 pb^{-1} of good data are obtained for the CC category by including runs marked bad for detector components such as plug calorimeter, XFT, and a portion of a COT superlayer. Some of these miscellaneous runs are also added to the central conversion list.

Table 6.2 summarizes the good runs included for each diphoton category, and also shows the corresponding integrated luminosity.

Table 6.2: Integrated luminosity ($\mathcal{L}$) used for each diphoton category. The text provides an explanation of the data that is added or subtracted from the version 44 photon GRL.

$\mathcal{L}$ (fb^{-1})	CC	CP	C′C	C′P
v44 photon GRL	9.48	9.09	9.48	9.09
CES Runs	+0.24	+0.24	+0.24	+0.24
Misc. Runs	+0.29	+0	+0.20	+0
COT Compromised	–0	–0	–0.06	–0.06
Total	10.0	9.34	9.87	9.28

6.6 Diphoton Selection

Chapters 4 and 5 described in detail how each type of photon is reconstructed from detector information. Photon candidates that pass the full photon selection for one of these types are considered tight photons since they pass strict photon identification requirements.

Of the events that pass the vertex requirements of Section 6.4, we begin the diphoton selection with those that are included in the CC photon GRL. For each event, we then search for the two highest E_T photon candidates that pass either the tight central or tight plug photon ID requirements. This selection provides either a CC or CP event. If a CP event is found, it is required to be included in the CP GRL. If neither a CC or CP photon pair is found in the event, then we search for the highest E_T tight central or plug photon to be paired with the highest E_T tight central conversion photon. If a C′C (C′P) event is found in this way, it is required to be included in the C′C (C′P) GRL. This selection provides the four independent diphoton categories. The selection is also inclusive, which means we search for two photons but other objects (including more photons) may also be in the event.

6.7 Diphoton Mass

The kinematic variables for single photons were provided in Chapters 4 and 5. Once two photons have been reconstructed, the diphoton mass is calculated as

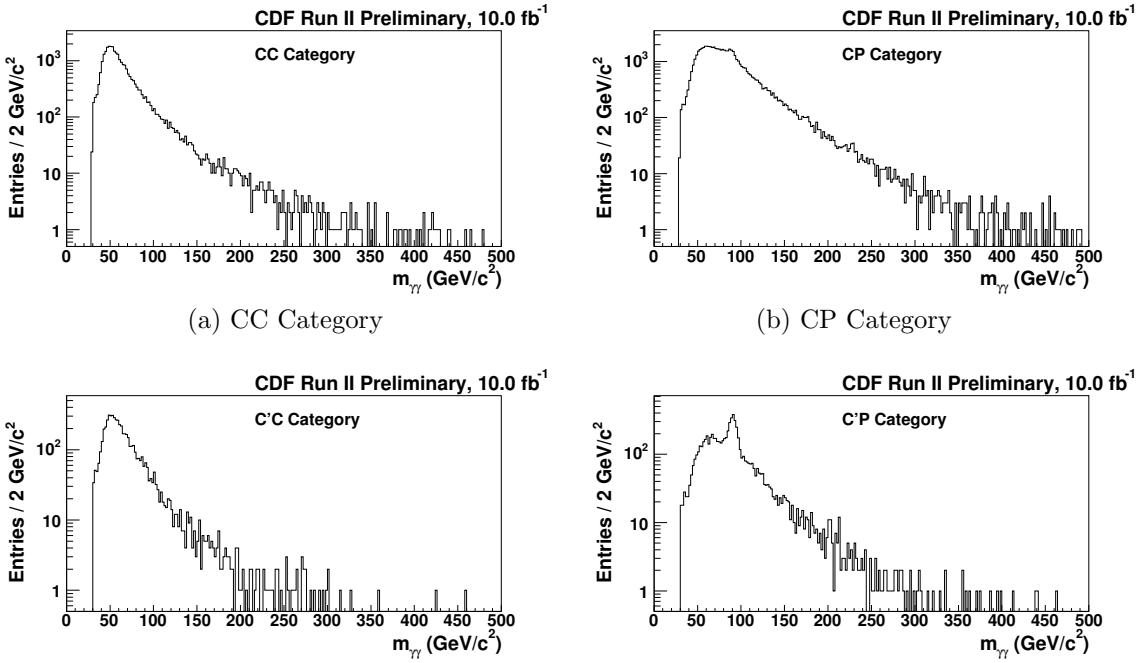
$$m_{\gamma\gamma}^2 = 4E_1E_2\sin^2(\alpha/2) \tag{6.1}$$

where E_i is the energy of a single photon with $i = 1, 2$ and α is the angle between the two photons. This angle is obtained from

$$\cos\alpha = \frac{\vec{p_1}\cdot\vec{p_2}}{p_1p_2}, \tag{6.2}$$

where $\vec{p_i}$ is the vector momentum of a single photon and p_i is the magnitude of the momentum.

The diphoton mass shapes in the data for each category are provided in Figure 6.1. Many SM processes other than potential $H \to \gamma\gamma$ events pass the diphoton selection described in this chapter. These backgrounds are next discussed.



(a) CC Category

(b) CP Category

(c) C′C Category

(d) C′P Category

Figure 6.1: The $m_{\gamma\gamma}$ distribution in the data for each diphoton category.

Chapter 7

Background Model

In order to search for a $H \to \gamma\gamma$ signal in the data, an understanding of known SM processes that pass the diphoton selection is necessary. In particular, we obtain a prediction of the $m_{\gamma\gamma}$ shape of these background processes and the number of such events in the region where we search for $H \to \gamma\gamma$ signal. For this analysis, we do not model each background component separately, but instead take advantage of the $m_{\gamma\gamma}$ resolution of the Higgs boson decaying to photons. We use a data-driven background model where the $m_{\gamma\gamma}$ shape and normalization are obtained from a fit made to the region of the data where we do not search for signal, called the sideband region of the data.

For the rest of this chapter, a discussion of the background composition is provided first, followed by a description of the fitting method and corresponding systematic uncertainties.

7.1 Background Composition

7.1.1 Prompt Photon Pairs

At a rate several orders of magnitude higher than the SM $H \to \gamma\gamma$ prediction, pairs of photons at the Tevatron are produced directly from QCD interactions in the hard

scattering process. This background to the $H \to \gamma\gamma$ analysis is called the prompt (or direct) diphoton background, and since the photon identification from Chapters 4 and 5 cannot distinguish photons originating from QCD interactions and those originating from $H \to \gamma\gamma$ decays, this is considered an irreducible background.[1]

The leading-order (LO) production of prompt diphotons at the Tevatron is from quark-antiquark scattering, as shown in Figure 7.1 (a). Higher-order contributions [66]

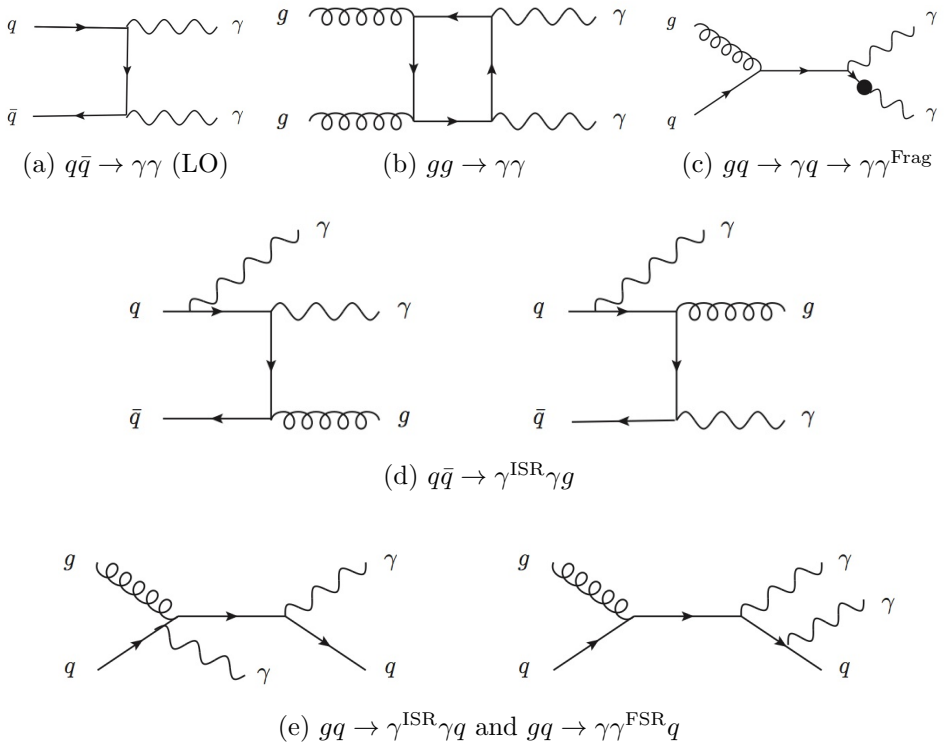


Figure 7.1: Diagrams [28, 66] of prompt diphoton production at hadron colliders. The LO process at the Tevatron is quark-antiquark annihilation in (a). The black circle shown in (c) represents a fragmentation process of a final-state quark radiating a photon, which takes most of its energy and which radiates at very small angle. The diagrams for (c) and (e) are time-like processes, however there are also the equivalent space-like processes.

[1] Though we have not done so for this analysis, it is possible to take advantage of some kinematic differences to better distinguish $H \to \gamma\gamma$ events from the prompt diphoton background. For example, our preliminary studies show that a gain in sensitivity of about 5% or more may be achieved by dividing events into bins of $p_T^{\gamma\gamma}$. Further gain would be expected by the use of a multivariate discriminant based on kinematic differences in several variables.

come from gluon fusion, as shown in Figure 7.1 (b), fragmentation of a photon off of a final-state quark, as shown in Figure 7.1 (c), and either quark-antiquark or quark-gluon scattering with a photon radiated from an initial- or final-state quark, as shown in Figures 7.1 (d) and (e).

The prompt diphoton background has a smoothly falling $m_{\gamma\gamma}$ distribution for each diphoton category, as shown in Figure 7.2.

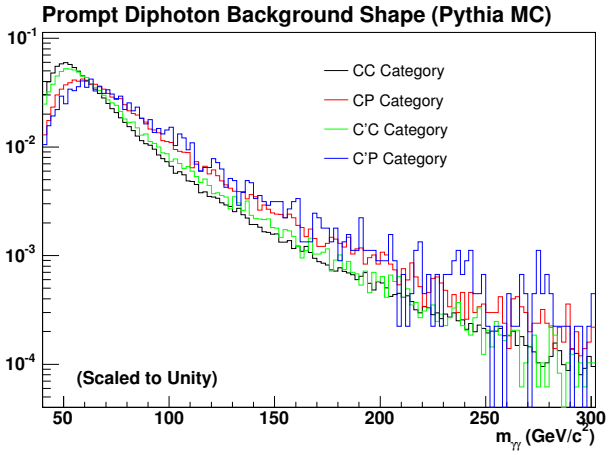


Figure 7.2: The prompt diphoton background shape for each diphoton category is shown, obtained from a $\gamma\gamma$ PYTHIA sample including the LO diagrams (a) and (b) of Figure 7.1, but also modified to include diagrams (d) and (e) [66]. Each shape is normalized to unit area.

7.1.2 $\gamma +$ Jet and Dijet Events

Another significant background occurs from events in which one or two jets fake the signature of a photon in the detector, passing the photon ID selection. As discussed in previous chapters, this jet background is most often due to a neutral meson decaying into a pair of collinear photons. The main processes that contribute to this background [67, 68] are shown in Figure 7.3. The first row of figures shows the production of a photon along with either a quark or gluon; these processes contribute to the $\gamma +$ jet background. The remaining diagrams show different ways in which two jets are produced, which contribute to the dijet background.

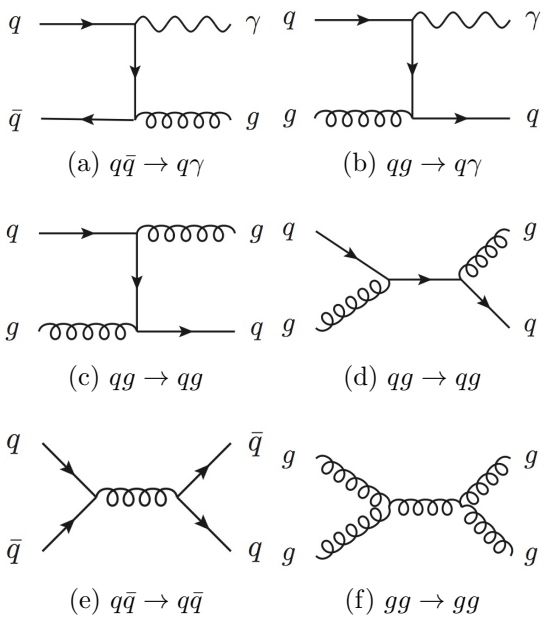


Figure 7.3: The main diagrams for γ+jet production (first row) and dijet production (second and third rows) at the Tevatron, which can be misreconstructed as diphoton events in the data [28].

The γ + jet and dijet fake backgrounds are reducible by optimizing the photon ID selection to better identify true photons and reject jet backgrounds, as discussed in Chapters 4 and 5. Like the prompt diphoton background, this fake background has a smoothly falling $m_{\gamma\gamma}$ distribution for each diphoton category. Figure 7.4 shows this shape for the CC category.

7.1.3 Drell-Yan

A smaller fake background comes from the Drell-Yan process (Figure 7.5), where a virtual photon or Z boson is produced from $q\bar{q}$ annihilation. The photon or Z boson then decays to an e^+e^- pair, which is reconstructed as a pair of photons.

In the central region, the probability of a single electron passing the standard central photon ID selection at CDF has been measured to be about 1% [69]. Although we use a NN ID for this analysis, the same electron track removal selection is applied, and therefore, the electron fake rate remains on the order of 1%. This probability

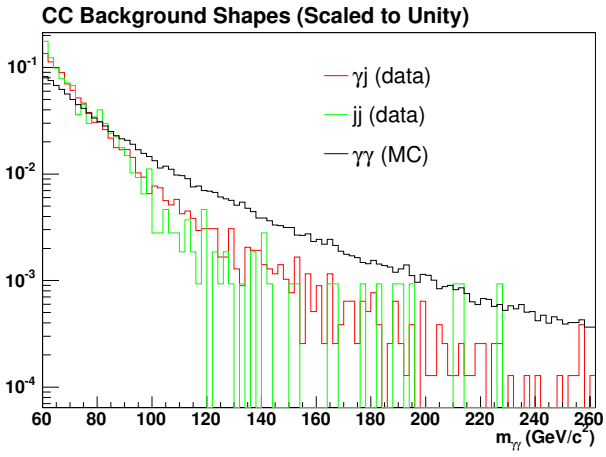
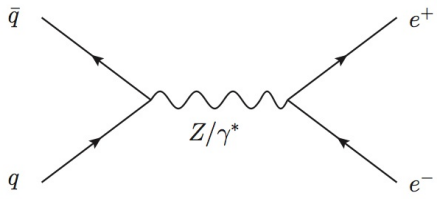


Figure 7.4: As an example, the shape of the $\gamma +$ jet and dijet backgrounds in the CC data are shown, normalized to unit area. The enriched fake data samples are obtained by selecting diphoton events in the same way as that for the CC diphoton data, except one or both photons have a NN value < 0.3 rather than both passing the standard > 0.74 requirement. The prompt CC diphoton background shape from Figure 7.2 is also shown for comparison.



(a) $q\bar{q} \to Z/\gamma^* \to e^+e^-$

Figure 7.5: Diagram of the Drell-Yan process.

corresponds to the probability that an electron will radiate an energetic photon via bremsstrahlung such that the electron track p_T is less than $\sim$2 GeV/c, a p_T small enough to pass the track isolation requirement. In this case, the radiated photon is reconstructed as though it were a prompt photon. The probability for this to occur twice, for two central photons, is small and is on the order of 0.01%. The tracking reconstruction efficiency in the central region is very good at CDF and the fake rate from failing to reconstruct an electron's track is negligible.

In the plug region, there is more material for an electron to travel through and the probability for an electron to produce a photon via bremsstrahlung is therefore

higher. Moreover, there is less coverage for tracking and an additional contribution to the electron background arises from the failure to reconstruct the electron's track. For the CP and C′P categories, the larger fake rate from the plug leg leads to a bump near 91 GeV/c^2 in the data from $Z \to e^+e^-$ decays.

The fake rate for central conversions is much smaller than that for plug photons, which can be observed from the lack of a bump at the Z pole for C′C events. The good track reconstruction efficiency in the central region contributes to this, along with the trident veto cut applied in the conversion ID selection discussed in Chapter 5.

Events from fake electrons have a small contribution to the background in the signal regions relative to the prompt diphoton, $\gamma +$ jet, and dijet backgrounds. They do, however, contribute significantly to the lower mass sideband region of the data in the CP and C′P categories (below about 100 GeV/c^2), and good modeling of this sideband is necessary when making fits to the data to predict the background yield in the signal regions.

7.2 Fitting Method

The background prediction for each Higgs boson mass hypothesis is obtained from a fit made to the data, where we exclude from the fit a signal region centered on each mass point. The size of the excluded region is based on the width of the reconstructed $H \to \gamma\gamma$ diphoton mass shape, which is about 3 GeV/c^2 as shown later in Figures 8.5–8.7. A 12 GeV/c^2 window is, therefore, selected in order to retain ∼95% of the signal.

7.2.1 Prompt Diphoton, $\gamma +$ Jet, and Dijet Backgrounds

The smooth portion of the diphoton data in each category is fit to a polynomial multiplied by the sum of two exponentials, where the degree of one polynomial is a parameter of the fit. The parametrized function using six parameters p_0–p_5 is

$$f_{\rm s}(m) = (p_0 x^{p_1} + p_2 x^{0.2})(p_3 e^{-p_3 x} + p_4 p_5 e^{-p_5 x}) \tag{7.1}$$

where $x = m_{\gamma\gamma} - 30$ GeV/c^2 (which is due to a threshold produced by removing events below this mass). This function is found to model well the smooth portion of the data distribution for each diphoton category, which is mostly composed of the prompt diphoton background and the fake $\gamma +$ jet and dijet backgrounds, with only a small contribution from electrons faking a photon.

7.2.2 Z Background in the CP and C′P Categories

We first model the $Z \to e^+e^-$ contribution using a PYTHIA MC sample in order to predict the $m_{\gamma\gamma}$ shape with its mean and resolution. Generated events from this sample are required to pass the same diphoton selection as that applied in the data. Figure 7.6 (Figure 7.7) shows the resulting CP (C′P) $m_{\gamma\gamma}$ distributions fit to both a Breit-Wigner and Gaussian function. The Breit-Wigner function provides a better fit and a more reasonable χ^2. It is described by the following function:

$$f_{\mathrm{Z}}(m) = \frac{p_0 p_1}{2m^2\left[(m-p_2)^2 + \frac{p_1^2}{4}\right]}, \tag{7.2}$$

where m is the diphoton mass, p_0 is a parameter used for normalization, and p_1 and p_2 are the width and mean of the Z pole, respectively.

We use this same function when fitting to the Z contribution in the data, where the mean and width parameters are initially set to the values predicted from the fits using the Breit-Wigner function in Figures 7.6 and 7.7. The parameters are a part of the fit to the data; however, we apply bounds to constrain the fit to the Z pole. The mean parameter[2] is initially set to 88.6 GeV/c^2 for the CP category and 89.9 GeV/c^2 for the C′P category, and it is allowed to fluctuate within a boundary of ± 2 GeV/c^2. The width parameter is set to 6.9 GeV/c^2 for the CP category and 6.7 GeV/c^2 for the

[2] There is an observed shift in the Z mean for both categories from the accepted mass of 91.2 GeV/c^2 [3]. This feature is suspected to be from electrons faking photons in Z events and would not be seen in a $H \to \gamma\gamma$ peak. The reason is that a large contribution of the Z background is from events in which one of the electrons radiates a hard photon through bremsstrahlung and it is the photon that is reconstructed rather than the electron. The post-bremsstrahlung electron track still carries a small bit of energy, however, so the photon that is reconstructed has less energy than the original electron. This results in a shift in the reconstructed Z mean from the expected value.

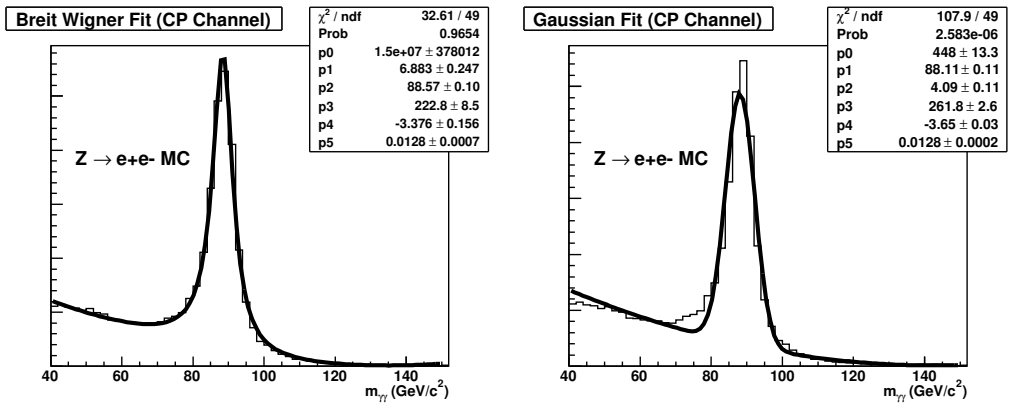


Figure 7.6: MC simulation of $Z \to e^+e^-$ events that pass the diphoton selection for the CP category. The left and right plots show the $m_{\gamma\gamma}$ distribution fit to a Breit-Wigner and Gaussian function, respectively. The former better describes the shape and is used when fitting to the Z contribution in the data. In the left plot, the parameters p_0–p_2 correspond to those described in Equation (7.2).

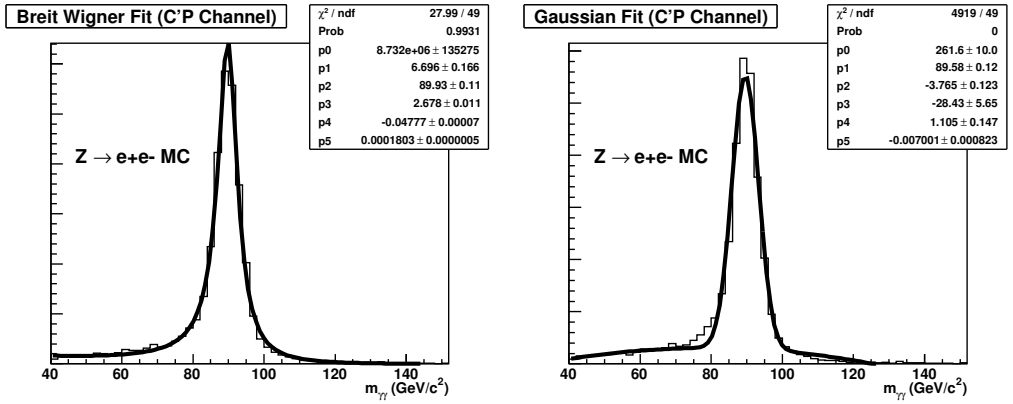


Figure 7.7: MC simulation of $Z \to e^+e^-$ events that pass the diphoton selection for the C′P category. The left and right plots show the $m_{\gamma\gamma}$ distribution fit to a Breit-Wigner and Gaussian function, respectively. The former better describes the shape and is used when fitting to the Z contribution in the data. In the left plot, the parameters p_0–p_2 correspond to those described in Equation (7.2).

C′P category, and it is allowed to fluctuate between zero and twice its value. Putting bounds on the parameters confines the fit to a small region around the Z boson, but allows some freedom for fluctuations in the data when reconstructing the particle's mean and width.

7.2.3 Fitting Minimization

The fit to the data is performed using the `TH1F::Fit` routine from the `ROOT` software, which uses the function minimization methods in the `MINUIT` package [70]. Given a histogram and a fit function such as $f_{\rm s}(m)$ and/or $f_{\rm Z}(m)$, `MINUIT` constructs a binned likelihood function (which we request) assuming the data bin counts are described by a Poisson probability distribution. Then for j parameters used for the fit function, the binned likelihood (L) is given by the product of likelihoods for each bin,

$$L(p_1, p_2, p_3, ..., p_j) = \prod_{i=1}^{N} \frac{f_b^{n_b} e^{-f_b}}{n_b!}, \tag{7.3}$$

where N is the number of bins considered, n_b is the data yield in the i$^{\rm th}$ bin, and f_b is the value of the fit function at that bin. For a particular set of parameters, `MINUIT` calculates the value of the negative log likelihood. The package finds the set of parameters for which the negative log likelihood is minimized, and the result is the best fit to the measured data.[3]

7.2.4 Resulting Fits for Each Category

For fitting to the $m_{\gamma\gamma}$ data, we use $f_{\rm s}(m)$ for the CC and C$'$C categories and the sum of $f_{\rm Z}(m)$ and $f_{\rm s}(m)$ for the CP and C$'$P categories, with an initial set of parameters provided for the minimization described in Section 7.2.3. The corresponding fits for a Higgs boson mass hypothesis of 115 GeV/c^2 are provided in Figure 7.8, where we fit to the data using the range $m_{\gamma\gamma}$ = 60–240 GeV/c^2, excluding the signal region around 115 GeV/c^2. The fit is interpolated into the signal region in order to predict the background shape and expected number of events where we search for the Higgs boson signal. The interpolated fits for $m_H = 115$ GeV/c^2 are shown in the upper plots in each subfigure of Figure 7.9. A background histogram with the same binning as the data is obtained from these fits and used to compare the background yield to that

[3] Fitting discussed in this paper is generally done using the same methods (though a χ^2 fit may be used instead). It is only described in more detail here so that the background estimation method is better understood, along with systematic uncertainties discussed later.

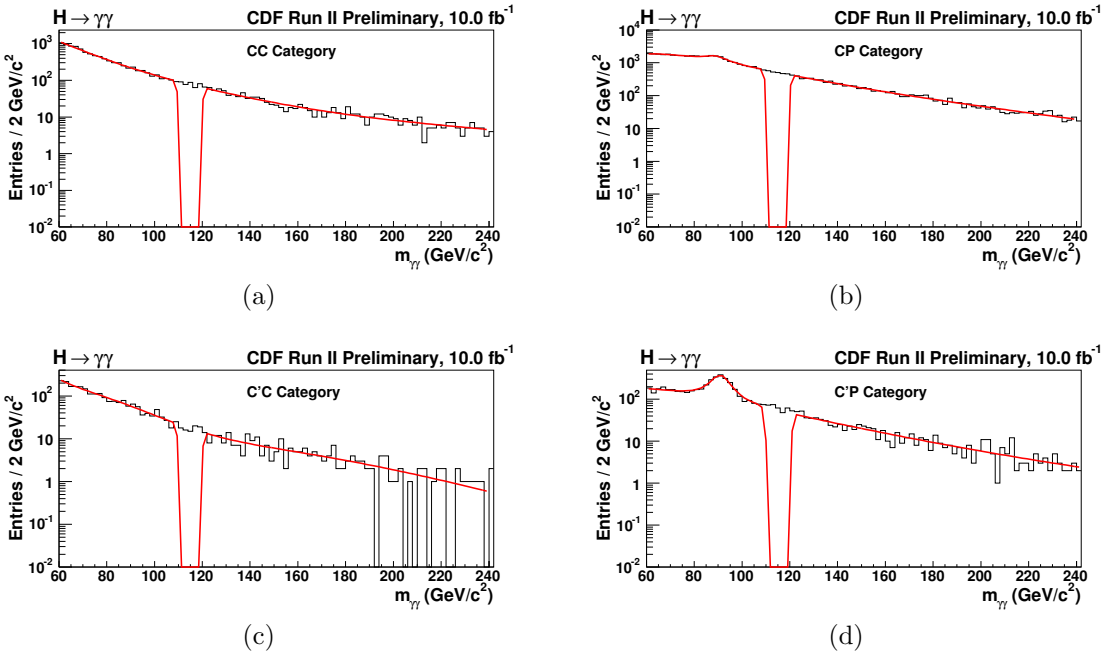


Figure 7.8: Each figure shows a fit (red line) made to the sideband region of the data for each diphoton category using $f_{\rm s}(m)$ for the CC and C′C categories and the sum of $f_{\rm Z}(m)$ and $f_{\rm s}(m)$ for the CP and C′P categories. The two vertical lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 115 GeV/c^2.

of the data. This is shown in the lower residual plots in each subfigure of Figure 7.9, obtained from the difference in the data and background expectations, divided by the Poisson statistical error from the background expectation. The y axis of the residual plot, therefore, provides an indication of how many standard deviations a data bin differs from the background prediction. In the signal region centered at 115 GeV/c^2, there is no significant excess of data events relative to the background.

We repeat this process for each Higgs boson mass hypothesis and there is, therefore, a separate fit made for each mass tested. These fits for each diphoton category are provided in Appendix D, along with the corresponding residual plots. As with the mass of 115 GeV/c^2, we find no evidence of a resonance in the corresponding signal regions of the data relative to the background prediction. The resulting background yields in the 12 GeV/c^2 signal region for each mass and diphoton category are provided in Chapter 10.

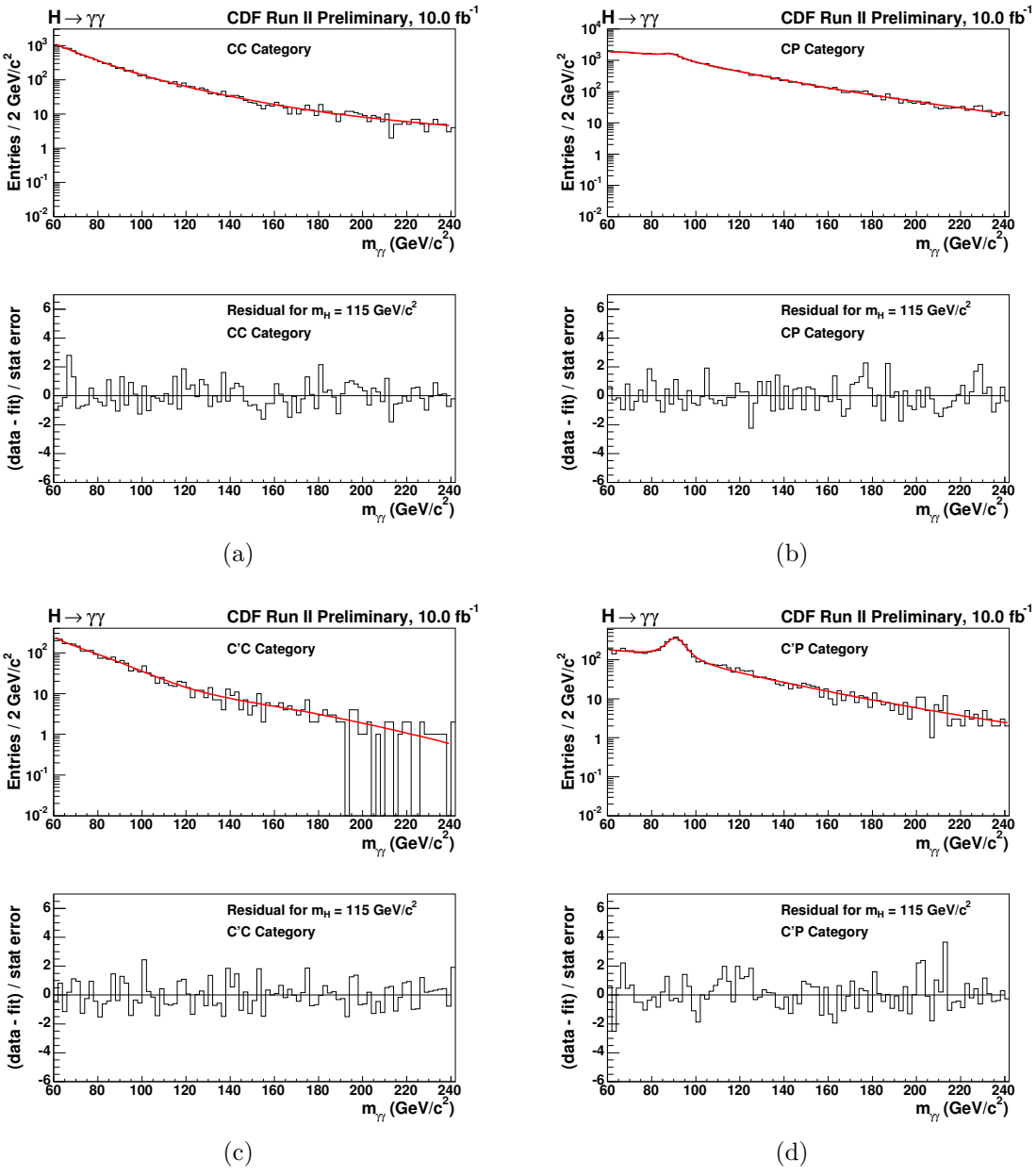


Figure 7.9: The upper plot in each subfigure shows the sideband fit from Figure 7.8 interpolated into the 115 GeV/c^2 signal region, for each diphoton category. The residual in the lower plot of each subfigure shows the data yield minus the background yield, divided by the statistical error from the background yield.

7.3 Systematic Uncertainty

7.3.1 Parameter Variation

Systematic uncertainties on the background yields are obtained by randomly varying the parameter values from the best fit based on their errors, uniformly within ± 1 standard deviation. To ensure that a smeared set of parameters provides a reasonable fit, the negative log likelihood value (Equation (7.3)) is computed with the new set of varied parameters and compared to the value obtained from the best fit parameters. If the difference between these values is greater than one, then we reject the trial.[4]

For all other cases, we obtain a new prediction of the background expectation and compare it to the background yield from the best fit. Of all of these trials, the highest and lowest yields relative to the nominal background prediction are used to determine an uncertainty:

$$\Delta b = \frac{b_{\text{highest}} - b_{\text{lowest}}}{2b_0} \tag{7.4}$$

where each yield is obtained from the 12 GeV/c^2 signal region centered on the Higgs boson mass hypothesis. These are applied as fractional uncertainties on the $m_{\gamma\gamma}$ background prediction. The uncertainties are shown in Table 7.1 for each diphoton category and mass signal region. Categories or regions of the data with more events constrain the fit better and, therefore, have a smaller uncertainty.

7.3.2 Model Dependence

It is possible that there is a model dependence in the predicted background yields due to the choice of a fit function. We address this by studying the effect on the background yield when a different fit function is used. For this purpose, we replace

[4] We ignore these trials because we are interested in studying cases that have a negative log likelihood value near the minimum. For minimizing with likelihood functions, trials could better be rejected if the difference in the negative log likelihood value for a trial relative to the minimum was a half unit rather than one. (See the statistics section of Reference [3].) We have selected a more conservative boundary, however, because the resulting background rate uncertainties encompass both the parameter uncertainty of the nominal fit function and the uncertainty in the choice of this particular fit function. The latter is discussed more in Section 7.3.2.

Table 7.1: Background rate uncertainties (in %) applied to each diphoton category and for each Higgs boson mass hypothesis.

m_H (GeV/c^2)	100	105	110	115	120	125	130	135	140	145	150	Average
CC	1.2	1.7	2.2	2.6	3.5	3.4	2.8	3.1	3.5	3.4	3.9	2.8
CP	1.3	1.0	0.6	0.7	0.9	0.6	1.0	0.7	0.8	0.8	1.0	0.9
C′C	4.4	5.2	6.0	5.2	5.3	5.4	5.9	6.7	7.4	7.6	8.2	6.1
C′P	3.2	3.3	3.5	3.6	4.6	3.0	2.2	2.7	3.7	3.0	3.2	3.3

the function that models the smooth portion of the data, $f_{\rm s}(m)$, with a 2$^{\rm nd}$ degree polynomial multiplied by an exponential,

$$f'_{\rm s}(m) = (p_0 + p_1 m + p_2 m^2) e^{-p_3 m}, \tag{7.5}$$

where m is again the diphoton mass. As before, the Breit-Wigner function is added to this for the CP category, however, we leave the C′P category out of this study due to difficulty finding a reasonable fit. Table 7.2 shows the percent difference between the background yields obtained using the fit function above relative to that obtained using the nominal function.

Table 7.2: Background yield differences obtained by replacing $f_{\rm s}(m)$ with $f'_{\rm s}(m)$. Differences are given in %, relative to the background yield obtained using $f_{\rm s}(m)$.

m_H (GeV/c^2)	100	125	150
CC	–1.1	–0.3	1.7
CP	0.6	0.7	0.4
C′C	–1.3	3.7	–5.0

A second approach is taken by adding polynomial terms of different degrees to the fit function being used. For example, when fitting to the CC data we have tested $g(m) + p_n(m-120)^n$, where $g(m)$ is either $f_{\rm s}(m)$ or $f'_{\rm s}(m)$, m is the diphoton mass, n is the integer order of the polynomial, and p_n is a normalization fit parameter. This technique is applied to all categories with n = 1, 2, and 3, using both $f_{\rm s}(m)$ and

$f'_s(m)$ for $g(m)$. In essentially every trial the background expectation for each mass region varies from the nominal prediction given in Section 7.2.4 by <1%.

Based on the results of these two studies, the background expectation due to a different fit function is either considered insignificant, or it is within the uncertainties already obtained from Section 7.3.1. Therefore, we apply only the values shown in Table 7.1 as the systematic uncertainties on the background rate.

Chapter 8

Signal Model

A signal template is built in order to test the SM $H \to \gamma\gamma$ hypothesis in the data. This template consists of the predicted $m_{\gamma\gamma}$ shape for each production type, the normalization of these shapes to the number of expected $H \to \gamma\gamma$ signal events, and systematic uncertainties where appropriate. The signal template is constructed from the production cross sections and branching ratios from Chapter 1, in addition to information obtained from the MC simulation of $H \to \gamma\gamma$ events in the CDF detector. Corrections and/or systematic uncertainties are applied to the MC prediction based on known differences between the simulation and data.

This chapter begins with a description of the MC simulation of $H \to \gamma\gamma$ events and then describes the $m_{\gamma\gamma}$ shape and the normalization of this shape. The next chapter describes the corresponding systematic uncertainties.

8.1 MC Samples

To simulate $H \to \gamma\gamma$ events, we use Pythia version 6.2 [71, 72], a MC generator used in high energy physics to simulate collision events between two incoming particles and the set of outgoing particles from the collision. The event possibilities and kinematics are based on leading-order (LO) theoretical calculations (of Feynman diagrams). Pythia models the hard interaction of interest such as $gg \to H \to \gamma\gamma$,

followed by the fragmentation and decay of particles. The fragmentation and decay process continues until stable particles are formed. Examples of this include quarks and gluons fragmenting (i.e. hadronizing) into jets and a Z boson decaying to objects such as a lepton pair.

The stable particles are then passed through a full simulation of the detector material, CDFSIM [73], which is based on a software package called GEANT [74]. This software is used to simulate physics processes due to elementary particle interactions with the nuclei of detector material, including energy loss of particles (dE/dx), electrons radiating photons due to bremsstrahlung, and photons converting to electron pairs. The showering of particles in the electromagnetic and hadron calorimeters is also simulated (using GFLASH [75]), in addition to the electronic readouts from the detector due to particle interactions. TRIGSIM [73] is used to simulate the trigger data and the corresponding digital response from the detector. Simulated events are analyzed with the same CDF reconstruction software that is used for the collision data, and physics objects are constructed based on vertex, tracking, and calorimeter variables.

The PYTHIA MC generator also incorporates knowledge of parton distribution functions, initial- and final- state radiation, the underlying event, and pileup. Each of these are explained in more detail in Sections 8.1.1–8.1.4, which is largely based on the description given in the PYTHIA manual [72] and Reference [76].

8.1.1 Parton Distribution Functions

In the center of mass frame for a $p\bar{p}$ collision, both the proton and antiproton have a total energy of $E_{\rm CM} = 980$ GeV. The two partons of the hard interaction, however, contain only a portion of the beam energy. Their four-momenta can then be written as $E_{\rm CM}(x_1, 0, 0, x_1)$ and $E_{\rm CM}(x_2, 0, 0, -x_2)$, with the mass of each parton treated as negligible. The probability density function for finding a parton with a momentum fraction x for a given momentum transfer Q^2 is called the parton distribution function (PDF). As an example, the probability of finding a gluon with x_1 and another gluon with x_2 that interact with momentum transfer Q^2 is $P_{g_1}(x_1, Q^2) \cdot P_{g_2}(x_2, Q^2)$ [77],

where the first (second) probability in this product is the PDF value for the first (second) gluon. The PYTHIA event generator relies on a set of PDFs that are determined from experimental inputs. Of the available PDF datasets [78, 79], CTEQ5L [80] was used for the production of signal MC samples.

8.1.2 Initial- and Final-State Radiation

Particles that are colored can radiate gluons and particles that are charged can radiate photons. This alters the configuration of an event such that corrections to the primary interaction are necessary. Initial-state radiation (ISR) is radiation from an initial-state particle before the collision and is observed as additional particles in the detector that are not produced by the primary interaction. In gluon fusion, two gluons interact to produce the Higgs boson, which can then decay to a pair of photons. An example of ISR occurs when one of the gluons radiates a separate gluon — in addition to two reconstructed photons, the separate gluon will fragment and be reconstructed as a jet. Final-state radiation (FSR) is radiation from a final-state particle after the collision; an additional object will also be reconstructed, but this time it has taken some of the energy from one of the final-state particles produced. PYTHIA has a set of parameters that can be adjusted to control the amount of ISR/FSR in an event. The default CDF values are used for generation of the signal MC samples (Tune A [81]).

8.1.3 Underlying Event

Interactions in hadron colliders are often classified as either "hard" or "soft", where the hard interaction generally refers to the parton-parton scattering that results in a set of (or at least one) high p_T objects and the corresponding ISR and FSR radiation (Figure 8.1). All of the remaining objects from soft interactions surrounding the primary scattering process are considered the underlying event (UE) and consist of beam-beam remnants and objects associated with multiple parton interactions. Beam-beam remnants occur in hadron colliders and are due to the fact that protons and antiprotons are composite particles. The hard interaction occurs between a

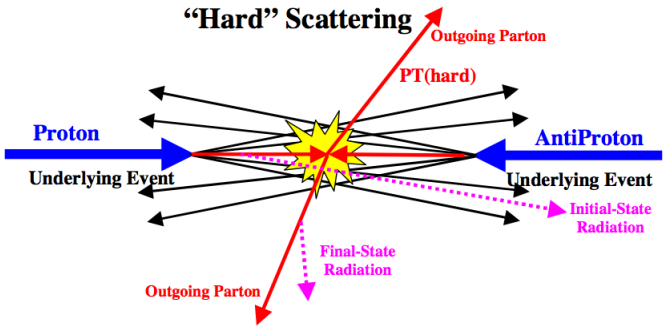


Figure 8.1: A schematic [76] of a hard scattering interaction between a pair of partons from the proton and antiproton, indicated by the red lines pointing towards one another; the resulting high p_T (hard) particles produced by the collision have red lines pointing outward away from the collision vertex. As an example, ISR is shown coming from the parton in the proton and FSR is shown coming from one of the outgoing particles. All other objects shown are the soft $p\bar{p}$ (beam) remnants, which are considered the underlying event.

parton in the proton and a parton in the antiproton; the remaining objects are the remnants and contribute particles to the underlying event with relatively small p_T. Hard or semi-hard scattering can also occur between one or more pairs of partons other than those from the primary interaction. These are called multiple parton interactions (MPI) and the resulting objects from the additional interactions contribute to the event activity and are considered part of the underlying event (Figure 8.2).

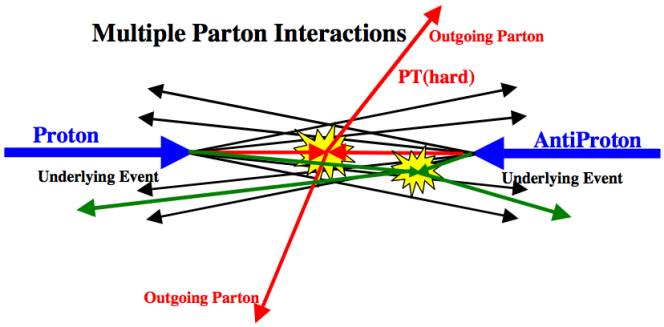


Figure 8.2: The same schematic and description as Figure 8.1 [76], except now a second pair of partons are shown having interacted, indicated by the green arrows moving away from the second vertex. Both beam remnants and multiple parton interactions contribute to the underlying event of the hard scattering process.

The PYTHIA parameters that control the behavior of the underlying event are "tuned" in such a way that the MC simulation better fits measured distributions from Tevatron data, such as the p_T spectrum from jets or the p_T spectrum from lepton pairs produced by Z boson decays [76, 82, 81]. We use the standard CDF UE tune, Tune A [81].

8.1.4 Pileup

PYTHIA events have been described thus far as a hard interaction along with its associated UE and ISR/FSR. For a single bunch crossing, however, there is a finite probability that more than one proton-antiproton collision occurs. This is called pileup. Like MPI, this complicates event reconstruction for two reasons: (i) final-state objects can overlap in the detector and (ii) there is a higher probability of misreconstructing the primary vertex corresponding to the hard interaction of interest. More such interactions means a higher number of reconstructed vertices to choose from.

PYTHIA generates several events, one after the other, to simulate the pileup effect [72]. The number of such events generated for the $H \to \gamma\gamma$ signal simulation may be different from that in the data. As will be described in Section 8.2, we later correct the MC prediction to the data based on the number of reconstructed vertices.

8.1.5 Generated Samples

We generate approximately 300,000 simulated $H \to \gamma\gamma$ events including all physics described in this section, for masses between 100 and 150 GeV/c^2 in 5 GeV/c^2 steps. This is done for each production method for a total of 33 separate signal samples (11 mass points and 3 production mechanisms). For systematic uncertainty studies, two additional samples are also generated for each production method with increased ISR/FSR and decreased ISR/FSR. All uncertainties on the signal template will be described in more detail in the next chapter.

It is known that the LO prediction used in PYTHIA does not model well the transverse momentum of the diphoton system ($p_T^{\gamma\gamma}$). Since our analysis is inclusive,

however, and events are not divided based on $p_T^{\gamma\gamma}$, correcting PYTHIA to a higher-order calculation has little effect on the signal modeling. In particular, the efficiency of signal events passing the diphoton selection and the $m_{\gamma\gamma}$ shape for GF events have been studied after correcting the generated signal from PYTHIA to a higher-order $p_T^{\gamma\gamma}$ prediction from the `HqT` program [83, 84, 85]. Based on the results of this study, which is described in Appendix C, we have maintained the PYTHIA modeling and not corrected GF signal events to the `HqT` prediction.

8.2 $m_{\gamma\gamma}$ Shape

We obtain the diphoton mass shape for $H \to \gamma\gamma$ signal events from the MC simulation. This technique is validated by a comparison of the reconstructed mass shape of $Z \to e^+e^-$ decays in both the data and the MC simulation. Based on this comparison, corrections are made to the EM energy scale such that the MC simulation reproduces the expected Z mean and width observed in the data. These energy corrections are also applied to the reconstruction of photons and were discussed in Chapter 4. After this correction, a small difference remains in the reconstructed $H \to \gamma\gamma$ diphoton mass relative to the generated mass. This is expected to be due to differences in the electron energy scale versus the photon energy scale [86] and we make an additional correction of about 99.2% to the reconstructed mass values for diphoton events in both the data and the $H \to \gamma\gamma$ MC simulation.

An understanding of the $m_{\gamma\gamma}$ shape is important because the more narrow the diphoton mass, the stronger the ability to discriminate a signal from the smoothly falling background. Furthermore, the limit calculation described in Chapter 11 is based on the binned $m_{\gamma\gamma}$ distribution centered on the Higgs boson hypothesis; the greatest sensitivity from this calculation is obtained from the central $m_{\gamma\gamma}$ bins and a more narrow mass shape improves this sensitivity.

As described in Sections 4.1 and 5.3, the four-momentum of each photon is determined by the reconstructed energy, polar angle θ, and azimuthal angle ϕ. The diphoton mass is obtained from the addition of these four-momenta. For Higgs boson

masses studied in this analysis, the corresponding mass width is dominated by the experimental resolution (the natural width of the Higgs boson is negligible [87]). In particular, the resolution is limited by the energy measurement and the ability to correctly identify the vertex of the primary collision. The EM energy is measured by the calorimeters, which have a specified energy resolution (see Chapter 3). Correctly identifying the z position of the primary vertex is important for the θ measurement for non-converting photons. The larger the number of reconstructed vertices from MPI and pileup, the higher the possibility of misidentifying the primary vertex.

Selection of the primary vertex is described in Section 6.4. We apply the same algorithm to events in the signal MC simulation. A study has been performed at CDF using Z decays to understand the effect of vertex misidentification on mass resolution [88]. In this study, the $Z \to e^+e^-$ invariant mass was formed from both data and MC samples using first the standard vertex algorithm and then using a modified vertex algorithm where the p_T of the electron tracks is removed. The resolutions were compared and the smearing effect was observed in both the data and the MC simulation. The results indicate that the smearing effect due to vertex misidentification is modeled well in the simulation.

We therefore correct the $m_{\gamma\gamma}$ shape obtained from the $H \to \gamma\gamma$ MC prediction based only on the difference in the number of events with $N_{\rm vtx}$ reconstructed vertices as observed in the diphoton data ($n^{\rm data}_{N_{\rm vtx}}$) compared to that generated for the signal MC samples ($n^{\rm MC}_{N_{\rm vtx}}$). This gives a set of weights ($w_{N_{\rm vtx}}$) from the ratio

$$w_{N_{\rm vtx}} = \frac{n^{\rm data}_{N_{\rm vtx}}}{n^{\rm MC}_{N_{\rm vtx}}}. \tag{8.1}$$

Then, for each MC event with $N_{\rm vtx}$, a weight of $w_{N_{\rm vtx}}$ is applied when adding entries to the $m_{\gamma\gamma}$ distribution. The weighting has the effect of better correcting the $m_{\gamma\gamma}$ shape to what we would expect in the data.[1]

[1] Since there is only a small fraction of events with higher reconstructed vertices, and since MC samples are generated with a similar $N_{\rm vtx}$ distribution to the data, this process has only a slight effect. The resolution for CC and CP events increases by about 3%, and this has about a 1% or less effect on the limit results calculated in Chapter 11. As we will soon describe, the C′C and C′P categories are not affected by this process.

We perform a few studies using the $H \to \gamma\gamma$ MC samples in order to better understand the $m_{\gamma\gamma}$ resolution dependence on vertex misidentification; however, no further corrections are made. The resolution studies described in the next sections are only to better understand the effect. The first study compares the resolutions as a function of $N_{\rm vtx}$. The second study is similar, except that truth information is used in the MC simulation to identify events with a misreconstructed primary vertex.

8.2.1 Resolution Dependence on $N_{\rm vtx}$

The number of vertices in an event is dependent on both the underlying event and pileup. The latter has a stronger effect for events with higher instantaneous luminosity, as described in Section 8.1.4. We study the resolution dependence here.

Table 8.1 shows the $m_{\gamma\gamma}$ signal resolutions for a Higgs boson mass of 115 GeV/c^2, for the most sensitive diphoton category. The resolution for the CC category smears significantly for higher $N_{\rm vtx}$; however, the percentage of events in the data with higher $N_{\rm vtx}$ is much smaller. From these results, we observe that the $m_{\gamma\gamma}$ resolution due to the energy resolution of the calorimeters is about 2.7 GeV/c^2 for $m_H = 115$ GeV/c^2, and we expect a wider resolution when including events with $N_{\rm vtx} > 1$. The effective resolution is based on a weighted average of all events. We also observe that the GF

Table 8.1: For $m_H = 115$ GeV/c^2, the signal resolution (in GeV/c^2) for the CC category is shown, determined from a Gaussian fit made to the $m_{\gamma\gamma}$ distribution. The widths are provided for each production mode as a function of $N_{\rm vtx}$.

	% of Events	$m_{\gamma\gamma}$ Resolution (GeV/c^2)		
$N_{\rm vtx}$	Diphoton Data	GF	VH	VBF
1	33.1	2.7	2.7	2.7
2	34.3	2.9	2.7	2.8
3	21.2	3.2	2.8	2.8
4	8.3	3.4	2.7	2.9
$\geq$5	3.1	3.7	3.0	3.0

sample is more sensitive to this effect and that the VH and VBF samples are only slightly sensitive to this. This is to be expected since in VH production, the Higgs boson is produced with a W or Z boson and in VBF production there are two quarks present. These other objects decay to final-state particles that leave high p_T tracks, which help better identify the primary vertex of the event. In GF production, there is often a track from an ISR particle, however, not for every event.

Since GF is the dominant production mode, we next focus on the GF sample and look at each diphoton category (Table 8.2). The $m_{\gamma\gamma}$ shapes for each category for a Higgs boson mass of 115 GeV/c^2 are shown in Figure 8.3. The resolutions are compared for events with $N_{\text{vtx}} = 1$ and $N_{\text{vtx}} \geq 4$. We notice that the CP category behaves similarly to the CC category, but that events with a conversion are not affected by the smearing. The high p_T tracks from the electron pair in the C′C and C′P categories are included in the vertex algorithm and, therefore, significantly improve the likelihood of selecting the correct vertex.

Table 8.2: For $m_H = 115$ GeV/c^2, the signal resolution (in GeV/c^2) for the GF production is shown, determined from a Gaussian fit made to the $m_{\gamma\gamma}$ distribution. The widths are provided for each diphoton category as a function of N_{vtx}.

	% of Events in	$m_{\gamma\gamma}$ Resolution (GeV/c^2)			
N_{vtx}	Diphoton Data	CC	CP	C′C	C′P
1	33.1	2.7	2.5	2.6	2.7
2	34.3	2.9	2.8	2.7	2.6
3	21.2	3.2	3.2	2.6	2.7
4	8.3	3.4	3.3	2.6	2.7
≥5	3.1	3.7	3.9	2.6	2.5

8.2.2 Resolution Dependence on Vertex Identification

We next estimate the fraction of signal events where the wrong vertex is reconstructed and show the smeared $m_{\gamma\gamma}$ resolution for these events. We use the known MC-generated data to compare the true z-position (z_{true}) of the vertex to the reconstructed

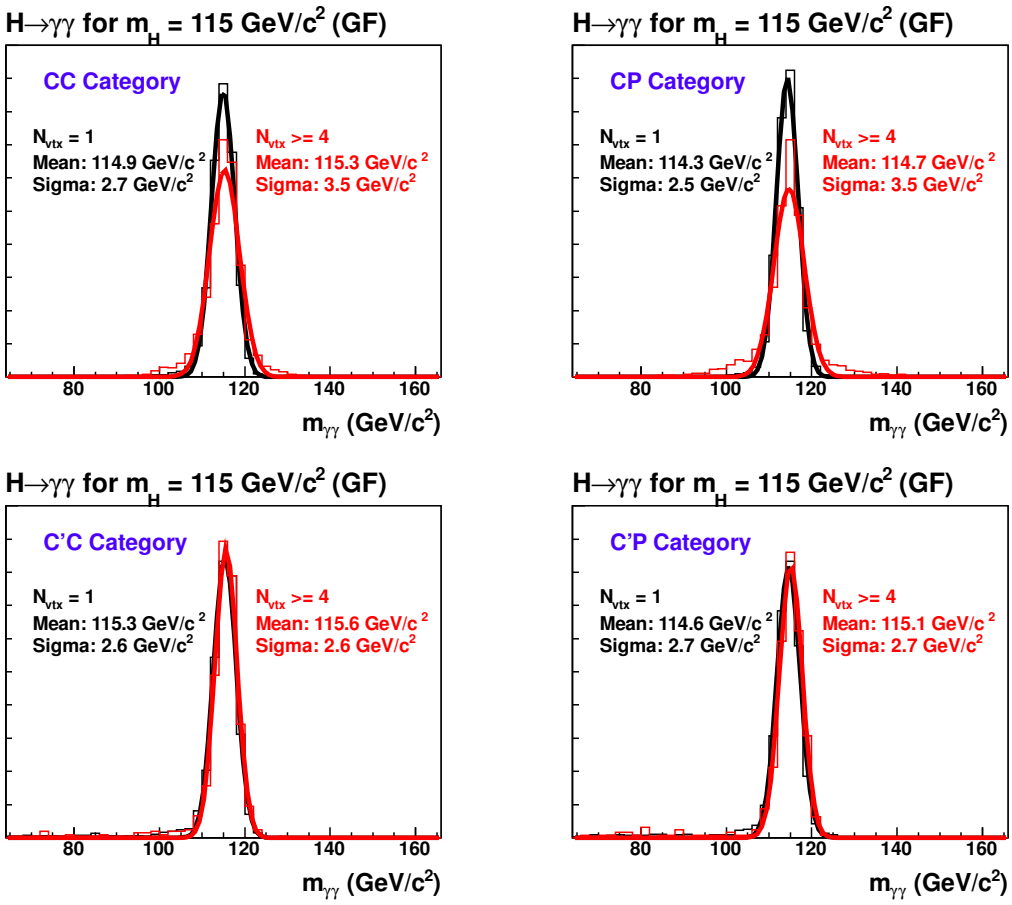


Figure 8.3: For $m_H = 115$ GeV/c^2, the signal resolution (in GeV/c^2) for GF production is shown, determined from a Gaussian fit made to the $m_{\gamma\gamma}$ distribution.

z-position ($z_{\rm reco}$). We then define events with the correct vertex chosen to be those where $|z_{\rm true} - z_{\rm reco}| < 5$ cm (based on the observed vertex reconstruction resolution in the data). We consider all other events to have selected the wrong vertex.

Tables 8.3 and 8.4 show the results for GF and VH production, respectively, where the percentage of wrong vertex events shown in the right four columns is relative to the total number of events with $N_{\rm vtx}$. The percentage relative to all events is shown in the bottom row of the tables, assuming the $N_{\rm vtx}$ distribution of the data. In the C′C and C′P categories, a negligible percentage of events is found to have misidentified the primary vertex. As observed in the previous section, CC and CP events are the only categories significantly smeared by vertex misidentification, and more so for

Table 8.3: The percentage of data events with a given number of vertices (N_{vtx}) is shown in the left two columns. Of GF events with N_{vtx} and $m_H = 115$ GeV/c^2, the fraction predicted to have the wrong vertex reconstructed is obtained from the signal MC simulation. These values are shown in the right four columns (in %), where the statistics were too small for some entries to be provided (shown by a –).

		% of GF Events for each N_{vtx} with $\lvert z_{\text{true}} - z_{\text{reco}} \rvert > 5$			
N_{vtx}	% of Events in Diphoton Data	CC	CP	C′C	C′P
1	30.3%	2.6%	2.7%	0.0%	0.0%
2	34.5%	17.1%	18.3%	0.0%	0.1%
3	22.8%	26.2%	27.6%	0.4%	0.9%
4	9.0%	34.1%	34.1%	0.2%	0.7%
5	2.7%	37.6%	41.8%	0.0%	1.1%
6	0.6%	40.7%	45.5%	–	–
7	0.1%	47.1%	46.6%	–	–
8	0.1%	50.0%	60.0%	–	–
Total	100.0%	17.2%	18.0%	0.1%	0.3%

GF events. For the data sample used in this analysis, we estimate about 17–18% of CC and CP events in the GF sample to have a wrong vertex, about 4–6% in VH events, and about 8.5–9.5% in VBF. (A table is not shown for VBF.) Again, the GF mechanism is a resonant production process; tracks are usually present due to ISR, but not for all events. VH events are suspected to have the smallest fraction of events with a misidentified vertex due to the high p_T leptons often produced by the W or Z boson decay.

The resolution is compared for correct- and wrong-vertex events in Figure 8.4 for only the CC and CP categories and for each production method. For $m_H = 115$ GeV/c^2, CC (CP) events with the primary vertex correctly reconstructed have a $m_{\gamma\gamma}$ resolution of about 2.7 (2.5) GeV/c^2; those with a wrong vertex reconstructed have a $m_{\gamma\gamma}$ resolution of about 6.7 (11) GeV/c^2.

Table 8.4: The percentage of data events with a given number of vertices (N_{vtx}) is shown in the left two columns. Of VH events with N_{vtx} and $m_H = 115$ GeV/c^2, the fraction predicted to have the wrong vertex reconstructed is obtained from the signal MC simulation. These values are shown in the right four columns (in %), where the statistics were too small for some entries to be provided (shown by a –).

		% of VH Events for each N_{vtx} with $\lvert z_{\text{true}} - z_{\text{reco}} \rvert > 5$			
N_{vtx}	% of Events in Diphoton Data	CC	CP	C′C	C′P
1	30.3%	1.1%	1.5%	0.0%	0.0%
2	34.5%	4.6%	5.6%	0.1%	0.1%
3	22.8%	6.6%	8.3%	0.1%	0.4%
4	9.0%	7.8%	10.4%	0.4%	0.9%
5	2.7%	12.2%	13.8%	0.7%	1.5%
6	0.6%	13.4%	13.1%	–	–
7	0.1%	13.7%	11.0%	–	–
8	0.1%	14.3%	28.6%	–	–
Total	100.0%	4.7%	5.7%	0.1%	0.2%

8.2.3 Final Shapes

The shapes used for each diphoton category and production mechanism are provided in Figures 8.5–8.7. These shapes are obtained from the Pythia MC prediction and reweighted to the N_{vtx} distribution of the data as previously described. The effective mass resolutions vary depending on the production process and diphoton category; however, each is less than 3 GeV/c^2.

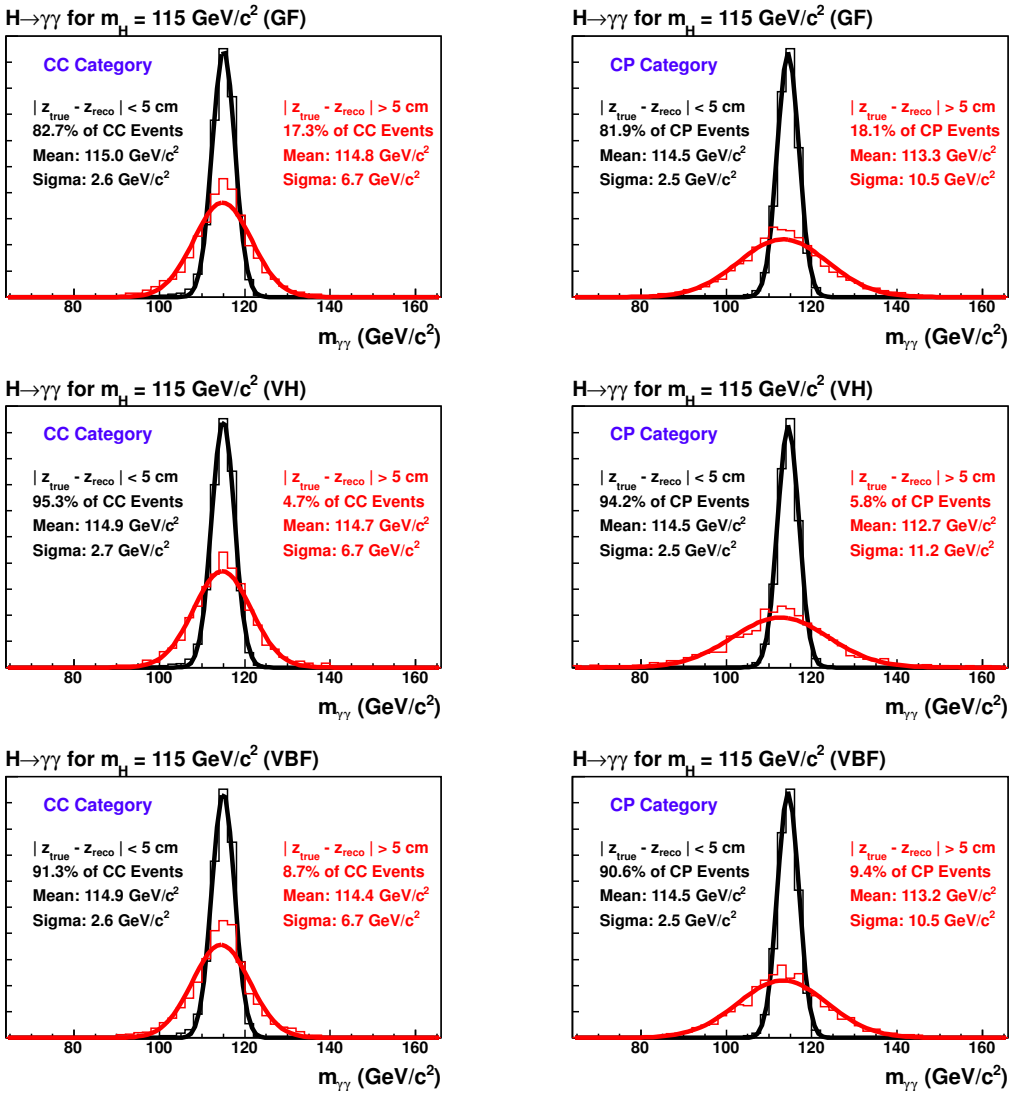


Figure 8.4: For an example Higgs boson mass of $m_H = 115$ GeV/c^2, the GF, VH, and VBF $H \to \gamma\gamma$ diphoton mass shapes are shown for events where a correct and incorrect vertex are reconstructed. These are provided for CC and CP categories only since the effect is negligible for C′C and C′P events.

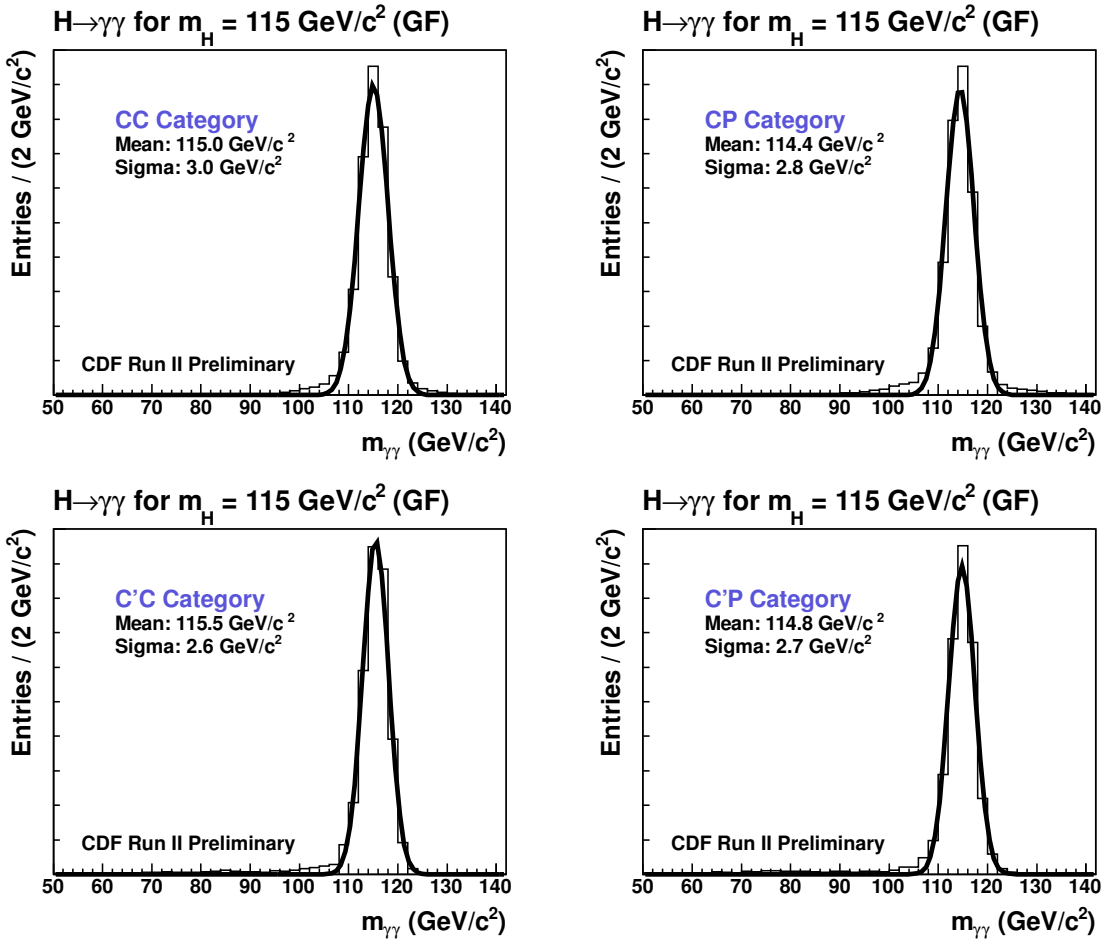


Figure 8.5: The $m_{\gamma\gamma}$ distributions for GF events in each diphoton category along with Gaussian fits made to these distributions. The shapes are obtained from the PYTHIA MC prediction and reweighted to the $N_{\rm vtx}$ distribution of the data.

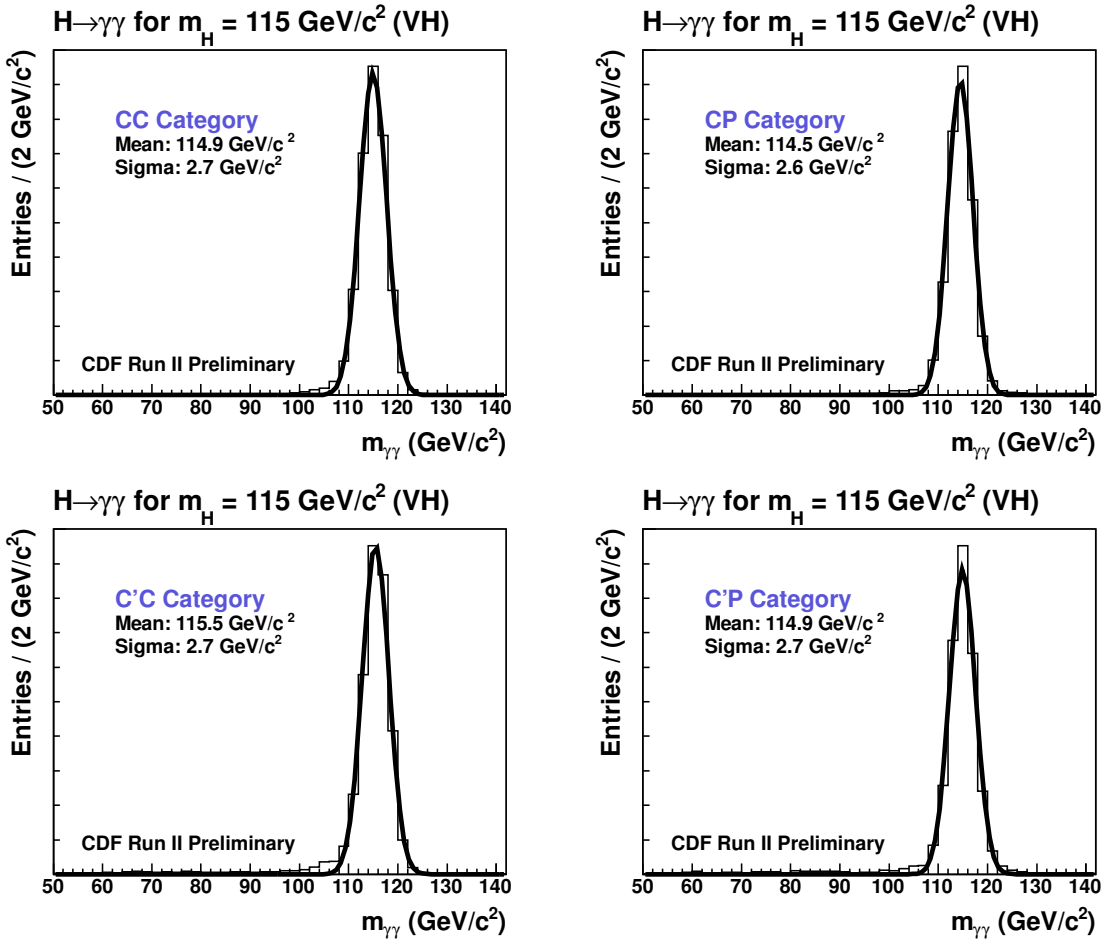


Figure 8.6: The $m_{\gamma\gamma}$ distributions for VH events in each diphoton category along with Gaussian fits made to these distributions. The shapes are obtained from the PYTHIA MC prediction and reweighted to the $N_{\rm vtx}$ distribution of the data.

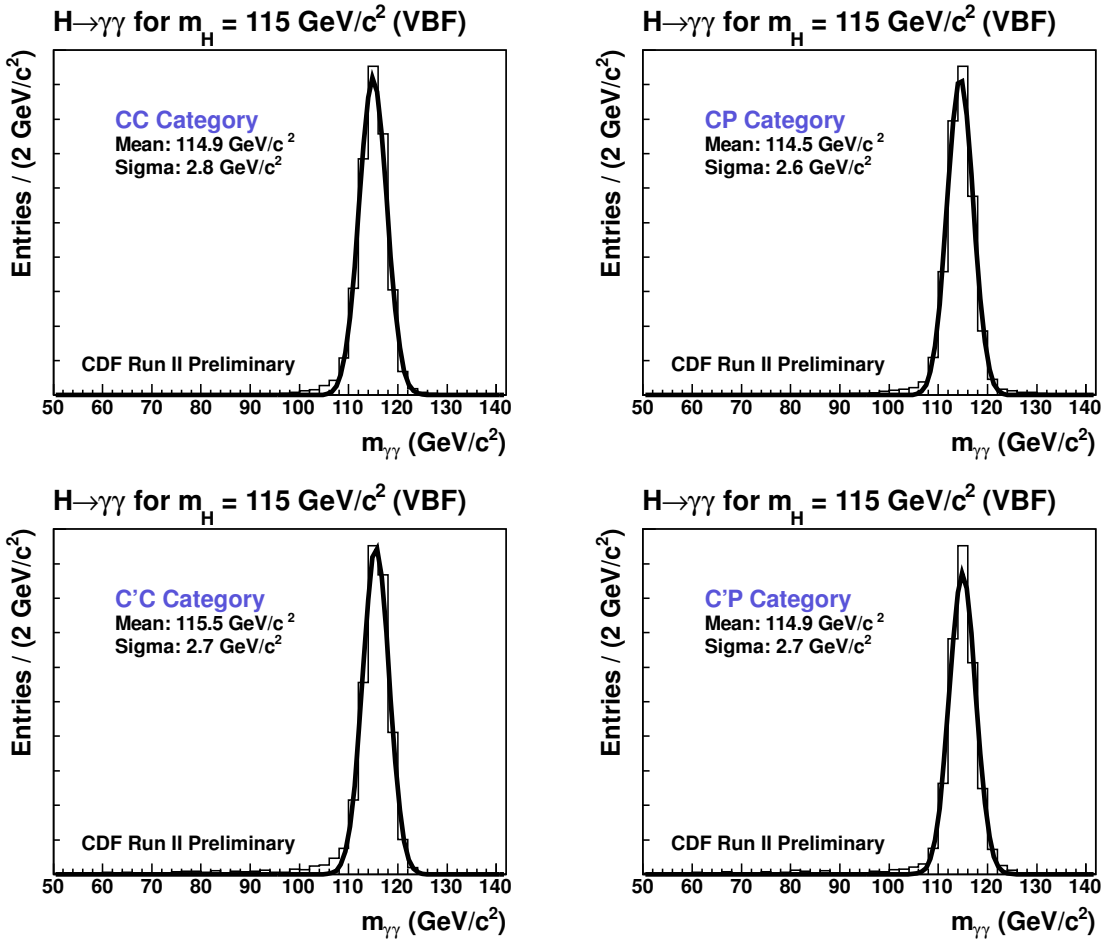


Figure 8.7: The $m_{\gamma\gamma}$ distributions for VBF events in each diphoton category along with Gaussian fits made to these distributions. The shapes are obtained from the PYTHIA MC prediction and reweighted to the $N_{\rm vtx}$ distribution of the data.

8.3 Normalization

As discussed in Chapter 1 and in the previous section, we consider three Higgs boson production mechanisms in this analysis: gluon fusion, associated production with a W or Z boson, and vector boson fusion. For a given theory model including each of these mechanisms, the expected $H \to \gamma\gamma$ cross section can be calculated as the product of the total production cross section and the diphoton branching ratio, $(\sigma_{\rm GF} + \sigma_{\rm VH} + \sigma_{\rm VBF}) \times B(H \to \gamma\gamma)$. The SM prediction for these values was provided in Table 1.1 for each Higgs boson mass hypothesis between 100 and 150 GeV/c^2. For a mass of 115 GeV/c^2 and $\mathcal{L} = 10.0$ fb^{-1} for example, we would expect $\sigma \times B \times \mathcal{L} = 3.35 \times 10.0 = 33.5$ $H \to \gamma\gamma$ events produced.

For each requirement that we make in the event selection, however, only a fraction of the produced $H \to \gamma\gamma$ signal events fall into a well-instrumented region of the CDF detector, pass trigger requirements, and pass the diphoton ID selection. In order to estimate the number of total signal events produced ($N_{\rm produced}$) that are actually reconstructed and make it into our data sample ($N_{\rm reco}$), the efficiency of each requirement is studied in detail. The net efficiency ($\epsilon_{\rm net}$) forms the ratio

$$\epsilon_{\rm net} \;=\; \frac{N_{\rm reco}}{\sigma B \mathcal{L}} = \frac{N_{\rm reco}}{N_{\rm produced}} \tag{8.2}$$

and is the product of the individual consecutive selection efficiencies such that

$$\epsilon_{\rm net} \;=\; \frac{N_{\rm zvtx}}{N_{\rm produced}} \cdot \frac{N_{\rm zvtx+ID_{\gamma\gamma}}}{N_{\rm zvtx}} \cdot \frac{N_{\rm zvtx+ID_{\gamma\gamma}+trig}}{N_{\rm zvtx+ID_{\gamma\gamma}}}. \tag{8.3}$$

In Equation (8.3), $N_{\rm zvtx}$ refers to the number of events produced that have a primary vertex with a z position within 60 cm of the center of the detector (described in Chapter 6), $N_{\rm zvtx+ID_{\gamma\gamma}}$ refers to the number of these events that are additionally within the detector acceptance and pass the diphoton ID selection, and $N_{\rm zvtx+ID_{\gamma\gamma}+trig}$ refers to the number of these events that additionally pass the trigger selection. The full event selection consists of these requirements.[2] The first ratio is the efficiency of the z vertex requirement ($\epsilon_{\rm zvtx}$), the second ratio is the efficiency for identifying a

[2] Note that the good run requirement described in Chapter 6 is only applicable to the data and is therefore excluded here.

diphoton pair in the CDF detector using our photon ID requirements ($\epsilon_{\gamma\gamma}$), and the third ratio is the efficiency of the trigger given our diphoton selection ($\epsilon_{\rm trig}$). We can then rewrite Equation (8.3) as

$$\epsilon_{\rm net} \;=\; \epsilon_{\rm zvtx} \cdot \epsilon_{\gamma\gamma} \cdot \epsilon_{\rm trig}. \tag{8.4}$$

We use both data and simulation to either measure or estimate each of these efficiencies, which are described in more detail in the following sections.

8.3.1 z Vertex Efficiency

As described in Chapter 6, the z position of the primary event vertex is required to fall within 60 cm of the center of the detector, which is the well-instrumented (fiducial) region for vertex identification. This cut is not fully efficient, however, and reduces the effective luminosity of the dataset. A luminosity correction is, therefore, applied which is obtained from minimum bias data.[3] A luminosity-weighted z vertex distribution is obtained from this dataset and is fit to the expected $d\mathcal{L}(z)/dz$ profile. The efficiency of the cut is then taken as

$$\epsilon_{\rm zvtx} \;=\; \epsilon(|z| < 60 \text{ cm}) = \frac{\displaystyle\int_{-60}^{+60} [d\mathcal{L}(z)/dz]\, dz}{\displaystyle\int_{-\infty}^{+\infty} [d\mathcal{L}(z)/dz]\, dz} \tag{8.5}$$

This measurement is performed by the Joint Physics Group at CDF and is obtained separately for different data-taking periods. For the data we use in this analysis, the average z vertex efficiency was found to be 97.43% with a systematic uncertainty of 0.07%. We apply this value for $\epsilon_{\rm zvtx}$ to the normalization of all $H \to \gamma\gamma$ signal MC samples.

[3] Ideally, minimum bias data are the set of events you would get from a totally inclusive trigger. In practice, it is just data from a more unbiased trigger.

8.3.2 Diphoton Efficiency and Correction Factors

The second ratio of Equation (8.3) was simplified to $\epsilon_{\gamma\gamma}$ and it is often thought of as having two components. The first component is called the detector acceptance ($A_{\rm det}$). It is defined as the the fraction of $H \to \gamma\gamma$ events passing the z vertex requirement that additionally pass the diphoton selection of Section 6.6 when only the E_T and fiducial requirements of the photon ID selection are applied. For events that pass this selection, the second component ($\epsilon_{\gamma\gamma}^{\rm ID}$) is the efficiency for selecting two photons when applying the remaining cuts of the photon ID selection. The combined efficiency for identifying a diphoton pair is then $\epsilon_{\gamma\gamma} = A_{\rm det} \cdot \epsilon_{\gamma\gamma}^{\rm ID}$, and is called the diphoton efficiency in this paper.

The diphoton efficiency is obtained from the signal MC samples with the same diphoton selection that is used for the data (see Section 6.6). For each category and mass point, this efficiency is determined from the ratio of the number of events that pass the full diphoton selection to the number of generated MC events that pass the $|z| < 60$ cm cut. These values, labeled $\epsilon_{\gamma\gamma}^{\rm MC}$, are provided in Table 8.5 and shown graphically in Figure 8.8.

In order to better describe the $H \to \gamma\gamma$ signal response in the data, we correct the efficiencies from the simulation ($\epsilon_{\gamma\gamma}^{\rm MC}$) with data-MC scale factors (SF) for central and plug photons. These are obtained by comparing photon efficiencies from the MC prediction with those measured in the data. The method for obtaining these photon ID efficiencies using $Z \to e^+e^-$ decays was described in Section 4.3, where the results are summarized in Table 8.6 for both central and plug photons as a function of the number of reconstructed vertices ($N_{\rm vtx}$). The efficiencies obtained from the data and the MC prediction are weighted according to the $N_{\rm vtx}$ distribution of the diphoton data and $H \to \gamma\gamma$ MC samples, respectively. The MC scale factor is taken from the ratio of these weighted average efficiencies.

The weighted average efficiency obtained from the MC prediction is calculated as

$$\epsilon_{\rm MC}^{\rm Avg} = \frac{\Sigma_{N_{\rm vtx}}(n_{N_{\rm vtx}} \times \epsilon_{N_{\rm vtx}})}{n_{\rm tot}}$$

Table 8.5: Diphoton efficiency for $H \to \gamma\gamma$ signal events ($\epsilon_{\gamma\gamma}^{\mathrm{MC}}$) as defined in the text, for each channel and Higgs boson mass hypothesis. Values are calculated from the MC simulation for each production mechanism separately.

m_H (GeV/c^2)		100	105	110	115	120	125	130	135	140	145	150
CC	GF	12.3	12.5	12.6	12.8	12.9	13.0	13.1	13.3	13.4	13.7	13.7
	VH	12.7	12.7	13.0	13.1	13.1	13.3	13.4	13.4	13.6	13.6	13.7
	VBF	13.8	13.9	14.0	14.2	14.2	14.3	14.4	14.5	14.7	14.7	14.9
CP	GF	16.3	16.4	16.7	16.7	16.8	16.8	16.8	16.7	16.8	16.8	16.7
	VH	14.5	14.6	14.9	15.1	15.2	15.2	15.3	15.4	15.4	15.5	15.5
	VBF	14.6	14.9	15.1	15.3	15.5	15.6	15.6	15.7	15.7	15.6	15.7
C′C	GF	2.8	2.9	2.8	2.9	3.0	3.0	3.0	3.0	3.0	3.0	3.1
	VH	2.7	2.8	2.8	2.8	2.9	2.9	2.9	2.8	2.9	2.9	2.9
	VBF	3.1	3.1	3.2	3.1	3.2	3.1	3.3	3.2	3.2	3.2	3.3
C′P	GF	1.8	1.8	1.8	1.8	1.9	1.9	1.9	1.9	1.9	1.9	1.9
	VH	1.5	1.6	1.6	1.6	1.6	1.7	1.6	1.6	1.6	1.6	1.6
	VBF	1.6	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7

Table 8.6: Central and plug photon ID efficiencies ($\epsilon_{N_{\mathrm{vtx}}}$) as a function of N_{vtx}, obtained from both the data and the MC prediction as described in Chapter 4.

	N_{vtx}	1	2	3	4	5
Central $\epsilon_{N_{\mathrm{vtx}}}$	p0 – p17 $Z \to e^+e^-$ Data	0.8716	0.8563	0.8285	0.8031	0.7339
	p18 – p38 $Z \to e^+e^-$ Data	0.8711	0.8302	0.8017	0.7594	0.7340
	$Z \to e^+e^-$ MC	0.8968	0.8809	0.8595	0.8390	0.8153
Plug $\epsilon_{N_{\mathrm{vtx}}}$	p0 – p17 $Z \to e^+e^-$ Data	0.8284	0.7562	0.6897	0.6184	0.5283
	p18 – p38 $Z \to e^+e^-$ Data	0.7983	0.7121	0.6330	0.5627	0.5147
	$Z \to e^+e^-$ MC	0.8454	0.7977	0.7464	0.6960	0.6477

where $n_{N_{\mathrm{vtx}}}$ is the number of diphoton MC events with N_{vtx} reconstructed vertices, provided in Table 8.7; the total number of diphoton events in the MC sample is given by n_{tot}; and the photon ID efficiency for events with N_{vtx} reconstructed vertices is given by $\epsilon_{N_{\mathrm{vtx}}}$, provided in Table 8.6. The efficiency measured from the data is similar, but takes into account the different photon ID efficiencies obtained for two separate

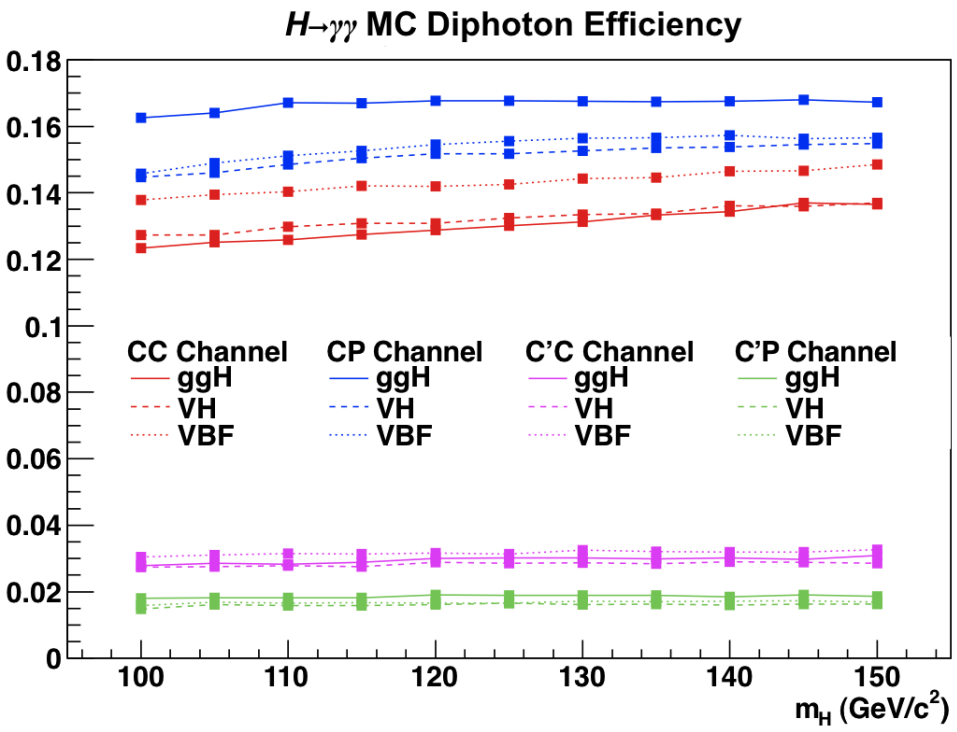


Figure 8.8: Diphoton efficiency for $H \to \gamma\gamma$ MC signal events ($\epsilon_{\gamma\gamma}^{\mathrm{MC}}$) as defined in the text, for each category and Higgs boson mass hypothesis. Values are calculated from the MC simulation for each production mechanism separately.

Table 8.7: The number of entries $n_{N_{\mathrm{vtx}}}$ in the data and generated MC sample with N_{vtx} reconstructed vertices. These values define the N_{vtx} distribution of both the diphoton data and signal MC samples and are used to weight the efficiencies in Table 8.6.

	N_{vtx}	1	2	3	4	5
$n_{N_{\mathrm{vtx}}}$	p0 – p17 Diphoton Data	15111	11501	5659	2043	763
	p18 – p38 Diphoton Data	14707	23007	16839	7027	1792
	$H \to \gamma\gamma$ Signal MC	43732	29594	13326	5308	1694

data-taking period ranges. This weighted average efficiency is calculated as

$$\epsilon_{\mathrm{data}}^{\mathrm{Avg}} = \frac{\Sigma_{N_{\mathrm{vtx}}}(n_{N_{\mathrm{vtx}}} \times \epsilon_{N_{\mathrm{vtx}}})_{\text{p0-p17}} + \Sigma_{N_{\mathrm{vtx}}}(n_{N_{\mathrm{vtx}}} \times \epsilon_{N_{\mathrm{vtx}}})_{\text{p18-38}}}{n_{\mathrm{tot}}}$$

where $n_{N_{\text{vtx}}}$ is the number of diphoton data events with N_{vtx} reconstructed vertices, provided in Table 8.7; the total number of diphoton events in the data sample is given by n_{tot}; and the photon ID efficiency for events with N_{vtx} reconstructed vertices is given by $\epsilon_{N_{\text{vtx}}}$, provided in Table 8.6.

For central photons, the N_{vtx}-weighted photon ID efficiency is found to be 83.26% from the data and 88.17% from the MC prediction. This gives a data-MC scale factor of 94.4% for central photons. For plug photons the weighted ID efficiency is 71.55% from the data and 80.42% from the MC prediction. This gives a data-MC scale factor of 89.0% for plug photons.

We then correct the diphoton efficiencies obtained from the $H \to \gamma\gamma$ MC samples (Table 8.5) by applying the appropriate scale factor once for each photon leg. As an example for CC events, the data-MC correction ($C_{\gamma\gamma}$) would be $0.944 \times 0.944 = 0.891$. The correction factors for each diphoton category are given in Table 8.8, where we do not obtain and apply a scale factor for conversion photons, but will instead later apply a systematic uncertainty. With these corrections defined, the diphoton efficiency from Equation (8.4) is given by

$$\epsilon_{\gamma\gamma} = \epsilon_{\gamma\gamma}^{\text{MC}} \cdot C_{\gamma\gamma}. \tag{8.6}$$

Table 8.8: Values are used to correct the diphoton efficiency obtained from the MC prediction ($\epsilon_{\gamma\gamma}^{\text{MC}}$) to what is expected in the data.

Diphoton Category	CC	CP	C′C	C′P
$C_{\gamma\gamma}$	0.891	0.840	0.944	0.890

8.3.3 Trigger Efficiency

The `PHOTON_25_ISO` trigger selection was described in Chapter 6. If there are $H \to \gamma\gamma$ events being produced in the data, it is possible that there is a set of events that pass our diphoton selection, but would not be collected by the trigger. The simulation of signal events does not directly take this into account, however, and the diphoton efficiencies calculated in the previous section exclude this information. In order to

accurately describe how signal events would behave in the data, we must obtain a measure of the trigger efficiency for our diphoton selection; of the events that pass our diphoton selection, the trigger efficiency is the fraction that would additionally pass the trigger selection.

For this analysis, we obtain a net trigger efficiency for a particular Higgs boson test mass from the MC simulation using CDF's trigger simulation software, TRIGSIM. In Chapter 9, systematic uncertainties are described for this method, based on differences between the simulation and the data. For each mass hypothesis, the net efficiency is determined by first counting the number of signal events in the MC simulation that pass the full diphoton selection for the denominator, and then counting the number of these events where at least one photon leg passes the trigger selection in Table 6.1 for the numerator. The details on using the simulation to determine whether a photon leg would pass the trigger selection are described later in Chapter 9.

Alternatively, we could apply the trigger cuts per event to the simulated trigger data for each Higgs MC sample and then only fill distributions if the trigger selection is passed (in addition to the other diphoton selection). This is important if the shape of a particular variable differs for events that would fail the trigger from events that would pass. Our analysis is dependent on the mass shape alone, however, and this distribution has been compared for events that pass only the diphoton selection and for events that additionally pass the simulated trigger requirements. The comparison demonstrates that the shape differences are insignificant and the results presented here use the net trigger efficiency obtained for each test mass when normalizing the signal mass templates.

The trigger efficiency is obtained for each Higgs boson mass hypothesis using the gluon fusion $H \to \gamma\gamma$ MC samples (GF is the most dominant production mode). The resulting values applied per mass are given for each diphoton category in Table 8.9, where the average over the masses is also shown.

The trigger efficiency for the CC category is essentially 100%. Both the CP and C′C categories have trigger efficiencies of about 98–99%, though the CP efficiency

Table 8.9: Trigger efficiencies (%) for each mass and channel obtained from the GF simulation.

m_H (GeV/c^2)	100	105	110	115	120	125	130	135	140	145	150	Average
CC	99.9	99.9	100	99.9	99.9	99.9	100	100	100	100	100	100
CP	95.9	96.5	97.5	98.0	98.6	98.9	99.1	99.4	99.5	99.7	99.7	98.4
C′C	98.1	98.1	98.4	98.1	98.5	98.7	98.6	98.7	98.7	98.9	99.0	98.5
C′P	90.3	91.8	92.4	93.6	94.6	95.4	95.9	96.5	97.4	97.3	97.4	94.8

varies more over the range of mass hypotheses. The trigger efficiency for the C′P category is between 90 and 97%.

8.3.4 Expected Signal Events

Equipped with values for the z vertex efficiency, diphoton efficiency, and trigger efficiency from Sections 8.3.1–8.3.3, we are now prepared to obtain the expected number of signal events that make it into our data sample. The net efficiency from Equation (8.4) becomes

$$\begin{aligned} \epsilon_{\rm net} &= \epsilon_{\rm zvtx} \cdot \epsilon_{\gamma\gamma} \cdot \epsilon_{\rm trig} \\ &= \epsilon_{\rm zvtx} \cdot \epsilon_{\gamma\gamma}^{\rm MC} \cdot C_{\gamma\gamma} \cdot \epsilon_{\rm trig} \\ &\equiv \epsilon A. \end{aligned} \tag{8.7}$$

A prediction for the number of SM $H \to \gamma\gamma$ signal events for each production mode and diphoton category in the full $m_{\gamma\gamma}$ range is obtained from

$$s = \sigma B \mathcal{L} \epsilon A. \tag{8.8}$$

The total number of signal events for a particular category ($s_{\rm cat}$) is taken from the sum of these values across the three production modes:

$$s_{\rm cat} = \sum_{j=1}^{3} s_j = \sum_{j=1}^{3} \sigma_j B \mathcal{L}^{\rm cat} (\epsilon A)_j^{\rm cat} \tag{8.9}$$

where the σB values are not dependent on the category, but all other values are. For each category, the net efficiencies and total predicted signal events over the full $m_{\gamma\gamma}$ range are given in Table 8.10.

Table 8.10: The product of the signal efficiency and acceptance (ϵA) are shown for each production mode, for each Higgs boson test hypothesis, and each diphoton category. The last column in each subtable is obtained from Equation (8.9). The values provided are for the full diphoton mass range.

CC Channel	Full $m_{\gamma\gamma}$ Range			10.0 fb^{-1}
m_H	ϵA (%)			Signal
(GeV/c^2)	GF	VH	VBF	Event Yields
100	10.7	11.0	12.0	4.1
105	10.8	11.0	12.1	4.0
110	10.9	11.3	12.2	3.9
115	11.1	11.4	12.3	3.7
120	11.2	11.3	12.3	3.5
125	11.3	11.5	12.4	3.2
130	11.4	11.6	12.5	2.8
135	11.6	11.6	12.6	2.4
140	11.7	11.8	12.7	2.0
145	11.9	11.8	12.7	1.5
150	11.9	11.9	12.9	1.1

(a)

CP Channel	Full $m_{\gamma\gamma}$ Range			9.3 fb^{-1}
m_H	ϵA (%)			H_{SM}
(GeV/c^2)	GF	VH	VBF	Signal
100	12.8	11.4	11.5	4.4
105	13.0	11.6	11.8	4.4
110	13.4	11.9	12.1	4.3
115	13.4	12.1	12.3	4.1
120	13.5	12.2	12.5	3.9
125	13.6	12.3	12.6	3.5
130	13.6	12.4	12.7	3.1
135	13.6	12.5	12.7	2.6
140	13.6	12.5	12.8	2.1
145	13.7	12.6	12.7	1.6
150	13.6	12.6	12.8	1.2

(b)

C′C Channel	Full $m_{\gamma\gamma}$ Range			9.9 fb^{-1}
m_H	ϵA (%)			H_{SM}
(GeV/c^2)	GF	VH	VBF	Signal
100	2.5	2.4	2.7	0.93
105	2.5	2.5	2.8	0.92
110	2.5	2.5	2.8	0.89
115	2.6	2.5	2.8	0.85
120	2.7	2.6	2.9	0.83
125	2.7	2.6	2.8	0.75
130	2.7	2.6	2.9	0.66
135	2.7	2.6	2.9	0.55
140	2.7	2.6	2.9	0.45
145	2.7	2.6	2.9	0.34
150	2.8	2.6	3.0	0.26

(c)

C′P Channel	Full $m_{\gamma\gamma}$ Range			9.3 fb^{-1}
m_H	ϵA (%)			H_{SM}
(GeV/c^2)	GF	VH	VBF	Signal
100	1.4	1.1	1.2	0.47
105	1.4	1.3	1.3	0.48
110	1.5	1.3	1.3	0.47
115	1.5	1.3	1.3	0.44
120	1.6	1.3	1.4	0.43
125	1.6	1.4	1.4	0.40
130	1.6	1.3	1.4	0.35
135	1.6	1.4	1.4	0.29
140	1.6	1.3	1.5	0.23
145	1.6	1.4	1.5	0.19
150	1.6	1.4	1.4	0.13

(d)

Chapter 9

Systematic Uncertainty on the Signal Model

The previous chapter described the $H \to \gamma\gamma$ signal model, including a calculation of the expected number of signal events in each diphoton category. This chapter describes sources of systematic uncertainty on the signal event yields.

9.1 Luminosity

The CDF luminosity measurement is determined from the CLC as described in Chapter 3. The largest contribution to the uncertainty on this CDF luminosity measurement is from the CLC acceptance for inelastic processes, though other contributions include the uncertainty on the inelastic $p\bar{p}$ cross section and detector stability. The total uncertainty obtained by the Luminosity Group at CDF is 6% [89, 90].

9.2 Production Cross Section

The cross sections for the GF, VH, and VBF production mechanisms were given in Table 1.1. We apply uncertainties to these theoretical predictions as described in Reference [91], which follows the prescription given in References [92] and [87]. An

uncertainty of 14%, 7%, and 5% is assumed for all masses for the GF, VH, and VBF cross sections, respectively.

9.3 Parton Distribution Functions

The theoretical uncertainty on the parton distribution functions (Section 8.1.1) in the proton and antiproton are propagated as an uncertainty in the $H \to \gamma\gamma$ diphoton efficiency. The uncertainty could be obtained by generating many different MC samples using variations of the central PDF set used. We instead use the common practice of generating a single MC sample and then reweighting the sample based on differences between the central PDF set and numerous PDF error sets. There are many available PDF error sets that each use a different number. We use CTEQ6.1 [93], which contains 40 error sets. The central set is based on a best fit from 20 free parameters and the error sets are upward and downward variations of these parameters. A description of the uncertainty calculation is given here, which is based on explanations from References [72] and [77].

The PDF value P for the parton flavor f_p ($f_{\bar{p}}$) from the proton (antiproton) is dependent on the parton's momentum fraction x_1 (x_2). This PDF value is also dependent on the momentum transfer Q^2 between the two partons. For a given PDF set, the probability of finding f_p with x_1 and $f_{\bar{p}}$ with x_2 for Q^2 is given by the product of the PDF values

$$P_{f_p}(x_1, Q^2) \cdot P_{f_{\bar{p}}}(x_2, Q^2). \tag{9.1}$$

For the i$^{\text{th}}$ PDF error set, an event n is weighted from the ratio of the corresponding probability to the probability obtained from the central 0$^{\text{th}}$ PDF set:

$$w_n^i = \frac{P_{f_p}^i(x_1, Q^2) \cdot P_{f_{\bar{p}}}^i(x_2, Q^2)}{P_{f_p}^0(x_1, Q^2) \cdot P_{f_{\bar{p}}}^0(x_2, Q^2)} \tag{9.2}$$

For the central PDF set, the weight w_n^0 for all events is equal to one. For each signal event, we obtain the four-momentum for the incoming and outgoing particles from generator-level information in the MC sample for a given event. The value of Q is taken to be equal to the Higgs boson mass m_H for GF events, $\hat{p}_T$ in VH events, and

approximated as m_H in VBF events. The momentum fractions are obtained from $x_{1,2}^2 = \hat{s}e^{\pm 2y}/s$ with $\sqrt{s} = 1960$ GeV and with $y = \frac{1}{2}\ln(\frac{E+p_z}{E-p_z})$, where E and p_z are the total energy and z-momentum of the incoming parton-parton system. (See Section 7.2 of Reference [72] for definitions of $\hat{s}$ and $\hat{p}_T$ and for more detail on some of these relationships.)

A diphoton efficiency from the MC simulation is determined as in Section 8.3.2, except now the efficiency is weighted based on the i$^{\text{th}}$ PDF error set as

$$\epsilon_{\gamma\gamma}^{i} = \frac{N_{\text{zvtx+ID}_{\gamma\gamma}}^{i}}{N_{\text{zvtx}}^{i}} = \frac{(\sum_n w_n^i)_{\text{zvtx+ID}_{\gamma\gamma}}}{(\sum_n w_n^i)_{\text{zvtx}}}, \tag{9.3}$$

where the denominator is calculated from the sum of the weights w_n^i for all events that pass the z vertex requirement and the numerator is the sum of the weights w_n^i for events that additionally pass the diphoton selection. The diphoton efficiency for the i$^{\text{th}}$ of 40 PDF sets is then compared to the efficiency obtained from the central PDF set to obtain a systematic uncertainty. We use Equations (3) and (4) of Reference [77] to obtain an uncertainty based on pairs of PDF eigenvectors (up and down), where the 20 resulting upward errors from the corresponding PDF error sets are summed in quadrature, and similarly for the downward errors. PDF uncertainties were obtained from CTEQ6.1 error sets [93] using the LHAPDF interface [79].

The resulting PDF uncertainties for three Higgs boson mass hypotheses are shown in Table 9.1 for each signal sample and diphoton category. For the CP and C′P categories, average uncertainties for each signal sample are within about 2% and we apply this as a uncertainty for all Higgs boson masses and production methods. This is also true for the CC and C′C categories, except for the GF production mechanism, which suggests about a 5.2–5.4% uncertainty. We instead consider that the VH and VBF mechanisms contribute about 25% to the Higgs boson signal expectation and apply a single uncertainty of 5% to all mass points and production methods for the CC and C′C categories.

Table 9.1: PDF upward and downward uncertainties are shown along with their average (Avg.), in %, for the SM Higgs boson selection. They are provided for three different Higgs boson mass points for each signal model.

	m_H	GF			VH			VBF		
Category	(GeV/c^2)	Up	Down	Avg.	Up	Down	Avg.	Up	Down	Avg.
CC	115	5.5	5.0	5.3	1.2	1.4	1.3	1.0	1.6	1.3
	125	5.7	5.0	5.4	1.2	1.4	1.3	1.1	1.7	1.4
	135	5.7	5.0	5.4	1.2	1.3	1.3	1.0	1.8	1.4
CP	115	0.9	1.7	1.3	1.0	0.8	0.9	0.9	0.5	0.7
	125	1.3	2.2	1.8	1.0	0.9	1.0	1.0	0.6	0.8
	135	1.7	2.7	2.2	1.1	0.9	1.0	1.2	0.7	1.0
C′C	115	5.5	4.9	5.2	1.0	1.3	1.2	1.0	1.5	1.3
	125	5.5	4.9	5.2	1.0	1.3	1.2	1.0	1.7	1.4
	135	5.5	4.9	5.2	1.1	1.3	1.2	0.9	1.5	1.2
C′P	115	1.2	2.0	1.6	1.0	0.9	1.0	0.7	0.6	0.7
	125	1.5	2.4	2.0	1.1	0.9	1.0	1.0	0.6	0.8
	135	2.0	3.0	2.5	1.1	1.0	1.1	1.4	0.9	1.2

9.4 Initial- and Final-State Radiation

The parameters in the Pythia MC generator that control the amount of parton showering are varied to obtain the uncertainty in the amount of initial- and final-state radiation. Separate MC samples are generated for one mass hypothesis with less and more ISR/FSR for all three production mechanisms. The diphoton efficiency is then recalculated and compared to the central efficiency obtained with the default ISR/FSR parameter values. We obtain uncertainties from

$$\Delta\epsilon_{\gamma\gamma} = \frac{|\epsilon_{\gamma\gamma}^{\text{more}} - \epsilon_{\gamma\gamma}^{\text{less}}|}{2\epsilon_{\gamma\gamma}^{0}} \tag{9.4}$$

where $\epsilon_{\gamma\gamma}^{\text{more}}$ is the diphoton efficiency from more ISR/FSR, $\epsilon_{\gamma\gamma}^{\text{less}}$ is the diphoton efficiency from less ISR/FSR and $\epsilon_{\gamma\gamma}^{0}$ is the central efficiency. The resulting uncertainties are shown in Table 9.2 for each signal production mechanism. In choosing the uncertainty to apply, weight was given to the GF sample, which is the dominant production

Table 9.2: ISR/FSR $\Delta\epsilon_{\gamma\gamma}$ uncertainties (%) obtained from Equation (9.4).

	Category			
Production Mechanism	CC	CP	C′C	C′P
GF	2.6	3.5	1.7	5.4
VH	3.4	3.9	3.6	3.8
VBF	4.9	4.6	5.5	6.2

mechanism. For each Higgs boson mass and production process, we apply an uncertainty of 3%, 4%, 2%, and 5% to the CC, CP, C′C, and C′P categories, respectively.

9.5 Trigger Efficiency (CC and CP)

The trigger selection was described in Chapter 6. For $H \to \gamma\gamma$ events, TRIGSIM is used to obtain and apply a trigger efficiency for each Higgs boson mass hypothesis as described in Chapter 8. The purpose of this section and the next is to describe the uncertainty applied to this trigger efficiency, based on differences between the data and TRIGSIM simulation. The uncertainty for the CC and CP categories is described in this section and that for the C′C and C′P categories is described in Section 9.6.

For the uncertainty for CC and CP events, we measure the `PHOTON_25_ISO` trigger efficiency in data and compare the results to that obtained from TRIGSIM. For this study, we again take advantage of the similarities between electrons and photons in the calorimeter. $Z \to e^+e^-$ events are selected in a similar manner as Section 4.3, using the same data[1] and MC samples. We then reconstruct a pair of electrons from the tight photon-like ID requirements given in Table 4.3 for central electrons and given in Table 4.4 for plug electrons.

To ensure a pure sample of electrons to study for the trigger efficiency, we require at least one tight central electron of the $Z \to e^+e^-$ decay to be used as the tag leg.

[1] We collect electron data from the `ELECTRON_CENTRAL_18` trigger, which only slightly overlaps with the photon trigger in the Had/Em requirement.

The second electron leg may either be a tight central or tight plug electron, and this probe leg is used to measure the trigger efficiency. The invariant mass of the electron pair is required to have $81 < M_{ee} < 101$ GeV/c^2 around the Z pole in order to reduce backgrounds from fake electrons.

The Stntuple software used in the analysis contains a block of trigger variables with almost[2] the exact L2 and L3 quantities that were used for trigger decisions during run time. For L2 values, the offline electron is matched to the associated L2 trigger-level electron by selecting the L2 EM (or isolation) cluster closest in $\eta - \phi$ space and then obtaining the corresponding E_T, Had/EM, and calorimeter isolation quantities. The L3 values are stored and obtained for E_T, Had/Em, and calorimeter isolation energy. The trigger decision is determined per electron leg by applying the `PHOTON_25_ISO` selection in Table 6.1 using these trigger-level variables just defined. In the $Z \to e^+e^-$ MC sample, we apply an identical procedure, except TRIGSIM is used to simulate the trigger clustering and corresponding trigger variables.

We begin with an explanation of the PEM trigger efficiency. For events with a tight central and tight plug photon, the PEM trigger efficiency is obtained from

$$\epsilon_{\rm trig}^{\rm PEM} = \frac{N_{\rm tight+trig}}{N_{\rm tight}} \tag{9.5}$$

where $N_{\rm tight}$ is the number of events in which the probe electron leg passes tight plug cuts and $N_{\rm tight+trig}$ is the number of events in which the probe leg additionally passes the trigger requirements. As in Section 4.3, we use a modified formula to calculate the efficiency in the central region (see also Appendix A). This formula assumes the two legs are identical, and we therefore require that the tag electron pass the trigger selection in addition to the tight photon ID selection just for the CEM trigger efficiency. The modified efficiency formula is

$$\epsilon_{\rm trig}^{\rm CEM} = \frac{2N_{\rm tight+trig}}{N_{\rm tight} + N_{\rm tight+trig}} \tag{9.6}$$

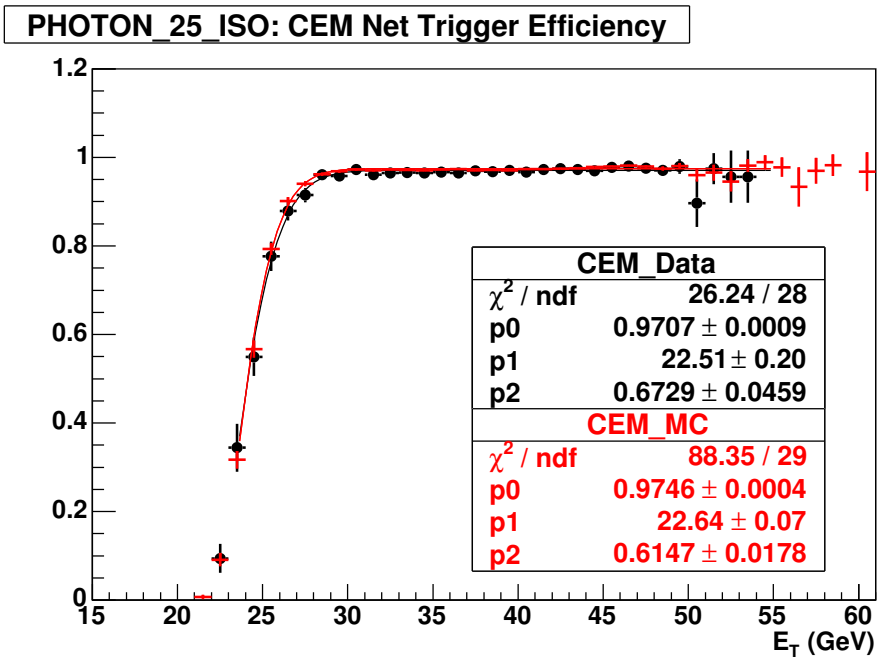
[2] Only the L3 χ^2 was not stored for all data-taking periods, however, the offline value is not expected to be different enough to change the results presented here.

where N_{tight} is the number of events in which the probe electron leg passes tight central cuts and $N_{\text{tight+trig}}$ is the number of events in which the probe leg additionally passes the trigger requirements.

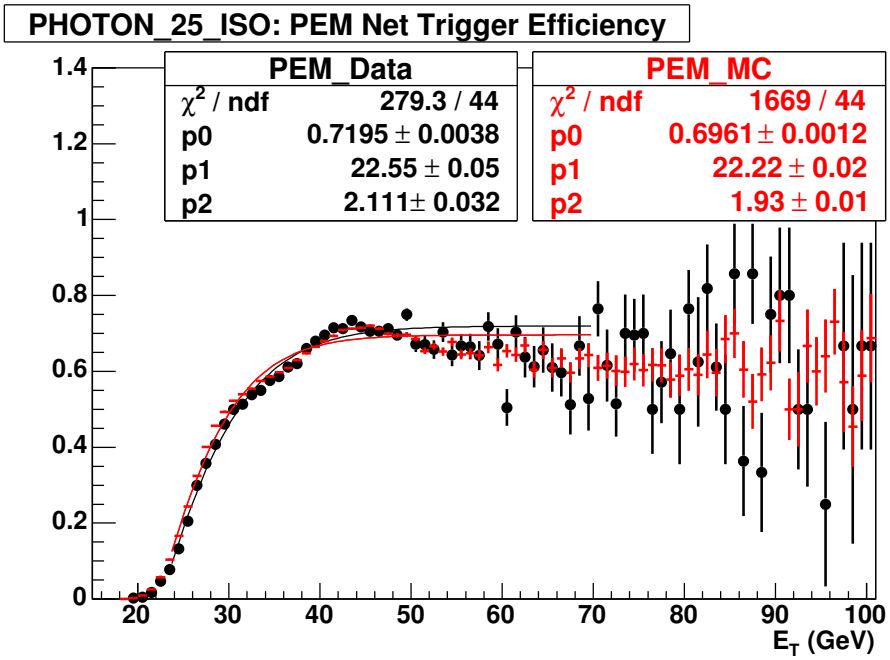
For both the data and MC prediction, the resulting trigger efficiency for central and plug electrons is shown in Figure 9.1 as a function of the probe leg E_T. (The loss in efficiency for both central and plug photons is found to almost entirely come from the L2 isolation requirement, and somewhat from L3 E_T for plug photons. This effect is also found to increase as a function of data-taking period.) The MC simulation is found to well model the trigger response, and we use this study to constrain the uncertainty applied to our trigger efficiencies. We take the difference in plateaus between the data and the MC prediction as a measure of the uncertainty in the MC simulation of the trigger efficiency. For central photons, the difference in the p_0 parameter is 0.4% and the difference for plug photons is 3.3%. These values are dependent on the statistics of the measurement, however, and may fluctuate depending on the sample used, and we therefore round these to 1% and 4%, respectively.

Events collected using the `PHOTON_25_ISO` trigger depend on a single photon passing the trigger selection. For a diphoton sample, events are selected if either leg passes the trigger requirements. The relevant uncertainty to apply is obtained by weighting the uncertainties determined for single central and plug photons (1% and 4%, respectively) depending on the fraction of events where each type of photon passes the trigger selection. We use a Higgs MC sample with a mass of 100 GeV/c^2 to approximate this fraction.

For CC events collected using our trigger, there is always at least one central photon leg that passes the trigger cuts. We therefore apply a 1% uncertainty on the trigger efficiency for CC photon events from the uncertainty obtained for a single photon. For the CP category, it is found that the central photon leg passes the trigger requirements for approximately 92% of the MC events. (For a small portion of these the plug leg also does as well, however, the central leg alone can be used to obtain an uncertainty for these events.) For the remaining 8% of events, the central photon leg does not pass the trigger requirements, however, the plug photon leg does. A



(a)



(b)

Figure 9.1: Net trigger efficiency for (a) central and (b) plug photons as a function of E_T, as measured from the data (black) and the MC prediction using TRIGSIM (red).

weighted efficiency is then taken as $0.01 \times 0.92 + 0.04 \times 0.08 = 0.0124$, and an upper bound is obtained by rounding the weighted uncertainty up to 1.3%.

9.6 Trigger Efficiency (C′C and C′P)

Although the results shown in the previous section indicate that, for central and plug non-converting photons, TRIGSIM is effective in modeling the trigger response, conversions are more complicated and not expected to be modeled quite as well. There is not a pure sample of conversion photons readily available in the data that are in a p_T range relevant for the $H \to \gamma\gamma$ analysis, and it is therefore hard to constrain the trigger efficiency for conversions.

We take a conservative approach for C′C and C′P events by basing the trigger efficiency uncertainty on the inefficiency in the trigger efficiency. Table 8.9 shows that the average trigger efficiency for C′C events for all masses is 98.5%, and therefore a 1.5% uncertainty is applied to the C′C category. The average efficiency for C′P events is 94.8% and a 6% uncertainty is applied.

9.7 z Vertex Efficiency

The uncertainty on the z vertex efficiency described in Section 8.3.1 is 0.07%, obtained from the Joint Physics Group at CDF.

9.8 Energy Scale

An energy scale uncertainty on the diphoton efficiency was checked by varying the reconstructed central and plug energy scales up and down by 1% and recalculating the efficiency. The resulting uncertainties are very small for central, plug, and central conversion photons. We apply an uncertainty of 0.1% for each central photon leg and 0.8% for each plug photon leg, and have considered the uncertainty for conversions negligible compared to the conversion ID uncertainty.

9.9 Photon ID Efficiency

From Chapter 4, there are four sources of uncertainty considered for central and plug photons: fits made to the Z boson mass shapes in the data and the MC simulation, data-taking period/run dependence, the difference between the electron and photon ID efficiency, and material uncertainty for removing conversions. The study describing the conversion ID efficiency is given in Chapter 5.

9.10 Summary of Uncertainties

A summary of all uncertainties on the signal expectation is given in Table 9.3.

Table 9.3: Summary of systematic uncertainties applied to the Higgs boson signal prediction.

	Systematic Uncertainties on Signal (%)			
	CC	CP	C′C	C′P
Luminosity	6	6	6	6
$\sigma_{\mathrm{GF}}/\sigma_{\mathrm{VH}}/\sigma_{\mathrm{VBF}}$	14/ 7/ 5	14/ 7/ 5	14/ 7/ 5	14/ 7/ 5
PDF	5	2	5	2
ISR/FSR	3	4	2	5
Energy Scale	0.2	0.8	0.1	0.8
Trigger Efficiency	1	1.3	1.5	6
z Vertex	0.07	0.07	0.07	0.07
Conversion ID	–	–	7	7
Material Uncertainty	0.4	3.0	0.2	3.0
Photon/Electron ID	1.0	2.8	1.0	2.6
Run Dependence	3.0	2.5	1.5	2.0
Data/MC fits	0.4	0.8	1.5	2.0

Chapter 10

Event Yields

A description of the diphoton selection for the data has now been described, along with the corresponding background and $H \to \gamma\gamma$ signal templates. Before moving on, we pause to collect some of the results from previous chapters. In particular, the number of events for the data, background, and $H \to \gamma\gamma$ signal are summarized in Tables 10.1–10.4. Each of the values in these tables is based on a 12 GeV/c^2 signal region centered on the mass hypothesis being tested, allowing a 2 GeV/c^2 overlap between signal regions.

For each mass hypotheses, the signal efficiency multiplied by the detector acceptance (ϵA) is shown as a percentage of the total number of $H \to \gamma\gamma$ decays for each production mechanism (GF, VH, and VBF). These values, along with the cross sections and branching ratios provided in Table 1.1, were used to obtain the predicted number of SM Higgs boson signal events. This was described in Section 8.3.4. The data event yields are obtained from the diphoton selection of Chapter 6. Integrated luminosities for each channel are also discussed in this chapter and are provided in each table. The background predictions, obtained from the description given in Chapter 7, are also provided. The final column in each table is the number of signal events divided by the square root of the number of background events ($s/\sqrt{b}$).

Table 10.1: Summary of the data, signal, and background event yields for the CC category. Values are provided for the 12 GeV/c^2 signal region centered on each Higgs boson mass hypothesis. See the main text for a description of each quantity.

CC Category (10 fb^{-1})							
m_H	ϵA (%)			Event Yields			
(GeV/c^2)	GF	VH	VBF	Signal	Background	Data	$s/\sqrt{b}$
100	9.9	10.2	11.0	3.8	857	840	0.13
105	9.9	10.1	11.1	3.7	725	688	0.14
110	9.9	10.3	11.1	3.6	566	561	0.15
115	10.0	10.2	11.1	3.4	479	491	0.15
120	10.1	10.2	11.1	3.2	380	413	0.16
125	10.0	10.2	11.0	2.8	332	372	0.16
130	10.1	10.3	11.1	2.5	276	279	0.15
135	10.1	10.1	11.0	2.1	244	246	0.13
140	10.2	10.3	11.1	1.7	196	216	0.12
145	10.3	10.2	11.0	1.3	177	181	0.10
150	10.2	10.2	11.1	1.0	156	132	0.08

Table 10.2: Summary of the data, signal, and background event yields for the CP category. Values are provided for the 12 GeV/c^2 signal region centered on each Higgs boson mass hypothesis. See the main text for a description of each quantity.

CP Category (9.3 fb^{-1})							
m_H	ϵA (%)			Event Yields			
(GeV/c^2)	GF	VH	VBF	Signal	Background	Data	$s/\sqrt{b}$
100	11.4	10.1	10.2	3.9	5427	5378	0.05
105	11.6	10.3	10.5	3.9	4524	4535	0.06
110	11.8	10.5	10.7	3.8	3636	3651	0.06
115	11.8	10.6	10.8	3.6	3179	3155	0.06
120	11.9	10.7	10.9	3.4	2610	2587	0.07
125	11.8	10.7	11.0	3.0	2295	2284	0.06
130	11.7	10.7	10.9	2.6	1901	1857	0.06
135	11.7	10.7	11.0	2.2	1667	1696	0.05
140	11.6	10.6	10.9	1.8	1392	1386	0.05
145	11.7	10.7	10.8	1.4	1236	1219	0.04
150	11.5	10.6	10.7	1.0	1037	1040	0.03

Table 10.3: Summary of the data, signal, and background event yields for the C′C category. Values are provided for the 12 GeV/c^2 signal region centered on each Higgs boson mass hypothesis. See the main text for a description of each quantity.

C′C Category (9.9 fb^{-1})							
m_H	ϵA (%)			Event Yields			
(GeV/c^2)	GF	VH	VBF	Signal	Background	Data	$s/\sqrt{b}$
100	2.3	2.3	2.5	0.9	214	216	0.06
105	2.3	2.2	2.5	0.8	172	184	0.06
110	2.3	2.3	2.6	0.8	142	119	0.07
115	2.3	2.2	2.5	0.8	111	113	0.07
120	2.5	2.4	2.6	0.8	86.3	88	0.08
125	2.4	2.3	2.5	0.7	77.6	73	0.08
130	2.4	2.3	2.6	0.6	61.9	60	0.07
135	2.4	2.3	2.5	0.5	54.2	53	0.07
140	2.4	2.3	2.6	0.4	45.8	48	0.06
145	2.3	2.2	2.5	0.3	42.3	47	0.05
150	2.5	2.3	2.6	0.2	38.7	31	0.04

Table 10.4: Summary of the data, signal, and background event yields for the C′P category. Values are provided for the 12 GeV/c^2 signal region centered on each Higgs boson mass hypothesis. See the main text for a description of each quantity.

C′P Category (9.3 fb^{-1})							
m_H	ϵA (%)			Event Yields			
(GeV/c^2)	GF	VH	VBF	Signal	Background	Data	$s/\sqrt{b}$
100	1.3	1.0	1.1	0.4	793	799	0.02
105	1.3	1.2	1.2	0.4	576	529	0.02
110	1.3	1.1	1.2	0.4	405	439	0.02
115	1.3	1.1	1.2	0.4	345	392	0.02
120	1.4	1.2	1.2	0.4	280	328	0.02
125	1.4	1.2	1.2	0.4	254	289	0.02
130	1.4	1.2	1.3	0.3	216	223	0.02
135	1.4	1.2	1.3	0.3	197	181	0.02
140	1.4	1.2	1.3	0.2	169	143	0.02
145	1.4	1.2	1.3	0.2	148	132	0.01
150	1.4	1.2	1.2	0.1	122	131	0.01

Chapter 11

Limit Calculation and Results

11.1 Introduction

As we saw from the results in Section 7.2.4, no evidence of a signal is observed in the data relative to the background prediction. We then seek to report a measure of how sensitive the data are to a SM $H \to \gamma\gamma$ observation by setting an upper limit on the number of signal events that may be present, given the results of the data.

For a simple example of what we mean by a limit, consider the total background prediction in the six $m_{\gamma\gamma}$ bins centered on a mass of 115 GeV/c^2 in the CC category. This is 479 events, which was provided in Table 10.1. If this prediction were perfectly accurate, then in the absence of signal, we would expect the data to fluctuate about the predicted mean of 479 based on the Poisson error $\sigma = \sqrt{\mu} = \sqrt{479} \approx 21.9$. About 68% of the time we would expect the data yield to be within 1σ of the background prediction (479 ± 21.9) and about 95% of the time we would expect the data yield to be within 2σ of the background prediction (479 ± 43.8).

Now suppose there were a theory that predicted >43.8 signal events in this $m_{\gamma\gamma}$ window, which is above the $+2\sigma$ background prediction. Using a simplistic approach, we might use this value of 43.8 to set an upper limit on the number of signal events we expect to be sensitive to observing, with a greater than 95% credibility level (C.L.). At CDF, using 10 fb^{-1} of data, we expect only about 3.4 SM $H \to \gamma\gamma$ signal events in

this $m_{\gamma\gamma}$ window. Using this simple calculation, the corresponding upper limit on the number of signal events we're sensitive to is then about a factor $f = 43.8/3.4 = 12.9$ times higher than the SM prediction. This limit is an approximate measure of the expected sensitivity of the CC category to the $H \to \gamma\gamma$ process for the 115 GeV/c^2 mass hypothesis. In the absence of a real signal, we would expect to obtain a similar result when performing the calculation from the data. Were we to obtain a sensitivity from the data such that $f < 1$, we would be able to exclude the SM Higgs boson hypothesis for a mass of 115 GeV/c^2 (with greater than a 95% C.L.) because we were sensitive enough to observe such a signal, but we did not.

11.2 The Limit Calculator

For the formal calculation of limits, we use a Bayesian approach to set a 95% C.L. upper limit on the $H \to \gamma\gamma$ cross section that the diphoton data from the CDF detector is sensitive to, $\sigma \times B(H \to \gamma\gamma)$, relative to the SM prediction.[1] These limits are calculated for Higgs boson mass hypotheses between 100 and 150 GeV/c^2, in 5 GeV/c^2 steps, and are produced using binned $m_{\gamma\gamma}$ data from all four categories (CC, CP, C′C, and C′P), incorporating systematic uncertainties on both the background and SM signal predictions. We additionally compare observed limits obtained from the data with limits we would expect if the data matched the background prediction (i.e. no real signal were present).

The `mclimit_csm` code used for obtaining limits can be obtained from Reference [94] and is described in References [95], [96], and [97]. In particular, we have used the `bayes_heinrich_withexpect` routine to obtain both the observed and expected limits. This limit calculator is described in Reference [95], and the method for treating uncertainties that we apply in our analysis (rate uncertainties) is additionally discussed in Reference [96]. For the rest of the chapter, a detailed description of how

[1] In this context, it is understood that σ refers to a cross section rather than a standard deviation. As a reminder, the $H \to \gamma\gamma$ cross section is the number of $H \to \gamma\gamma$ signal events produced per fb^{-1} of data.

the observed and expected limits are calculated is provided, followed by the results of both calculations.

11.3 Observed Limits

Suppose we expect, on average, μ events in our data sample. Then the probability that we observe n events, given μ, is determined by a Poisson distribution[2] as:

$$p(n|\mu) = \frac{\mu^n e^{-\mu}}{n!}. \tag{11.1}$$

The data are divided, however, into $m_{\gamma\gamma}$ bins of width 2 GeV/c^2 for each of the four diphoton categories. For the calculation, we retain six bins centered on the Higgs mass hypothesis being tested, forming a 12 GeV/c^2 signal region. In the absence of systematic uncertainties, the joint (binned) likelihood for a single category would be taken as the product of likelihoods for each $m_{\gamma\gamma}$ bin i, $L(\bar{n}|\bar{\mu}) = \prod_{i=1}^{6} \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}$, where $\bar{n}$ defines the set of observations n_i and $\bar{\mu}$ defines the set of parameters μ_i. We can similarly obtain the joint likelihood for all four categories by extending the bin index, i, to run over each of the $6 \times 4 = 24$ total bins:

$$L(\bar{n}|\bar{\mu}) = \prod_{i=1}^{24} \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}. \tag{11.2}$$

The Poisson probability of obtaining the set of observed results n_i given the set of expected means μ_i is determined by this likelihood distribution.[3]

The expected means μ_i can have both a signal and background contribution such that $\mu_i = s_i + b_i$, where the the total signal contribution that we search for in this analysis would be from three individual production processes: gluon fusion,

[2] The number of events in our data sample is formally distributed based on a binomial distribution because the number of $p\bar{p}$ collisions is finite. The number of particles in proton and antiproton bunches is very large, however, and the number of collisions is very large. Furthermore, the probability of one of these events passing the diphoton selection is very small. In this limit, the binomial distribution becomes a Poisson distribution.

[3] This is for our combined limits using all four categories. Note that for limits using a single category alone, we run over only the six $m_{\gamma\gamma}$ bins corresponding to that category.

associated production, and VBF. We then let $\mu_i = \sum_{j=1}^{3} s_{ij} + b_i$ and continue to tailor the calculation for the SM $H \to \gamma\gamma$ analysis by searching for a signal with (i) the same shapes as that predicted by the theory for each production method, and (ii) the same relative rates between the processes as that predicted by the SM. This allows us to set a limit on the total rate of signal allowed by the data, $s = \sum_{ij} s_{ij}$, as some factor of the SM prediction, $s^{\rm SM} = \sum_{ij} s_{ij}^{\rm SM}$. We define the parameter $f = s/s^{\rm SM}$, with $s_{ij} = f s_{ij}^{\rm SM}$. The expected mean for each bin then becomes $\mu_i = \sum_{j=1}^{3} f s_{ij}^{\rm SM} + b_i$, and we will set a limit on f given the results of the data.

We accomplish this using a Bayesian technique, starting with Bayes' theorem:

$$p(\bar{\mu}|\bar{n}) = \frac{p(\bar{n}|\bar{\mu})\pi(\bar{\mu})}{p(\bar{n})} = \frac{L(\bar{n}|\bar{\mu})\pi(\bar{\mu})}{\int_{-\infty}^{\infty} p(\bar{n}|\bar{\mu}')\pi(\bar{\mu}')d\bar{\mu}'} \tag{11.3}$$

where $p(\bar{n}|\bar{\mu})$ is the likelihood function $L(\bar{n}|\bar{\mu})$ from Equation (11.2), $\pi(\bar{\mu})$ is the probability density function of $\bar{\mu}$ before the results of the data are observed (called the "prior" probability) and $p(\bar{\mu}|\bar{n})$ is the updated probability distribution for $\bar{\mu}$ after obtaining the results of the data (called the "posterior" probability). The denominators shown are constant and provide normalization, which allows us to write Equation (11.3) as

$$p(\bar{\mu}|\bar{n}) = cL(\bar{n}|\bar{\mu})\pi(\bar{\mu}), \tag{11.4}$$

In terms of f and the set of predicted signal and background parameters, $\bar{s}$ and $\bar{b}$ (and where s from now on is understood to mean $s^{\rm SM}$),

$$p(f, \bar{s}, \bar{b}|\bar{n}) \;=\; c\prod_{i=1}^{24} \frac{(f\sum_{j=1}^{3} s_{ij} + b_i)^{n_i} e^{-(f\sum_{j=1}^{3} s_{ij} + b_i)}}{n_i!}\pi(f, \bar{s}, \bar{b}). \tag{11.5}$$

The parameter f is assumed to be independent of the background and signal estimations, but the background and signal estimations are also independent of one another for this analysis since we obtain them from different methods in subsidiary measurements — those for the background come from fits made to the data and those for the signal come from the MC and theory predictions. We can then rewrite the joint prior as $\pi(f, \bar{s}, \bar{b}) = \pi(f)\pi(\bar{s})\pi(\bar{b})$. The marginal posterior density for f can be obtained

by integrating over the set of signal and background priors:

$$p(f|\bar{n}) \;=\; c\,\pi(f)\iint\left[\prod_{i=1}^{24}\frac{(f\sum_{j=1}^{3}s_{ij}+b_i)^{n_i}e^{-(f\sum_{j=1}^{3}s_{ij}+b_i)}}{n_i!}\right]\pi(\bar{s})\,\pi(\bar{b})\;d\bar{s}\;d\bar{b}. \quad (11.6)$$

For each of the 24 bins (4 categories with 6 bins each), the integrals would be taken over the three types of signal contributions and one background contribution for a total of 96 integrals.

Instead of doing each of these integrals analytically, we apply MC integration by using the signal and background priors to generate random variations of the central estimates for the SM signal and background expectations. The central predictions for a particular category are varied by looping over the relevant set of systematic uncertainties and applying consecutive Gaussian factors of mean one and width 1σ corresponding to the fractional errors given in Chapter 9 for the SM signal and Chapter 7 for the background prediction.[4] Each of these random variations considers correlations and produces a new set of signal and background predictions, within the systematic uncertainties applied. This process is performed M times to produce a finite Bayesian prior-ensemble — M different sets of the 96 signal and background parameters. We then rewrite the marginalized posterior for f from Equation (11.6) as

$$p(f|\bar{n}) \;=\; \frac{1}{N}\,\pi(f)\sum_{k=1}^{M}\left[\prod_{i=1}^{24}\frac{(f\sum_{j=1}^{3}s_{ijk}+b_{ik})^{n_i}e^{-(f\sum_{j=1}^{3}s_{ijk}+b_{ik})}}{n_i!}\right] \quad (11.7)$$

where N is a normalization constant.

We now have a probability density function for f, given the results of the data, however, we have yet to select a prior distribution for f which will contain our *a priori* knowledge of f before we have performed the analysis. We know that the amount of signal will not be negative; however, we do not prefer any other amount over another. An (improper) flat prior is, therefore, chosen:

$$\pi(f) = \begin{cases} 0 & \text{for } f < 0 \\ a & \text{for } f \geq 0 \end{cases} \quad (11.8)$$

[4] These are truncated Gaussians in order to keep the total signal and background predictions for a bin greater than zero.

where a is just a constant that we will absorb into the normalization factor N from Equation (11.7). With this prior, we now set a Bayesian 95% C.L. limit such that 95% of the posterior density for f falls below the limit:

$$0.95 = \int_0^{f_{95}} p(f|\bar{n})df \tag{11.9}$$

$$= \int_0^{f_{95}} \frac{1}{N}\sum_{k=1}^{M}\left[\prod_{i=1}^{24}\frac{(f\sum_{j=1}^{3}s_{ijk}+b_{ik})^{n_i}e^{-(f\sum_{j=1}^{3}s_{ijk}+b_{ik})}}{n_i!}\right]df \tag{11.10}$$

As an example, the posterior density for a Higgs boson mass hypothesis of 115 GeV/c^2, obtained with $M = 10,000$ different systematic samplings, is shown in Figure 11.1. The observed 95% C.L. limit of 12.7 for this mass is shown by the blue line, where 95% of the posterior falls below this limit. Since, in our analysis, we only use 6 $m_{\gamma\gamma}$ bins per category, the calculation of Equation (11.10) is relatively fast compared to other analyses, only taking a few minutes (on a MacBook with a Core 2 processor at 2.1 GHz).

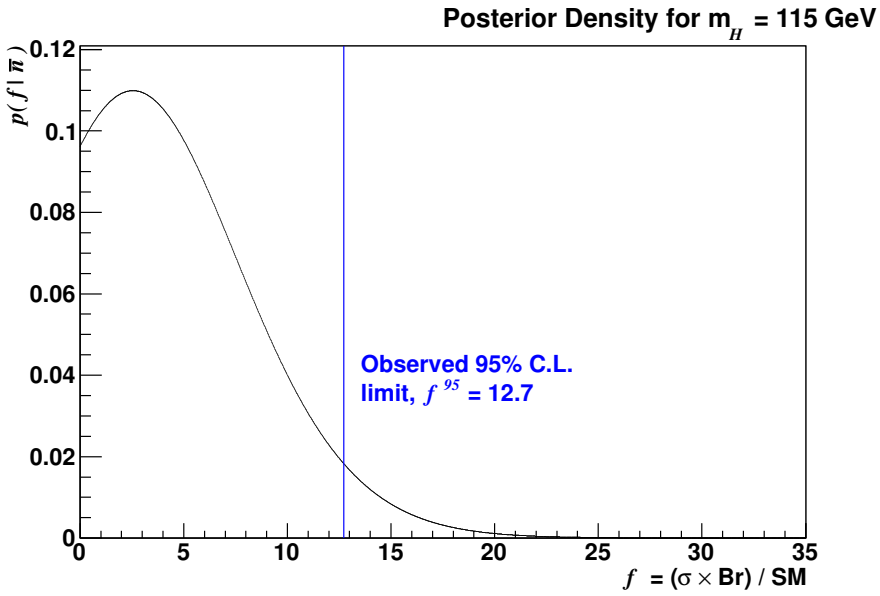


Figure 11.1: The posterior density for $f = \sigma \times B(H \to \gamma\gamma)/\text{SM}$ shows the observed limit set by the data of 12.7 times the SM prediction for a Higgs mass hypothesis of 115 GeV/c^2. This limit is indicated by the blue line, where 95% of the density is below this limit. This posterior density was obtained using all four diphoton categories.

11.4 Expected Limits

We now compare the results obtained from the data with the results we would expect if the limit were obtained using background predictions only (no data). This is achieved by performing an identical calculation as described in the previous section with two modifications described here.

11.4.1 Pseudodata

We replace the 24 bins of real data with 24 values of "pseudodata." The set of pseudodata is obtained from the background prediction by fluctuating the background expectations in the exact same manner as described in the previous section — by applying consecutive Gaussian factors based on the background systematic uncertainties. The only thing that changes about Equation (11.10) is that we replace the values of n_i with the pseudodata d_i. We can then obtain a limit from

$$0.95 = \int_0^{f_{95}} \frac{1}{N} \sum_{k=1}^{M} \left[\prod_{i=1}^{24} \frac{(f \sum_{j=1}^{3} s_{ijk} + b_{ik})^{d_i} e^{-(f \sum_{j=1}^{3} s_{ijk} + b_{ik})}}{d_i!} \right] df \qquad (11.11)$$

just as we did for the observed limit, which has a similar calculation speed of a few minutes. The process of replacing the data with pseudodata and calculating a corresponding limit is called a pseudoexperiment.

11.4.2 Pseudoexperiments

We then repeat the above step a sufficient number of times (on the order of 10^3) so as to produce a 95% C.L. limit space demonstrating where we would expect the observed limit, in the absence of a real signal, to lie. The median of these trials is called the expected limit. The region where 68% of these pseudoexperiments lie around the median is the 1σ expected region, and the region where 95% of them lie around the median is the 2σ expected region.

As an example, Figure 11.2 shows the distribution of expected limits from 15,000 pseudoexperiments for a Higgs boson mass of 115 GeV/c^2 using all four categories.

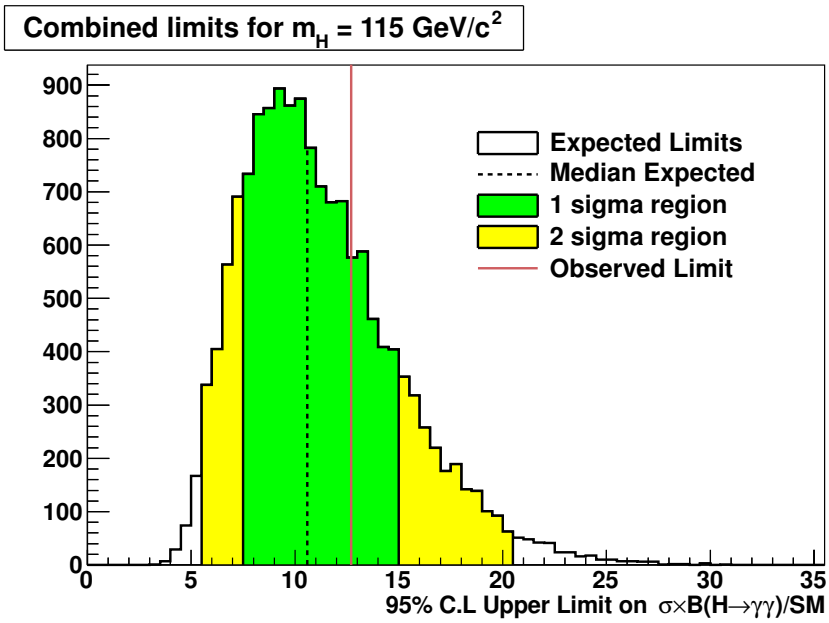


Figure 11.2: For $m_H = 115$ GeV/c^2 and using all four categories, a distribution of 95% upper credibility level limits on the production cross section multiplied by the $H \to \gamma\gamma$ branching ratio relative to the SM prediction. The expected limit distribution is shown with the median, $\pm 1\sigma$, and $\pm 2\sigma$ regions. The observed limit is also shown as the vertical line (from Figure 11.1).

The median of these limits is shown, along with the $\pm 1\sigma$ and $\pm 2\sigma$ regions. The observed is also shown, where a local p-value could be obtained from the fraction of the expected limit distribution above the observed limit [98].

Running a sufficient number of pseudoexperiments is what takes the longest amount of time for the limit results presented. We take advantage of a large network of computer farms residing at Fermilab and designated for CDF use. For this analysis, running 2000 pseudoexperiments, divided into 100 sections on the computer farm, for example, takes roughly half a day or less. The results presented here used up to five times this amount, taking a couple of days.

11.5 Limit Results

11.5.1 Limits for Each Diphoton Category

In order to understand the significance of each category to the combined results, limits are first provided for each category alone. These are shown in Table 11.1 and Figure 11.3, where f^{95} is now fully written out as the 95% C.L. upper limit on the diphoton $H \to \gamma\gamma$ cross section relative to the SM theory prediction. The observed limit for each category is compared to the median expected limit and the $\pm 1\sigma$ and $\pm 2\sigma$ expected limit regions.

The CC category provides the greatest expected sensitivity with expected limits of about 13 times the SM theory prediction ($13 \times$ SM) for $m_H = 115$ GeV/c^2. Due to the implementation of the NN photon ID selection for central photons, this is an improvement of about 9% relative to limits obtained using the standard central photon ID selection. The C′C and CP categories have a similar expected sensitivity of about $27 \times$ SM and $29 \times$ SM at $m_H = 115$ GeV/c^2, respectively. The C′P category by itself obtains expected limits near $100 \times$ SM.

For a single category, an intuitive way to see the limit from that category alone is provided in Figure 11.4 for the 115 GeV/c^2 Higgs boson mass hypothesis. The subfigures for each row show invariant mass shapes for the CC, CP, C′C, and C′P categories, respectively, where the whole mass range is given in the left column and the figure is zoomed in for the right column. The grey region shows the background prediction and the data are given as points. The scale of the $H \to \gamma\gamma$ signal to the expected (red) and observed (dashed line) limits indicates the minimum rate of signal events needed in order to observe a signal in the data above the background prediction at a 95% C.L.

In Section 11.1, a simple calculation was performed using the background expectation and its statistical uncertainty to approximate the expected limit for $m_H =$ 115 GeV/c^2 in the CC category only. Notice that this calculation demonstrates that the limit results for this mass and category are reasonable, having been obtained from the full Bayesian procedure, which used the binned $m_{\gamma\gamma}$ distribution and takes into

account systematic uncertainties and correlations in the background and SM signal predictions.[5]

Table 11.1: 95% C.L. upper limits on the production cross section multiplied by the $H \to \gamma\gamma$ branching ratio, relative to the SM prediction, calculated for each diphoton category alone.

CC Category				CDF Run II Preliminary, 10 fb^{-1}		
m_H	95% C.L. Limit/σ(SM)$\times B(H \to \gamma\gamma)$					
(GeV/c^2)	-2σ	-1σ	Median Exp	$+1\sigma$	$+2\sigma$	Observed
100	7.5	10.2	14.4	20.1	27.5	10.7
105	7.2	9.8	13.8	19.3	26.3	10.2
110	7.0	9.5	13.2	18.4	24.9	11.2
115	6.9	9.4	12.9	18.4	25.2	15.6
120	7.0	9.6	13.4	18.8	25.9	22.1
125	7.2	9.6	13.4	18.9	26.1	20.1
130	7.3	9.9	13.8	19.7	26.9	14.6
135	8.1	11.0	15.3	21.7	29.8	16.7
140	9.3	12.5	17.5	24.7	34.0	26.7
145	11.3	15.1	21.0	29.7	40.9	24.8
150	15.0	20.2	28.2	40.0	55.4	15.5

(a) CC Channel

CP Category				CDF Run II Preliminary, 10 fb^{-1}		
m_H	95% C.L. Limit/σ(SM)$\times B(H \to \gamma\gamma)$					
(GeV/c^2)	-2σ	-1σ	Median Exp	$+1\sigma$	$+2\sigma$	Observed
100	20.6	27.8	38.9	54.5	74.7	28.5
105	17.8	23.9	33.7	47.1	64.3	50.6
110	15.7	21.1	29.5	41.0	56.6	25.0
115	15.5	20.9	29.2	40.9	55.6	25.0
120	15.5	21.0	29.4	41.0	56.3	35.2
125	15.7	21.1	29.5	41.3	56.2	19.2
130	17.4	23.5	32.9	46.0	63.0	34.4
135	18.7	25.2	35.2	49.3	67.1	41.4
140	21.8	29.4	41.0	57.4	78.6	40.8
145	26.3	35.5	49.6	69.4	94.7	48.6
150	34.3	46.4	64.7	90.5	124	67.0

(b) CP Channel

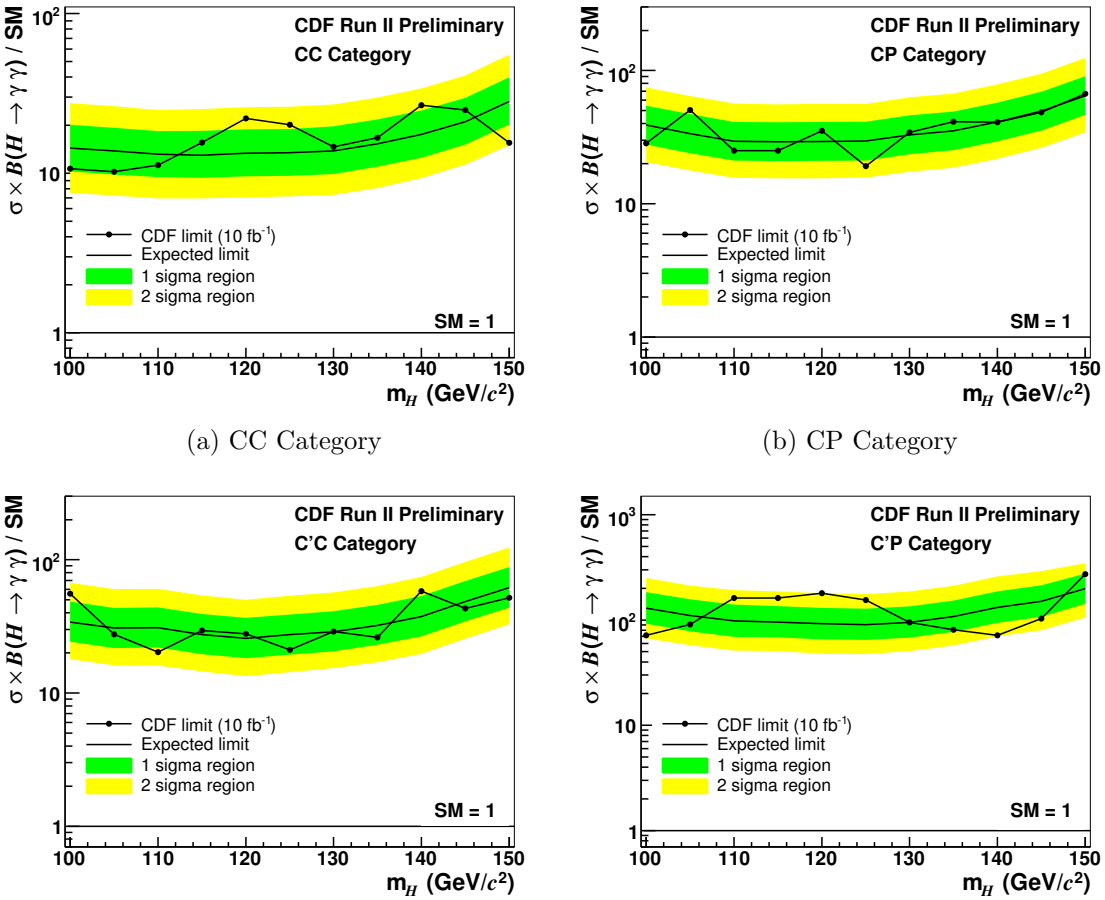
C′C Category				CDF Run II Preliminary, 10 fb^{-1}		
m_H	95% C.L. Limit/σ(SM)$\times B(H \to \gamma\gamma)$					
(GeV/c^2)	-2σ	-1σ	Median Exp	$+1\sigma$	$+2\sigma$	Observed
100	18.0	24.3	34.1	48.5	67.1	55.6
105	16.1	21.8	30.8	43.5	60.1	27.6
110	16.1	21.9	30.9	43.9	60.5	20.3
115	14.5	19.6	27.4	39.1	54.0	29.4
120	13.4	18.2	25.7	36.6	50.1	27.9
125	14.4	19.4	27.4	38.8	53.7	21.1
130	15.4	20.5	28.8	41.0	56.9	29.0
135	17.0	23.0	32.1	45.9	63.5	26.2
140	19.7	26.6	37.5	53.4	74.4	58.3
145	25.5	34.5	48.6	69.2	96.3	43.2
150	32.9	43.8	61.8	88.2	124	51.8

(c) C′C Channel

C′P Category				CDF Run II Preliminary, 10 fb^{-1}		
m_H	95% C.L. Limit/σ(SM)$\times B(H \to \gamma\gamma)$					
(GeV/c^2)	-2σ	-1σ	Median Exp	$+1\sigma$	$+2\sigma$	Observed
100	67.1	92.3	130	184	250	71.4
105	57.7	78.4	110	157	212	90.6
110	51.2	69.2	98.3	140	192	162
115	50.6	68.4	95.7	136	187	162
120	47.7	65.3	92.3	130	180	180
125	47.4	64.7	90.5	128	176	155
130	50.8	68.0	94.8	134	186	95.0
135	57.1	77.3	108	153	212	80.7
140	70.1	93.8	131	187	259	71.5
145	79.9	107	151	214	291	103
150	105	142	198	276	346	273

(d) C′P Channel

[5] Generally speaking, the simple calculation demonstrates that the true limits are in the right ball park. It just so happens that the example provided using the back-of-the-envelope method gave an identical expected sensitivity of 12.9 $\times$ SM as the Bayesian expected limit for this mass and category.



(a) CC Category

(b) CP Category

(c) C′C Category

(d) C′P Category

Figure 11.3: 95% C.L. upper limits on the production cross section multiplied by the $H \to \gamma\gamma$ branching ratio, relative to the SM prediction, calculated for each diphoton category alone.

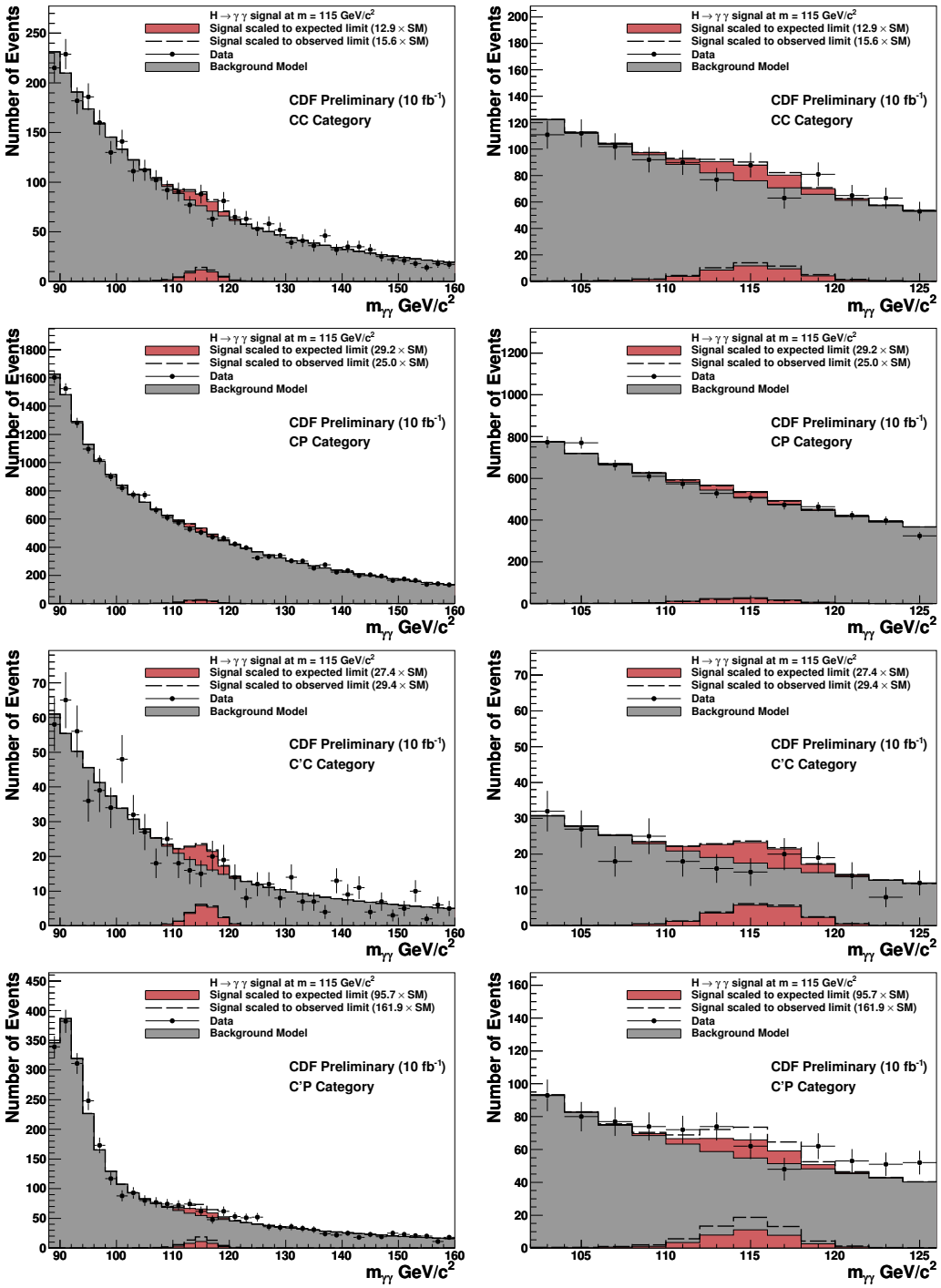


Figure 11.4: Invariant mass distribution over (left) the whole mass range and (right) zoomed in, for an example theoretical SM Higgs mass at 115 GeV/c^2. The signal is shown scaled to the expected and observed limits obtained from the respective category alone.

11.5.2 Combined Limits

In Figure 11.5, the combined observed and expected limits using all four categories are compared to the results obtained from each category alone. The CC category contributes the majority of the sensitivity (roughly 80–85%). Recovering central photon conversions provides an approximate 13% improvement to the CC limits with the inclusion of the C′C category. The CP category has similar limits as the C′C category, and contributes most of the remainder of the sensitivity in the combined limits. The C′P category contributes only about 1–2%.

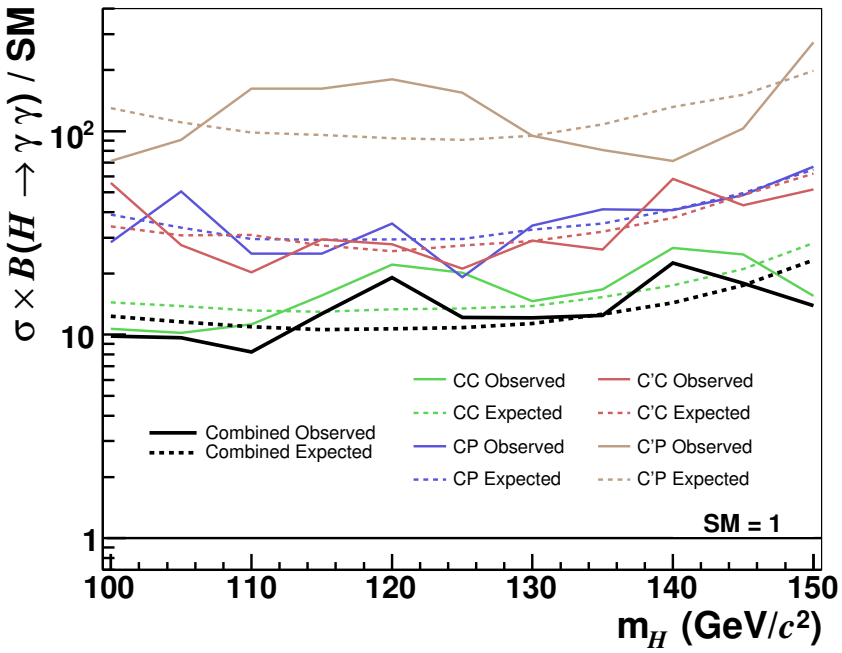


Figure 11.5: 95% C.L. upper limits on the production cross section multiplied by the $H \to \gamma\gamma$ branching ratio, relative to the SM prediction. Observed and expected limits are shown for each category calculated alone and for the combined limit using all four categories.

The combined observed and expected limits with the $\pm 1\sigma$ and $\pm 2\sigma$ expected limit regions are given in Table 11.2 and shown graphically in Figure 11.6. The greatest sensitivity is obtained for Higgs boson mass hypotheses between 110 and 125 GeV/c^2, where the expected limits are about 10.8 $\times$ SM. Observed limits are found to agree with the expected limits within the 2σ expectation band. These results demonstrate

Table 11.2: 95% C.L. upper limits on the production cross section multiplied by the $H \to \gamma\gamma$ branching ratio, relative to the SM prediction, for categories combined.

m_H	95% C.L. Limit/σ(SM)$\times B(H \to \gamma\gamma)$					
(GeV/c^2)	-2σ	-1σ	Median Exp	$+1\sigma$	$+2\sigma$	Observed
100	6.5	8.8	12.3	17.2	23.4	9.9
105	6.2	8.3	11.6	16.3	22.2	9.7
110	5.8	7.8	10.9	15.4	20.7	8.2
115	5.6	7.6	10.6	14.9	20.3	12.7
120	5.6	7.6	10.7	15.1	20.6	19.0
125	5.7	7.8	10.8	15.2	20.9	12.2
130	6.0	8.1	11.4	16.1	22.2	12.1
135	6.6	8.9	12.6	17.8	24.2	12.5
140	7.6	10.2	14.4	20.3	28.0	22.6
145	9.2	12.5	17.4	24.5	33.6	17.9
150	12.3	16.5	23.2	32.7	44.7	13.9

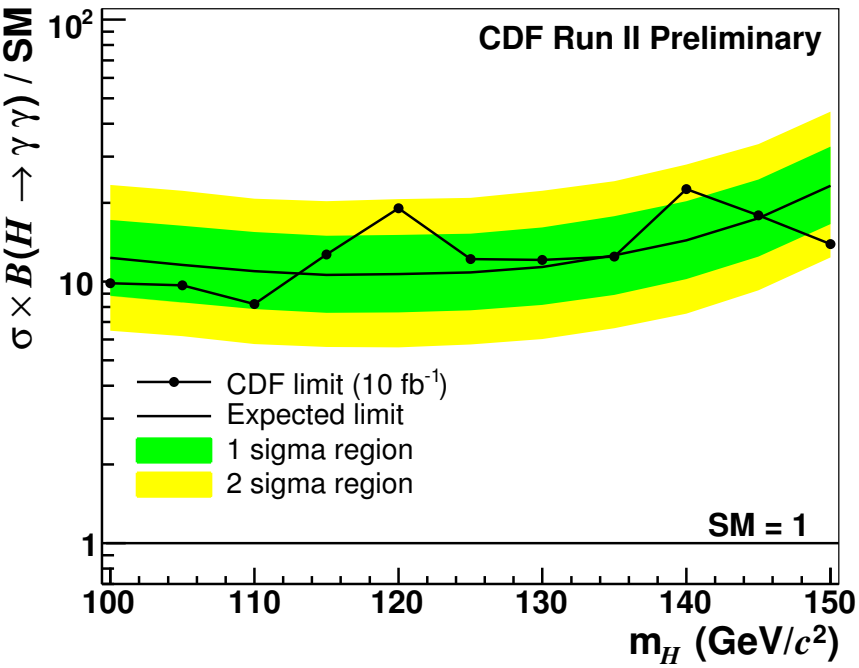


Figure 11.6: 95% C.L. upper limits on the production cross section multiplied by the $H \to \gamma\gamma$ branching ratio, relative to the SM prediction, for categories combined.

the sensitivity of the CDF diphoton data ($\mathcal{L} = 10$ fb^{-1}) to a SM $H \to \gamma\gamma$ observation based on the diphoton mass spectrum using central, plug, and central conversion photons.

These results have been combined with those from the D0 experiment and the corresponding limits are provided in Figure 11.7. The expected sensitivity from the combined CDF and D0 diphoton data is about $7 \times$ SM for $m_H = 125$ GeV/c^2. The limits presented in this chapter furthermore contribute to the overall Higgs boson search at CDF and the Tevatron. The have been included in the March 2012 limit combinations, which are based on results from multiple search channels in both CDF and D0 data [31, 99]. The diphoton channel contributes sensitivity particularly for $110 < m_H < 130$ GeV/c^2.

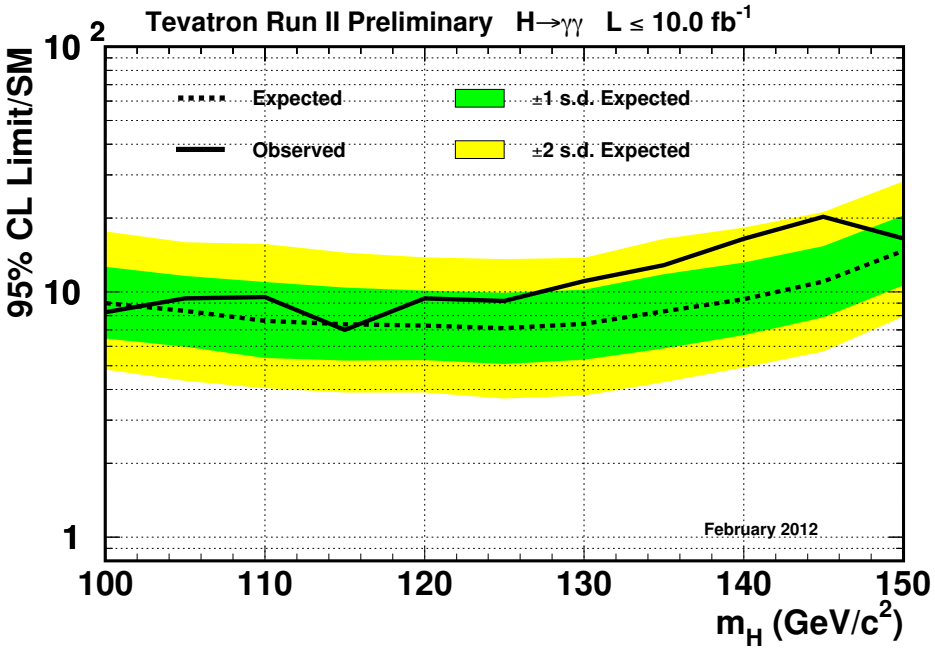


Figure 11.7: 95% upper credibility level limits on cross sections times branching fraction relative to the SM prediction, for combined results from the CDF and D0 $H \to \gamma\gamma$ searches.

Chapter 12

Conclusions

We have presented a search for the Standard Model (SM) Higgs boson decaying into a pair of photons ($H \to \gamma\gamma$). The diphoton data are produced in $p\bar{p}$ collisions at the Fermilab Tevatron with a center of mass energy of 1.96 TeV and collected by the CDF Run II detector. The results are based on searching for a SM $H \to \gamma\gamma$ resonance in the diphoton mass spectrum using data corresponding to an integrated luminosity of 10 fb^{-1}.

The previous CDF analysis in this channel searched for a beyond-the-Standard-Model (BSM) Higgs boson using central and plug photons reconstructed from the standard CDF photon identification [100]. The results presented in this thesis also use central and plug photons, and apply a modified event selection in order to optimize the search for the SM scenario. We furthermore improve upon the previous analysis by recovering central photons that have converted to an e^+e^- pair and implementing an improved technique for identifying central photons with the use of a neutral network (NN). Many aspects of this measurement are original contributions by the author, particularly developing an algorithm for reconstructing central conversion photons. Other contributions include updating the analysis to use the central NN selection and developing a method for modeling the Z boson background for diphoton categories that contain a plug photon. Though not described in this dissertation, the BSM $H \to \gamma\gamma$ analysis was updated at CDF [101] alongside this analysis, and was found

to benefit greatly from these contributions. The techniques described in this thesis have been published using a smaller set of data, applied to both the SM and BSM analyses [34].

For the full Run II data collected at CDF, we find no evidence for a resonance in the diphoton mass spectrum. We then set 95% C.L. upper limits on the SM Higgs boson cross section multiplied by the $H \to \gamma\gamma$ branching ratio. Compared to using the standard central photon selection, we find an improvement in limit sensitivity of 9% with the implementation of the NN selection. The gain obtained from recovering central conversions is about 13%. Using central, plug, and conversion photons, we find an observed (expected) limit of 12.2 (10.8) times the SM theory prediction for a Higgs boson mass of 125 GeV/c^2. The observed (expected) limits range between 8.2 and 22.6 (10.6 and 23.2) times the SM theory prediction for $100 < m_H < 150$ GeV/c^2.

A search for the SM Higgs boson in the diphoton channel was not an analysis initially predicted to have sufficient sensitivity at the Tevatron to be worth pursuing. This analysis, however, has produced limits considerably better than expected. The results presented in this thesis are combined with the results from other analyses and are included in the final CDF and Tevatron combined limits on the SM Higgs boson cross section. The greatest contribution to these limits from the diphoton search channel is obtained for $110 < m_H < 130$ GeV/c^2, with the CDF and D0 combined $H \to \gamma\gamma$ searches reaching a sensitivity in this region of about $7 \times$ SM. These are valuable results as the LHC continues this year to search for the Higgs boson in this channel.

Appendix A

Central Photon ID Efficiency Formula

In this appendix, a derivation is provided of Equation 4.5, based on References [102] and [103]. This formula was used to calculate the efficiency of the central photon ID selection (see Section 4.3).

Figure A.1 shows the categorization of events in which both electron legs from a Z boson decay pass the loose (L) central photon-like selection of Table 4.3. The initial number of such events is denoted as N_{LL}. In the tree, the symbol P (F) indicates passing (failing) a particular cut. We then wish to obtain the cumulative efficiency for an electron passing the selection at least up through the i$^{\text{th}}$ cut. We also obtain the efficiency formula for an electron passing all of the tight cuts (T) in Table 4.3. The efficiencies up through the i$^{\text{th}}$ cut and tight cuts are given by ϵ_i and ϵ_T, respectively. The ratio ϵ_T/ϵ_i is referred to as ϵ'_T.

The events of interest to the efficiency study are those contained within the small boxes in the last row of the diagram. These are events for which at least one electron leg passes photon-like cuts, denoted here as N_{TL}. The category represented by bold font indicates the number of N_{TL} events for which the second electron leg passes cuts up through at least the i$^{\text{th}}$ cut, denoted here as N_{Ti}. The number of these events for which both electrons pass the tight photon-like ID cuts are denoted as N_{TT}

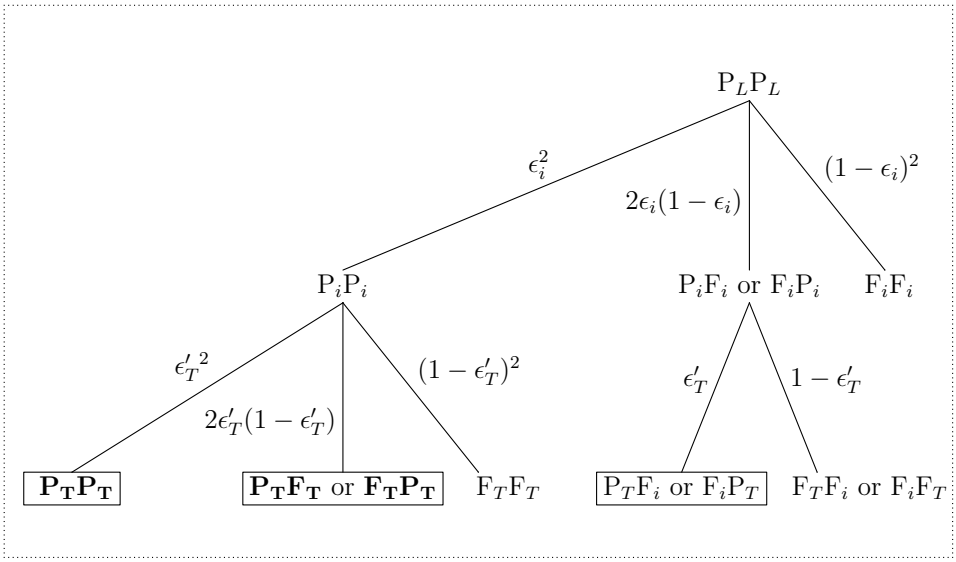


Figure A.1: The categorization of events based on ϵ_i and ϵ_T, relative to events that contain two loose central electrons.

(the lower-left box in the diagram). From the figure we then develop the following equations:

$$N_{TT} = N_{LL}\ \epsilon_i^2\ \epsilon_T^{'\,2} = N_{LL}\ \epsilon_T^2 \tag{A.1}$$

$$\begin{aligned} N_{Ti} &= N_{TT} + 2N_{LL}\ \epsilon_i^2\ \epsilon_T'(1-\epsilon_T') \\ &= N_{TT} + 2N_{LL}\ \epsilon_i\ \epsilon_T(1-\frac{\epsilon_T}{\epsilon_i}) \end{aligned} \tag{A.2}$$

$$\begin{aligned} N_{TL} &= N_{Ti} + 2N_{LL}\epsilon_i(1-\epsilon_i)\epsilon_T' \\ &= N_{Ti} + 2N_{LL}(1-\epsilon_i)\epsilon_T \end{aligned} \tag{A.3}$$

Solving for ϵ_i, we find that the cumulative efficiency for the i$^{\text{th}}$ cut of the central photon-like selection is:

$$\epsilon_i = \frac{N_{Ti} + N_{TT}}{N_{TL} + N_{TT}}. \tag{A.4}$$

This is Equation 4.5 of Section 4.3. We can also solve for ϵ_T, the cumulative efficiency for an electron passing the tight photon-like selection:

$$\epsilon_T = \frac{2N_{TT}}{N_{TL} + N_{TT}}. \tag{A.5}$$

With a modification made to the definition of a "loose" and "tight" electron, this is the formula used to calculate the trigger efficiency described in Section 9.5 (Equation 9.6).

Appendix B

Sep and Radius of Conversion Calculation

As discussed in Section 3.4, if the signed impact parameter d_0 and the track charge $q = \pm 1$ are of the same sign, then the helix curve does not contain the z axis and the radial distance from the z axis to the helix center is $\rho = r + |d_0|$ where $r = \frac{1}{2|c_0|}$. Figure B.1 shows an example of this. If d_0 and q are of opposite sign, then the helix curve contains the z axis and $\rho = r - |d_0|$. Figure 3.15 showed an example of this. We can write a general equation describing both of these scenarios from

$$\rho = r + qd_0. \tag{B.1}$$

The (x,y) center of the track helix is next calculated from $x = \rho\cos\beta$ and $y = \rho\sin\beta$ where β is the angle between the x axis and the line formed from the origin to the helix center. Figure B.1 shows this angle for a positively-charged track where it is seen that $\beta = \phi_0 + \frac{\pi}{2}$ with $0 < \phi_0 < 2\pi$. For $q = +1$, the (x, y) position can then be written as $x = -\rho\sin\phi_0$ and $y = +\rho\cos\phi_0$. This position is similarly obtained for negatively-charged tracks, except now $\beta = \phi_0 - \frac{\pi}{2}$ and $x = +\rho\sin\phi_0$ and $y = -\rho\cos\phi_0$. The (x, y) center for any track is written more compactly as

$$x = -q\rho\sin\phi_0, \qquad y = q\rho\cos\phi_0. \tag{B.2}$$

Equations (B.1) and (B.2), respectively, provide the radius and (x, y) center of a general helix track in the CDF detector [50].

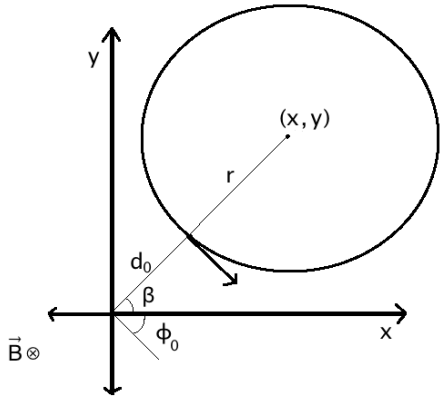


Figure B.1: The helical track for a positively-charged particle in the $r-\phi$ plane. The CDF magnetic field points along the negative z axis. For $q > 0$, the angle $\beta = \phi_0 + \frac{\pi}{2}$ with $0 < \phi_0 < 2\pi$. The figure is not drawn to scale relative to the detector geometry, but rather demonstrates the calculations performed based on track parameters.

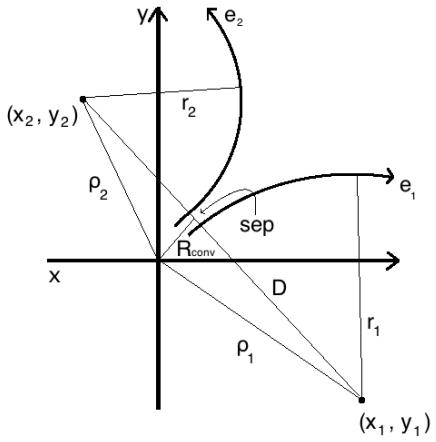


Figure B.2: A diagram is shown for two tracks resulting from a conversion. The sep variable is the distance between the two tracks in the $r-\phi$ plane at the location where they are parallel. The radial distance from the origin to midpoint of the sep segment is defined as the radius of the conversion, $R_{\rm conv}$. The figure is not drawn to scale relative to the detector geometry, but rather demonstrates the calculations performed based on track parameters.

Figure B.2 shows a diagram for two oppositely signed tracks produced from a common point at the radius of the conversion $R_{\rm conv}$. The distance between the two helix centers is

$$D = \sqrt{\Delta x^2 + \Delta y^2} \quad \text{with} \quad \Delta x = x_2 - x_1 \quad \text{and} \quad \Delta y = y_2 - y_1. \tag{B.3}$$

The $r - \phi$ distance between the two oppositely signed tracks where they are parallel is then calculated as

$$\text{sep} = D - r_1 - r_2. \tag{B.4}$$

For same signed tracks, $\text{sep} = r_> - D - r_<$ where $r_>$ ($r_<$) is the larger (smaller) of r_1 and r_2.

The calculation of R_{conv} is based on forming a right triangle from the two helix centers and then using similar triangles. This is shown in Figure B.3 where the conversion origin is

$$x_{\text{conv}} = x_1 + (r_1 + \frac{1}{2}\text{sep})\frac{\Delta x}{D}, \quad y_{\text{conv}} = y_1 + (r_1 + \frac{1}{2}\text{sep})\frac{\Delta y}{D}. \tag{B.5}$$

The radius of the conversion is then

$$R_{\text{conv}} = \sqrt{x_{\text{conv}}^2 + y_{\text{conv}}^2} \tag{B.6}$$

where R_{conv} is given a positive (negative) sign if the two tracks point away from (toward) the origin.

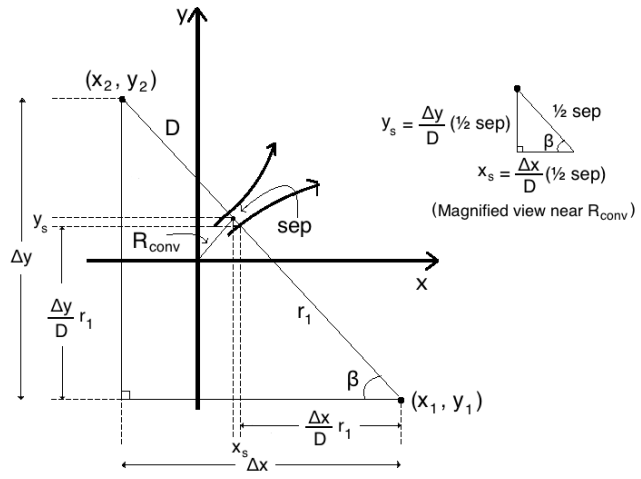


Figure B.3: Figure B.2 is redrawn focusing on the geometry used for calculating R_{conv}. The subfigure at right is a magnified view of the region near the conversion origin. The figure is not drawn to scale relative to the detector geometry, but rather demonstrates the calculations performed based on track parameters.

Appendix C

Effect of Reweighting GF to HqT

In this appendix, we show the comparison of the LO PYTHIA prediction of the Higgs boson p_T spectrum for GF events to the prediction at NNLL + NLO accuracy obtained from the HqT program. In particular, we compare the diphoton efficiency and $m_{\gamma\gamma}$ shape differences from PYTHIA to those obtained from PYTHIA after GF signal events are reweighted to the HqT prediction. Figure C.1 shows the Higgs p_T spectrum from PYTHIA compared to that predicted by HqT, where the right plot is the ratio of the two. Weights are obtained from the plot on the right as a function of the Higgs p_T in 2 GeV/c intervals. The PYTHIA sample is then weighted on an event-by-event basis to compare the resulting diphoton efficiency and $m_{\gamma\gamma}$ shapes to those described in Chapter 8.

As an example, Figure C.2 shows a comparison of the $m_{\gamma\gamma}$ shape for the CC category before and after reweighting. The effect is negligible for this category as well as other categories used in the analysis (CP, C′C, C′P). Moreover, the diphoton efficiencies are identical to the nominal efficiencies obtained in Chapter 8. For an inclusive $H \to \gamma\gamma$ analysis, which does not distinguish events based on the number of jets or on $p_T^{\gamma\gamma}$, the effect of reweighting events is negligible and the final results presented do not include these effects.

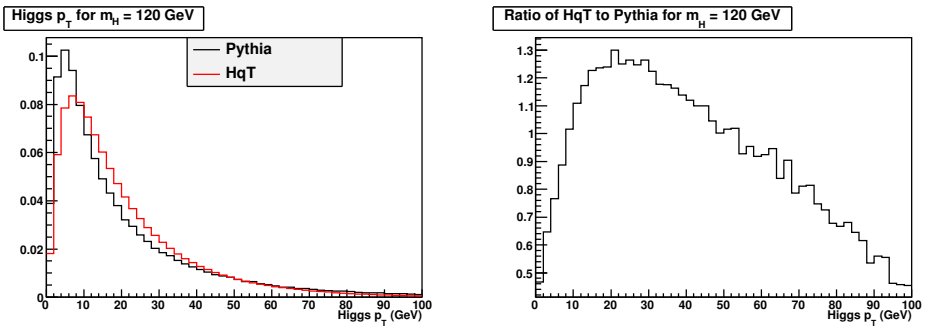


Figure C.1: Left: The Higgs boson p_T spectrum from HqT compared to that from PYTHIA, for a mass of 120 GeV/c^2. The integral of each histogram is normalized to 1.0. Right: The ratio of the two histograms on the left; the spectrum from HqT is divided by the spectrum from PYTHIA.

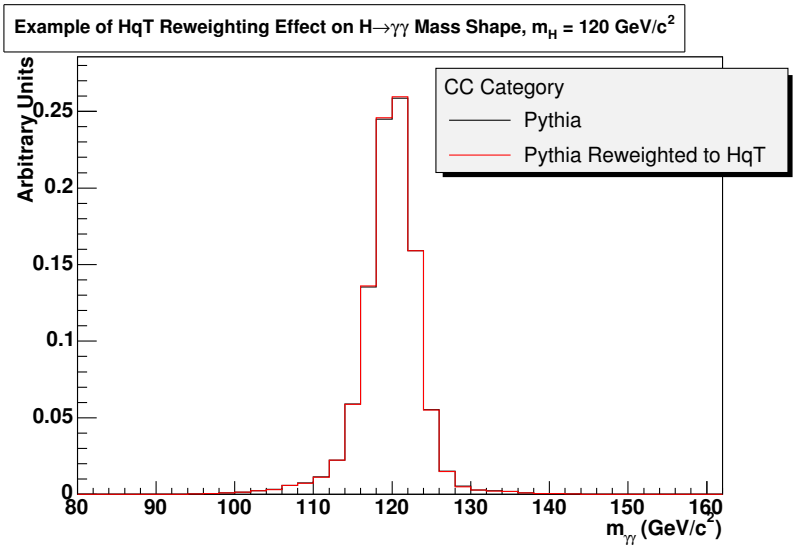


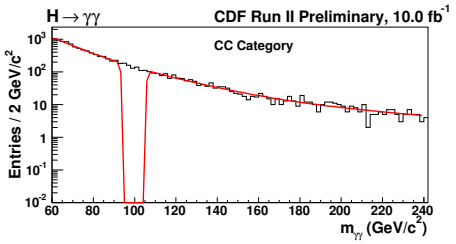
Figure C.2: Effect of HqT reweighting on the $m_{\gamma\gamma}$ shape (GF signal events). An example is shown for a mass of 120 GeV/c^2 in the CC category.

Appendix D

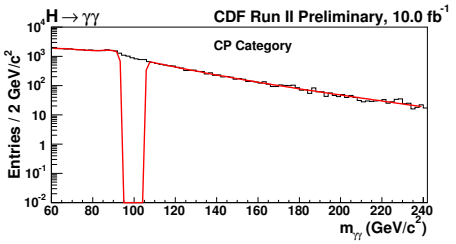
Background Fits for Each Mass Hypothesis

In this section, the background fit results are provided for each Higgs boson mass hypothesis and diphoton category, as described in Section 7.2.4. Each page shows the CC category results in the upper left, the CP category results in the upper right, the C′C category results in the lower left, and the C′P category results in the lower right. There are eleven pages, starting with the mass of 100 GeV/c^2 and going up to a mass of 150 GeV/c^2. The results are provided in this format such that scrolling through pages provides an indication of how the fit for a single category slightly changes for different mass hypotheses due to the different window excluded from the fit.

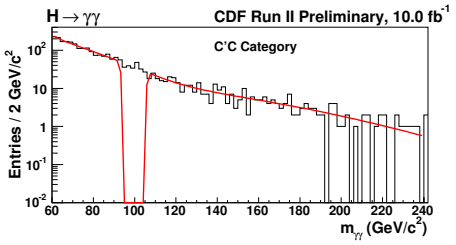
D.1 $m_H = 100\ \mathrm{GeV/c^2}$



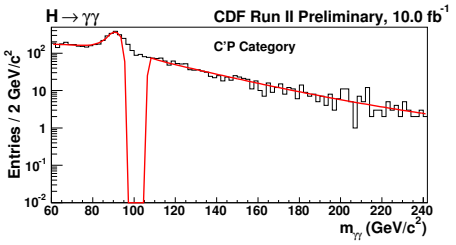
(a) CC Category



(b) CP Category



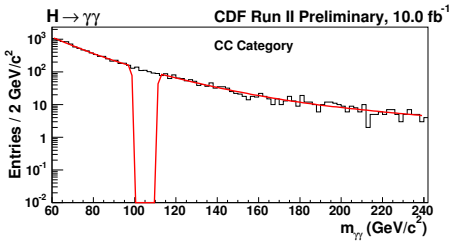
(c) C′C Category



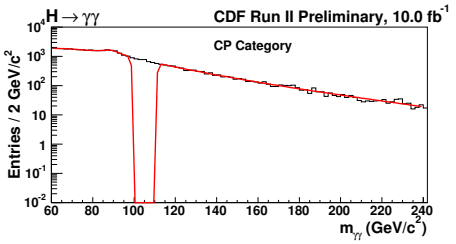
(d) C′P Category

Figure D.1: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 100 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

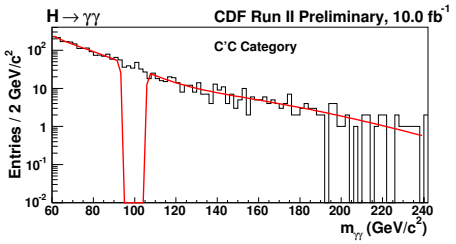
D.2 $m_H = 105$ GeV/c^2



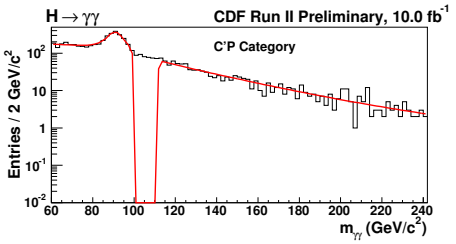
(a) CC Category



(b) CP Category



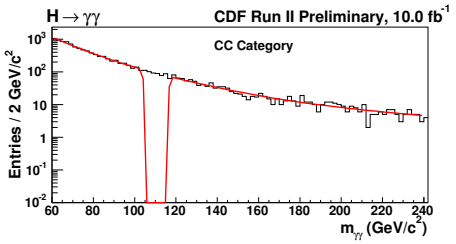
(c) C′C Category



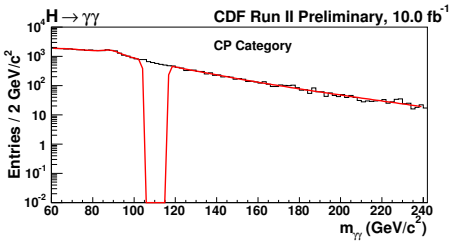
(d) C′P Category

Figure D.2: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 105 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

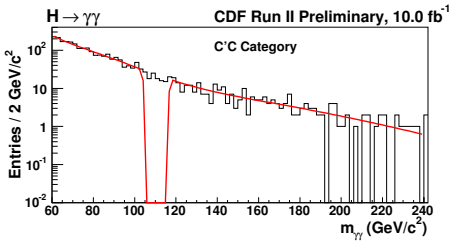
D.3 $m_H = 110$ GeV/c^2



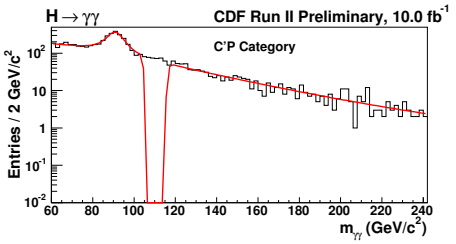
(a) CC Category



(b) CP Category



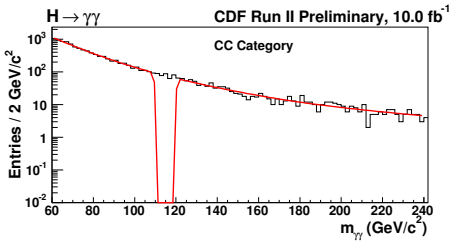
(c) C′C Category



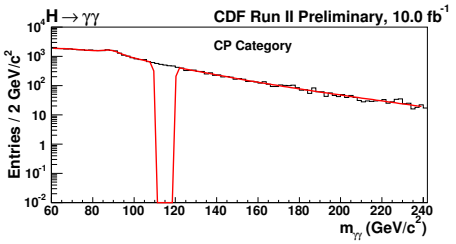
(d) C′P Category

Figure D.3: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 110 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

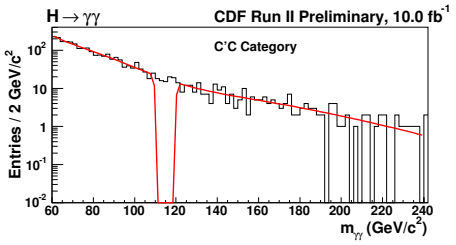
D.4 $m_H = 115$ GeV/c^2



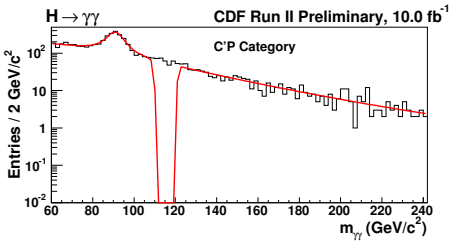
(a) CC Category



(b) CP Category



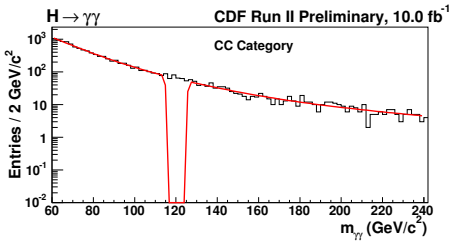
(c) C′C Category



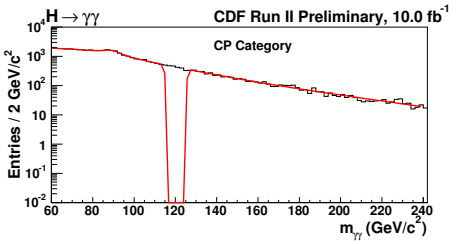
(d) C′P Category

Figure D.4: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 115 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

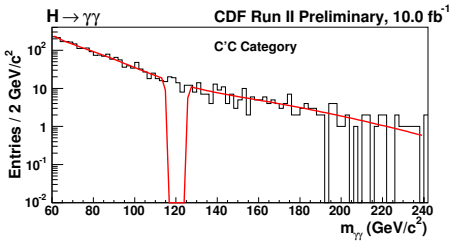
D.5 $m_H = 120$ GeV/c^2



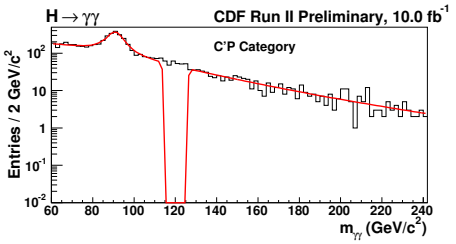
(a) CC Category



(b) CP Category



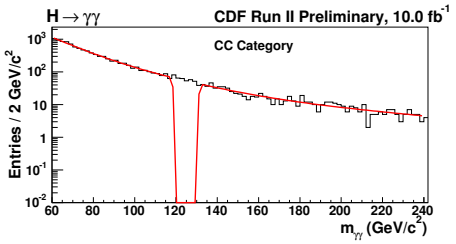
(c) C′C Category



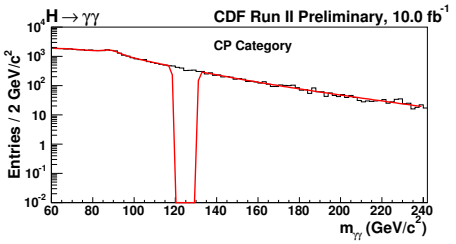
(d) C′P Category

Figure D.5: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 120 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

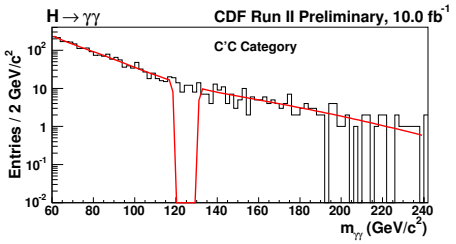
D.6 $m_H = 125\ \mathrm{GeV/c^2}$



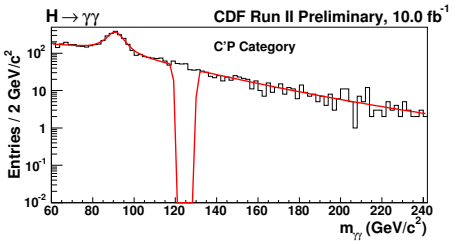
(a) CC Category



(b) CP Category



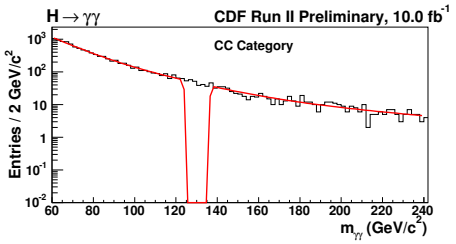
(c) C′C Category



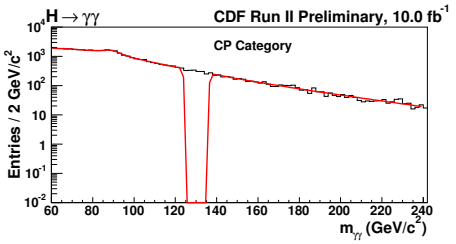
(d) C′P Category

Figure D.6: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 125 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

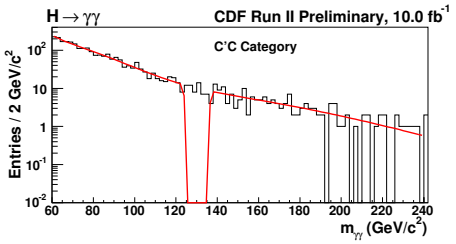
D.7 $m_H = 130\ \mathrm{GeV/c^2}$



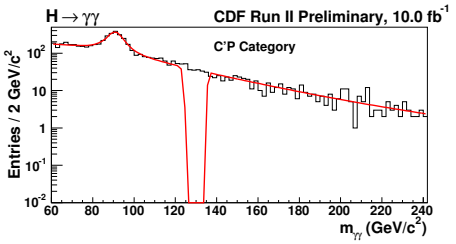
(a) CC Category



(b) CP Category



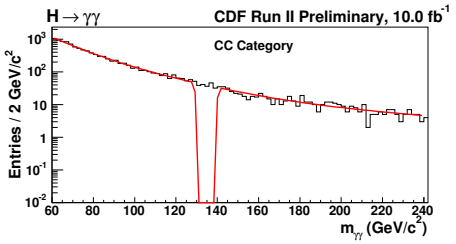
(c) C′C Category



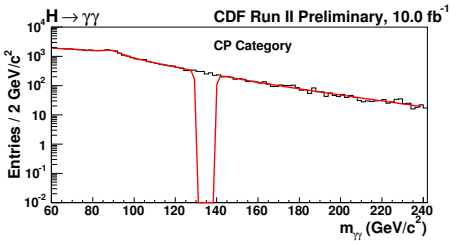
(d) C′P Category

Figure D.7: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 130 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

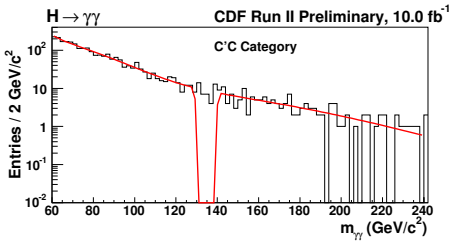
D.8 $m_H = 135\ \mathrm{GeV/c^2}$



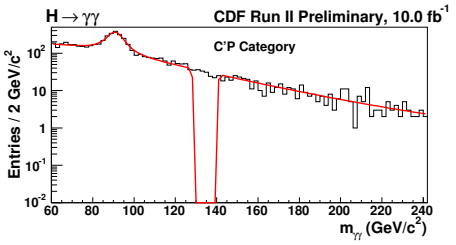
(a) CC Category



(b) CP Category



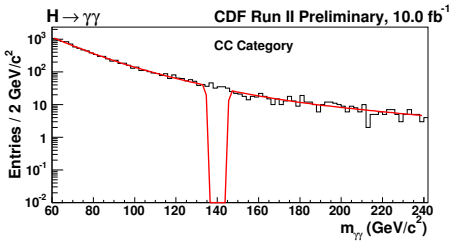
(c) C′C Category



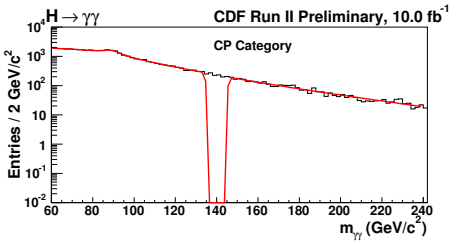
(d) C′P Category

Figure D.8: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 135 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

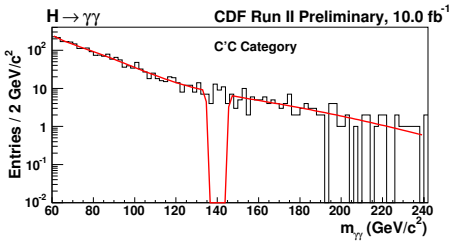
D.9 $m_H = 140$ GeV/c^2



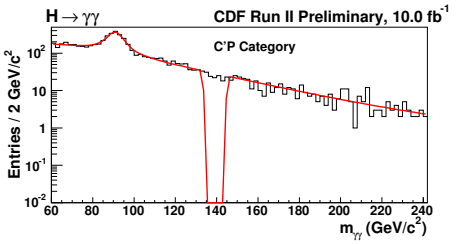
(a) CC Category



(b) CP Category



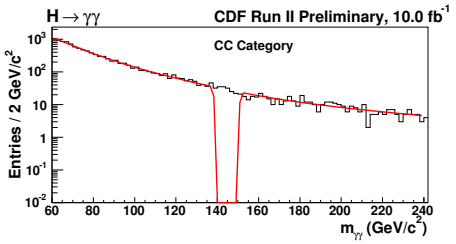
(c) C′C Category



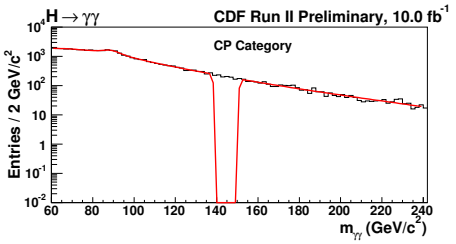
(d) C′P Category

Figure D.9: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 140 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

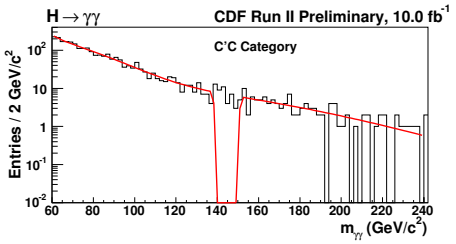
D.10 $m_H = 145\ \mathrm{GeV/c^2}$



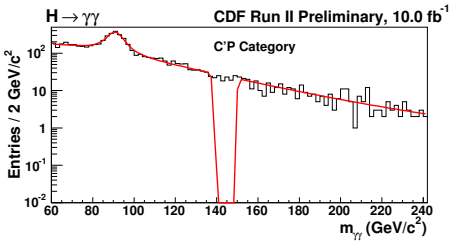
(a) CC Category



(b) CP Category



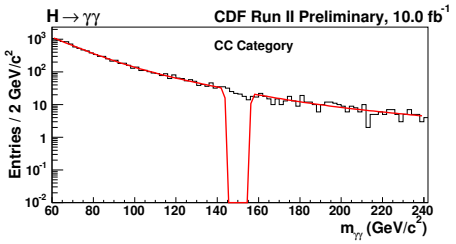
(c) C′C Category



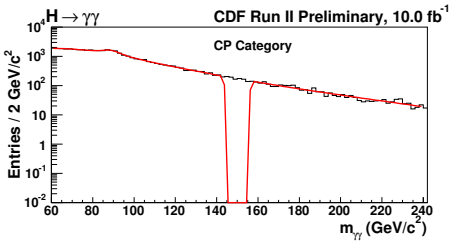
(d) C′P Category

Figure D.10: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 145 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

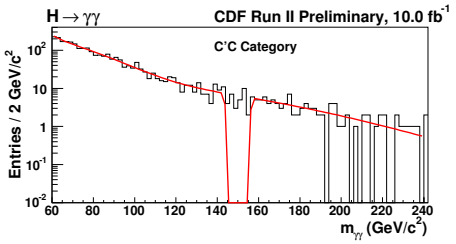
D.11 $m_H = 150$ GeV/c^2



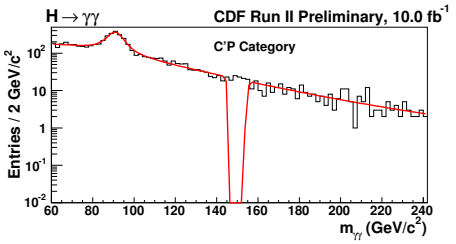
(a) CC Category



(b) CP Category



(c) C′C Category



(d) C′P Category

Figure D.11: Each subfigure shows a fit (red line) made to the sideband region of the data for each diphoton category using Equation (7.1) for the CC and C′C categories and the sum of Equations (7.1) and (7.2) for the CP and C′P categories. The two vertical red lines in each figure indicate the signal window excluded from the fit, here for a Higgs boson mass of 150 GeV/c^2. The bottom plot in each subfigure is obtained from first interpolating the fit into the signal region in order to obtain the background prediction in the signal region. The residual shown is then the resulting data yield minus the background yield, divided by the statistical error from the background yield.

Appendix E

The CDF Collaboration

The Collider Detector at Fermilab (CDF) is an experiment located at Fermilab outside of Chicago, IL. The data used for the results presented in this thesis were collected by the CDF II detector. All members of the collaboration are included as authors in publications because of their contribution to data collection, developing analysis techniques and software, and review of publication materials. There are roughly 475 scientists that are members of this collaboration, from 60 institutions around the world. The full author list is provided on the next three pages.

T. Aaltonen,[21] B. Álvarez González[z],[9] S. Amerio,[40] D. Amidei,[32] A. Anastassov[x],[15] A. Annovi,[17] J. Antos,[12] G. Apollinari,[15] J.A. Appel,[15] T. Arisawa,[54] A. Artikov,[13] J. Asaadi,[49] W. Ashmanskas,[15] B. Auerbach,[57] A. Aurisano,[49] F. Azfar,[39] W. Badgett,[15] T. Bae,[25] A. Barbaro-Galtieri,[26] V.E. Barnes,[44] B.A. Barnett,[23] P. Barria[hh],[42] P. Bartos,[12] M. Bauce[ff],[40] F. Bedeschi,[42] S. Behari,[23] G. Bellettini[gg],[42] J. Bellinger,[56] D. Benjamin,[14] A. Beretvas,[15] A. Bhatti,[46] D. Bisello[ff],[40] I. Bizjak,[28] K.R. Bland,[5] B. Blumenfeld,[23] A. Bocci,[14] A. Bodek,[45] D. Bortoletto,[44] J. Boudreau,[43] A. Boveia,[11] L. Brigliadori[ee],[6] C. Bromberg,[33] E. Brucken,[21] J. Budagov,[13] H.S. Budd,[45] K. Burkett,[15] G. Busetto[ff],[40] P. Bussey,[19] A. Buzatu,[31] A. Calamba,[10] C. Calancha,[29] S. Camarda,[4] M. Campanelli,[28] M. Campbell,[32] F. Canelli,[11,15] B. Carls,[22] D. Carlsmith,[56] R. Carosi,[42] S. Carrillo[m],[16] S. Carron,[15] B. Casal[k],[9] M. Casarsa,[50] A. Castro[ee],[6] P. Catastini,[20] D. Cauz,[50] V. Cavaliere,[22] M. Cavalli-Sforza,[4] A. Cerri[f],[26] L. Cerrito[s],[28] Y.C. Chen,[1] M. Chertok,[7] G. Chiarelli,[42] G. Chlachidze,[15] F. Chlebana,[15] K. Cho,[25] D. Chokheli,[13] W.H. Chung,[56] Y.S. Chung,[45] M.A. Ciocci[hh],[42] A. Clark,[18] C. Clarke,[55] G. Compostella[ff],[40] M.E. Convery,[15] J. Conway,[7] M.Corbo,[15] M. Cordelli,[17] C.A. Cox,[7] D.J. Cox,[7] F. Crescioli[gg],[42] J. Cuevas[z],[9] R. Culbertson,[15] D. Dagenhart,[15] N. d'Ascenzo[w],[15] M. Datta,[15] P. de Barbaro,[45] M. Dell'Orso[gg],[42] L. Demortier,[46] M. Deninno,[6] F. Devoto,[21] M. d'Errico[ff],[40] A. Di Canto[gg],[42] B. Di Ruzza,[15] J.R. Dittmann,[5] M. D'Onofrio,[27] S. Donati[gg],[42] P. Dong,[15] M. Dorigo,[50] T. Dorigo,[40] K. Ebina,[54] A. Elagin,[49] A. Eppig,[32] R. Erbacher,[7] S. Errede,[22] N. Ershaidat[dd],[15] R. Eusebi,[49] S. Farrington,[39] M. Feindt,[24] J.P. Fernandez,[29] R. Field,[16] G. Flanagan[u],[15] R. Forrest,[7] M.J. Frank,[5] M. Franklin,[20] J.C. Freeman,[15] Y. Funakoshi,[54] I. Furic,[16] M. Gallinaro,[46] J.E. Garcia,[18] A.F. Garfinkel,[44] P. Garosi[hh],[42] H. Gerberich,[22] E. Gerchtein,[15] S. Giagu,[47] V. Giakoumopoulou,[3] P. Giannetti,[42] K. Gibson,[43] C.M. Ginsburg,[15] N. Giokaris,[3] P. Giromini,[17] G. Giurgiu,[23] V. Glagolev,[13] D. Glenzinski,[15] M. Gold,[35] D. Goldin,[49] N. Goldschmidt,[16] A. Golossanov,[15] G. Gomez,[9] G. Gomez-Ceballos,[30] M. Goncharov,[30] O. González,[29] I. Gorelov,[35] A.T. Goshaw,[14] K. Goulianos,[46] S. Grinstein,[4] C. Grosso-Pilcher,[11] R.C. Group[53],[15] J. Guimaraes da Costa,[20] S.R. Hahn,[15] E. Halkiadakis,[48] A. Hamaguchi,[38] J.Y. Han,[45] F. Happacher,[17] K. Hara,[51] D. Hare,[48] M. Hare,[52] R.F. Harr,[55] K. Hatakeyama,[5] C. Hays,[39] M. Heck,[24] J. Heinrich,[41] M. Herndon,[56] S. Hewamanage,[5] A. Hocker,[15] W. Hopkins[g],[15] D. Horn,[24] S. Hou,[1] R.E. Hughes,[36] M. Hurwitz,[11] U. Husemann,[57] N. Hussain,[31] M. Hussein,[33] J. Huston,[33] G. Introzzi,[42] M. Iori[jj],[47] A. Ivanov[p],[7] E. James,[15] D. Jang,[10] B. Jayatilaka,[14] E.J. Jeon,[25] S. Jindariani,[15] M. Jones,[44] K.K. Joo,[25] S.Y. Jun,[10] T.R. Junk,[15] T. Kamon[25],[49] P.E. Karchin,[55] A. Kasmi,[5] Y. Kato[o],[38] W. Ketchum,[11] J. Keung,[41] V. Khotilovich,[49] B. Kilminster,[15] D.H. Kim,[25] H.S. Kim,[25] J.E. Kim,[25] M.J. Kim,[17] S.B. Kim,[25] S.H. Kim,[51] Y.K. Kim,[11] Y.J. Kim,[25] N. Kimura,[54] M. Kirby,[15] S. Klimenko,[16] K. Knoepfel,[15] K. Kondo[*],[54] D.J. Kong,[25] J. Konigsberg,[16] A.V. Kotwal,[14] M. Kreps,[24] J. Kroll,[41] D. Krop,[11] M. Kruse,[14] V. Krutelyov[c],[49] T. Kuhr,[24] M. Kurata,[51] S. Kwang,[11] A.T. Laasanen,[44] S. Lami,[42] S. Lammel,[15] M. Lancaster,[28] R.L. Lander,[7] K. Lannon[y],[36] A. Lath,[48] G. Latino[hh],[42] T. LeCompte,[2] E. Lee,[49] H.S. Lee[q],[11] J.S. Lee,[25] S.W. Lee[bb],[49] S. Leo[gg],[42] S. Leone,[42] J.D. Lewis,[15] A. Limosani[t],[14] C.-J. Lin,[26] M. Lindgren,[15] E. Lipeles,[41] A. Lister,[18] D.O. Litvintsev,[15] C. Liu,[43] H. Liu,[53] Q. Liu,[44] T. Liu,[15] S. Lockwitz,[57] A. Loginov,[57] D. Lucchesi[ff],[40] J. Lueck,[24] P. Lujan,[26] P. Lukens,[15] G. Lungu,[46] J. Lys,[26] R. Lysak[e],[12] R. Madrak,[15] K. Maeshima,[15] P. Maestro[hh],[42] S. Malik,[46] G. Manca[a],[27] A. Manousakis-Katsikakis,[3] F. Margaroli,[47] C. Marino,[24] M. Martínez,[4] P. Mastrandrea,[47] K. Matera,[22] M.E. Mattson,[55] A. Mazzacane,[15] P. Mazzanti,[6] K.S. McFarland,[45] P. McIntyre,[49] R. McNulty[j],[27] A. Mehta,[27] P. Mehtala,[21] C. Mesropian,[46] T. Miao,[15] D. Mietlicki,[32] A. Mitra,[1] H. Miyake,[51] S. Moed,[15] N. Moggi,[6] M.N. Mondragon[m],[15] C.S. Moon,[25] R. Moore,[15] M.J. Morello[ii],[42] J. Morlock,[24] P. Movilla Fernandez,[15] A. Mukherjee,[15] Th. Muller,[24] P. Murat,[15] M. Mussini[ee],[6] J. Nachtman[n],[15] Y. Nagai,[51] J. Naganoma,[54] I. Nakano,[37] A. Napier,[52] J. Nett,[49] C. Neu,[53] M.S. Neubauer,[22] J. Nielsen[d],[26] L. Nodulman,[2] S.Y. Noh,[25] O. Norniella,[22] L. Oakes,[39] S.H. Oh,[14] Y.D. Oh,[25] I. Oksuzian,[53] T. Okusawa,[38] R. Orava,[21] L. Ortolan,[4] S. Pagan Griso[ff],[40] C. Pagliarone,[50] E. Palencia[f],[9] V. Papadimitriou,[15] A.A. Paramonov,[2] J. Patrick,[15] G. Pauletta[kk],[50] M. Paulini,[10] C. Paus,[30] D.E. Pellett,[7] A. Penzo,[50] T.J. Phillips,[14] G. Piacentino,[42] E. Pianori,[41] J. Pilot,[36] K. Pitts,[22] C. Plager,[8] L. Pondrom,[56] S. Poprocki[g],[15] K. Potamianos,[44] F. Prokoshin[cc],[13] A. Pranko,[26] F. Ptohos[h],[17] G. Punzi[gg],[42] A. Rahaman,[43] V. Ramakrishnan,[56] N. Ranjan,[44] I. Redondo,[29] P. Renton,[39] M. Rescigno,[47] T. Riddick,[28] F. Rimondi[ee],[6] L. Ristori[42],[15] A. Robson,[19] T. Rodrigo,[9] T. Rodriguez,[41] E. Rogers,[22] S. Rolli[i],[52] R. Roser,[15] F. Ruffini[hh],[42] A. Ruiz,[9] J. Russ,[10] V. Rusu,[15] A. Safonov,[49] W.K. Sakumoto,[45] Y. Sakurai,[54] L. Santi[kk],[50] K. Sato,[51] V. Saveliev[w],[15] A. Savoy-Navarro[aa],[15] P. Schlabach,[15] A. Schmidt,[24] E.E. Schmidt,[15] T. Schwarz,[15] L. Scodellaro,[9] A. Scribano[hh],[42] F. Scuri,[42] S. Seidel,[35] Y. Seiya,[38] A. Semenov,[13] F. Sforza[hh],[42] S.Z. Shalhout,[7] T. Shears,[27] P.F. Shepard,[43] M. Shimojima[v],[51] M. Shochet,[11] I. Shreyber-Tecker,[34] A. Simonenko,[13] P. Sinervo,[31] K. Sliwa,[52] J.R. Smith,[7] F.D. Snider,[15] A. Soha,[15] V. Sorin,[4] H. Song,[43] P. Squillacioti[hh],[42] M. Stancari,[15] R. St. Denis,[19] B. Stelzer,[31] O. Stelzer-Chilton,[31] D. Stentz[x],[15] J. Strologas,[35] G.L. Strycker,[32] Y. Sudo,[51] A. Sukhanov,[15]

I. Suslov,[13] K. Takemasa,[51] Y. Takeuchi,[51] J. Tang,[11] M. Tecchio,[32] P.K. Teng,[1] J. Thom[g],[15] J. Thome,[10] G.A. Thompson,[22] E. Thomson,[41] D. Toback,[49] S. Tokar,[12] K. Tollefson,[33] T. Tomura,[51] D. Tonelli,[15] S. Torre,[17] D. Torretta,[15] P. Totaro,[40] M. Trovato[ii],[42] F. Ukegawa,[51] S. Uozumi,[25] A. Varganov,[32] F. Vázquez[m],[16] G. Velev,[15] C. Vellidis,[15] M. Vidal,[44] I. Vila,[9] R. Vilar,[9] J. Vizán,[9] M. Vogel,[35] G. Volpi,[17] P. Wagner,[41] R.L. Wagner,[15] T. Wakisaka,[38] R. Wallny,[8] S.M. Wang,[1] A. Warburton,[31] D. Waters,[28] W.C. Wester III,[15] D. Whiteson[b],[41] A.B. Wicklund,[2] E. Wicklund,[15] S. Wilbur,[11] F. Wick,[24] H.H. Williams,[41] J.S. Wilson,[36] P. Wilson,[15] B.L. Winer,[36] P. Wittich[g],[15] S. Wolbers,[15] H. Wolfe,[36] T. Wright,[32] X. Wu,[18] Z. Wu,[5] K. Yamamoto,[38] D. Yamato,[38] T. Yang,[15] U.K. Yang[r],[11] Y.C. Yang,[25] W.-M. Yao,[26] G.P. Yeh,[15] K. Yi[n],[15] J. Yoh,[15] K. Yorita,[54] T. Yoshida[l],[38] G.B. Yu,[14] I. Yu,[25] S.S. Yu,[15] J.C. Yun,[15] A. Zanetti,[50] Y. Zeng,[14] C. Zhou,[14] and S. Zucchelli[ee6]

(CDF Collaboration[†])

[1] *Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China*
[2] *Argonne National Laboratory, Argonne, Illinois 60439, USA*
[3] *University of Athens, 157 71 Athens, Greece*
[4] *Institut de Fisica d'Altes Energies, ICREA, Universitat Autonoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain*
[5] *Baylor University, Waco, Texas 76798, USA*
[6] *Istituto Nazionale di Fisica Nucleare Bologna, [ee] University of Bologna, I-40127 Bologna, Italy*
[7] *University of California, Davis, Davis, California 95616, USA*
[8] *University of California, Los Angeles, Los Angeles, California 90024, USA*
[9] *Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain*
[10] *Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA*
[11] *Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA*
[12] *Comenius University, 842 48 Bratislava, Slovakia; Institute of Experimental Physics, 040 01 Kosice, Slovakia*
[13] *Joint Institute for Nuclear Research, RU-141980 Dubna, Russia*
[14] *Duke University, Durham, North Carolina 27708, USA*
[15] *Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
[16] *University of Florida, Gainesville, Florida 32611, USA*
[17] *Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy*
[18] *University of Geneva, CH-1211 Geneva 4, Switzerland*
[19] *Glasgow University, Glasgow G12 8QQ, United Kingdom*
[20] *Harvard University, Cambridge, Massachusetts 02138, USA*
[21] *Division of High Energy Physics, Department of Physics, University of Helsinki and Helsinki Institute of Physics, FIN-00014, Helsinki, Finland*
[22] *University of Illinois, Urbana, Illinois 61801, USA*
[23] *The Johns Hopkins University, Baltimore, Maryland 21218, USA*
[24] *Institut für Experimentelle Kernphysik, Karlsruhe Institute of Technology, D-76131 Karlsruhe, Germany*
[25] *Center for High Energy Physics: Kyungpook National University, Daegu 702-701, Korea; Seoul National University, Seoul 151-742, Korea; Sungkyunkwan University, Suwon 440-746, Korea; Korea Institute of Science and Technology Information, Daejeon 305-806, Korea; Chonnam National University, Gwangju 500-757, Korea; Chonbuk National University, Jeonju 561-756, Korea*
[26] *Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[27] *University of Liverpool, Liverpool L69 7ZE, United Kingdom*
[28] *University College London, London WC1E 6BT, United Kingdom*
[29] *Centro de Investigaciones Energeticas Medioambientales y Tecnologicas, E-28040 Madrid, Spain*
[30] *Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[31] *Institute of Particle Physics: McGill University, Montréal, Québec, Canada H3A 2T8; Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6; University of Toronto, Toronto, Ontario, Canada M5S 1A7; and TRIUMF, Vancouver, British Columbia, Canada V6T 2A3*
[32] *University of Michigan, Ann Arbor, Michigan 48109, USA*
[33] *Michigan State University, East Lansing, Michigan 48824, USA*
[34] *Institution for Theoretical and Experimental Physics, ITEP, Moscow 117259, Russia*
[35] *University of New Mexico, Albuquerque, New Mexico 87131, USA*
[36] *The Ohio State University, Columbus, Ohio 43210, USA*
[37] *Okayama University, Okayama 700-8530, Japan*
[38] *Osaka City University, Osaka 588, Japan*
[39] *University of Oxford, Oxford OX1 3RH, United Kingdom*
[40] *Istituto Nazionale di Fisica Nucleare, Sezione di Padova-Trento, [ff] University of Padova, I-35131 Padova, Italy*
[41] *University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*

[42]*Istituto Nazionale di Fisica Nucleare Pisa,* [gg]*University of Pisa,*
[hh]*University of Siena and* [ii]*Scuola Normale Superiore, I-56127 Pisa, Italy*
[43]*University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*
[44]*Purdue University, West Lafayette, Indiana 47907, USA*
[45]*University of Rochester, Rochester, New York 14627, USA*
[46]*The Rockefeller University, New York, New York 10065, USA*
[47]*Istituto Nazionale di Fisica Nucleare, Sezione di Roma 1,*
[jj]*Sapienza Università di Roma, I-00185 Roma, Italy*
[48]*Rutgers University, Piscataway, New Jersey 08855, USA*
[49]*Texas A&M University, College Station, Texas 77843, USA*
[50]*Istituto Nazionale di Fisica Nucleare Trieste/Udine,*
I-34100 Trieste, [kk]*University of Udine, I-33100 Udine, Italy*
[51]*University of Tsukuba, Tsukuba, Ibaraki 305, Japan*
[52]*Tufts University, Medford, Massachusetts 02155, USA*
[53]*University of Virginia, Charlottesville, Virginia 22906, USA*
[54]*Waseda University, Tokyo 169, Japan*
[55]*Wayne State University, Detroit, Michigan 48201, USA*
[56]*University of Wisconsin, Madison, Wisconsin 53706, USA*
[57]*Yale University, New Haven, Connecticut 06520, USA*

PACS numbers:

*Deceased
†With visitors from [a]Istituto Nazionale di Fisica Nucleare, Sezione di Cagliari, 09042 Monserrato (Cagliari), Italy, [b]University of CA Irvine, Irvine, CA 92697, USA, [c]University of CA Santa Barbara, Santa Barbara, CA 93106, USA, [d]University of CA Santa Cruz, Santa Cruz, CA 95064, USA, [e]Institute of Physics, Academy of Sciences of the Czech Republic, Czech Republic, [f]CERN, CH-1211 Geneva, Switzerland, [g]Cornell University, Ithaca, NY 14853, USA, [h]University of Cyprus, Nicosia CY-1678, Cyprus, [i]Office of Science, U.S. Department of Energy, Washington, DC 20585, USA, [j]University College Dublin, Dublin 4, Ireland, [k]ETH, 8092 Zurich, Switzerland, [l]University of Fukui, Fukui City, Fukui Prefecture, Japan 910-0017, [m]Universidad Iberoamericana, Mexico D.F., Mexico, [n]University of Iowa, Iowa City, IA 52242, USA, [o]Kinki University, Higashi-Osaka City, Japan 577-8502, [p]Kansas State University, Manhattan, KS 66506, USA, [q]Korea University, Seoul, 136-713, Korea, [r]University of Manchester, Manchester M13 9PL, United Kingdom, [s]Queen Mary, University of London, London, E1 4NS, United Kingdom, [t]University of Melbourne, Victoria 3010, Australia, [u]Muons, Inc., Batavia, IL 60510, USA, [v]Nagasaki Institute of Applied Science, Nagasaki, Japan, [w]National Research Nuclear University, Moscow, Russia, [x]Northwestern University, Evanston, IL 60208, USA, [y]University of Notre Dame, Notre Dame, IN 46556, USA, [z]Universidad de Oviedo, E-33007 Oviedo, Spain, [aa]CNRS-IN2P3, Paris, F-75205 France, [bb]Texas Tech University, Lubbock, TX 79609, USA, [cc]Universidad Tecnica Federico Santa Maria, 110v Valparaiso, Chile, [dd]Yarmouk University, Irbid 211-63, Jordan,

Appendix F

Biography

Karen Renee Bland was born on September 15th, 1981 in Raleigh, NC. The first six years of her life were spent in Cary, NC with her loving parents, Bill and Diane Bland, and her older sister, Kristy. Her family then moved to the suburbs of Washington, DC to live in Woodbridge, VA. She graduated from Woodbridge Senior High School in 1999. That year, she began her undergraduate studies at James Madison University (JMU), in Harrisonburg, VA. She majored in physics with minors in math, astronomy, and Spanish. During her time as an undergraduate, she spent a semester studying Spanish in Salamanca, Spain. She also spent each summer of college performing research. These research experiences led to studies in materials science, experimental neutrino physics (with the AMANDA collaboration), and radio astronomy. After graduating from JMU in December 2003, she gained teaching experience in private schools in VA. She spent a semester teaching algebra and pre-calculus to 10th and 11th grade students. She then spent two years teaching science, math, and Spanish to 4th–8th grade students at Redeemer Classical School in Harrisonburg, VA. She began her graduate studies at Baylor University in Waco, TX in the fall of 2006.

Bibliography

[1] Excerpt from http://en.wikipedia.org/wiki/File: Standard_Model_of_Elementary_Particles.svg, Retrieved May 2012.

[2] Fermilab Inquiring Minds, http://www.fnal.gov/pub/inquiring/physics/index.html.

[3] K. Nakamura *et al.* (Particle Data Group), "Review of particle physics," J. Phys. G **37**, 075021 (2010), 2011 partial update for 2012 found at PDG website http://pdg.lbl.gov/.

[4] PBS NOVA, http://www.pbs.org/wgbh/nova/physics/elegant-universe.html.

[5] Excerpt from http://en.wikipedia.org/wiki/File:Elementary_particle_interactions.svg, Retrieved May 2012.

[6] L. Lederman and D. Teresi, *The God Particle: If the Universe is the Answer, What is the Question?*, (First Mariner Books, New York, 2006).

[7] D. Griffiths, *Introduction to Elementary Particles*, 2nd ed. (Wiley-VCH, Weinheim, Germany, 2008).

[8] H. Weyl, "A new extension of the theory of relativity," Ann. Phys. (Germany) **59**, 101 (1919).

[9] W. Pauli, "Relativistic field theories of elementary particles," Rev. Mod. Phys. **13**, 203 (1941).

[10] S. S. Schweber, *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga*, (Princeton University Press, New Jersey, 1994).

[11] C. N. Yang and R. L. Mills, "Conservation of isotopic spin and isotopic gauge invariance," Phys. Rev. **96**, 191 (1954).

[12] M. Han and Y. Nambu, "Three-triplet model with double SU(3) symmetry," Phys. Rev. **139**, B1006 (1965).

[13] O. W. Greenburg, "Spin and unitary spin independence in a paraquark model of baryons and mesons," Phys. Rev. Lett. **13**, 598 (1964).

[14] S. Weinberg, "A model of leptons," Phys. Rev. Lett. **19**, 1264 (1967).

[15] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak interactions with lepton-hadron symmetry," Phys. Rev. D **2**, 1285 (1970).

[16] A. Salem, in *Elementary Particle Theory: Relativistic Groups and Analyticity (Nobel Symposium No. 8)*, edited by N. Svartholm (Almquist and Wiksell, Stockholm, 1968), p. 367.

[17] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," Phys. Rev. Lett. **13**, 508 (1964).

[18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles," Phys. Rev. Lett. **13**, 585 (1964).

[19] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," Phys. Rev. Lett. **13**, 321 (1964).

[20] M. Robinson, *Symmetry and the Standard Model: Mathematics and Particle Physics*, (Springer, New York, 2011).

[21] F. Halzen and A. D. Marin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, (John Wiley & Sons Inc, New York, 1984).

[22] Excerpt from http://en.wikipedia.org/wiki/File:Mecanismo_de_Higgs_PH.png, Retrieved May 2012.

[23] T. Aaltonen *et al.* (CDF Collaboration), "Search for a fermiophobic higgs boson decaying into diphotons in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV," Phys. Rev. Lett. **103**, 061803 (2009).

[24] Feynman diagrams modified by the author from the original version made by R. C. Group.

[25] Standard Model Higgs Cross Sections at Hadron Colliders, http://maltoni.home.cern.ch/maltoni/TeV4LHC/SM.html.

[26] Tevatron New Phenomina and Higgs Working Group (CDF and D0 Collaborations), "Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb-1 of Data," arXiv:1107.5518, 2011.

[27] A. Djouadi, J. Kalinowski, and M. Spira, "HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension," Comput. Phys. Commun. **108**, 56 (1998).

[28] JaxoDraw, http://jaxodraw.sourceforge.net/.

[29] M. Baak *et al.* (Gfitter Group), "Updated Status of the Global Electroweak Fit and Constraints on New Physics," arXiv:1107.0975, 2011.

[30] M. Baak *et al.* (Gfitter Group), Updated results at: http://gfitter.desy.de/.

[31] Tevatron New Phenomina and Higgs Working Group (CDF and D0 Collaborations), "Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data," arXiv:1203.3774, 2012.

[32] ATLAS Collaboration, "An Update to the Combined Search for the Standard Model Higgs boson with the ATLAS detector at the LHC Using Up to 4.9 fb^{-1} of pp Collision Data at $\sqrt{s} = 7$ TeV," ATLAS-CONF-2012-019, 2012 (unpublished).

[33] S. Chatrchyan *et al.* (CMS Collaboration), "Combined Results of Searches for the Standard Model Higgs Boson in pp Collisions at $\sqrt{s} = 7$ TeV," arXiv:1202.1488, 2012.

[34] T. Aaltonen *et al.* (CDF Collaboration), "Search for a Higgs boson in the diphoton final state in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV," Phys. Rev. Lett. **108**, 011801 (2012).

[35] Fermilab Tevatron Interactive Timeline, http://www.fnal.gov/pub/tevatron/.

[36] CERN, "LHC Sets New World Record," CERN Press Release, Nov. 30, 2009, http://press.web.cern.ch.

[37] Fermilab Accelerator Operations Rookie Books, http://www-bdnew.fnal.gov/operations/rookie_books/rbooks.html.

[38] A. Valishev, "Tevatron Accelerator Physics and Operation Highlights," arXiv:1202.5525, 2012.

[39] N. Goldschmidt (CDF Collaboration), "$p\bar{p}$ Cross Sections for the Tevatron Energy Scan," [Fermilab Report No. FERMILAB-FN-0929, 2011].

[40] The CDF Experiment, http://www-cdf.fnal.gov.

[41] M. Herndon, "CDF Experience," LHC / Tevatron Tracking Workshop, 2004.

[42] A. Affolder *et al.* (CDF Collaboration), "CDF central outer tracker," Nucl. Instrum. Meth. A **526**, 249 (2004).

[43] P. T. Lukens (CDF Collaboration), "The CDF IIb Detector Technical Design Report," [Fermilab Report No. FERMILAB-TM-2198, etc., 2003].

[44] R. Blair *et al.* (CDF Collaboration), "The CDF-II Detector: Technical Design Report," [Fermilab Report No. FERMILAB-DESIGN-1996-01, 1996].

[45] J. Elias *et al.*, "Luminosity monitor based on Cherenkov counters for $p\bar{p}$ colliders," Nucl. Instrum. Methods Phys. Res. Sect. A **441**, 366 (2000).

[46] D. Acosta *et al.*, "The performance of the CDF luminosity monitor," Nucl. Instrum. and Methods Phys. Res. Sect. A **494**, 57 (2002).

[47] C. Hays *et al.*, "Introduction to PadTrack," CDF Internal Note 6044, 2002 (unpublished).

[48] C. Hays *et al.*, "The PAD Track Collection in 5.3," CDF Internal Note 7304, 2004 (unpublished).

[49] C. Hays *et al.*, "Inside-out tracking at CDF," Nucl. Instrum. and Methods Phys. Res. Sec. A **538**, 249 (2005).

[50] H. Wenzel, "Tracking in the SVX," CDF Internal Note 1790, 1998 (unpublished).

[51] Jean-Francois Arguin and Beate Heinemann, "Underlying Event Corrections for Run II," CDF Internal Note 6239, 2003 (unpublished).

[52] R. Wagner, "Electron Identification for Run II: Algorithms," CDF Internal Note 5456, 2003 (unpublished).

[53] M. Riveline, "CES Clustering in Run II," CDF Internal Note 5863, 2002 (unpublished).

[54] G. Pauletta (CDF Collaboration), "Operation and performance of the CDF calorimeters," J. Phys.: Conference Series **160**, 012007 (2009).

[55] K. Yasuoka, S. Mikamo, T. Kamon, and A. Yamashita (CDF Collaboration), "Response maps of the CDF central electromagnetic calorimeter with electrons," Nucl. Instrum. Meth. A **267**, 315 (1988).

[56] R. Wagner *et al.* (CDF Collaboration), "Cosmic ray test of the CDF central calorimeters," Nucl. Instrum. Meth. A **267**, 330 (1988).

[57] J. Rauch, "Implementing a Multivariate Analysis to Identify Plug Photons in the Search for the Higgs Boson at CDF," Baylor University REU Program, 2011 (unpublished).

[58] A. Hoecker *et al.*, "TMVA - Toolkit for multivariate data analysis," POS ACAT **040** (2007).

[59] J. Ray, "Training a Multivariate Photon ID Against Jets," Poster, 2011, http://www-cdf.fnal.gov/~jamieray/100831_Poster4.pdf.

[60] J. Ray, R. C. Group, and R. Culbertson, "Training a Multivariate Photon ID Against Jets," CDF Internal Note 10282, 2010 (unpublished).

[61] C. R. Hall, "Measurement of the Isolated Direct Photon Cross Section with Conversions in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ GeV," PhD thesis, Harvard University, [Fermilab Report No. FERMILAB-THESIS-2002-40, 2007].

[62] E. Halkaidakis, C. Hays, M. Tecchio, and W. Yao, "A Conversion Removal Algorithm for the 2003 Winter Conferences," CDF Internal Note 6250, 2003 (unpublished).

[63] J. P. Chou, M. Franklin, and S. Grinstein, "Conversion Backgrounds For the SLT_e Tagger," CDF Internal Note 8788, 2007 (unpublished).

[64] A. Attal and A. Canepa, "Photon Conversion Removal Efficiency," CDF Internal Note 8073, 2006 (unpublished).

[65] B. Kilminster *et al.* (private communication).

[66] T. Aaltonen *et al.* (CDF Collaboration), "Measurement of the cross section for prompt isolated diphoton production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," Phys. Rev. D **84**, 052006 (2011).

[67] S. Hewamanage, "Search for Anomalous Production of Photon + Jets + Missing Transverse Energy in $p\bar{p}$ Collisions at Center of Mass Energy = 1.96 TeV Using the CDF II Detector," PhD thesis, Baylor University, [Fermilab Report No. FERMILAB-THESIS-2011-49, 2011].

[68] X. Bu, "$H \to \gamma\gamma$ Search and Direct Photon Pair Production Differential Cross Section Measurement at D0," PhD thesis, University of Science and Technology of China, [Fermilab Report No. FERMILAB-THESIS-2010-29, 2010].

[69] R. Culbertson, A. Pronko, and S.-S. E. Yu, "The Probability of an Electron Faking an Isolated Prompt Photon in the CEM," CDF Public Note 8220, 2006 (unpublished).

[70] F. James, "MINUIT - Function Minimization and Error Analysis Reference Manual," CERN Program Library Long Writeup D506.

[71] T. Sjostrand *et al.*, "High-energy-physics event generation with PYTHIA 6.1," Comput. Phys. Commun. **135**, 238 (2001).

[72] T. Sjostrand, L. Lonnblad, and S. Mrenna, "PYTHIA 6.2: Physics and Manual," arXiv:0108264, 2001.

[73] CDF Simulation, http://www-cdf.fnal.gov/cdfsim/cdfsim_main.html.

[74] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, "GEANT3," [CERN Report No. CERN-DD/EE/84-1, 1987].

[75] G. Grindhammer, M. Rudowicz, and S. Peters, "The fast simulation of electromagnetic and hadronic showers," Nucl. Instrum. Meth. A **290**, 469 (1990).

[76] T. Affolder *et al.* (CDF Collaboration), "Charged jet evolution and the underlying event in $p\bar{p}$ collisions at 1.8 TeV," Phys. Rev. D **65**, 092002 (2002).

[77] D. Bourilkov and R. C. Group and M. R. Whalley, "LHAPDF: PDF use from the Tevatron to the LHC," arXiv:0605240, 2006.

[78] M. R. Whalley and D. Bourilkov and R. C. Group, "The Les Houches accord PDFs (LHAPDF) and LHAGLUE," arXiv:0508110, 2005.

[79] LHAPDF, http://projects.hepforge.org/lhapdf/.

[80] H. L. Lai *et al.* (CTEQ), "Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions," Eur. Phys. J. C **12**, 375 (2000).

[81] T. Aaltonen *et al.* (CDF Collaboration), "Studying the underlying event in Drell-Yan and high transverse momentum jet production at the Tevatron," Phys. Rev. D **82**, 034001 (2010).

[82] R. Field and R. C. Group (CDF Collaboration), "PYTHIA Tune A, HERWIG, and JIMMY in Run 2 at CDF," arxiv:0510198, 2005.

[83] G. Bozzi, S. Catani, D. de Florian, and M. Grazzini, "The q(T) spectrum of the Higgs boson at the LHC in QCD perturbation theory," Phys. Lett. B **564**, 65 (2003).

[84] G. Bozzi, S. Catani, D. de Florian, and M. Grazzini, "Transverse-momentum resummation and the spectrum of the Higgs boson at the LHC," Nucl. Phys. B **737**, 73 (2006).

[85] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini, "Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC," JHEP **1111**, 064 (2011).

[86] S.-S. E. Yu and R. Culbertson, "Studies of Electron and Photon Energy Scales," CDF Internal Note 9591, 2008 (unpublished).

[87] S. Dittmaier *et al.* (LHC Higgs Cross Section Working Group), "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables," arXiv:1101.0593, 2011.

[88] R. C. Group (private communication).

[89] V. Papadimitriou, "Luminosity Determination at the Tevatron," arXiv:1106.5182, 2011.

[90] CDF Luminosity Group, http://www-cdf.fnal.gov/~konigsb/lum_official_page.html.

[91] Tevatron New Phenomina and Higgs Working Group (CDF and D0 Collaborations), "Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb^{-1} of Data," arXiv:1107.5518, 2011.

[92] M. Botje *et al.*, "The PDF4LHC Working Group Interim Recommendations," arXiv:1101.0538, 2011.

[93] D. Stump *et al.*, "Inclusive jet production, parton distributions, and the search for new physics," JHEP **0310**, 046 (2003).

[94] T. Junk, MClimit Software, http://www-cdf.fnal.gov/~trj/mclimit/production/mclimit.html.

[95] J. Heinrich, "Bayesian Limit Software: Multi-Channel with Correlated Backgrounds and Efficiencies," CDF Public Note 7587, 2005 (unpublished).

[96] T. Junk, "Building a More General χ^2," CDF Public Note 7904, 2007 (unpublished).

[97] T. Junk, "Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties," CDF Public Note 8128, 2007 (unpublished).

[98] LEP Working Group for Higgs Boson Searches (ALEPH, DELPHI, L3, and OPAL Collaborations), "Search for the standard model Higgs boson at LEP," Phys. Lett. B **565**, 61 (2003).

[99] The CDF Higgs Discovery Group, "Combination of CDFs searches for the standard model Higgs boson with up to 10.0 fb^{-1} of data," CDF Public Note 10804, 2012 (unpublished).

[100] T. Aaltonen *et al.* (CDF Collaboration), "Search for a fermiophobic Higgs boson decaying into diphotons in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," Phys. Rev. Lett. **103**, 061803 (2009).

[101] The CDF Collaboration, "Search for a Fermiophobic Higgs Boson in the Di-photon Final State Using 10.0 fb^{-1} of CDF Data," CDF Public Note 10731, 2012 (unpublished).

[102] R. C. Group, R. Culbertson, and J. Ray, "Photon Efficiency Scale Factors (up to p12)," CDF Internal Note 9429, 2008 (unpublished).

[103] J. Dittmann (private communication).