OBSERVATION OF A HIGGS BOSON DECAYING TO PHOTONS USING MULTIVARIATE TECHNIQUES AT CMS

By

RISHI PATEL

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosphy

Graduate Program in Physics and Astronomy

written under the direction of

Prof. Yuri Gershtein

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2014

ABSTRACT OF THE DISSERTATION

Observation of a Higgs Boson Decaying to Photons using Multivariate Techniques at CMS

By RISHI PATEL

Dissertation Director:

Prof. Yuri Gershtein

The results for the observation of a Higgs boson decaying into two photons with the full Run 1 dataset from the Large Hadron Collider is presented. The analysis is performed with the CMS detector at the Large Hadron Collider from pp collisions at centre of mass energy of 7 TeV and 8 TeV, which correspond to integrated luminosities of $5.1fb^{-1}$ and $19.7fb^{-1}$ respectively. The sequence of the analysis from vertex selection, photon energy reconstruction, photon identification and finally event selection is done using multivariate techniques. The Higgs boson is observed as a significant excess of events in the diphoton invariant mass spectrum from 100 to 180 GeV. The best fit signal strength $\mu = \sigma_{obs}/\sigma_{SM}$ gives the compatibility of the observed cross-section with the predicted cross-section from the Standard Model. In addition, the analysis accounts for different Higgs production mechanisms to target the Higgs couplings to the top quark and also the W and Z vector bosons.

Acknowledgments

First I would like to acknowledge the work done by the $H \to \gamma\gamma$ working group, in which I participated. The work is described in detail in [51] and [52], which gives the technical information about the Higgs to two photon analysis at CMS. The published results are available in citeLegacyPap

It needs to be said that a Ph.D. program is privelege for a certain few. Anyone who says that graduate school is hell has not read the news or seen a good arthouse film to be familiar with the darker shades of the human condition. Perhaps now emerging from the surface after a long dive, I can see how lucky I have been to be a phycisist. I owe gratitude to many people.

First to the set the ship asail was Prof. Dave Toback from Texas $A\&M$ who considered me to be 'a diamond in the rough' and also owe thanks to the Rutgers high energy physics group for taking his recommendation to heart.

Also a very special thanks goes to Yuri Gershtein, my advisor, who kept connecting me back to that Freudian expansive feeling that keeps me and many other scientists hooked to this profession. For future scientists, the best metaphor that I can think of is from a long prose piece by Paul Bowles about midnight in the Sahara as a psychologically transforming addiction. There is everything the stillness, the isolation, feeling insignificant and confronting your ignorance, and then molding your mind to be more plyable. Even now, I can still feel myself being drawn back. But life never slips into mediocity as long as you remember to 'go all the way' as Bukowski said in *Roll the Dice*.

This is not to say that the demands are not easy and one will not slip into dark places. It also gets hard on the people you care about (sorry P.) Then of course there is the self-questioning, that in retrospect has some very simple answers if you ask the right questions. Instead of asking am I brainy enough or am I cut out for this, I probably should have kept in sight the 'exploratory impulse' that can never be quieted. As fun as it is to slowly unfold a problem and arrive at a solution, it is even more fun to think of all the things that are still left to know. There is a great quote by Bulgakov: "What would the world be without shadows". He meant both the mysteries of the world and the weaknesses of man–both make life interesting.

Finally it was a great pleasure to be a part of the LHC experiment and work with the CMS collaboration. I owe thanks to the National Science Foundation for funding my stay in Switzerland

which in itself was transformative. It is my personal wish that the United States preserves its long-standing committment to science and fundamental research. Every scientific endeavor is a unique canvas which illuminates the universe with only the colors provided. Perhaps the limits of the imagination will be extended for a new generation of creative scientists.

Dedication

To my parents, my good friends and all the roots of wonderful people that pushed me out of the ground and into the sky to be a budding scientist.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

At the beginning of this thesis, it seems appropriate to start with a very early point in the universe though the very beginning of the universe still remains mysterious. The prevailing model of the early universe is the big bang, where a very hot dense medium cools to form stable systems like atoms and nuclei. Particle physics attempts to understand the basic ingredients, the particles, and their interactions in the early universe to try to describe the most fundamental laws of physics. Nobel Prize winner, Leon Lederman phrased this in the title of his book: "If the Universe is the answer, then what are the questions?" It is possible that all the currently known forces are manifestations of a fundamental force. As the universe cooled the fundamental force became differentiated into separate forces, thus hiding their unity. The same can be true for the known set of fundamental particles. Each has a unique mass that is essential for formation of stable systems of particles. It is possible that all particles in the early universe are massless and acquire mass as the universe evolved. For massless particles, none of the beautiful structure of the universe would exist, like galaxies, planets or even atoms and the particles would fly apart and never coalesce. There might not be anymore questions (and also no one to ask them)! Instead, there is the observed Universe with an important question: how do particles acquire mass? The Standard Model has, to date, given a description of the properties and interactions of known particles, and has been verified to a high degree of precision by the results of accelerator experiments. An important missing piece in the Standard Model predicts a scalar field, the Higgs field, gives an important ingredient in answering the question of how particles acquire mass. The main evidence of the Higgs field is the Higgs boson, and it will be shown to be produced in the largest particle accelerator, the Large Hadron Collider, and detected at the Compact Muon Solenoid detector when it decays to a pair of photons.

Chapter 2 will describe the very successful Standard Model of particle physics that motivates the existence of a Higgs field with an associated particle, the Higgs Boson. One of the largest successes of the model is to unify two fundamental forces: the Weak force and the Electromagnetic force. However, the unification seemed to require only massless particles associated to the fields. This contradicts the experimental evidence of three massive particles associated with the Weak force: W^+, W^-, Z^0 as well as the properties of the weak force which rely on massive particles to facilitate

the weak interaction. This motivated the need for a mass generating mechanism that arises by spontaneously breaking the unification of the Electromagnetic and Weak force. This symmetry breaking results in a Higgs field with an associated Higgs boson, that couples to W^+, W^-, Z^0 and other massive particles while some particles massless like the photon. Physically, this results in a field with a non-zero vacuum state that exists throughout the universe and massive particles arise by interacting with the Higgs particles of the Higgs field.

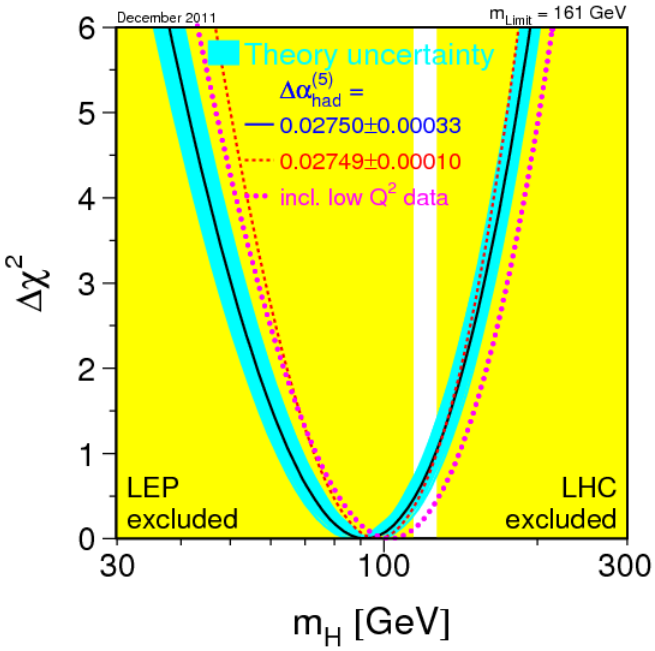


Figure 1.1: LEP and the LHC exclusion of Higgs Masses allowing only a narrow window around 125 GeV. (LEP Electroweak Working Group, http://lepewwg.web.cern.ch/LEPEWWG/)

Discovery of the Higgs boson is the main evidence for the existence of a Higgs field. The Higgs boson is theorized to be produced in high energy proton-proton collisions from the colliding quarks and gluons. However, the mass of the Higgs boson is not predicted by the Standard Model. Large self-interactions can make the Higgs boson very massive, and large quantum corrections from virtual particles can lower the mass significantly as well. The first set of searches for the Higgs boson were conducted at the e^+e^- collider, LEP, at CERN. The three important processes for producing the Higgs boson were for a light Higgs, less than the Z boson mass: $Z \to H + f\overline{f}$, and for a heavier Higgs it is produced in association with a Z boson: $e^+e^- \to Z + H$. High precision electroweak data along with the mass of the top-quark made it possible to constrain the mass of the Higgs boson around the most likely value of 94^{+29}_{-24}GeV. By the year 2000, the LEP collider was able to exclude a Higgs

mass $M_H > 114.4$GeV at 95% CL. Figure 1.1 shows the exclusion of Higgs masses as a function of the χ^2 for the precision electroweak data by both LEP and also the Large Hadron Collider. The Large Hadron Collider experiment is designed with a major emphasis on the discovery of the Higgs boson and extending the range of observable Higgs masses, which allows the exclusion of a large mass range of potential Higgs mass leaving only a narrow window where it is observed. The focus of this thesis is the predicted decay of a Higgs boson to two photons. Standard model particles are detected by particle detectors surrounding the proton-proton collision point.

Chapter 3 will describe the Large Hadron Collider and the Compact Muon Solenoid detector. The Higgs decay to massless photons is facilitated by loop diagrams with virtual particles, and has a much smaller branching fraction compared to decays to heavy particles. However, the background processes have a smoothly falling non-resonant diphoton mass spectrum while the mass resonance from the photons of the Higgs decay give a sharp peak. This gives a good signal-to-background ratio for detecting the Higgs boson as a mass resonance in selected diphoton events.

Chapter 4 describes how photons are reconstructed from the detector information with particular emphasis on the photon energy measurement which drives the width of the signal peak. A major gain is made by using multivariate techniques to derive an energy correction based on simulated photons from Monte Carlo. A good criteria for identification of the photons from the collision point is also key to suppress as much of the background as possible from photons originating from the decay of mesons. Reconstructing conversions of photons into e^+e^- allow signal photons to be associated to a trajectory which can be used to find the Higgs decay point. Chapter 5 describes how the diphoton mass is reconstructed for selected events. Selected events are classified to indicate their signal-to-background ratio and the expected signal model and background model are constructed as probability density functions in each category. The final statistical treatment gives a measure of the production rate of the Higgs boson decaying to photons at a particular mass value. Chapter 6 will summarize the method and the results.

Chapter 2
Theory

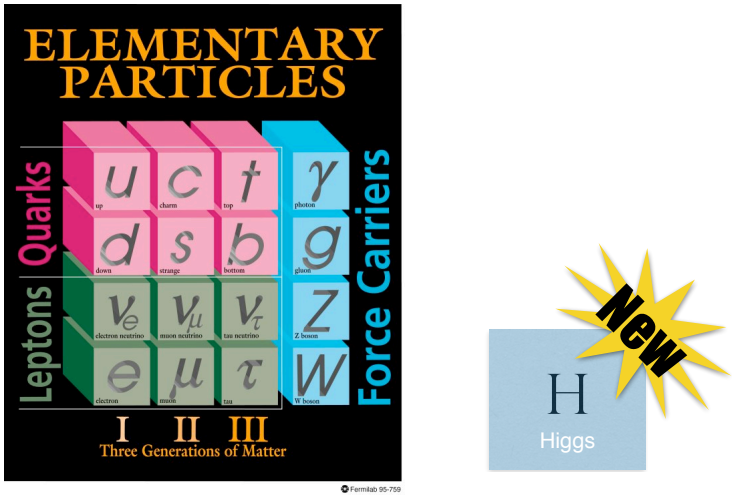
The Standard Model is a quantum field theory that provides a remarkably precise description of the interactions of the known elementary particles. This chapter will give a brief description of the fundamental forces. These fundamental forces are described by quantum fields of fundamental particles. The first level of categorization in the Standard Model is by spin. According to the Standard Model, the universe is made of elementary spin $\frac{1}{2}$ particles, *fermions*, that interact by exchanging integral spin particles, *bosons*. The quantum fields of bosons are invariant under a continuous set of transforms called gauge transforms. Requiring gauge invariance gives important physical observables for the quantum field. A major triumph of the Standard Model is to unify two of the fundamental forces the electromagnetic force and the weak force as manifestations of a single electro-weak force. However, the theory needs to accommodate the mass of the weak force carriers and also preserve gauge invariance. An additional Higgs field, spontaneously breaks the symmetry between the two forces and a as result of "hiding" the symmetry gives rise to massive bosons $W^{\pm}$, Z, as well as a massless photon. This chapter will describe the theoretical prediction of this Higgs field and how it spontaneously breaks electroweak symmetry. This chapter will also explain the processes by which the Higgs boson, the quanta of the Higgs field, can be produced in high energy proton-proton collisions and the process by which it can decay to two photons. Discovery of the Higgs boson and the measurement of its properties confirms the existence of a Higgs field and provides evidence of the electro-weak symmetry breaking mechanism.

2.1 Standard Model and the Higgs Mechanism

2.1.1 Elementary Particles and Fundamental Forces

There are four fundamental interactions observed in nature. Firstly, gravity is not considered in the Standard Model because at the particle level gravity is insignificant compared to the three other forces. The photon is the quanta of the electromagnetic interaction between charged particles. Quarks are the subconstituents of protons and neutrons, as well as shorter lived mesons. Gluons mediate the strong force that bind quarks into nucleons and mesons. Photons are massless and

have infinite lifetime, so act on infinite range with 1/r behavior for potential energy. Gluons are also massless and may be expected to have the same behavior as photons, but have an additional principle of confinement due to the effect of color. Color is an additional quantum state for quarks. The color of a quark can be changed by emitting or absorbing a gluon. Confinement requires that quarks and gluons are bound into color neutral configurations. Due to confinement, violently disintegrating hadrons (as in proton-proton collisions at relativistic speeds) will not result in completely free quarks or gluons. Instead, quarks will radiate gluons and gluons will radiate other gluons or quark anti-quark pairs. These particles will reassemble into a final state of color neutral particles. Section 2.2 will show how the Higgs can be produced from quarks and gluons after a proton-proton collision. The third fundamental force is the weak force which is responsible for the radioactive decay of atoms. The weak force can slowly pull apart large nuclei into smaller nuclei in radioactive elements and also participates in the fusing of small nuclei into larger nuclei in energetic systems like the sun. Table 2.1 shows the properties of fundamental forces and their associated particles. The electromagnetic force has infinite range, and due to the principle of confinement the strong force has range confined to a typical nuclear radius. The strengths of the forces is normalized to the strong force. The weak force is aptly named to have the smallest strength and is confined to about 1% of the radius of the nucleus.



(a) The Standard Model Particles (b) The Higgs Boson

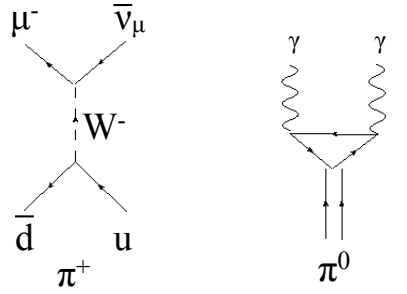
Figure 2.1: The Standard Model particles including the Higgs

The three fundamental forces can be observed in the system of an atom. The electrons are electromagnetically attracted to nuclei. Atoms can absorb photons to give electron excitations or emit photons from electron de-excitations. The strong force binds the quarks into protons and

Force	Carrier	Spin	Strength	Range(m)	Mass(GeV/c^2)
Strong	gluon	1	1	10^{-15}	0
EM	photon	1	1/137	∞	0
Weak	$W^{\pm}, Z^0$	1	10^{-16}	10^{-18}	80.4, 91.2

Table 2.1: The properties of the three fundamental forces described by the Standard Model.

neutrons to form a nucleus. The weak force is capable of changing one flavored quark into a different flavored quark while emitting a lepton in processes like beta decay in the nucleus. The weak force is unique in that it has three force mediators that have non-zero mass. The mass of the bosons give the weak force the important feature that it is short ranged. By the uncertainty principle: $\frac{i\hbar}{\Delta E} \approx \frac{\hbar}{Mc^2} \approx 10^{-3} fm$ confining the force to very small distances of about 1000 times smaller than a hydrogen nucleus. The mass of the bosons is larger than of most of the fermions. The mass enters into the particle-propagator used to compute decay amplitudes and causes a slow rate for weak decays proportional to $\frac{1}{M_W^2}$. An example is shown in Figure 2.2, the neutral pion decays electromagnetically to photons at a rapid rate with a lifetime of 8.4×10^{-17}seconds while the charged pions, which decay weakly, have a much longer lifetime of 2.6×10^{-8} seconds since the decay amplitude is inversely proportional to the mass of the W boson. The properties of the W and Z bosons have important consequences, so the theory of how these bosons acquire mass is a major cornerstone of the Standard Model. This will be described in Section 2.1.3.



(a) π^0 decay (b) π^+ decay

Figure 2.2: Neutral pion lifetime compared to the charged pion lifetime

The Standard Model can be thought of as a multi-lever universe machine. The three main levers are the strengths of the three fundamental forces. What would happen to the stability of the nucleus if the weak force was comparable to the strong force? However, at very high energies, $> 10^{14}$ GeV, in the early universe one can imagine that all three levers are level so the three forces become manifestations of a single force. As the universe cools and inflates the levers in the multi-lever 'universe machine' become set to what we observe now and what the Standard Model predicts. It is fascinating to think of the evolution of a single force at high energies governing all interactions

that becomes separated into forces that govern seemingly different phenomena. These separate forces give rise to structures like the atom and nuclei when the levers are set to what is observed. To understand this evolution, one needs to formalize the states of a large system of interacting relativistic particles. [40] A quantum field theory ties together essential theories of particles: particles are point-like and distinguished by a set of quantum numbers, they can be created and destroyed, and their interactions should be relativisitically invariant (they do not depend on inertial frame). The Dirac equation is the the first example of such theory.

The Dirac equation describes a relativistic quantum electron with spin 1/2. The Dirac field $\Psi(x)$ is relativistically invariant when transforming the four-vector x. Quantizing this field allows for particles to be created and destroyed within the Dirac field and leads to a quantum field theory. The dynamics of the field are described by the Dirac equation:

$$\mathcal{L} = \overline{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi$$

Like a classical field it consists of a kinetic and potential term. However, the Euler-Lagrange equation described the continuous trajectory of a particle. The Dirac equation describes the particles being created and destroyed in a field where their interactions are kept invariant in every inertial frame. For example, an electromagnetic interaction can be described as a quantum state of a photon which is the quanta of an electromagnetic field. The Standard Model is a quantum field theory where particles interact with one another through interaction fields via an associated force mediating particle. The particles that source the interaction fields are the fermions. The quanta of the interaction fields are bosons. This gives the first level of categorization in the Standard Model.

The first level of categorization of elementary particles is by their spin. An important theorem that comes from the Dirac equation is the spin statistics theorem that connects the statistics a particle obeys with the spin of the particle. The fermions are defined to have half-integral spin. Bosons are defined to have integral spin. For a system of identical fermions, the wave function switches sign when two particles are exchanged (anti-commutation relation). This manifests in the familiar Pauli exclusion principle that two fermions cannot occupy the same quantum state, and leads to the stability of the electron shells in the atom. Bosons follow Bose-Einstein statistics, where the wave function commutes, and so indistinguishable particles can have identical quantum states. In a ground state, many bosons can aggregate into a condensate. An example of this phenomena is in superconductors where electron pairs become a condensate of cooper-pairs in the ground state. The pair acts like a spin-less boson [3]. This condensate ground state can interact with photons to give them an effective mass, and provides an important inspiration for the Higgs mechanism and the Higgs field which will be shown in Section 2.1.3.

Fermions can be further divided further based on the interactions in which they participate. Leptons interact only through the weak and electromagnetic interactions. Quarks interact via the weak, electromagnetic, and strong force. The six quarks and six leptons can be ranked into 3 generations based on mass. The u, d, quark pair and the electron lepton form the first generation which are the stable constituents of atoms. The 2nd generation c,s quarks have larger mass and exist within unstable mesons that decays via the weak interaction. Likewise the 2nd generation lepton, the muon is unstable with a finite lifetime: 2.2 μs. The third generation quarks b,t have even shorter lifetimes as does the 3rd generation lepton τ. Figure 2.1 summarizes the known bosons and fermions.

2.1.2 Gauge Field Theory

An important guiding principle for both classical fields and quantum fields is symmetry. Hermann Weyl very poetically described the attractiveness of this principle "as wide or narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection." [53] In the next section, it will be shown that even the breaking of a symmetry can lead to order! Symmetry of a field can be discovered by transforming the fields under a particular operation, and observing whether or not the dynamics of the field are unchanged. Noether's theorem states that such a transform will lead to a conserved quantity. For a field ϕ that transforms as $\phi(x) \Rightarrow \phi + \delta\phi$ if the transformation is a symmetry it will leave the Lagrangian unchanged so that:

$$\mathcal{L}(\phi) - \mathcal{L}(\phi + \delta\phi) = 0$$

Expanding about the infintesimal transform of the field:

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta(\partial_\mu\phi)$$

Using the identity:

$$\delta(\partial_\mu\phi) = \partial_\mu(\phi + \delta\phi) - \partial_\mu\phi = \partial_\mu(\delta\phi)$$

and the Euler Lagrange equation:

$$\frac{\partial\mathcal{L}}{\partial\phi} = \partial_\mu\frac{\partial\mathcal{L}}{\partial_\mu\phi}$$

The equation for $\delta\mathcal{L}$ simplifies with chain rule for differentiation:

$$\partial_\mu\left(\frac{\mathcal{L}}{\partial(\partial_\mu\phi)}\right)\delta\phi + \frac{\mathcal{L}}{\partial(\partial_\mu\phi)}\partial_\mu\delta\phi = \partial_\mu\left(\frac{\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) = 0$$

So $\partial_\mu\left(\frac{\mathcal{L}}{\partial(\partial_\mu\phi)}\right) = \partial_\mu j^\mu = 0$ gives a conserved current j^μ for a symmetric transform of the field ϕ. One example of a symmetry transform is rotation, and a rotationally invariant Lagrangian will lead

to conservation of momentum. Rotation is an example of a continuous symmetry that can be done infinitesimally. There are also discrete symmetries, the most common of which is bilateral symmetry between left and right. For particles the handed-ness is formalized in the projection operators:

$$P_R = \frac{1+\gamma_5}{2}$$

$$P_L = \frac{1-\gamma_5}{2}$$

The Standard Model is built out of symmetry groups: $U(1) \otimes SU(2) \otimes SU(3)$ each with its own conserved quantity. The symmetries in each group and symmetry breaking has important consequences for the foundations of the theory. Here emphasis is given on the electroweak symmetry in the first two groups: $U(1) \otimes SU(2)$. The consequences of U(1) gauge symmetry will be demonstrated for the quantum electrodynamic field. Of the transforms in U(1) of the form: $\psi \Rightarrow \psi' = e^{-i\chi(x)}\psi$ a particular set of gauge transforms gives important physical observables like the charge and conserved current.

The Standard Model also needs to address when symmetries are not preserved. The mass term in the Dirac equation $m\psi\overline{\psi}$ will not be parity invariant: $m\psi\overline{\psi} = m\left(\overline{\psi}_L\psi_R - \overline{\psi}_R\psi_L\right)$. The photons which are massless preserve parity in electromagnetic interactions, so the left-handed and right handed electromagnetic interactions occur with equal probability. The weak interactions do not have this property so this symmetry of handedness needs to be broken. This section will demonstrate the importance of local gauge symmetry in U(1) for quantum electrodynamics while the next section will show how the same procedure fails for the weak interactions where the handed-ness of particles and massive gauge bosons needs to be accommodated. These problems are addressed by the Standard Model which unifies the weak force with the electromagnetic into a single symmetry group $U(1) \otimes SU(2)$ and then breaks the symmetry to give the massive weak force carriers.

A hint of a particular kind of symmetry comes from classical electrodynamics. A particular electric potential can define a field, but this potential need not be unique. Given an electric potential $A^\mu = \left(\Phi(x,t), \overrightarrow{A}(x,t)\right)$, Maxwell's equation remain unchanged if the electric potential is transformed as:

$$A^\mu \longrightarrow A^{\mu\prime} = \left(\Phi(x,t) - \frac{\partial\chi(x,t)}{\partial t}, \overrightarrow{A}(x,t) + \nabla\chi(x)\right)$$

The scalar function $\chi(x,t)$ is the gauge function, and including it as above in the electromagnetic potential gives the same magnetic fields and electric fields. Applying a Gauge condition on $\chi(x,t)$ can give some convenient relations, for example requiring the Columb gauge with $\nabla\overrightarrow{A} = 0$ gives the $\Delta\Phi = \rho/\epsilon_0$ where ρ is the charge density and so the scalar potential can be specified by integrating the charge density over a volume. Another important Gauge condition is Lorentz gauge, where $\partial_\mu A^\mu = 0$ which keeps the wave functions of the potentials propagating at the speed of light. For

quantum fields this gauge invariance becomes essential and has important physical consequences. The relative phases of wave functions can be measured but the absolute phase is immeasurable and arbitrary. The gauge function is now not a global function, but local and so the phase of a wave function can change from point to point. Gauge invariance should fix the phase transform from point to point. Under an arbitrary U(1) transform:

$$\psi \Rightarrow \psi' = e^{-i\chi(x)}\psi$$

Inputting this into the Dirac equation:

$$\mathcal{L}' = \overline{\psi}'(i\gamma^\mu\partial_\mu - m)\psi'$$

$$= \mathcal{L} - \overline{\psi}\gamma^\mu\psi\partial\chi$$

The Dirac equation is not gauge invariant, so consider a vector field A_μ that modifies the potential and represents a charged particle field coupling to an electric field with strength e:

$$\mathcal{L} = \overline{\psi}'(i\gamma^\mu\partial_\mu - m - eA_\mu)\psi$$

Now under a particular gauge transformation of the vector field and the fermion field:

$$\psi \Rightarrow \psi' = e^{-ie\chi(x)}\psi$$

$$A \Rightarrow A' = A + \partial\chi$$

The gauge function now depends on the charge and the electric potential now transforms with the gauge function. Inputting these relations into the Dirac field gives:

$$\mathcal{L}' = \overline{\psi}'(\imath\gamma^\mu\partial_\mu - m - eA_\mu)\psi'$$

$$= \overline{\psi}(\imath\gamma^\mu\partial_\mu - e\partial\chi - m - eA_\mu + e\partial_\mu\chi)\psi$$

$$= \mathcal{L}$$

The above shows how requiring gauge invariance leads to physical predictions from the Lagrangian. Charged fermions couple to a vector field, the photon field, with field strength e, the charge of the fermion. Also this interaction term $e\overline{\psi}\gamma^\mu\overline{\psi}A_\mu$ gives a conserved current $e\overline{\psi}\gamma^\mu\overline{\psi} = ej^\mu$ which gives the same interaction energy as in electromagnetism:$ej^\mu A_\mu = e\Phi - \overrightarrow{j}\cdot\overrightarrow{A}$ To show the full power of the theoretical predictions it is useful to rewrite the equation in terms of a field-tensor: $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ and all of Maxwell's equations can be written as $\partial_\mu F^{\mu\nu} = j^\nu$. Conservation of charge is embodied as $\partial_\nu j^\nu = 0$. The full quantum electromagnetic field can be written as:

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu(\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - (J^\mu A_\mu)$$

The first term is the Dirac equation for a fermion in free space, the second term is a field tensor that describes the photon field, and the final term is the interaction between the fermion field and the boson field. The third term yields the Maxwell equations, so gives the electric field sourced by the charged fermion. In terms of a quantum field, the electric field appears as a consequence of requiring the Dirac equation to be gauge invariant. The gauge invariance is fascilitated by including a gauge boson field in the equation. Note a mass term in the vector field would disrupt the gauge invariance: $eA_\mu + m_\gamma$ so the vector field has massless quanta: the photons.

2.1.3 Electroweak Symmetry Breaking

In Section 2.1.2, it was shown that the Lagrangian of quantum electrodynamics is invariant under a group U(1) of gauge transformations. Yang and Mills proposed that this symmetry argument can be generalized from local phase invariance to invariance under any continuous symmetry group. [54] This section will focus on a major accomplishment of the Standard Model to unify the electromagnetic and weak force based on the gauge group: $U(1)\otimes SU(2)$, and allow for this symmetry to be spontaneously broken to allow for 3 massive gauge bosons along with the massless photon. This broken symmetry predicts the massive bosons couple to a scalar field to acquire mass. This field has an associated gauge boson: the Higgs Boson. The Higgs field also has Yukawa couplings to fermions which generates their masses. This section will start by describing the unbroken $U(1) \otimes SU(2)$ group and how it conflicts with the observed properties of the weak interaction. Then the Higgs mechanism will be described that generate the masses of the gauge bosons and fermions.

SU(2) is a symmetry group with the symmetry transformation given by isospin operators with the commutation relation: $[T_i, T_j] = i\epsilon_{ijk}T_k$. The free field ψ will transform as doublets, so ψ will represent a generation of quarks or leptons. The 'S' represents that the special property that the 2x2 matrix transformations of the field always have unit determinant. The isospin operators are written as the Pauli matrices $T_i = \frac{1}{2}\sigma_i$ The free field Lagrangian again is:

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi$$

As in the U(1) case, the field must be invariant under a local gauge transformation:

$$\psi \rightarrow [1 - \imath g\alpha_i(x) \cdot T_i]\,\psi$$

The constant g is analagous to the electric charge, but represents coupling to the weak field and $\alpha(x)$ is a three-vector in isospin space multiplied by the 2x2 isospin transformations. The free field Lagrangian can be written conveniently in terms of a covariant derivative D_μ.

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu D_\mu - m)\psi$$

$$D_\mu = \partial_\mu + \imath g W_\mu \cdot T$$

The above implies the existence of an isospin triplet of three gauge fields $W_\mu = \left(W^0, W^1, W^2\right)$ that transform simultaneously as:

$$W_\mu(x) \rightarrow W_\mu(x) + \partial_\mu \alpha(x) + g\alpha(x) \times W_\mu(x)$$

The transformation is more complicated than in the U(1) case since the isospin operators of SU(2) do not commute. The gauge field part of the Lagrangian is written in terms of a field tensor:

$$\mathcal{L} = -\frac{1}{4} W_{\mu\nu} W^{\mu\nu}$$

$$W^i_{\mu\nu} \equiv \partial_\mu W^i_\nu - \partial_\nu - g\left(W_\mu \times W_\nu\right)^i$$

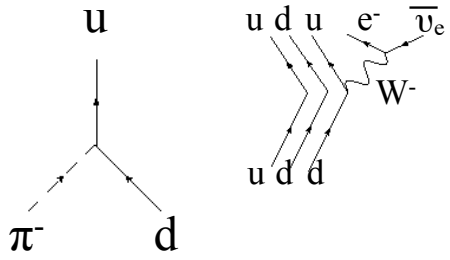
The full SU(2) gauge invariant Lagrangian is:

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \frac{1}{2} g \sum_i \left(\overline{\psi}\gamma^\mu T_i \psi\right) W^i_\mu - \frac{1}{4} \sum_i W_{\mu\nu} W^{\mu\nu}$$

As in the U(1) case for QED, the first term is the free field Lagrangian, the second term is the interaction of the fermion isospin current with the W fields and the final term is the gauge boson field energy. The conserved current of this gauge model has 3 components $J^{\mu i} = \overline{\psi}\gamma^\mu T^i \psi + \left(W^\mu \nu \times W_\nu\right)^i$ predicts three conserved charges corresponding to the three isospin generators. The isospin triplet gives three gauge bosons: $W^\pm_\mu = \frac{1}{\sqrt{2}}\left(W_{1\mu} \mp \imath W_{1\mu}\right)$, $W^0_\mu = W_{3\mu}$

Heisenberg thought of the isospin as two states of the same particle. [37] In particular, the proton and the neutron differed only by a sign in the 3rd isospin component: the proton $|\frac{1}{2}, \frac{1}{2}\rangle$ and the neutron $|\frac{1}{2}, -\frac{1}{2}\rangle$, just as an electron's state can be specified as spin up or spin down. A family of strongly interacting particles, pions, each in a distinct isospin state: $\left(\pi^+, \pi^0, \pi^-\right)$ could facilitate changing neutrons to protons and vice-versa conserving the T_3 and the charge. For families of particles a formula can be used to relate the charge and the isospin to a new quantity called the Hypercharge (Y): $Q = T_3 + \frac{1}{2}Y$, known as the Gell-Mann Nishijima formula. For the pion family, Y=0, so the charge has the same value as the isospin component. For the proton and neutron, Y is 1 to relate the isopin and the charge. This relation becomes important for the unification of the electromagnetic and weak interactions in defining the weak hypercharge. The weak force has a family of 3 force carriers: W^+, W^-, Z^0 which mediate the force. For example, the neutron decays to proton via W^- which changes the up quark to down quark and emits an e^- and $\overline{\nu}_e$. Figure 2.3a. shows the theorized strong force version of an up quark changing to a down quark via a pion. The pion was thought to be a special massless mediator that changes a family of quarks, i.e up to down and vice-versa. The up and down can be thought of isospin states of the same particle. However, this particular view of the exact symmetry breaks down because the pion is not massless and the

up and down quarks do not have identical masses. Figure 2.3b. shows the real force mediator that converts one flavor of quark into another by the weak force while emitting a family of leptons.



(a) Yukawa Pion Scattering (b) Weak Charged Current

Figure 2.3: Yukawa pion scattering and the Weak Charged Current

The above Lagrangian from the SU(2) isospin group does not account for two very important features of the weak interaction. The above equation gives identical couplings to left and right fermions despite that weak interactions do not conserve parity. In fact, parity is maximally violated in weak interactions. Secondly, the above gauge invariance requires massless gauge bosons as was the case for the photon field. However, the observation of the weak gauge bosons at CERN shows that the mass of the neutral Z is $91.2 GeV/c^2$ while the charged bosons $W^{\pm}$ is $80.4 GeV/c^2$. The appropriate Lagrangian should take into account the handed-ness of the fermions and also allow the gauge bosons to have mass.

To generate the left and right handed structure of weak interaction, the fermion fields are defined in terms of the parity projection operators:

$$\psi_R = (1 + \gamma_5)\,\psi$$

$$\psi_L = (1 - \gamma_5)\,\psi$$

The SU(2) group is combined with the U(1) group: $SU_{Iso}(2) \otimes U_Y(1)$ which will have four vector gauge bosons. The SU(2) part will have the weak isospin and the U(1) which is invariant under phase transforms will have the weak hypercharge symmetry. The right-handed particles are represented as singlets instead of doublets, they will all have total isospin of zero, and will transform non-trivially only in $U_Y(1)$ a subgroup of U(1) where hypercharge is conserved. This gives the observed result that only left handed particles participate in the weak interaction and the left handed particles will have doublet representations that transform in the full group: $SU_{Iso}(2) \otimes U_Y(1)$. This group will give rise to four gauge fields: an isospin triplet from $SU_{Iso}(2)$ with the field tensor $W^i_{\mu\nu}$ and a singlet from $U_Y(1)$ with $B_{\mu\nu}$ with the same form as the electromagnetic field tensor.

Defining the covariant derivative as $D_\mu = \partial_\mu + igW_\mu \cdot T + ig'\frac{1}{2}B_\mu Y$ and also the vector field for $U_Y(1)$ with the same form as the photon field: $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ the Lagrangian can be written as:

$$\mathcal{L} = -\frac{1}{4}W^{\mu\nu}W_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + \overline{\psi}\imath\gamma^\mu D_\mu\psi$$

The above Lagrangian is invariant under local gauge transformations for $SU_{Iso}(2)$ and $U_Y(1)$ independently since the field terms will transform differently depending on their handedness. The gauge transformations in SU(2) are:

$$\psi_L \to \left[1 - ig\alpha\left(x\right)\cdot T\right]\psi_L$$

$$W_\nu \to \left[W_\nu + \partial_\mu\alpha\left(x\right) + g\alpha\left(x\right)\times W_\mu\right]$$

The ψ_R are left unchanged in the SU(2) gauge transform. In U(1), both handed wave functions change:

$$\psi_L \to \left[1 - ig'\frac{1}{2}Y\beta\left(x\right)\right]\psi_L$$

$$\psi_R \to \left[1 - ig'\frac{1}{2}Y\beta\left(x\right)\right]\psi_R$$

$$B_\mu \to B_\mu + \partial_\mu\left(x\right)$$

To unify the electromagnetic and weak force, the neutral term of the covariant derivative $\imath gW_3\cdot T_3 + \imath g'\frac{1}{2}B_\mu Y$ must relate to the electromagnetic term $\imath eQA$, so the W_3 and B fields must be a linear combination of A the photon field (in the previous section) and another neutral field Z:

$$\begin{bmatrix} W_3 \\ B \end{bmatrix} = \begin{bmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{bmatrix} \times \begin{bmatrix} Z \\ A \end{bmatrix}$$

The angle θ_W is a parameter of the theory, the weak mixing angle. The above equality gives:

$$\imath gW_3T_3 + \imath g'\frac{1}{2}BY = \imath A\left[g\sin\theta_W T_3 + g'cos\theta_W\frac{1}{2}Y\right] + \imath Z\left[g\cos\theta_W T_3 - g'\sin\theta_W\frac{1}{2}Y\right]$$

For the term with the field A in front to be the electromagnetic field component: $\imath eQA = \imath e\left(T_3 + \frac{1}{2}Y\right)$ the constants g, g' must relate to the electric charge:

$$g = \frac{e}{\sin\theta_W}$$

$$g' = \frac{e}{\cos\theta_W}$$

The covariant derivative now includes a Z term:

$$D_\mu^Z = ig_Z Z_\mu(T_3 - x_W Q)$$

$$g_Z = \frac{e}{\sin\theta_W\cos\theta_W}$$

$$x_W = \sin^2\theta_W$$

The interaction term of the Lagrangian will give four conserved currents: $\overline{\psi}\imath\gamma^\mu D_\mu\psi$ which can be written in terms of isospin lowering and raising operators as: $T^\pm = \frac{T_1 \pm \imath T_2}{\sqrt{2}}$. One of the interaction currents must be the electromagnetic conserved current. The other three currents are given by the left handed isopin operators (which vanish on ψ_R):

$$\overline{\psi}\imath\gamma^\mu D_\mu\psi = eJ_e^\mu mA_\mu + \frac{g}{\sqrt{2}}\left(J_L^{+\mu}W_\mu^+ + J_L^{-\mu}\right) + g_Z J_Z^\mu Z_\mu$$

$$J_e^\mu m = \sqrt{2}\overline{\psi}\gamma^\mu Q\psi$$

$$J_Z^\mu = \overline{\psi}\gamma^\mu\left[T_3^L - x_W Q\right]\psi$$

$$J_{\pm W}^\mu = \sqrt{2}\overline{\psi}\gamma^\mu T_\pm^L\psi$$

The most critical problem of the above Lagrangian is that it relies on massless weak force mediators. A mass term for the weak force carriers would not be gauge invariant. It was stated earlier that gauge invariance leads to important physical consequences, but more importantly a gauge field can be renormalized. [49] Major hints for generating mass came from condensed matter physics. Bardeen, Cooper and Schrieffer described how a photon acquires mass in a superconducting medium [3]. Spin up and spin down electron pairs form a bounded state at low temperatures: Cooper pair. These bound states will act like bosons that aggregate in the ground state to give a non-zero vacuum expectation: $\langle 0|H|0\rangle = V$ where V is given by the sum over the wave functions of Cooper pairs in the ground state. P.W. Andersen then showed how the photons acquire mass by interacting with the electric field of Cooper pairs. [31]

In free space photons are described by the $U_{EM}(1)$ gauge field with a massless photon, here the vacuum corresponds to zero momentum and energy of the electromagnetic field and so also of the photon. Anderson used an analogy between a Dirac field of electrons and a fermi gas of electrons in a plasma. Note this system does not need be relativistic. Cooper pairs in the ground state form a 'plasma matter field' with a non-zero expectation value for the vacuum, and excitations for this field are thermal excitations of the cooper pair that can create a non-zero residual electromagnetic field. The free photon field couples to this 'plasma matter field' to give a massive vector field that has photons that propagate with an effective mass proportional to: $\frac{4\pi Ne^2}{M_e}$ which is called the plasma frequency. The effective mass scales with the density of Cooper-pairs, N. In other words, the electromagnetic wave traveling inside the plasma now has become a new combination of free space oscillating electric and magnetic fields and also oscillations of the cooper pairs in the 'plasma matter field', and the combined wave in the medium travels slower than the speed of light.

For Quantum Fields, Nambu and Goldstone proved that when any continuous symmetry is broken it gives rise to a massless scalar field with an associated Goldstone boson. As shown in Figure 2.4, a zero vacuum state for a quadratic potential has a minimum at $(0, 0)$ which trivially respects any symmetry. A non-zero vacuum of the scalar field will not be invariant under the symmetry. This principle can be demonstrated using a complex scalar field $\Phi = (\phi_1 + \imath\phi_2)/\sqrt{2}$ with a Lagrangian [35]:

$$\mathcal{L} = \partial_\mu \overline{\Phi} \partial^\mu \Phi - m^2 \overline{\Phi} \Phi$$

If Φ is constant only the potential energy $m^2\overline{\Phi}\Phi$ contributes to the Lagrangian. This function has a global minimum: $(\phi_1, \phi_2) = (0, 0)$. However for a positive sign in front of the potential energy there is no global minimum since the potential is unbounded from below. The potential can give stable ground states by introducing two parameters λ, μ:

$$\mathcal{L} = \partial_\mu \overline{\Phi} \partial^\mu \Phi - V(\Phi, \lambda, \mu)$$

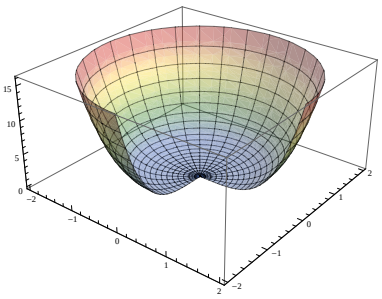
$$V(\Phi, \lambda, \mu) = \frac{1}{2}\lambda^2 |\overline{\Phi}\Phi|^4 - \frac{1}{2}\mu^2 |\overline{\Phi}\Phi|^2$$

Including the vacuum expectation value in terms of ϕ_1 defines it as an excitation above the vacuum expectation value $\langle|\phi_1|\rangle = \mu/\lambda$.

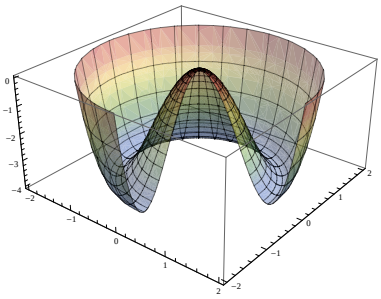
$$\phi_1' = \phi - \langle|\phi_1|\rangle$$

The ϕ_1' field will have a real mass term in the Lagrangian: $\mu^2 (\phi_1')^2$ while ϕ_2 is left massless. The shape of the potential is like the bottom of a wine bottle, the dregs collect in a circular well which corresponds to the vacuum expectation value as shown in Figure 2.4b. The shape of the potential in terms of the complex fields determines this behavior, ϕ_2 remains unchanged when moving along the circular equipotential which is shown in Figure 2.4c, in the mode defined as the Goldstone mode corresponding to a massless boson from symmetry breaking. The massive field ϕ_1 is defined as an excitation above the non-zero expectation value and acquires mass corresponding to the radius of the well as shown in Figure 2.4c labeled as the Higgs mode. The original $U(1)$ symmetry is lost because the vacuum expectation value of the ϕ_1 will not be invariant under the phase transform: $\Phi \to \Phi e^{i\alpha(x)}$. [35]

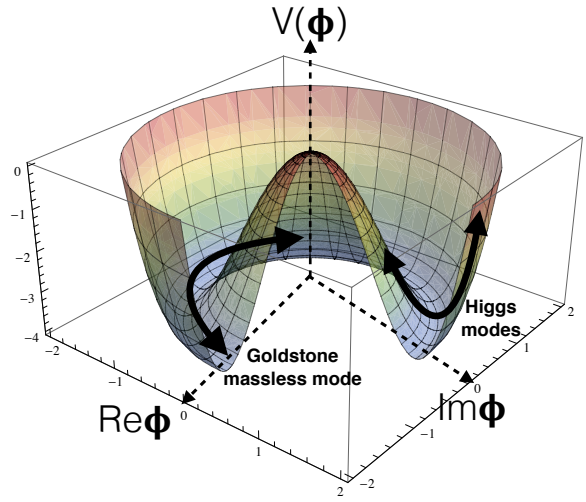
The Higgs mechanism will be the procedure to break the electroweak symmetry of $SU(2)\otimes$U(1). This will generate the masses of the W^+, W^-, Z^0 while leaving the photon massless and also give rise to an additional massive boson from the scalar field resulting from the symmetry breaking: Higgs boson. A similar procedure is applied to SU(2) in the Standard Model with the complex scalar doublet. The most important feature of this Higgs symmetry breaking mechanism, which was not present in the U(1) symmetry breaking, is that the gauge bosons that arise from the symmetry



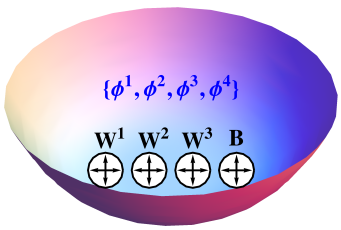
(a) Quadratic Scalar Potential



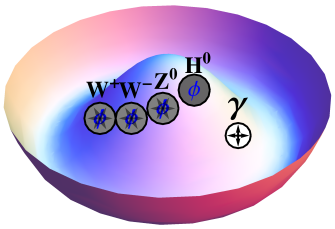
(b) Quartic Scalar Potential for Symmetry Breaking



(c) Massless and Massive modes



(d) Massless Particles in a Quadratic Potential



(e) Massive Particles with a Broken Symmetry

Figure 2.4: The first figure a scalar potential with a zero vacuum state defined as a paraboloid $|\Phi|^2$ in the complex plane. Introducing a symmetry breaking parameter gives a quartic potential with a non-zero vacuum state along the brim shown in the second figure. This quartic potential shaped like a Mexican hat gives one Goldstone mode for a massless particle and massive Higgs mode. The result is shown in the last two figures, a paraboloid potential would trap all particles in a massless state, but the symmetry breaking allows for a massless photon and massive vector bosons, and introduces a Higgs boson associated with the scalar potential.

breaking will not all be massless, only the photon remains massless while the W^+, W^-, Z^0 bosons acquire mass. Also the scalar field will have a massive boson associated with it, the Higgs boson. Instead of a single complex scalar field, SU(2) will have a complex scalar doublet like

$$\phi^+ \equiv (\phi_1 + \imath\phi_2)/\sqrt{2}$$

$$\phi^0 \equiv (\phi_3 + \imath\phi_4)/\sqrt{2}$$

The most general potential that is renormalizable is again written:

$$V(\Phi) = \mu^2 |\Phi|^2 + \lambda^4 |\Phi|^4$$

$$\mathcal{L} = \frac{1}{2}\partial_\mu \Phi \partial^\mu \phi - V(\phi)$$

The potential is at its minimum when $|\Phi|^2 = \phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 = \dfrac{\mu^2}{2\lambda}$ which is the vacuum expectation value or vev. As was done in the complex scalar field case one of the four scalar fields can be defined as an excitation above the vev: $\langle 0|\phi_3|0\rangle = \sqrt{\dfrac{\mu^2}{\lambda}} = v$ while the other fields have zero vacuum expectation value, then we can define $H(x) = \phi_3 - v$. This corresponds to a rotation of the field Φ in iso-spin space.

$$\Phi \equiv \begin{pmatrix} 0 \\ (v + H(x))/\sqrt{2} \end{pmatrix}$$

This particular rotation in the isospin space of SU(2)⊗U(1) has important consequences for the Higgs mechanism. The three additional fields $\phi_1\phi_2, \phi_4$ that would give rise to massless Goldstone bosons are eaten away by the rotation. Only one massless Goldstone boson remains corresponding to the photon as shown in Figure 2.4e, while the other massive bosons correspond to the Higgs modes. The Φ field has only a neutral component that does not couple to the electromagnetic field. The $SU_{Iso}(2) \otimes U_Y(1)$ symmetry will be broken by the non-zero vev, and there will now be three non-zero mass eigenstates, and only one massive boson from the excitations of the scalar field. This is seen by operating the covariant derivative: $D_\mu = \partial_\mu + \imath g W_\mu^i T^i + \imath g' \frac{1}{2} B_\mu Y$ on the vacuum state:

$$|D_\mu \langle 0|\Phi|0\rangle|^2$$

$$\frac{1}{8}\left| \begin{pmatrix} gW_\mu^3 + g'B_\mu & g\left(W_\mu^1 - \imath W_\mu^2\right) \\ g\left(W_\mu^1 + \imath W_\mu^2\right) & -g\left(W_\mu^3 + g'B_\mu\right) \end{pmatrix} \begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2$$

For the off diagonal terms it is useful to write $W^\pm \equiv \left(W^1 \mp \imath W^2\right)/\sqrt{2}$ for the solution:

$$= \frac{1}{8}v^2\left(gW_\mu^3 - g'B_\mu\right)^2 + 0\left(g'W_\mu^3 + gB_\mu\right)^2 + \left(\frac{1}{2}gv\right)^2 W_\mu^+ W^{\mu -}$$

Now we can to relate the above terms to the mass eigenstates of the Electroweak fields: $Z^0 = \left(gW_\mu^3 - g'B_\mu\right)$ gives the mass $M_Z = v^2\dfrac{\sqrt{g'^2 + g^2}}{2}$ and the last term above gives $M_W = \dfrac{1}{2}gv$. Note

also the photon field $A^{\mu} = gW_{\mu}^{3} + g'B_{\mu}$ corresponds to mass eigenstate of zero due to only the neutral component of Φ in the vacuum state.

The vacuum expectation value can be reverse engineered from the gauge boson masses: $v = 2M_W \sin\theta_W/e$. The above equation also has a physical scalar field as the Higgs field:

$$\frac{1}{2}(\partial H)^2 - V\left[\frac{1}{2}(v+H)^2\right]$$

$$= \frac{1}{2}(\partial H) - \frac{1}{2}\left(-2\mu^2\right)H^2 + \frac{1}{4}\mu^2 v^2\left[-1 + \frac{4H^3}{v^3} + \frac{H^4}{v^4}\right]$$

The above describes a physical field with a scalar boson:$M_H = \sqrt{-2\mu^2}$, the last terms describe self-couplings of the boson. The self-interactions can cause the Higgs boson mass to become very large but quantum corrections from virtual particle loops can decrease the mass so that it is observable.The Higgs boson mass is the only free parameter in the theory, and in this thesis it will be shown to be: $M_H = 124.73GeV$.

A Yukawa interaction between the Φ field and fermion fields will generate the masses of the fermions. For example, the electron mass can be generated with the Lagrangian:

$$\mathcal{L} = -G_F\left[\overline{eR}\left(\overline{\Phi}\ell_L\right) + \left(\overline{\ell_L}\Phi\right)e_R\right]$$

G_F is a coupling constant and a lepton doublet $\begin{pmatrix} \nu_e \\ e \end{pmatrix}$ with Φ doublet $\begin{pmatrix} - \\ \phi^0 \end{pmatrix}$ which gives:

$$\mathcal{L} = \left(G_e v/\sqrt{2}\right)\overline{e}e - \left(G_e\sqrt{2}\right)H\overline{e}e$$

$$M_F = \left(G_e v/\sqrt{2}\right)$$

and a coupling of the leptons to the Higgs with a strength proportional to the mass:$G_e = \sqrt{2}m_e/v = 2.9\times10^{-6}$. The above Lagrangian keeps the neutrinos massless, since there is no observed: ν_R. The quark masses can be generated in a similar way. Not just with one multiplet Φ but also a conjugate Higgs doublet with a neutral member in the upper component :

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}\overline{\Phi} = \begin{pmatrix} \phi^0 \\ \phi^- \end{pmatrix}$$

Again the charged components are zero for the vacuum. The field Φ and its conjugate $\overline{\Phi}$ operate on the upper and lower components of a quark doublet like u, d:

$$\mathcal{L}_Y^Q = -G_d\left(\overline{u},\overline{d}\right)\overline{\Phi}d_R - G_u\left(\overline{u},\overline{d}\right)\overline{\Phi}u_R + conjugate$$

Again the mass will be $M_Q = \left(G_Q v/\sqrt{2}\right)$ and is determined by the vacuum expectation value of the Higgs field.

This section has described how a scalar field is predicted by the Standard Model to give mass to fermions and massive gauge bosons. An SU(2) Yang-Mills Lagrangian predicts a triplet of gauge bosons, but requires them to be massless. The Standard Model electroweak Lagrangian takes into account the handedness of particles in $SU_{Iso}(2)\otimes U_Y(1)$. Using spontaneous symmetry breaking with a scalar potential in the electroweak Lagrangian generates the mass of the gauge bosons and adding Yukawa terms generates the mass of fermions. The symmetry breaking predicts a scalar field with a gauge boson, the Higgs boson. The free parameter in the above theory is the mass of the Higgs boson which will be described in this thesis. The next section will describe the production mechanisms and decay processes of the Higgs boson, which will allow it to be produced in proton-proton collisions and observe its decay into Standard Model particles, in particular two photons.

2.2 Higgs Production and Decays

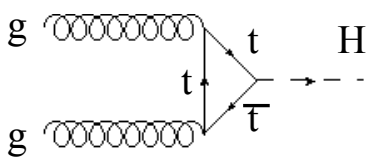
Section 2.1.3 described the theory of the existence of a Higgs field to which massive vector bosons and fermions couple to acquire mass. The main observable for this field will be the boson associated to it, whose mass is not predicted. This section will describe how a Higgs boson can be produced from Standard model particles in high energy proton-proton collisions and also how the Higgs decays to an observable final state. For this thesis, the decay mode investigated is the two photon decay. The production of the Higgs boson can be divided according to whether or not the Yukawa coupling to fermions exists or whether Higgs couples only to massive vector bosons (fermiophobic Higgs). For the Higgs coupling to fermions, the heaviest fermion, the top dominantly contributes to Higgs production. For the Higgs coupling to massive vector bosons, the Higgs is produced by the fusion of virtual vector bosons or produced along with a vector boson. This section will describe the main production mechanisms. The Higgs mass is the one free parameter of the theory. It determines the couplings to other particles, so determines the rate of decay and the rate of production. The Higgs does not couple to photons directly but instead decays with a virtual top loop at a low mass below 130GeV. This will be the important discovery channel of the Higgs for this thesis.

An upper limit on the Higgs mass can be postulated from the required behavior of the vector boson scattering amplitudes. An important property of the S-matrix of scattering amplitudes is that it is unitary which ensures that the total probability is one. Thus, for the electroweak theory to be renormalizable the scattering amplitudes must respect unitarity. A massive spin 1 particle with 4-momentum $q^\mu = \left(\frac{q}{M}, 0, 0, M\right)$ is described by polarization vectors:

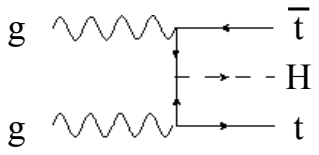
$$\varepsilon^{\pm 1} = \mp\left(0, 1, \pm\imath, 0\right)/\sqrt{2}$$

$$\varepsilon^{0} = \left(|q|, 0, 0, E\right)/M$$

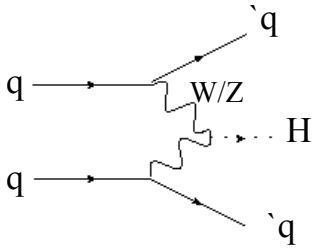
For massive vector bosons the contributions from longitudinally polarized vector bosons grows fast with energy. $\varepsilon_\mu^L \to \frac{q_\mu}{M}$ as $q \to \infty$ which will violate unitarity. However, if the Higgs is exchanged as a virtual gauge boson between the scattering vector bosons a delicate cancellation occurs that preserves renormalizability. At high energies the Higgs exchange diagrams dominate the scattering processes and preserves unitarity as longs as $M_H < 1TeV$. However, if this limit is greatly exceeded then new phenomena should emerge to restore renormalizability. [1]



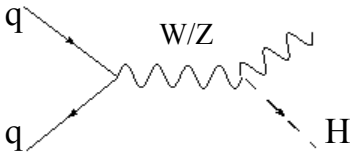
(a) $gg \to H$



(b) $gg \to ttH$



(c) $qq \to Hq'q'$



(d) $qq \to VH$

Figure 2.5: The four Higgs production mechanisms at the LHC

Roughly half of the momentum of a proton is carried by the gluons and the other half by quarks. One can imagine a possible mechanism for Higgs production in proton collisions to be a heavy

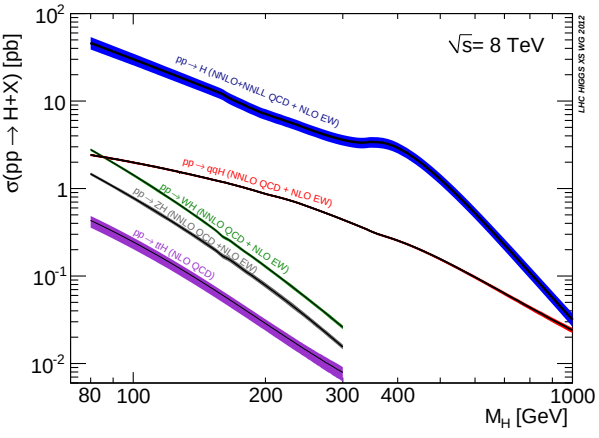
quark annihilating with a heavy anti-quark. The chance to find a heavy quark in a proton and simultaneously a heavy anti-quark in another is negligible. [34] The fraction of momentum of the gluons is large at high center-of- mass collision energy. The gluons do not couple to the Higgs directly because they are massless. However, the gluons do split into quark and anti-quark pairs during high energy collisions. The quarks from the gluons can induce coupling to the Higgs via a virtual loop of quarks [34]. The rate of production in these gluon-gluon processes depend on the Yukawa coupling of a quark to the Higgs, which is proportional to the quark mass and also on m_H^2/s where s is the square of the center-of-mass energy of the proton-proton collision. The coupling of the Higgs to the virtual particle loop scales with the ratio $\lambda_q = m_q^2/m_H^2$ [4], so the dominant contribution to the gluon-gluon process comes from the heaviest quark, the top quark. The top quark is ≈ 40 times heavier than the next heaviest bottom quark.This production is shown in Figure 2.5a. Also each gluon can radiate a top and anti-top pair where a set of top anti-top anhilates and produces a Higgs while the other set is in the final state. This is the Higgs production associated with top production shown in Figure 2.5. In this process it is possible to obtain a direct measure of the Yukawa coupling between top and the Higgs but in pp collisions it has the smallest production cross-section.

The precision of the rate of Higgs production in pp collisions requires a precise knowledge of the gluon distribution at low x, the fraction of parton momentum to the full proton momentum. The gluon-fusion Higgs production also requires knowledge of the QCD loop corrections to the heavy top loop. Each level of correction (next to leading order (NLO), next-to-next-to leading order NNLO) predicts a larger gluon-fusion Higgs production cross section and proves to be non-negligible. [1] The unknown higher order effects are the main source of theoretical uncertainty for gluon-fusion Higgs production. [1].

The quarks and anti-quarks can annihilate to produce vector bosons which couple to the Higgs. The Higgs can be produced by fusing or annihilating vector bosons. Quarks from the colliding proton beams can emit virtual W or Z bosons when they interact weakly. The W^+W^- virtual pair can annihilate to give a Higgs boson or the ZZ can fuse to give the Higgs. In the final state, the p_T of the Higgs is balanced by the quarks that emit the virtual W^+W^- or ZZ. For this production the scattered quarks are emitted in the forward direction along the initial pp beamline, which hadronize into jets as an additional signature for this Vector Boson fusion process. Also the transverse momentum of the Higgs is large for the W^+W^- annihilation process [9]. The measured rate of vector boson fusion offers a measure of the coupling between the Higgs and the W, Z. The vector boson fusion process, VBF, has the second highest production cross-section in pp collisions an order of magnitude lower than the gluon fusion process. This production mode has an appreciable cross-section over a range Higgs masses and becomes comparable to the gluon-fusion process at large

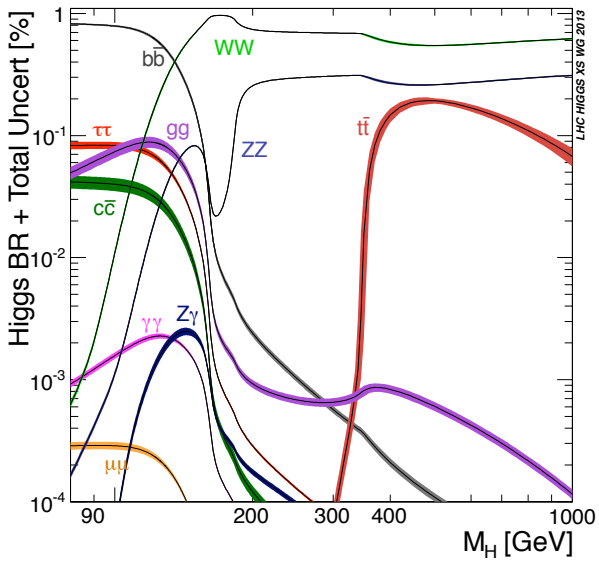
Higgs mass. [36]

A Higgs boson can also be radiated directly from a virtual vector boson. First a quark-antiquark pair annihilates to produce a vector boson W, Z which can then emit a bremsstrahlung-like emission Higgs boson before it decays. This would be the dominant process at an e^+e^- collider [4]. The QCD corrections for this process at NLO can be inferred from Drell-Yan production and are calculated to NNLO giving very small theoretical uncertainties. [1]. The rate of production for this process can be combined with the VBF to indicate the vector boson coupling to the Higgs.



[41]

(a) Higgs Production Cross Sections vs. Higgs mass



[41]

(b) Higgs Decay Widths

Figure 2.6: Higgs Production Cross Section and Branching fractions [41]

Since the photon is massless, as with the case of gluons, it does not couple directly to the Higgs but instead the coupling is generated with virtual loops of massive particles. The photon loop has

contributions from the heaviest quark and also from the W. For masses much larger than the Higgs boson mass, virtual particles that contribute to the loop do not decouple since their couplings to the Higgs boson grow with the mass. These decays are thus extremely interesting since their strength is sensitive to scales far beyond the Higgs boson mass and can be used as a possible probe for new charged and/or colored particles whose masses are generated by the Higgs mechanism and which are too heavy to be produced directly. [29] The suppression by electroweak or strong coupling make this loop decay important for M_H below 130 GeV. The partial decay width to photons can be parameterized by $\tau = M_H^2/4M_i^2$:

$$\Gamma\left(H \to \gamma\gamma\right) \propto \left| \sum_f N_c Q_f^2 A_{1/2}^H\left(\tau_f\right) + A_1^H\left(\tau_W\right) \right|^2$$

The fermions are summed over. Since the Higgs coupling to fermions scales with the fermion mass the top has the largest contribution for the fermion term. The fermion form factor: $A_{1/2}^H$ has opposite sign to the bosonic form factor A_1^H so the top loop interferes with the W loop contribution. [29] The W-loop gives the larger contribution since the Higgs coupling to the W is proportional to M_W^2 and the fermions Yukawa term scales with the fermion mass. The two main contributing diagrams are shown in Figure 2.8.

The Higgs decay width depends on the mass of the Higgs boson. The subject of the thesis will be the Higgs decay to photons which is possible for a low mass Higgs $M_H \lesssim 130$ GeV. For this mass range the mass resonance is very narrow $\Gamma_H < 10$ MeV. The main decay modes are to the heaviest available quark, the b quark anti-quark pair with a branching ratio of 90% followed by the next heaviest the charm. The τ lepton which is heavier than the strange quark gives a branching fraction 5%. [1] The loop induced decay to photons is rare with a small branching fraction less than a percent: 0.3%. The branching fractions for the Higgs decay are shown in Figure 2.6 and the total width is shown in Figure 2.7.

Despite the small contribution to the decay width, the $H \to \gamma\gamma$ channel is an attractive discovery channel for low mass Higgs. Unlike the more dominant decays like $b\bar{b}$ and $c\bar{c}$ this decay is not swamped by a large dijet background from pp collisions. The background is dominated by photon-pairs from QCD processes (irreducible background), and jet-jet and γ+jet events where jets are mis-identified as photons. The natural width of the Higgs less than 10MeV is spread out by the resolution of the detector to about 1GeV and can be observed as an excess on top of a the smoothly falling background of non-resonant photons.This will be the subject of this thesis and will be described in Chapter 5.

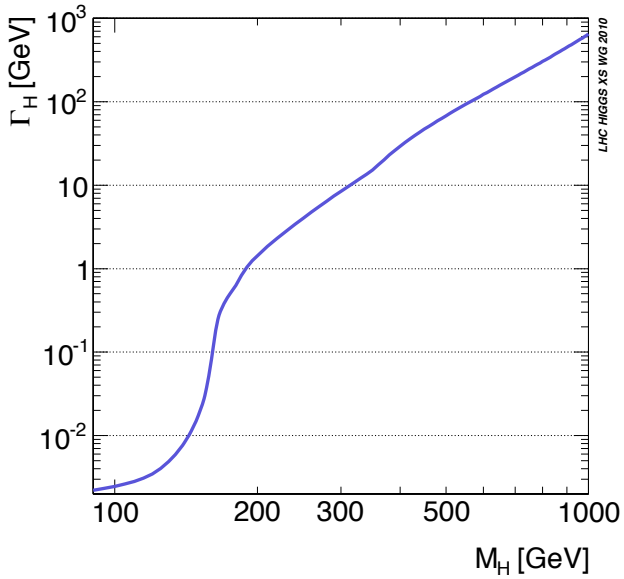


Figure 2.7: Higgs total decay width vs. Higgs Mass [41]

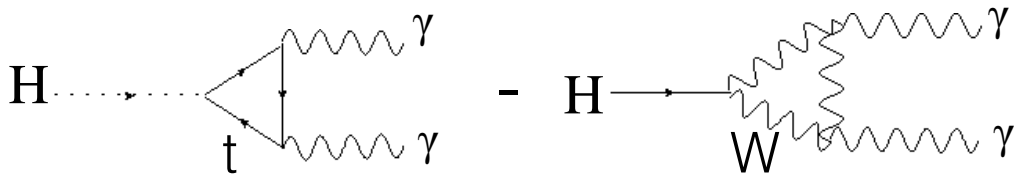


Figure 2.8: Higgs Decay to photons via a heavy quark loop and also an interfering contribution from a vector boson loop.

Chapter 3

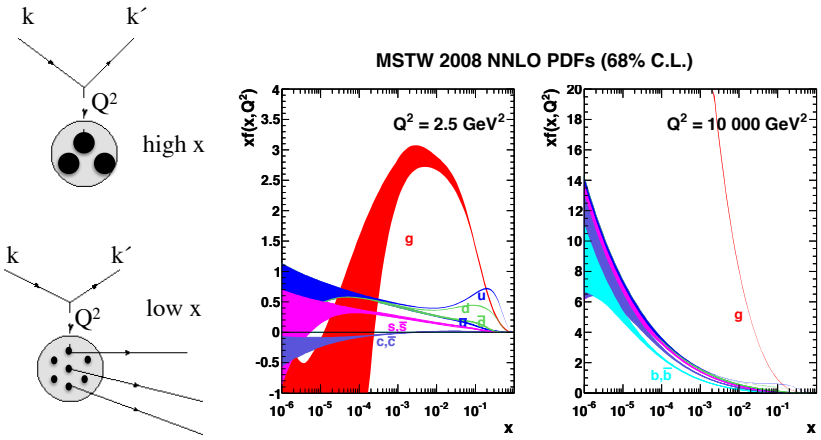
The Large Hadron Collider and the Compact Muon Solenoid Detector

Chapter 2 described the Higgs boson as an essential part of the Standard Model to explain electroweak symmetry breaking. Section 2.2 described how the Higgs boson can be produced and how it can decay to Standard Model particles. This chapter will describe the Large Hadron Collider (LHC), which is a proton-proton collider built with one major objective being to produce and observe the Higgs boson. The focus of this thesis will be the observation of the Higgs decaying to two photons in the Compact Muon Solenoid detector (CMS). Section 3.1 will describe the LHC which is to date the largest particle accelerator in the world. The LHC collides protons at a high center of mass energy and a rapid collision rate to produce the Higgs. Section 3.2 will describe the CMS detector with particular emphasis on the Electromagnetic Calorimeter (ECAL), which is the main tool in observing the Higgs decaying into photons and reconstructing the Higgs mass.

3.1 Large Hadron Collider

3.1.1 Proton-Proton Collisions

The structure of a proton can be probed in high energy lepton proton collisions. The momentum transfer $q^2 = p^2_{final} - (E_{final} - M_p)^2$ and the energy transfer $\nu = (E_{final} - M_p)$ can be used to define structure functions of the proton $F\left(q^2, \nu\right)$. At large q^2 and ν, if the structure function must remain finite, it can depend only on a finite ratio of the two quantities. The structure functions become functions of only the ratio $x = \frac{q^2}{2M_p\nu}$, which was found by Bjorken. [5] The lepton then interacts only with the point-like constituents of the proton that carry a fraction of the total proton momentum x. For $x = 1$ the lepton will have only elastically collide with the proton while $x < 1$ the lepton will probe the structure of the proton and it will disintegrate. The proton contains valence quarks uud and a sea of quark anti-quark pairs of all flavors and gluons. The two types collisions at high and low x are shown in Figure 3.1. At high x the lepton would interact mainly with the valence quarks while at low x it starts to probe the sea. Experimentally probing the proton disintegration and integrating the structure functions of the quarks and anti-quarks over x gives only 50% of the



(a) Elastic and Deep Inelastic Scattering

(b) Parton Distribution Function

Figure 3.1: Deep Inelastic Scattering and Parton Distribution Function at low $Q^2 = 2.5TeV$ and high $Q^2 = 10^4 TeV$ [44].

total proton mass, the remaining half is carried by the gluons. [46] Figure 3.1 shows the structure functions, $xF\left(x, Q^2\right)$ plotted vs. x the fraction of the total momentum imparted to a parton. At low x the gluons dominantly contribute, so dominantly spill out of the proton with fractional momentum x. As described in Section 2.2, the dominant production of the Higgs boson is from gluons. The large spill of gluons at low x, which carry a fraction of the total energy of the protons, is important for producing Higgs events across a large mass range.

The inelastic proton-proton collision cross-sections vs. center of mass energy show a large gap between the total inelastic and one of the more abundant cross-sections for a hard interaction that produces b-jets. This gap which is 60 to 70 mb is dominates the total inelastic cross-section of 100mb (see Figure 3.1) [28]. Most of the time in the proton-proton collision, the protons 'ooze' through each other and produce many soft particles uniformly distributed in rapidity [33]. The bulk of such events can be described as minimal bias events because they can be found with a minimal set of trigger requirements. The underlying event are the debris that accompany the hard process. The "pile-up" are the soft collisions from multiple proton interactions. Overlapping particles from these processes will have a large effect on measuring photon energy and the Higgs decay vertex. The number of interactions is measured to follow a Poisson probability with an average rate of 20 interactions per event at $\sqrt{s} = 8$TeV. It will be measured as the number of reconstructed interaction points in an event and also as the average particle density in the event. [10]

To produce a mass resonance in a proton-proton collision the momentum transfer between two partons (with fractional momentum x_1, x_2) is $\hat{s} = x_1 x_2 s$ where s is the square of the center of

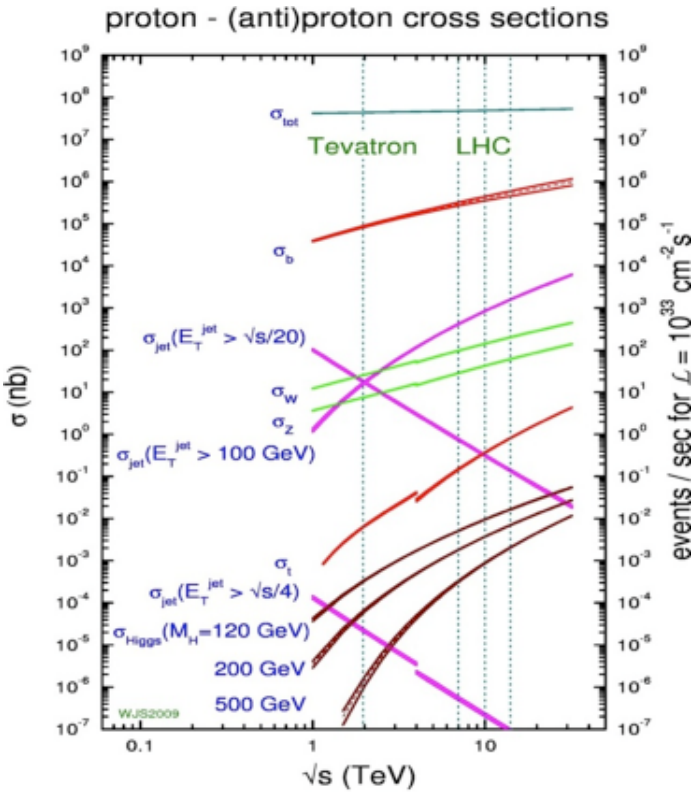


Figure 3.2: Total Inelastic Proton-Proton collision cross section

mass energy, then $\hat{s} \geq M^2$ to produce a particle of mass M. The energy of the new particle is $E = (x_1 + x_2)\sqrt{s}/2$ and the logitidunal momentum $p = (x_1 0 x_2)\sqrt{s}/2$, which gives the defintion of the rapidty y of the particles $e^y = \sqrt{\frac{x_1}{x_2}}$. For two partons of similar x $e^y \to 1$ and $y \to 0$, so the new particle will be produced at central rapidity close to perpendicular to the collision axis. While if one of the partons has significantly smaller momentum than the other, y can be a large positive or negative value so M will be produced longitudinally closer along the direction of the parton with larger momentum. [28] The CMS detector described in Section 3.2 is designed to cover a large rapidity range. The cylindrical geometry of the detector captures particles produced in softer interactions in the endcaps, while the harder interactions are captured in the barrel of the cylinder. The geometry of the cyldrincal detector can be described in terms of R and polar angle ϕ and also can be described in terms of an azimuthal angle and a polar angle: η, ϕ. The radius of the cylinder measured perpendicular to the beamline is $R = \sqrt{x^2 + y^2}$. The polar angle ϕ gives the angular position in the x-y plane. The pseudo-rapidity is computed from the angle along the $R - Z$ plane and is computed as $\eta = -\ln\left(\frac{\theta}{2}\right)$. This relates to the rapidity, y, described above in the relativistic limit $E \sim |p|$ so large rapidity corresponds to softer produced particles at large η with longitudinal momentum along the beamline. Particles from harder parton interactions will be produced more

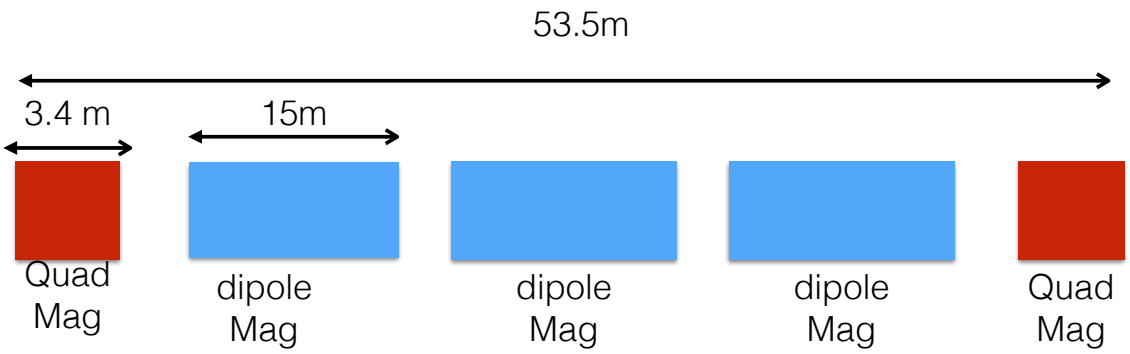
centrally in the detector at η closer to zero.

3.1.2 Proton Accelerator

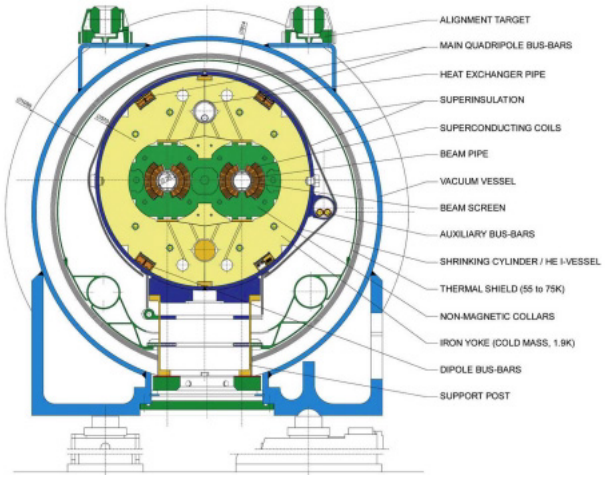
The key design objective of the LHC is to explore Standard Model and beyond Standard Model phenomena for pp collisions at a high center of mass energy and also a rapid collision rate. [6] The center of mass energy ($\sqrt{s}$) for pp collisions is not precise as in e^+e^- collisions where the annihilation goes entirely into producing new particles in the final state. Instead the increase of $\sqrt{s}$ scales the parton distributions functions so that there is a large probability of a hard process between partons as described in 3.1.1. A large $\sqrt{s}$ for the collision energy requires high powered dipole magnets to accelerate the protons in circular boosting loops. The spread of the beam about the orbit of the tunnel is minimized by using quadrapole magnets for focusing. [6] The advantage of accelerating protons as opposed to electrons is that the loss of energy in the beam due to synchrotron radiation is much smaller for protons. The energy loss per turn in the accelerator $P \propto \frac{1}{m^4}$ is greatly smaller for the proton $\left(\frac{m_e}{m_p}\right)^4 \sim 10^-13$. For $\sqrt{s} = 7TeV$ for pp collisions can be achieved in a reasonably sized ring with powerful dipole magnets. However, the power emitted by the proton beam is not negligible and can produce thermal energy which can affect the superconductivity of the magnets. So the power is emitted into a cryogenic liquid helium cooling system to keep the magnets at 1.9 K. [39] This section will describe the technology used to accelerate protons to the high center of mass energy of $7TeV$ which was later ramped up to $8TeV$ at the LHC.

The Large Hadron collider is built within a 27km tunnel [39]. The curved sections of the tunnel are used to install the bending dipole magnets, and the remaining sections are used to install the quadrapole focusing magnets. The periodic structure of the magnet installation is shown in Figure 3.3. The LHC has 8 arc sections that consist of 23 cells. Each cell has six dipole magnets to deflect the particles and two quadrapole magnets with opposite polarities to focus the beam in two planes to keep it in the beam orbit. The cell pattern is shown in Figure 3.3 as well as the fields of the two types of magnets. The oppositely circulating beams circulate in separate beam pipes installed in the twin aperture magnets as shown in Figure 3.3. The Figure also shows the bulk of structure required to maintain the thermal stability of the superconducting magnets. The magnets are made of Niobium-Titantanium which is superconducting at a set of operational parameters. A magnetic field of 8.4 Tesla requires a very low operating temperature at 1.9K, so the magnets are cooled with liquid Helium. The protons are boosted using the synchrotron principle, where a proton is boosted if its velocity is in phase with a radio frequency electromagnetic wave in an RF cavity. At the LHC this boosting is done in stages using the booster loops in the different rings shown in Fig. 3.4. The protons are boosted in each of circular loops. The protons cannot be accelerated from

zero momentum so they are injected from the LINAC with some velocity. Then in each booster loop the protons are accelerated per turn with the momentum kick shown in Fig. 3.4. The protons are boosted using the synchrotron principle, where a proton entering a radio frequency cavity is kicked by an oscillating electric field if it is not in phase with the radio frequency. The radio frequency is at 400MHz resulting in 2808 proton bunches in the LHC ring. This is shown in Figure 3.4, protons are grouped together into bunches with a spacing of 25ns. When bunches from counter-rotating beams are crossed, protons within the bunches collide.



(a) Magnet Structure

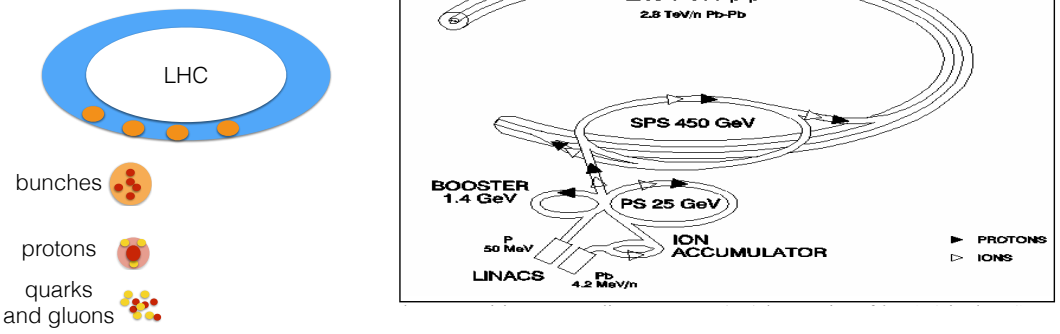


(b) Magnet Cross-Section

Figure 3.3: Periodic structure of quadrapole and dipole magnets installed in the LHC arcs. The cross-section shows two apertures at the center where the two counter rotating proton beams pass.

The number of events for any process can be given by $N = \mathcal{L}\sigma_{prod}$ where σ_{prod} is the Higgs production cross section and the Luminosity is given by $\mathcal{L} = \dfrac{f_{rev} n_b N^2}{\sigma_x \sigma_y} F\left(\phi, \sigma_x, \sigma_y, \sigma_s\right)$ and is determined by

the parameters of the proton beam. In the numerator f_{rev} is the revolution frequency 112246 Hz based on the tunnel circumference of the tunnel 26.8km. The number of bunches n_b which is 2808 bunches (particle packages) each spaced 25 ns apart and N is the number of particles in each bunch 1.15×10^{11} p per bunch. The denominator $\sigma_x\sigma_y$ is the beam size at the interaction point in the horizontal and vertical plane which are $16\mu m$. F is a geometric factor that accounts for the angle of crossing of the beams ϕ, the bunch size $\sigma_s = 7.5cm$ and the bunch spread $\sigma_x\sigma_y$ to give an interaction length. The crossing angle is small $320\mu Rad$ to give a large interaction length that can give multiple collisions in the same bunch crossing described in Section 3.1.1. [6] [7]



(a) LHC Booster Loops

(b) LHC Bunches

Figure 3.4: The LHC boosting loops that boost the proton beams to high energies by giving a 'kick' at a particular RF frequency. The RF frequency groups bunches of protons within the beam.

3.2 Compact Muon Solenoid

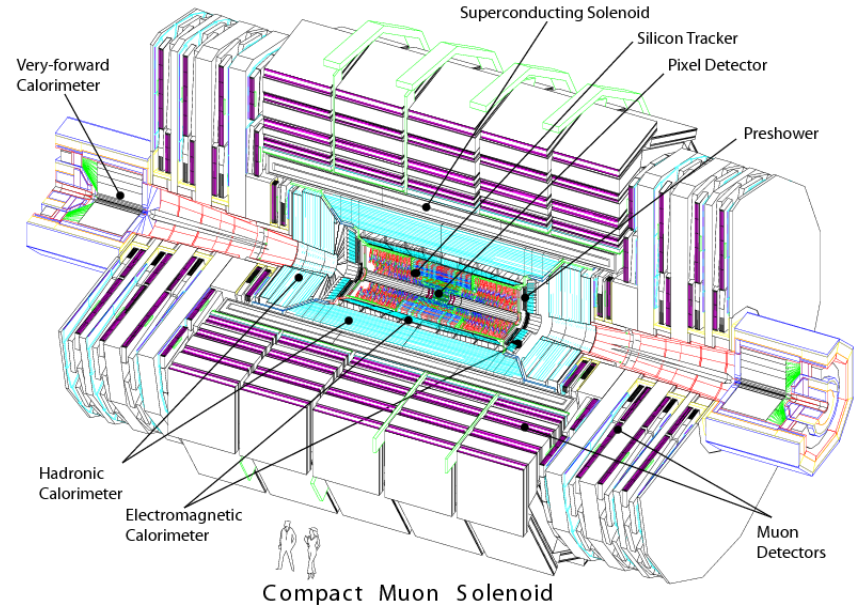
The Compact Muon Solenoid is one of the general purpose detectors at the LHC. This section will describe the sub-detectors in CMS which are responsible for reconstructing the 4-momentum of the final state particles. There will be particular emphasis on the electromagnetic calorimeter and the reconstruction of the photons, since the reconstructed Higgs mass will rely on the resolution of the photon energy. The tagging of production mechanisms will rely on multiple sub-detectors within CMS which will be described. Figure 3.5 shows an expanded view of the CMS detector. The detector is optimized to reconstruct the four-momentum of stable particles produced from the collision point. This requires a measurement of the trajectory of particles and their energy. At the heart of the detector is a 4 Tesla magnet that is 13m long with a 5.9m inner radius that provides a solenoidal field to pull apart the many charged particles produced in the multiple proton collisions recorded in an event. Within the solenoidal magnet is the silicon micro-strip tracker and the electromagnetic

calorimeter. The micros-strip tracker is used to reconstruct the three dimensional trajectories of charged particles within the magnetic field. The tracker also consists of silicon pixel detectors very close to the interaction point to give fine grained separation of tracks from the collisions. The electromagnetic calorimeter, ECAL, is made of transparent Lead-Tungstate crystals, $PbWO_4$ which capture electrons, positrons, and photons giving a measurement of their energy from the scintillations of the crystal . Surrounding the ECAL is the Hadronic Calorimeter which samples the energy of a cascade of hadronic particles. The cascades are caused by layers of brass and the energy is sampled by scintillators between the brass. Outside the solenoid are the muon chambers, which measure the trajectories of muons as they exit the detector. The description of the detector will be described in sections, Section 3.2.1 will describe the measurement of charged particles, Section 3.2.2 will describe ECAL, and Section 3.2.3 will describe the reconstruction of hadronic cascades.

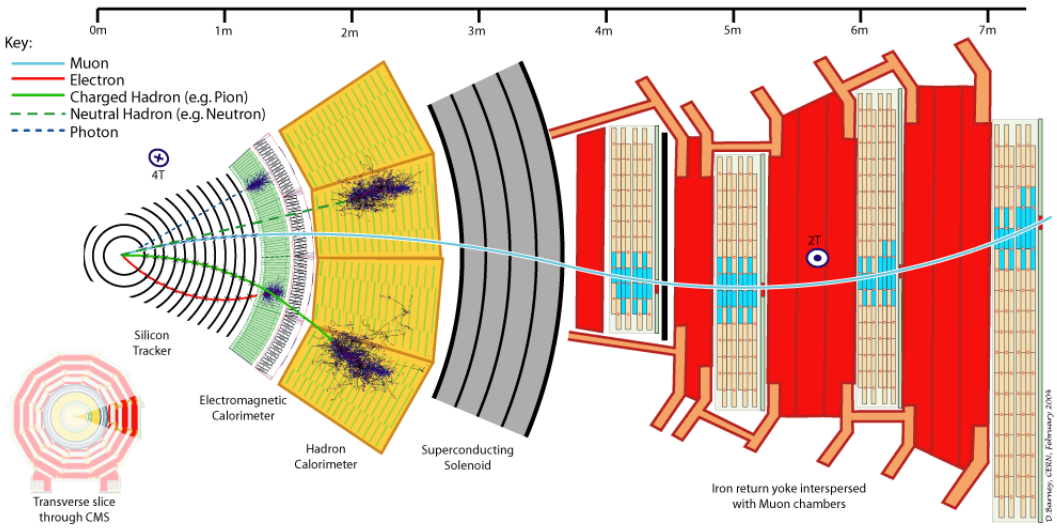
3.2.1 Charged Particle Detection

This section will describe the components of the CMS detector that separate charged particles and provide a measure of their momentum. The solenoidal super-conducting magnet at the heart of the detector pulls apart charged particles produced in the detector and also the bend of the particles in the field determines their charge and momentum. This bend is measured by tracking the trajectory of the particle as it bends in the field. The charged particle momentum resolution at CMS relies on a 4 Tesla super conducting solenoid. The solenoid is 13m long with a diameter of 5.9m. The bore of the magnet is large enough to house the tracker and the calorimeters. The return field of the solenoid is saturated in 1.5m of iron allowing for 4 muon stations to be created. The stations consist of layers of aluminum drift tubes in the barrel region and cathode strip chambers in the endcap region. In both the barrel and endcap regions a complementary system of resistive plate chambers is also added. The overall dimensions of the CMS detector are relatively compact: length 21.6 m and diameter of 14.6m, which is what the 'C' stands for! [17] One of the reasons for the compactness of the detector is the large field that pulls particles apart so that they can be detected without overlap in the sub-detectors. The field combined with the sub-detectors make it possible to use reconstruction algorithms like Particle Flow which will be described in Section 4.3. Combining the magnet with the large Silicon Tracker described in this section gives excellent pattern recognition of the trajectories of charged particles and so excellent momentum resolution.

The required performance of the magnet is determined by the required momentum resolution muon system to measure very narrow mass peaks like the Higgs. The muon momentum of narrow final states can have momentum of $\approx 1TeV/c$ and require a resolution of $\frac{\Delta P}{P} \approx 10\%$. [17] A large bending power on charged particles can be achieved with a modestly sized solenoid. Detection



(a) Expanded view of CMS



(b) Expanded view of CMS

Figure 3.5: The expanded view of the Compact Muon Solenoid Detector.

of charged particles makes use of their electromagnetic interaction when passing through matter. In material, particles can ionize the atoms of the material or can cause the emission of transition radiation. First the tracker will be described which uses silicon to detect transition radiation. The end of the section will describe the Muon chambers which make use of ionization detection.

The silicon tracker is used to measure the momentum of charged particles like the muon as well as shorter lived particles that are stopped in the calorimeters. The tracker is a semiconductor detector that captures transition radiation. Semiconductor detectors work as solid state ionization chambers. A silicon wafer is kept between two electrodes generating an electric field. A transitioning particle creates electron-hole pairs by exciting electrons from the valence band into the conducting band. The electrons can then create secondary electron-hole pairs. The number of electrons is larger than that from gas chambers, but the electrons need to be collected more rapidly before they recombine with holes. A diode with a p-n junction is formed by doping the silicon wafer with electron acceptor impurities (p-type) and also an n-type zone with electron doner impurities. A layer of charges form at the interface. Conducting electrons migrate to the acceptors on the p-side leaving holes in the n-type zone. This causes a surplus of negative charge on the p-side and a surplus of holes on the n-side that is countered by an electric field. The electric field pushes back electrons to the n-side creating a layer where free charged carriers are depleted. If the p-side is connected to a negative terminal and the n-side is connected to a positive terminal, a reverse bias voltage is applied which widens the depletion region as free charge carries move toward the terminals and away from the depletion zone. The strength of the electric field in the depletion zone will increase with the voltage applied. A charged particle transitioning through the material will create an avalanche of electron-hole pairs in the depletion region that can be collected before they recombine. The silicon microstrip detectors and pixel detectors make use of this principle. [38]

The solenoidal field bends charged particles in helical trajectories, so that softer low p_T particle are pulled into tight rings with small radius while high p_T particles form curved trajectories with radius proportional to the p_T. The momentum of a particle in the tracking region can be determined by a curved segment of a particle's trajectory along a length L in the tracker. For a granular tracker $r \gg L/2$ so that the length of the segment can be written $s = L^2/(8r)$ and the momentum is $p = \dfrac{0.3L^2B}{8s}$. [28] Since the error on the trajectory will be inversely proportional to the momentum $\delta s \propto \dfrac{1}{L^2B}$, so the best tracking system will be large with a large lever arm and a large magnetic field. The lever arm is optimized by having many tracking strips perpendicular to the direction of the field. The tracker is designed based on key physics performance issues summarized as follows:

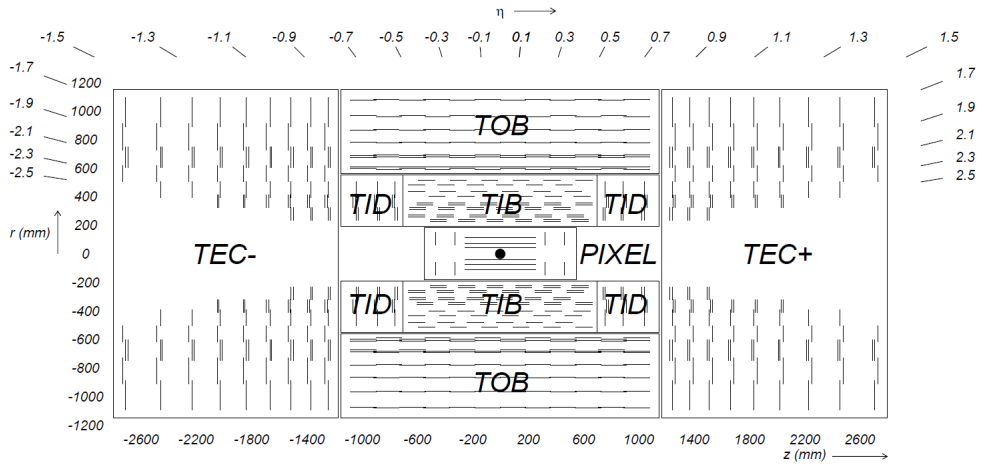
- Robust pattern recognition, despite a large particle flux that includes pile-up interactions.

- Precise measurement of momentum for all charged particles above 1GeV.
- Allow for rapid pattern recognition to trigger on charged particles like the electron or muon
- Provide some particle identification power

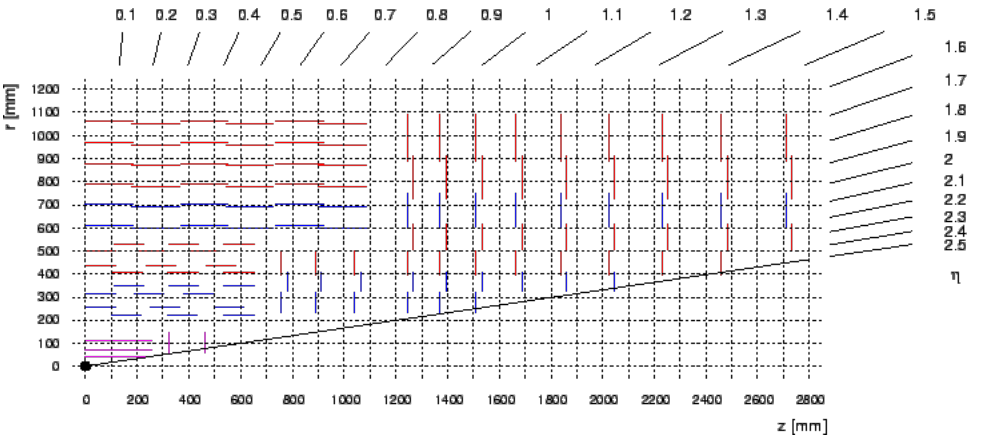
[28] The tracking system is designed to give a precise measurement of particle trajectories based on the expected level of particle flux. As a result, the occupancy of the track hits decreases steeply radially from the interaction point as r^2. Close to the interaction point many charged particles pass through a small area, so require very granular detection. The particle flux is largest at the near the interaction vertex $\approx 10^7$ particles/s at $r = 10cm$ where pixel detectors are used to measure their positions. Each pixel is $\approx 100 \times 150\mu m^2$ with an occupancy of about 10^{-4} per pixel per bunch crossing. For $r < 10cm$, pixel sensors are arranged in cylindrical pattern in three barrel layers at radii 4, 7, and 11 cm. The three layers give full 3D information about the curvature of the track close to the interaction point. In the forward regions, at z=34.5cm and z=46.5 there are pixel layers arranged in a disk pattern with radius extending from 6 to 15cm.

In the intermediate region, $20 < r < 55cm$ the particle flux is significantly smaller,microstrip detectors are used with a minimum cell size of $10cm \times 80\mu m$ with an occupancy of $\approx 2 \times 10^{-2}$ per detector per bunch crossing. The barrel part of the tracker is divided into two parts outside the pixel layers: Tracker Inner Detector, Tracker Outer Detector. The first 2 layers in the inner detector are made of stereo modules, which are "double-sided". Thin $320\mu m$ are glued together with tilt in the angle between them of 100 mrad. Each double sided module provides a measure of the $r-\phi$ and $r-z$ giving a single point resolution of between 23-34 μm in $r-\phi$ and 23 μm in z. The remaining layers are only single sided with only 2D information. In the outer region r $55cm$, the pitch of the strips is larger with a maximum cell size of $25 \times 180\mu m$ while still keeping the occupancy $\approx 1 \times 10^{-2}$ per strip per bunch crossing. The first 2 layers of the Outer barrel gives stereo hits with the same stereo angle as the Inner detector modules and the remaining layers are single sided modules. The last two layers in the outer module have thickness 500 μm. The forward region is divided into the Tracker End Cap and the Tracker Inner Disk which covers the gap between the Tracker Endcap and the Tracker Inner Barrel. The silicon microstrips in the forward region are arranged in a disk pattern. The first two rings and the fifth ring have stereo modules, and the remaining modules are single sided. The first 3 inner rings have $320\mu m$ and the remaining two have thickness of $500\mu m$. [17]

For the analysis in this thesis, the tracking will be essential to identify the Higgs decay vertex amongst other vertices that come from soft interactions. A track is built by combining hits to form the trajectory of a particle and requires a very rapid pattern recognition algorithm. The algorithm steps can be summarized as follows [17]:



(a) a)Tracker Geometry



(b) b)Tracker Strips

Figure 3.6: a) Shows the full geometry of the silicon strip tracker divided into several track layers. b) Shows the dimensions of the measurement measured in each layer: the pixels in magenta give a full 3D measurement, the blue lines indicate stereo layers of the tracker while red are the mono-layers in the outer region of the tracker.

- Group together strip hits in a layer of the Silicon detector, and compute the position and uncertainty based on the collection of hits.
- Choose the 'seed' hits which are the start of a particle's trajectory
- Build the trajectory of a track starting from the seed and searching for a consistent hit on the next layer.
- Resolve overlap between overlapping tracks and ambiguous hits
- Fit the track to give a smooth trajectory starting at the seed

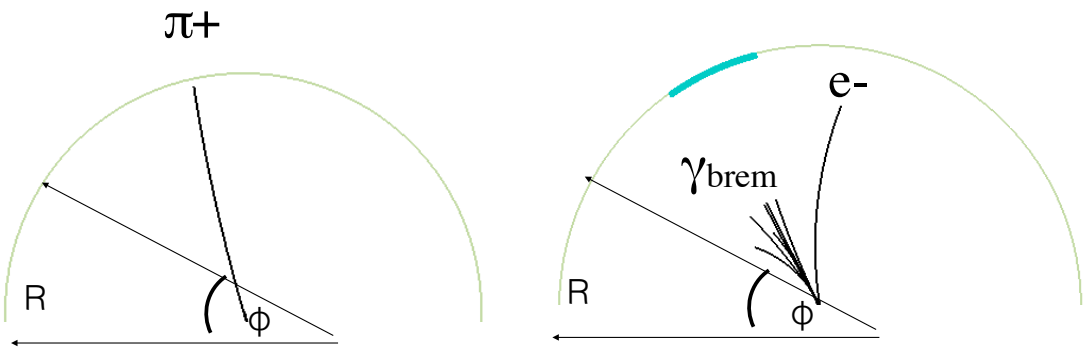


Figure 3.7: The Kalman Filter track shown on the left for the pion tracks which are roughly straight with a slight bend. The Gaussian Sum filter track for the less massive electron in contrast allows for a significant bend to account for the radiation of photons when the electron is bent in the magnetic field.

A heavy particle like a muon or a charged pion has very little p_T loss from bremsstrahlung radiation when transversing the tracker, so follow large helical trajectories. An electron however will have radiation loss from bremsstrahlung when it is bent in the magnetic field, and the p_T loss needs to be accounted for when building the track. The reconstruction of tracks with negligible p_T loss have a gaussian uncertainty at each track layer, while for more p_T loss the uncertainty is a mixture of several Gaussians. The two types of tracks, negligible bremsstrahlung or significant bremsstrahlung are fit by two different algorithms. The Kalman filter pattern recognition is used for negligible radiation loss and the Gaussian Sum Filter is used to account for radiation loss specifically to reconstruct electrons. The Gaussian Sum Filter pattern recognition algorithm accounts for the bremsstrahlung energy loss of the electrons along the track by using multiple gaussian predictors at each track layer. This gives a more accurate description of the electron momentum which can have a lot of energy loss along its trajectory. For each radiated photon, the helical radius of the track will change based on the energy of the photon radiated. At each layer, the Gaussian Sum Filter must account for a

tighter radius if there is an emitted bremsstrahlung photon. An example of the two types of tracks is shown in Figure 3.7, where the Gaussian Sum Filter is used to reconstruct the electron trajectory which radiates many photons. In contrast, the heavier pion track has only a slight bend given by the helix based on the p_T of the particle. [17]

To reconstruct an interaction point, or vertex, tracks are combined based on how consistently they can be associated to a point of origin in the detector. The reconstruction of vertices involves two steps: vertex finding and vertex reconstruction. Vertex finding is the task of finding the interaction point given a set of compatible tracks from an origin. Vertex finding takes into account the physics of the interaction point: the number of particles originating from the point or the displacement from the primary collision point. The primary vertex is considered to be the one from the hard collision in the event. For the primary vertex reconstruction, tracks are selected that are compatible with the beamline, and the tracks are clustered according to their z-position. Tracks within a cluster are then fit to determine the most compatible vertex. A Kalman filter fit, which is a global least squares fit, to determine the vertex position and its uncertainty based on track parameters like p_T, track fit quality χ^2, and the covariance matrix of tracks. The initial vertex fit and the upper tail of $P\left(\chi^2\right)$ can be used to indicate the goodness of the vertex fit. Outlying tracks, which can be poorly measured or possibly not associated to the vertex point, are down-weighted to to improve the vertex reconstruction. By default the vertex with the largest $\sum |Track_{pT}|^2$ is considered the primary vertex. For events with a large multiplicity of charged particle tracks like $t\bar{t}H$ is very high at 99% due to the multiple high p_T particles originating from the decay point. For events with low charged particle multiplicity like $H \to \gamma\gamma$ it decreases below 76% [32]

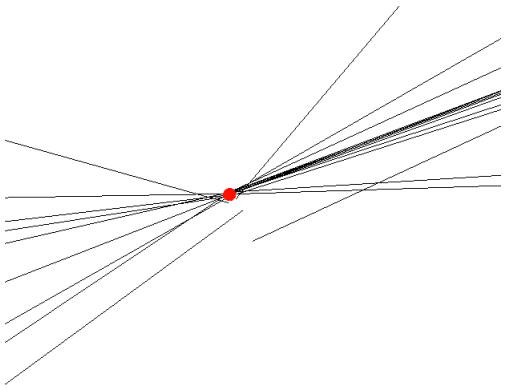
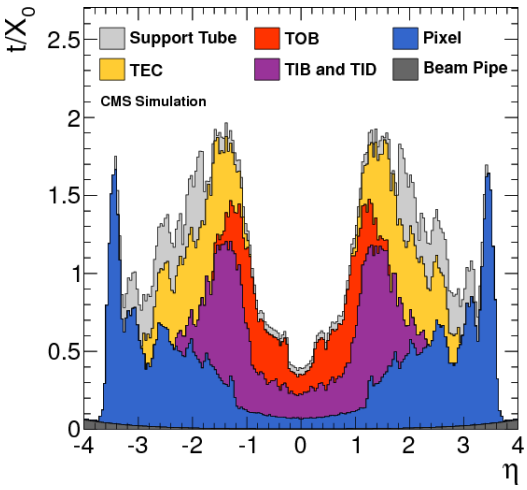


Figure 3.8: Vertex Reconstructed from tracks. This is an example of a primary vertex made from tracks that gives the highest $\sum track_{pt}$ in the whole event.

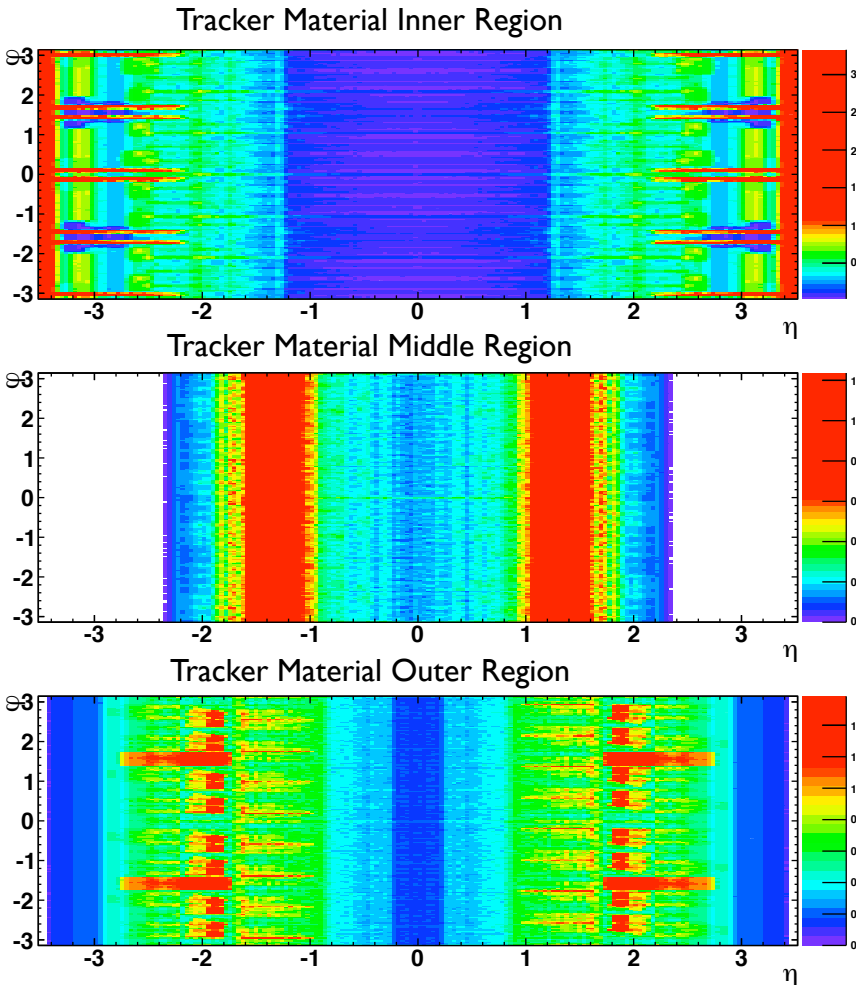
The entire silicon strip detector consists of almost 15 400 modules, which will be mounted on carbon-fibre structures and housed inside a temperature controlled outer support tube.[17] In addition to the amount of silicon tracking material, there is the support structure and the support tubes within the heavy tracker. The amount of material in the tracker which is in front of the Electromagnetic Calorimeter will result in many photon conversions in the tracking material. The tracker material budget in terms of the radiation length is shown in Figure. 3.9. The reconstruction of these conversions requires precise tracking information to reconstruct a vertex where the conversion takes place and to obtain the direction of the photon at that point. [17]

The other charged particle detector is the Muon chamber specifically designed to measure the muon's momentum. The muon detectors make use of gas ionization chambers. The simplest form an ionization chamber is a capacitor filled with an ionizing gas. Ionized particles will be collected by the field across the capacitor plates. The moving charges can be measured as a voltage pulse given by the number of charged carriers collected: $\Delta V = \frac{Ne}{Cd} v_d \Delta t$ where the numerator is N charged particles times the electron charge and the denominator is the capacitance times the distance between the plates. v_d is the drift velocity characteristic of the gas and Δt is the collection time, multiplied together this is the distance the charged carries must travel. For a cylindrical capacitor, with an inner anode radius of r_b and and an outer cathode radius of r_a the pulse for a particle at r_0 is $\Delta V = \frac{Ne}{C} \frac{\ln\left(\frac{r_a}{r_0}\right)}{\ln\left(\frac{r_a}{r_i}\right)}$. For very high fields, the electrons from ionization have enough energy to ionize other atoms in the gas and amplifies the pulse. The rise time of the pulse can be used to compute the position of the passing particle from the anode wire by using the drift velocity. The drift tube chambers consist of 60 tubes in a $2m \times 2.5m$ volume filled with ionizing gas and give a position point resolution of $\approx 100\mu m$. The Cathode-strip chambers consist of anode (+ charge) collecting electrons and cathode (- charge) collecting the positive ions. The cathodes wires and anode wires crossed to form a grid so by collecting the ions and electrons a 2D point can be quickly determined. The resistive plate chambers consist of parallel plate capacitors with highly resistive plates, the ionized electrons are read by metalic strips outside the resistive plates. [38]

Centrally produced muons are measured in three regions: starting in the inner tracker, then after the solenoid coil and finally in the return flux. The muon bending angle is determined at the exit of the 4 T coil using the interaction point measured from the start of the muon track. The resolution is dominated by multiple scattering in the material before the first muon station for momentum $p_T = 200GeV/c$ and the spatial resolution dominates. For the low momentum muons, the best resolution is given by the silicon tracker. For high momentum muons, the muon trajectory beyond the return yoke can be extrapolated to the beam line when there is negligible



(a) 1D in psuedo-rapidity

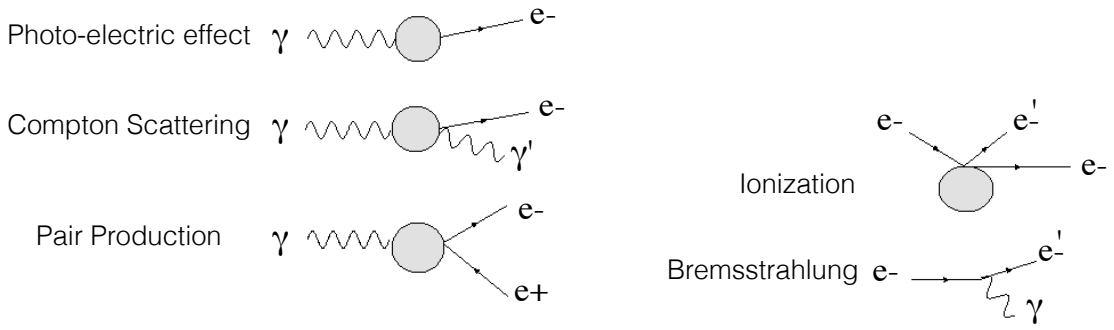


(b) 2D ϕ vs. η

Figure 3.9: The material budget in terms of the ECAL coordinate η and also a fine grained 2-dimension map of the material budget in η, ϕ which shows the granularity of the tracking material.

multiple scattering. The muon momentum resolution improves when combining the tracker and muon detector measurements. [17] Gaseous detectors are used to identify and measure muons, and the choice of technology depends on the coverage area of the detector and radiation environment. The main background for muons is the neutron induced background. At $|\eta| < 1.2$ this background is small and also the muon rate is low and the residual magnetic field is low. Drift tubes are used in this region. In the endcaps, $|\eta| < 2.4$, the opposite is true and cathode strip tubes are used in this region. Resistive plate chambers are used in both the barrel and endcap, which gives a rapid readout with good time resolution but worse position resolution. The RPC can therefore identify the correct bunch crossing and provides a rapid trigger to accept or reject the muon.[17]

3.2.2 Electromagnetic Calorimeter



(a) Photons interacting wtih matter

(b) Electrons interacting wtih matter

Figure 3.10: Photons interact with matter in several ways, but at high energies pair production becomes dominant. Likewise for electrons the bremsstrahlung process is dominant.

Electrons and photons interact electromagnetically when transversing matter as shown in Figure 3.10. At high energies above 10 MeV the interactions become dominated by Bremsstrahlung and electron-positron pair production. At high energies, the transitioning particles interact mainly with the nucleus of the material. Electrons decelerate in the electric field of the nucleus and emit a photon. Photons also interact with the nucleus and the energy of the photon converts to the rest mass of an electron-positron pair and their kinetic energy. For a relativistic electron $v \sim c$ the energy loss by ionization is given by the Bethe-Bloch function [38]:

$$\frac{-dE}{dx} = 4\pi N_0 \frac{Z}{A} r_e^2 m_e^2 c^2 \left[\ln\left(2mv^2\gamma^2\right)/I - I\right]$$

while the average energy loss by Bremsstrahlung is:

$$\frac{-dE}{dx} = 4\alpha N_0 \frac{Z}{A} r_e^2 E \ln\frac{183}{Z^{\frac{1}{3}}} = E/X_0$$

The ratio of the processes is approximately: $\frac{-dE}{dx}_{Brem} / \frac{-dE}{dx}_{ion} \sim \frac{ZE}{580MeV}$ and depends on the atomic number of the material and the energy of the electron. Similar for photons the dominant process for $E_\gamma > 2GeV$ is pair production. The photoelectric effect in atoms has a cross section $sigma \approx \frac{Z^n}{E^3}$ so this interaction becomes almost negligible at high energies. Similarly, the compton effect where a photon scatters against an electron $\gamma + e^- \Rightarrow \gamma + e^-$ in the atoms of the material, has a maximum transfer of energy to the electron in a head-on configuration $\theta = \pi$ given by $E_{Comp} = \frac{2E^2}{m_e c^2 + 2E}$ for $E_\gamma \gg m_e c^2$ this becomes a constant $2m_e^2/2 \approx 0.25MeV$. Pair production begins when $E_\gamma\ 2m_e c^2 \approx 1.0MeV$. For high energy photons the intensity of a beam of photons decreases exponentially by $I = I_0 \exp\left(\frac{-7x}{9X_0}\right)$ where X_0 is the radiation length depending on the properties of the material. The interaction of photons and electrons in matter is summarized in Figure 3.10. [38]

A chain of Bremsstrahlung and photon conversion interactions with the nucleus of the material leads to a cascade of photons, electrons, and positrons, called an electromagnetic shower. The

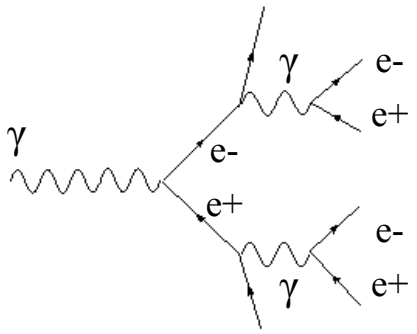


Figure 3.11: The development of a photon shower.

development of the shower is described in the following steps and the stages are shown in Figure 3.11.:

- For an incident photon with energy E_0 that traverses a length of X_0 has a 54% probability to convert.
- The resulting e^+e^- will each have on average energy of $E_0/2$.
- Once the e^+e^- have traversed a length of X_0, the charged particles will emit a photon of energy between $E_0/2$ and $E_0/2e$. At $2X_0$ the average particle multiplicity will be 4.
- Radiated photons will begin again the process in the first stage. [38]

After n-generations created through a thickness of nX_0 there are 2^n particles with average energy of $E_0/2^n$. The shower develops until a critical energy is reached where the electron bremsstrahlung energy loss is equal to the one from ionization which is given by $E_{crit} \sim \frac{580MeV}{Z}$, so large Z can give

greater stopping power for the shower. The incident particle energy will be related to the particle multiplicity at E_{crit} and also the penetration depth of the electrons and positrons in the material. A useful unit for gauging the shower containment in a material is the Moliere radius where 99% of the shower is contained in a cone of radius $3R_M$. [38]

Detectors to measure the incident EM particle energy from the resulting shower is called an Electromagnetic calorimeter (ECAL). Homogenous shower detectors attempt to contain the whole shower in the material, so they are made of large Z and small X_0 materials. To give the best energy resolution of the two photons from the Higgs decay of the CMS uses Lead Tungstate: $PbWO_4$ crystals for the ECAL. The crystals have small Moliere radius: $2.19cm$ and a also small radiation lengths: $X_0 = 0.89cm$ resulting in a fine grained homogenous calorimeter made of compact crystals. In the In the barrel, the crystal face is $\approx 22 \times 22mm^2$ and has a length corresponding of 230 mm corresponding to $26X_0$. In the endcap the crystal front face is $\approx 28.6 \times 28.6mm^2$ and a length of 220mm. This region also has two preshower planes made of a layer of lead absorber, which causes an electromagnetic shower, followed by silicon strip detectors arranged in a grid which detects the positions of the electron positron pairs. The purpose of this detector is discriminate single photon showers photons from meson decays to photons $\pi_0 \rightarrow \gamma\gamma$. [17]

The measurement of the energy of incident electrons and photons is given by the scintillation in the crystal. The relativistic shower electrons in traversing the crystal emit Cherenkov light about 80% of the light is emitted in 25ns, which is the same time as the LHC bunch crossing. The light output is low ≈ 4.5 photoelectrons per MeV and depends on a stable temperature at $\pm 0.05°C$. The crystals are designed to be radiation hard to withstand the radiation levels through out the run periods of the experiment. Impurities in the crystal cause color centers and absorption bands to form. The crystal then has a wavelength dependent transparency loss that can be tracked by a laser monitoring system. The known laser injection at wavelength of 440nm can be used to recalibrate the transparency loss to give a more accurate energy measurement.

Due to the low light yields, the rear end of every crystal must have sensitive photodetectors that are capable of operating in the 4 Tesla magnetic field. The Avalanche Photo-diodes for the barrel have a quantum efficiency of 75% while the Vacuum Photo-triodes have a quantum efficiency of 22% . Two APDs are glued to backs of the crystals each with an area of $5 \times 5mm^2$. The photodiodes are based on the p-n junctions at a reverse bias close to the breakdown voltage. An incident photon, in the junction results in an avalanche current in the device. The APDs have a gain of 50 with a sensitivity to ionizing radiation within an effective thickness of $6\mu m$ which corresponds to a minimum ionizing particle of $\approx 100MeV$. The endcap features Vacuum photo-triodes which are photomultipliers with a signal gain stage. A photon will strike a photocathode which emits an

electron by the photo-electric effect, and the electron is accelerated toward an anode which emits more electrons when struck. The anode is made of a very fine copper mesh which allows it to be operational in the 4 Tesla field. The VPT gives a gain of 10 in a 0 T field.[11]

For photon energies in the range 25 GeV to 500 GeV appropriate for the Higgs decay to photons, a gaussian energy resolution can be parameterized as:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{\sigma_n}{\sqrt{E}}\right)^2 + C^2$$

The first term is the stochastic term which accounts for the statistical fluctuations in the number of photoelectrons from the shower and fluctuations in the gain process. The second term accounts for electronic noise in the readout of each crystal as well as the energy from pileup. The constant term accounts for energy leakage from the back of the calorimeter and variation of the responses from different crystals which can depend on the thermal stability and transparency of the crystal. Since constant term is not scaled by energy its contribution dominates the total resolution. [22][11]

The calibration of the ECAL defines the ultimate performance of the CMS crystal calorimeter. In particular to exploit the excellent containment of the crystals and give well resolved photons at high energies, the constant term must be kept below 0.55%. [11] This requires a minimization of the transparency loss in the crystals which is achieved by using the laser monitoring system. Also a uniform energy response across the ECAL to give a stable measure of a mass resonance like the Higgs. This is achieved by calibrating one set of crystals with another another set or inter-calibration of crystals. Both techniques give a final set of online calibrations that is further improved after photon reconstruction when analyzing Higgs decays to photons in Section 4.4 and Section 4.7.

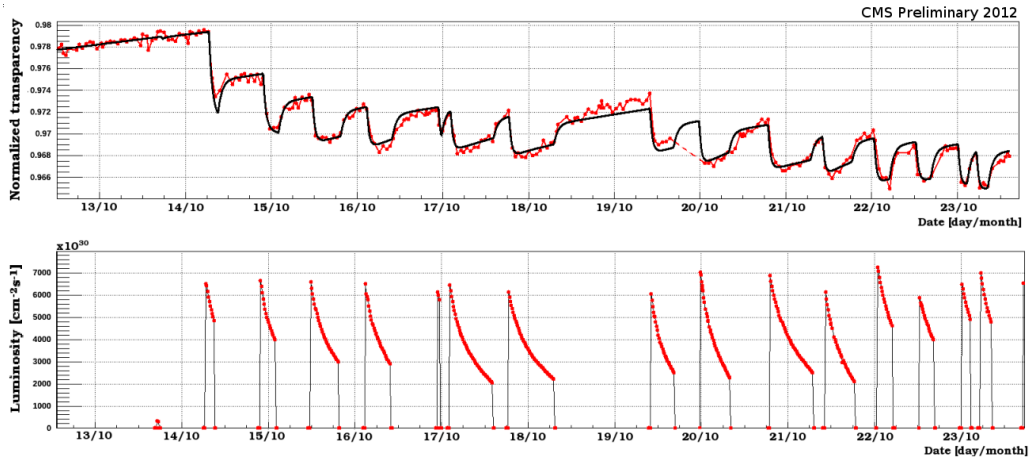


Figure 3.12: Loss of transparency of ECAL crystals (top) with the luminosity of a particular run period (bottom) [22]

Irradiated crystals suffer a transparency loss because they absorb radiation and the transparency

is recovered through a self-annealing process after a relaxation time. The relaxation time depends on the radiation dose a certain crystal absorbs. This results in a time-dependent energy response of the ECAL crystal. [11] The details of the mechanism of radiation absorption and the self annealing is given in [2]. The transparency of the crystals over time is shown in Figure 3.12 for different radiation doses. The ECAL light monitoring system is used to determine corrections, denoted by $S_i(t)$, to response changes in the ECAL. The laser light is injected through optical fibers in each EB and EE crystal through the front and rear surface respectively. Two laser wavelengths are used one peaked at $\lambda = 440$nm which is close to the scintillation emission spectrum and another in the infared range $\lambda = 796$ to test effects that do not depend on transparency loss. A power law is used to relate the changes in the laser light read at the photodetector to changes in the scintillation light:

$$\frac{S(t)}{S_0} = \left(\frac{R(t)}{R_0}\right)^{\alpha}$$

$S_i(t)$ is the time dependent channel response to scintillation light while S_0 is the initial response, likewise $R(t)$ and R_0 are the time dependent and initial response to laser light. The power α is determined from radiation studies on a limited set of crystals during beam tests. The spread in α was found to be 10% and arises from residual differences in transparency and different surface treatments of the crystals. The precision on α limits the precision of the response correction to 0.3% in the barrel and 0.5% in the endcap. The response corrections are tuned in data using meson decays: $\eta \to \gamma\gamma$ [22]

Inter-calibration involves a few steps to calibrate the energy response in groups of crystals at the same pseudorapidy η:

- A large sample of minimum bias events can be used to calibrate a ring of crystals in ϕ for a given η The total deposited energy should be the same in every crystal.

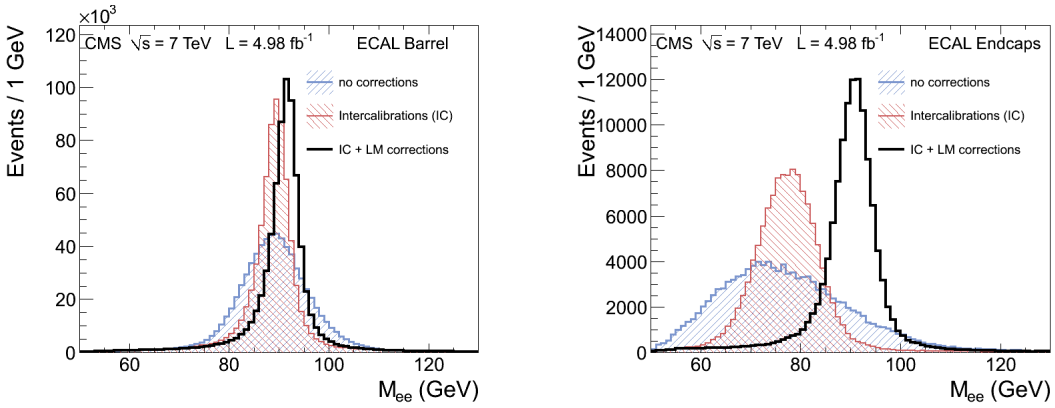
- The invariant mass of meson decays like π_0 and η into photons. The photon energy is the energy deposit in 3×3 crystals about a local energy maximum (the seed crystal). The intercalibration constants C_i for each crystal are chosen to correct the fitted mass value. The distribution of energy is fitted with a gaussian for the signal mass resonance and a fourth-order polynomial for the background.

- W-boson and Z-boson decays provide electrons for the inter-calibration procedure. The momentum measure of the track of the electron can be compared to the energy measured in the crystals. Inter-calibration constants in ϕ are derived accounting for module and supermodule gaps as well as different amounts of tracking material in η and ϕ. Crystals in the same ϕ ring are calibrated by using the E/p from Monte Carlo as reference distributions. [30]

The gaussian resolution of the ECAL is measured in electron beam tests with maximally contained showers in a 3x3 matrix of crystals without the inter-calibration corrections and without the tracker in front of the ECAL. The ideal resolution for the ECAL barrel:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{\sqrt{E}}\right)^2 + (0.3\%)^2$$

Fig. 3.13 shows the dominant contribution of the laser calibrations to the total calibration. The value of the constant term is limited by the precision of the α coefficient in the laser calibration. The addition of the tracker and the effect of pileup contaminating energy motivates the need for further improving the energy resolution with offline corrections using multivariate regression techniques described in Section 4.4. Section 4.6 will describe further improving the transparency corrections using the mass resonance of the Z peak from electron-pairs. The final indication of the energy resolution will be given after regressing the energy and scaling it per run period, and measuring the width of the Z peak in data as described in Section 4.7.

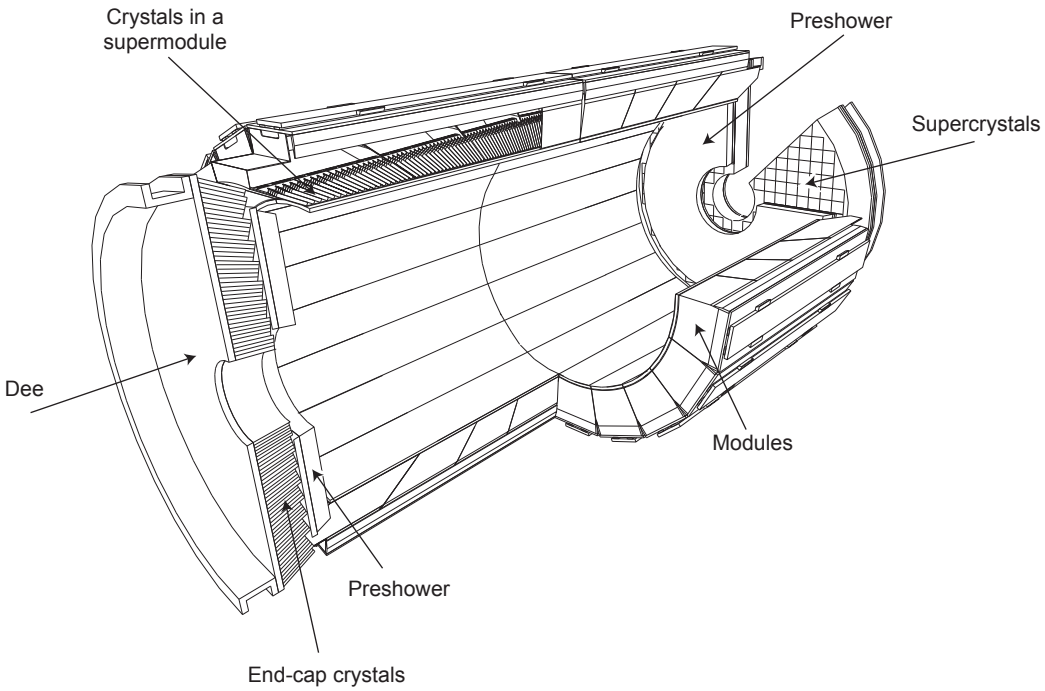


(a) ECAL Barrel

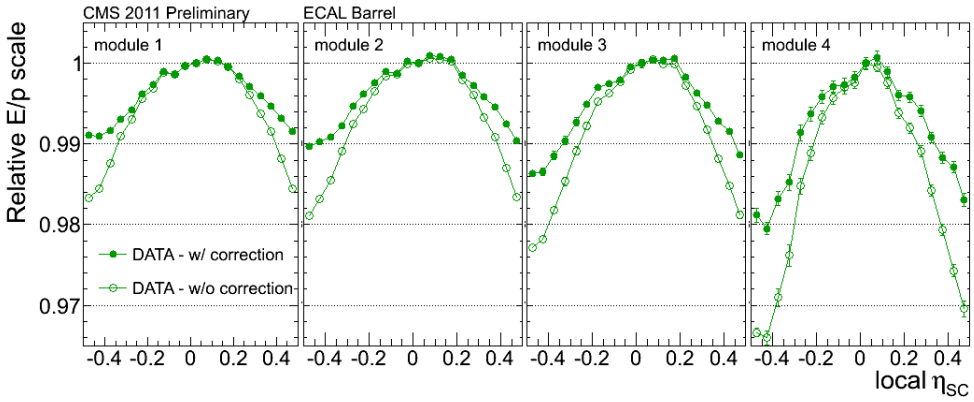
(b) ECAL Endcap

Figure 3.13: The effect of the two main calibrations of crystals: the intercalibration and the laser transparency correction. [22]

The geometry of the ECAL is divided into the two regions of the cylinder: barrel and the endcaps. The barrel covers a pseudorapidity region $|\eta| < 1.479$ and the endcap covers $1.55 < |\eta| < 3.0$ with a gap at the edges of the barrel. The endcaps also have an additional sub-detector the pre-shower at $1.653 < |\eta| < 2.6$. The preshower is designed to identify neutral pions in the endcap region. The preshower is a sampling calorimeter with two layers. Each layer consists of a lead absorber to trigger the shower and a silicon strip sensor to measure a hit pattern in one direction. The silicon strip sensors are oriented orthogonal to each other to give a 2D hit measurement. The energy deposited in the lead is estimated from the charge measurements in the silicon strip which are measured in terms of minimum ionizing particles, MIPs. A calibration function can be used to convert the measurement



(a) ECAL Geometry



(b) Shower Leakage across crystals

Figure 3.14: A diagram to show the cylindrical geometry of the ECAL divided into modules and supermodules. The boundaries across these regions is where parts of the shower can maximally leak out of the crystal and into the support structure. [22]

of MIPs in each layer to energy. The calibration function also gives the total energy measured in the endcap combining the pre-shower strips with the matched ECAL crystals behind the preshower.

Though the lead-tungstate crystals are compact, the density of the material is $8.3 g/cm^3$, so each crystal weighs 1.5kg. The ECAL is made of 75,848 crystals and requires a large support structure to hold them up in the barrel. Half of the barrel of the cylinder can be divided further into 18 supermodules of cylindrical slices. The supermodule weighs about 2 tons and is supported by a frame of steel and aluminum. Each of the rings of supermodules is divided into 4 wedges of crystals grouped into modules. The crystals are contained in a thin-walled glass-fibre alveola structures within submodules. The ECAL is not entirely homogenous because of the gaps and cracks that exist between, supermodules, modules, and individual crystals. In these gaps and cracks some portion of the EM shower will leak away. Figure 3.14 shows the ECAL geometry as well the amount of leakage on the edges of crystals in η. The endcaps of the ECAL are circular walls with a cartesian geometry with crystals grouped in 5x5 arrays within a carbon-fibre alveola supported by an aluminum backplate.

3.2.3 Hadronic Calorimeter and Jet Clustering

The energy measurement of the strongly interacting particles, or hadrons, will be of essential importance in reconstructing the vector-boson fusion Higgs where two scattered quarks are in the finals state in Section 5.7.1. This VBF Higgs analysis will use anti-k_T Particle flow jets as well as Fast Jets. The prior algorithm is used for the signal jets from the scattered quarks while the latter is used for removing energy deposits from pileup clustered as jets.

Hadrons can interact elastically or inelastically with the nuclei of a block of material. Inelastic collisions result in secondary hadrons which in turn give rise more hadrons when undergoing inelastic collisions which cascades into a hadronic shower. The scale for this spatial development is the nuclear absorption length $\lambda = A/\sigma N_0 \rho$, where ρ is the density of the material, A is the mass of the units in moles divided by N_0 Avogadro's number, and σ is the inelastic cross-section. The size of the hadronic shower can be large: in iron the typical shower length of the shower is 1m long and 0.6m in the transverse plane. [38] The CMS Hadronic Calorimeter, HCAL, is a sampling calorimeter made of sandwiches of absorbing material and scintillation material. For the barrel up to $|\eta| < 1.3$, the HCAL consists of 36 wedges of 20° in ϕ. The absorber plates in the wedge are made of brass alloy absorber plates interspersed with plastic scintillation tiles where the innermost and outermost layer are structurally strengthened by a layer of stainless steel. The total nuclear absorption length perpendicular to the beamline is 5.82λ and increases with the polar angle as $1/\sin\theta$ The ECAL also adds to the absorption length by 1.1λ.[45] In the central barrel region, the HCAL does not contain

the full hadronic shower so there is an extended calorimeter outside the solenoid, HCAL outer calorimeter. For the endcap region, the HCAL interlocks with the barrel region and is composed almost entirely out of brass absorber plates of thickness 78mm with scintillator plates of 3.7mm. [17] The total nuclear absorption lengths including the ECAL is 10λ. [45] The optical signals from the scintillating material are converted into electrical signals by hybrid photodiodes with a gain of $\approx$ 2000. Individual channels are tested with a collimated radiation source. The energy scale constants for the HCAL are obtained by combining test beam data taken with $e^{\pm}$, $\pi^{\pm}$, and muon beams. [17]

Jets which are streams of particles caused by hadronizing quarks are measured in several ways at CMS. The simplest is just to cluster together energy in a large cone from the energies measured in the HCAL channels and with the corresponding measurements in the 5x5 ECAL crystals (Calojets). The HCAL channel and the geometrically corresponding ECAL crystals are called towers. A four momentum can be associated to each tower above a threshold, taking the energy measurement and the direction from the tower position and the interaction point. The jet is created by clustering together the towers and summing the 4-momenta of each tower. If instead of the pure calorimeter based measurement, particles are reconstructed from the track, ECAL, and HCAL measurements via the Particle Flow reconstruction (further described in Section 4.3.1. Then the full granularity of the detector is exploited by accounting for the compatibility of different sub-detector measurements. Charged particles are reconstructed from tracks that can be extrapolated to ECAL and HCAL energy clusters and the 4-momenta is computed based on all three measurements. Charged hadrons within a jet are reconstructed in this way. Photons are reconstructed from ECAL clusters that are not associated to a track extrapolation. Likewise the energy deposits of HCAL clusters not associated to a track are reconstructed as neutral hadrons. The 4 momenta of these particles are summed to create a PF Jet within a specified cone and this greatly improves the energy and position measurements of the jet. [20]

The k_T algorithm clusters nearby particles into jets based on the distance between the particles and the E_T of the less energetic particle: $d_{ij} = min\{E_{Ti}^2, E_{Tj}^2\} * R_{ij}^2$. The anti-$k_T$ jet clustering algorithm is an improved clustering intended to cluster particles into jets whose shape is not influenced by soft radiation [8] which is a key priority for increasing pileup. Fast jet is a rapid calculation of the k_T algorithm clustering algorithm that clusters the nearest neighbors. The Fast Jets are used to identify clumps of energy deposits from the pileup based on jet shape and the vertex of the charged particles so that they can be removed from the event as in Section 5.7.1 . It is observed in simulation that Particle Flow charged hadrons make up $\approx 50\%$ of the total pileup contamination and the other half comes from the Particle Flow neutral particles. The charged hadrons that do not

originate from the primary interaction vertex can be rejected as well using the Fast Jet in which it is clustered. For the neutral particles, the shape of the jet becomes a handle since the overlapping pileup jets tend to be more diffuse. [15] This idea will be used in Section 5.7.1 for the vector-boson fusion Higgs tagging.

Chapter 4
Photon Reconstruction and Identification

Section 3.2.2 described how showers in single crystals and groups of crystals are calibrated to improve the ECAL energy resolution and to give a more uniform energy response across the calorimeter. This section will describe how the photons are reconstructed from groups of crystals and also how their energy is further improved at the reconstruction level. CMS uses a clustering procedure to identify crystals from the photon shower. The sum of the calibrated crystal energy together with the energy weighted position give the photon four-momentum. Section 4.1 will describe the standard version of photon reconstruction and Section 4.3.1 will describe an improved photon reconstruction procedure developed in parallel. This procedure of clustering crystals meets two particular challenges:

- There is energy loss when the photon showers early in the tracker and the e^+e^- bend away from the original photon direction and land outside the reconstruction window. The same problem of energy loss and containment also exists for electrons which will emit bremsstrahlung photons when they are bent in the magnetic field.

- The energy contamination from overlapping particles with the signal photon tends to cause the photon energy to be overestimated. The particles from pile up are described in Section 3.1.1 and are uniformly distributed in rapidity, so a reconstruction window with a larger area will be more susceptible to contamination.

To capture more photon fragments would require a larger reconstruction window which could give more contamination from overlapping particles as well. Thus the two challenges are linked, and the energy resolution depends on both minimizing the contamination from pileup and maximizing the energy containment of the clustered crystals.

The collection of energy from an electromagnetic shower in the fine grained calorimeter can be thought of as a pattern recognition procedure. The procedure must account for whether the shower is induced on the ECAL or within the tracking material. The basis of an energy clustering algorithm is to start from a local maximum energy clump, seed crystal, and add the energy measured in crystals around the seed based on a reconstruction window. A standard technique of photon reconstruction has been used for the Higgs search. This technique was developed during the design and testing

phase of the CMS detector, and the rapid reconstruction allows it to be used for high level triggering of photons and electrons as well.

A more refined energy clustering algorithm has been developed in parallel with the Higgs search to improve upon the energy captured in the clustering stage and simultaneously reduce the contamination of pileup energy. This technique includes as much of the displaced crystals from the seed crystal as possible. This technique is built within a global event reconstruction algorithm of other objects like electrons and jets. This provides a final set of very pure signal photons since the final step of the algorithm requires disambiguation of photons from similar objects like electrons or photons within jets.

Both techniques improve upon the online calibration described previously by training a multivariate analyzer on simulated photons. The simulation provides the true energy of a photon as well as the reconstructed energy from either energy clustering algorithm. The scale $\frac{E_{reco}}{E_{true}}$ is modeled as a response function of several variables that are correlated with energy loss and energy contamination using a Boosted Decision tree learning algorithm. The multivariate energy correction will be described in Section 4.4 and will be compared to standard techniques based on single variables.

Since the multivariate learning algorithm is trained on simulation, the final energy resolution must account for differences in the simulated ECAL response and the observed ECAL response in data. The final section, Section 4.7, will describe the procedure for determining the gaussian smear that needs to be applied to the Higgs mass given the resolution in data from the Z-boson mass.

4.1 Standard Photon Reconstruction

Two algorithms are used for the two different regions of the ECAL: the barrel $|\eta| < 1.479$ and the endcap $1.479 < |\eta| < 3.0$. The simplest case for energy reconstruction at CMS is for an unconverted photon which very neatly leaves 94% of its energy in 3×3 matrix of crystals about the seed crystal and almost all of it 99% in a 5×5 matrix of crystals. This can be seen in Figure 4.3 a. where a majority of the energy is in 3x3 crystals, and the remaining energy is a larger array of 5x5 crystals. A useful shower shape variable to separate converted and unconverted photons is $R9 = \frac{E_{3\times3}}{E_{reco}}$, which is above 94% for unconverted photons. With the large amount of material in front of the tracker, as shown in Figure 3.9, the probability for a photon to convert in the barrel is about 60% and for the endcap where there is more material in front it is above 80%. The majority of photons will fragment in the tracker and their charged particle constituents will leave a larger footprint on the ECAL as they bend in the magnetic field. The same is true for electrons that will radiate photons in the magnetic field. These secondary fragments from the initial electron or photon may land far

from the main shower cluster which is the center of the reconstruction window. One strategy for gathering the energy radiated from the EM object is to define a dynamic geometric reconstruction window to build a cluster of individual clusters from the fragments. This "cluster of clusters" is called the supercluster and defines the ECAL envelope of a photon or electron that radiates in the tracker.

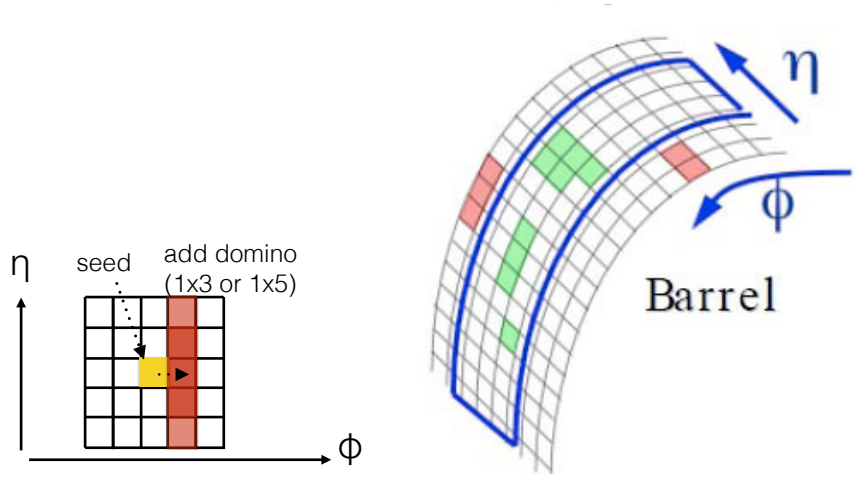
The standard superclustering procedure relies on the assumption that the field is purely axial so the EM radiation from an object will spread only in one direction ϕ and the clusters will be aligned in the other direction η. In Section 4.3.1, a refined dynamic envelope will be made without this assumption. The superclustering procedure for the barrel is the "Hybrid Algorithm" which is summarized in the following steps:

- Crystals are clustered together by adding neighboring crystals along η and ϕ scan lines until there is a rise in energy with respect to the previously added crystal or a crystal has no readout.
- These clusters become seeds if they pass an energy threshold of 1GeV
- A search window is built in steps around the seed cluster. Rectangular dominoes are added to the supercluster that are centered along the η-position of the seed. The domino can be an array of 1×3 crystals if the central crystal has energy below 1GeV or 1×5 if it is above. The larger domino accounts for larger showers. See Figure 4.1a.
- The dominoes are added until a cutoff condition is reached of 10 crystals from the seed. See Figure 4.1b.
- To eliminate some ECAL noise dominoes below 0.1 GeV and with a seed below 0.35 GeV are not included in the supercluster. Connected dominoes are clustered, and the sum of the cluster energy is summed to give the total energy of the supercluster. The position of the shower is found by the energy weighted sum of the basic cluster positions:

$$x = \frac{\sum x_i \cdot E_i^{clus}}{\sum E_i^{clus}}$$

Compared to the 3x3 square matrix for unconverted photons, the supercluster is a rectangle spread about the square 3x3 with a maximum η width and a dynamic ϕ width. A hybrid supercluster for the barrel is shown in Figure 4.2a. where the constituent clusters are neatly aligned along η.

The superclusters built in the ECAL endcap need to account for the different geometry than the barrel and also the two layers of preshower in front of the ECAL. The clustering of crystals is done by collecting in a fixed block of 5×5 crystals around a seed crystal. Also the preshower strip hits, are collected using the same clustering along η and ϕ scan lines. The preshower hits are matched to



(a) Adding Dominos (b) Hybrid Supercluster

Figure 4.1: The procedure for adding dominoes and the final shape of the hybrid supercluster.

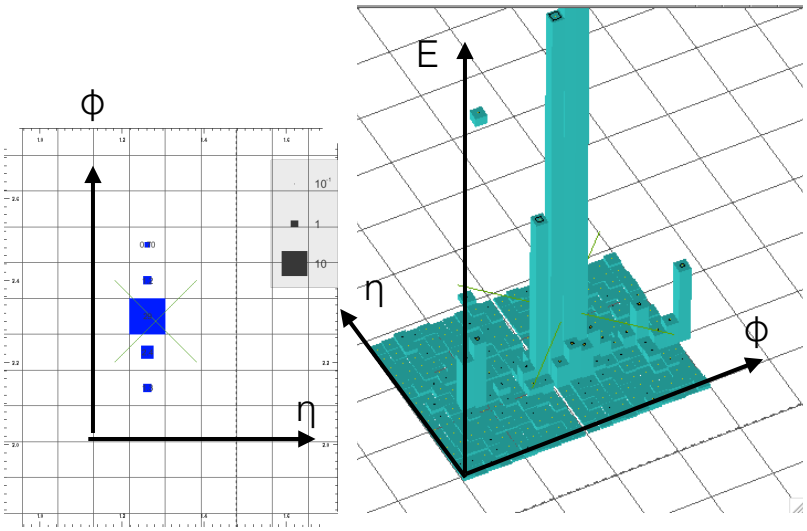
the 5×5 ECAL clusters by projecting the crystal cluster positions onto the two preshower planes. The energy of the supercluster is the sum of only the 5×5 ECAL clusters gathered, and the energy measured in the preshower layers is also saved.

The standard clustering algorithm is fast and simple, but will need energy corrections for various effects:

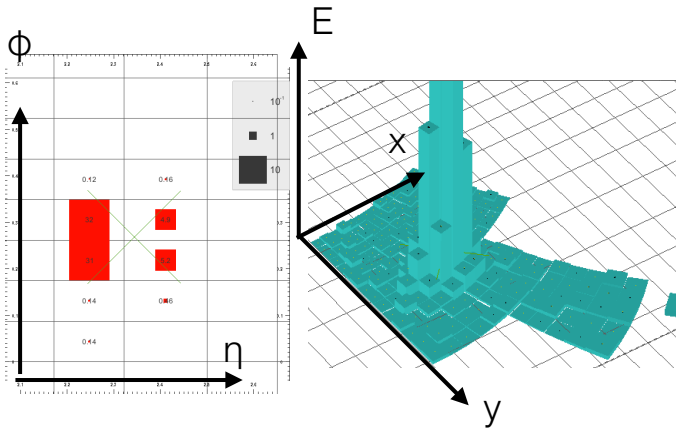
- Energy contamination from particles that land in the reconstruction window.
- Energy leakage from the fragments of converted photons that land outside the reconstruction window.
- Fraction of the EM shower that leaks into air gaps between crystals and also into the support structure that holds up modules and super-modules of crystals.

The corrections for these effects will be derived using simulation and will be described in Section 4.4. An important point is that the supercluster algorithms are the same for the electron and photon energy measurements. This allows electrons in data from Z decays to validate the simulation-based energy corrections applied to photons.

Also it is useful to reject a large number of superclusters that come from other objects that fake photons like a jet or from high p_T meson decays like π_0 and η. In order to do this, isolation cuts are an essential tool. A simple set of isolation cuts can be derived by summing the p_T measured by sub-detectors in a cone around the reconstructed photon. For the Higgs search loose cuts on these isolation values are applied to reject a majority of background photons. The following sub-detector measurements are used for isolation:



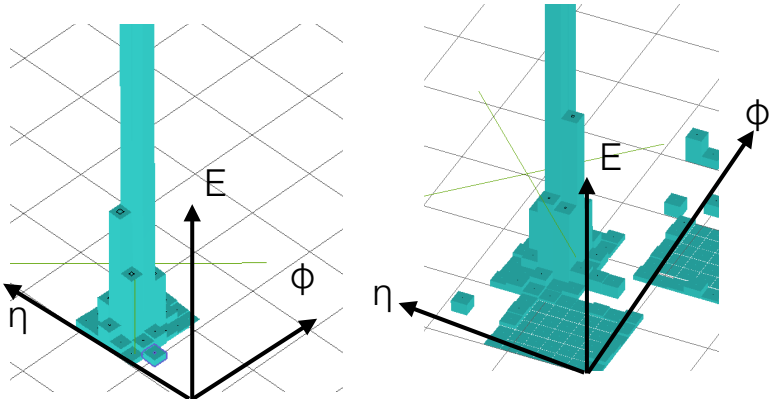
(a) Barrel Hybrid Supercluster



(b) Endcap Poly5x5 Supercluster

Figure 4.2: The 2 types of standard superclusters the hybrid supercluster for the barrel photons and the poly5x5 superclusters for the endcap.

- HCAL isolation: the sum HCAL E_T around the supercluster in an annular region $R_{inner} > 0.15$ and $R_{outer} < 0.3$ which gives a measure of hadronic energy around the photon.
- Track isolation: the sum of Track p_T around the supercluster in $R_{outer} < 0.3$ Electron tracks and conversion tracks are not included in the sum by excluding tracks that are $R < 0.04$ and also $|\Delta\eta| < 0.015$.
- ECAL isolation is not used, but it is the sum ECAL E_T in an annular region of $R_{inner} > 0.06$ and $R_{outer} < 0.3$. It would give a measure of the pileup energy contamination, but it was found not agree between data and simulation because of modeling of the noise.
- $\sigma_{i\eta i\eta}$ is a shower shape variable that gives an indication of the variation of energy in the *eta* direction.



(a) Photon Shower 3x3 crystals (b) Photons from pion decay

Figure 4.3: The photon shower is neatly contained in a matrix of crystals. The 2 photons from a pion decay give a different ECAL footprint with a larger spread in the η direction.

$\sigma^2_{i\eta i\eta} = \frac{\sum w_i \left(i\eta_i - i\eta_{seed}\right)^2}{\sum w_i}$ where the sum runs over a 5x5 matrix of crystals, the numerator subtracts the crystal index from the seed crystal index (number that denotes the position in η) and the weight is given by the expression: $w_i = max\left(0, 4.7 + \ln\frac{E_i}{E_{5x5}}\right)$. Thus it gives a weighted sum of the crystal spacing along η which should be narrow for a single photon shower, and broader for two photon showers that come from a meson decay. Figure 4.3 shows the differences between an electromagnetic shower from a single photon and a shower from two photons coming from a neutral pion decay. The pion decay has a characteristically larger spread along η from the two showers, and $\sigma_{i\eta i\eta}$ indicates this spread.

These variables will be used to define a pre-selection for $H \to \gamma\gamma$ events defined in Section 5.3. With appropriate cuts they reject a large amount of background that come from photons that do not originate from the collision point, but come from meson decays inside a jet of particles. The

remaining background will be rejected by using the Particle-based isolation described in Section 4.3 combined with the shower shape variables of the supercluster.

4.2 Reconstructing Conversions

In Section 3.2.1, it was described how the large material budget in front of the calorimeter causes a majority of photons to convert to e^+e^- pairs in the tracking material. The charged fragments of the converted photon provide important information about its trajectory. This information is essential for finding photons that likely come from the decay of the Higgs. This will be further described in Section 5.4 and will demonstrate how the choice of vertex can affect the reconstructed Higgs mass. This section will describe the reconstruction of conversions in the tracker. Two candidate electron tracks can be fit to a vertex which corresponds to a conversion. Even one leg of a converted photon can be identified as an electron candidate and used to determine the photon direction. Both techniques maximize the use of available tracking information to give precise resolution on the photon origin point.

The standard track reconstruction is optimized for the primary interaction point with the largest sum of track p_T. These tracks along the beam line lie very close to the collision point. Displaced interactions like those from photon conversion can occur anywhere within the tracking material. The main steps of track reconstruction are as follows:

- Hit reconstruction: clustering of hits in a pixel layer or strip layer, and estimating their position and its uncertainty.
- Determining seeds to start the trajectory
- Pattern recognition to determine the trajectory
- Resolving ambiguity when hits are shared between tracks
- Track Fitting

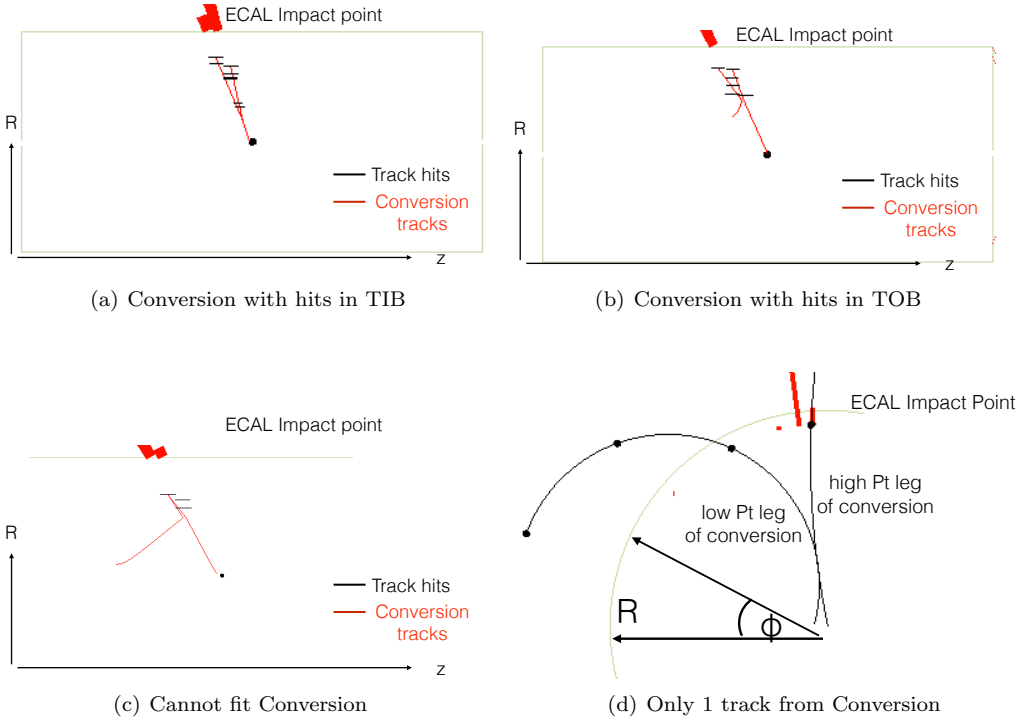
[17] For conversion tracks, the first step is preserved, but the other steps need to be modified depending on where the conversion takes place. Conversions in the granular pixel layers of the tracker give well-defined electron tracks from the 3D information from the pixel hits used to construct the GSF track described in Section 3.2.1. The hits in the pixel layer can be combined with the ECAL impact point of the electron to build a smooth trajectory with a well defined seed and endpoint. For two tracks originating in the pixel layer the conversion vertex is also well determined because the direction of the $e+$ and e^- will be measured well enough for them to be parallel at the conversion

point. For conversions in the outer tracking module, there may only be 2D hit positions from the single sided modules. These modules are also larger and more coarse, so the $e+$ and e^- tracks are difficult to disambiguate and will have a majority of shared hits. This can be seen in Figure 4.4 with conversions that are seeded at different radius in the tracker. Conversions seeded earlier can clearly disambiguate the closely overlapping tracks, and it becomes more difficult with only track hits in the outer layer without any stereo hits, which give both the $r-\phi$ and $r-z$ positions.

The seeding of conversion tracks is done in two ways: inward, outward. Inward seeding starts with any subcluster inside the supercluster as the starting point of the track. The assumption is that all identifiable conversions should land within the locale of the supercluster. The momentum of the track is assumed to be the energy deposited in the subcluster, this determines the bend where to look additional hits to reconstruct conversion tracks. If no hits are found in the outer tracker, the next inner tracking layer is searched for hits consistent with the track hypothesis. The seeds are fit using the Kalman fitter pattern recognition algorithm which uses only a single gaussian to predict energy loss. The Kalman fitter trajectories tend to have less bend since they give a more conservative energy loss for tracks. The χ^2 of the fit determines the goodness of the track fit: it is cut at $\chi^2 > 100$. Overlapping trajectories with more than 50% are disambiguated by keeping the tracks with more hits or the track with a smaller χ^2. [43]

The inward tracks are now used to find a potential conversion partner. The innermost hit of the track is taken as the conversion vertex and a neighboring subcluster to the inward track subcluster is used for the momentum measure of the partner track. Starting from the innermost hit a hypothesis trajectory is built that is parallel to the inward track and ends at the neighboring subcluster. This hypothesis is used to search for track hits to potentially find a partner conversion track with opposite charge to the inward track. [43]

Any combination of pairs of the reconstructed conversion is used to fit a conversion vertex. To boost the efficiency of finding early conversion vertices, Gaussian sum filter tracks are also added to create conversion pairs with Kalman filter tracks. The conversion vertex will have just two tracks which will be parallel at a the conversion point. The vertex is fit by minimizing the χ^2 using a Kalman filter fit to give the conversion point. The vertex fit depends on a set of vertex parameters: position, covariance matrix, and track parameters constrained by the vertex position and their covariances.[17]. The goodness of the fit is given by its χ^2 value and will vary depending on the quality of tracks. Also the conversion vertices should not be reconstructed for real decays from mesons or primary interaction points. Some selection needs to be applied to the candidate conversion tracks to reject other objects like pions from a decay or from the primary interaction point.



(a) Conversion with hits in TIB

(b) Conversion with hits in TOB

(c) Cannot fit Conversion

(d) Only 1 track from Conversion

Figure 4.4: Different cases for photon conversions in the tracker. The first case for a conversion in the inner detector produces a conversion pair in line with the photon direction. For the next two figures the track with fewer hits becomes more difficult to recontstruct. In the last case, only one track points along the photon direction while the low p_T track curls away.

The conversion reconstruction efficiency depends on the tracking efficiency which depends on the radius of conversion. At large radius the tracking modules are more coarse and there are only single sided modules without stereo information, so the polar angle of a conversion will be difficult to determine. The most well measured lever arm will be for conversions in the pixel layer or in the inner detector where there are stereo hits. For two electron tracks in the outer tracking module, the two tracks will have many overlapping hits, which can be disambiguated by one stereo hit. The overlap between the tracks makes it more difficult to determine the precise conversion vertex. For two electron tracks, in the granular pixel layer the hypothesis conversion vertices will lie within a small region on the layer of the pixel detector where the tracks start. [43]

Also a conversion can occur asymmetrically so that the a majority of the photon energy is carried by only one of the conversion legs. This is shown in Figure 4.4d. A single well measured high p_T track of the conversion is reconstructed, while the much lower p_T leg of the conversion bends far away from the supercluster. In this case the second leg can have a momentum too low to be reconstructed or too poorly reconstructed to fit a conversion vertex. The trajectory of the reconstructed track still has usable information to determine the photon direction. These additional conversion tracks are identified within the Particle Flow reconstruction using not just the track information but also information from the calorimeters to identify tracks as electrons or other charged particles. This will be further described in Section 4.3.1.

4.3 Particle Flow Reconstruction

The Particle Flow reconstruction, which for the CMS collaboration has been renamed the "Global Event Description", follows the philosophy of reconstructing and identifying all final state particles in an event: electrons, photons, muons, and charged pions and pions inside of jets. Any pp collision of interest is described as an event which consists of physics objects made of sub-detector measurements: tracks, calorimeter deposits, muon chamber hits etc. This motivates the need to understand combinations of different sub-detectors and follow the energy flow of the particle in each detector. Reconstructing a final set of stable particles gives an optimal physical description of an event. A global event makes analyses with the same final state more consistent: the same interpretation of a set of detector measurements. This section will briefly describe the Particle Flow algorithm with particular emphasis on the improved photon reconstruction. The input to the algorithm are the elements which are high purity tracks, granular ECAL clusters, and HCAL clusters. Each of these elements will be combined based on their geometric association in the detector. The final output of the Particle Flow is a list of individual particles in the event accounting for all the reconstructed

measurements in the subdetector. The list of candidates is used in this thesis, to predict the amount of activity around a candidate photon and decide if it is in a jet. The Particle Flow also gives the important observable of ρ which is the average particle density in an event. This is an important measure of the amount of contamination from pileup interactions.

Depending on the purity of a certain signal, a physics object can be more or less ambiguous. Isolated objects at high p_T will have unambiguous interpretation. In some kinematic regions however, the identification has more and more ambiguity. An electron can be hypothesized as a converted photon or even be a misidentified charged pion. Combining the track information with the energy measurements in the ECAL and HCAL can disambiguate these cases. Also each subdetector individually may have some resolution, but the combination of detectors can be calibrated to better the resolve the four-momentum of a particle. A pion's momentum from the track bend can be underestimated or over estimated. But if the track can be associated to ECAL deposit or an HCAL deposit the calorimeter measured energy can be used to calibrate the momentum and improve the resolution of the four-momenta of pions. This then propagates to improved resolution of reconstructed jets, the missing energy of an event, and also a list of fairly unambiguous stable particles in the event. Thus the Particle Flow, has been renamed the Global Event description, to give a single and complete interpretation of an event.

The CMS detector is well suited for Particle Flow because of its large tracker, uniform solenoidal magnetic field, and granular Electromagnetic calorimeter. The 3.8 Tesla field and the granularity of the tracker give a large reconstruction efficiency for charged particle tracks and a small fake down to a small momentum as described in Section 3.2.1. The Electromagnetic calorimeter with 2x2cm^2 crystals give a good energy resolution and also a well-determined ECAL position for impacting particles. The HCAL is 25 times more coarse than the ECAL, which does not allow for charged hadrons to be well separated from photons in jets. However combining the track trajectory through the ECAL and the HCAL allows for angular separation between overlapping particles. Also the direction of the track can be used to predict what amount of calorimeter energy should be associated to the charged particle and neighboring particles. To prevent energy mis-measurement, the tracks used in the algorithm must have the smallest number of 'fakes'. Figure 4.5 shows a slice of the CMS detector and the different kinds of final state particles that can be reconstructed within the Particle Flow algorithm. The sequence of reconstruction will be described in this section. [19]

Since a majority of events at CMS feature jet final states, the Particle Flow algorithm aimed first to disambiguate particles within a jet to improve the jet energy and p_T resolution. In Section 4.3.1, the reconstruction of photons will described as well as their disambiguation from the reconstructed jets. Two thirds of the energy of a jet is carried by charged constituents, which motivates the

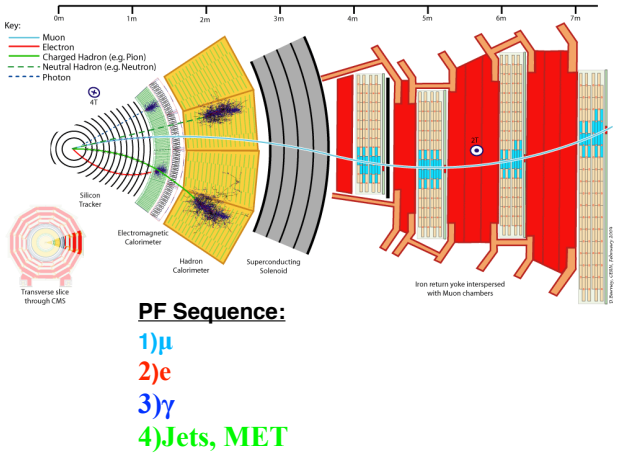


Figure 4.5: Slice of the CMS detector showing the length different particles traverse the detector and where particles are captured. [17]

need for excellent track identification. Track identification will also be important for reconstructing conversions. The iterative tracking procedure was aimed at lowering the number of fake tracks as much as possible while preserving a high efficiency for reconstructing low p_T tracks. [19] These tracks are also used outside of the Particle Flow algorithm by the CMS collaboration as high purity tracks. The high purity tracks are found using the following steps:

- To reduce the number of fake tracks, a tight seeding criteria is applied to track hits. These unambiguous track hits are fit to tracks.

- Removing the hits from the previous step, the seeding criteria is loosened and a new set of track hits is used to fit tracks. The looser seeding boosts the tracking efficiency with respect to the previous step. Removing the hits keeps the amount of fake tracks low due to reduced combinatorics for the pattern recognition. In three iterations of these two steps, more than 90% of charged hadron tracks in jets are reconstructed. [19]

- Finally, the constraint on the vertex of the track is loosened to allow for tracks associated to secondary vertices: particles from meson decays and photon conversions. This allows for the reconstruction of charged particles with only three track hits, p_T as loss as 150 MeV/c, and a vertex of more than 50cm from the beam line to be reconstructed while preserving a fake rte of the order of a percent. [19] This will be essential for conversion identification in Section 4.3.1.

[19] In order to trace the energy flow of particles through the calorimeters, the energy deposits on the calorimeter must be reconstructed to best separate overlapping EM showers in the ECAL or overlapping Hadronic cascades in the HCAL. Typical mesons in a jet like the charged and neutral

pions, will leave some deposit in the granular ECAL. ECAL clusters should disambiguate the two photons from the electromagnetic decay $\pi_0 \rightarrow \gamma\gamma$ and also charged particles that neighbor these secondary photons. The ECAL clusters should also be able to reconstruct low E_T energy deposits. The Moliere radius of the ECAL crystals is small at 2.2 cm and the front face of crystals in the barrel $2.2x2.2cm^2$ giving coverage of $0.0174rad$ in $\Delta\eta$ and $\Delta\phi$ giving fine grained 2D positions in $\eta - \phi$ of energy deposits.

Topological energy clusters are built using the following steps to exploit the granularity of the crystal calorimeter, and allow crystals to share energy between clusters:

- Seeds are selected from crystals above an energy threshold.
- Grow an energy cluster by adding nearest neighbor crystals (share at least a corner) as long as they are also above an energy threshold designed to reject noise. A single crystal can contribute energy to one or more energy clusters.
- Each crystal contributes energy with a weight based on its distance to the cluster position, the weighting factor of the ith crystal to the jth seed $\exp\left(-d_{ij}/R^2\right)$ where $R = 8$ is a damping factor to keep the cluster from engulfing a large amount of crystals. The energy measured by a crystal is split between several energy clusters based on how far it is the cluster position.
- The position of the cluster is re-evaluated by the weighted positions of all the contributing crystals where the energy weight $\log\left(E_{frac}/E_{total}\right)$. The position of the cluster is evaluating when adding each new crystal with the fraction in the previous step. As the crystals are added the cluster position changes, and so do the fractions in the previous step.
- The iteration between the previous two steps stops when the cluster position changes by an immeasurable amount.

These topological clusters result in as many ECAL clusters as seeds that are available. The result of this procedure is to have overlapping energy clusters where the energy measured by a crystal can split between several clusters. The position resolution of the energy cluster is not limited by the size of the crystal. This position resolution will exploited in Section 4.3.1 to measure an ECAL envelope for photons. Since each seed gives an ECAL cluster, noise from crystals could give rise to ECAL clusters causing "fake" energy deposits. To reduce the amount of this "fake" energy a seed threshold is applied to crystals that seed clusters and an energy threshold is applied to all crystals included in the ECAL clusters. These thresholds are listed in Table 4.1 for the ECAL barrel and the ECAL endcap. Because of the thresholds applied to the ECAL clusters, their energy will be smaller than those corresponding to photons in simulation. These "zero suppression" effects are

	EB	EE
Cell Energy Threshold (MeV)	80	300
Seed Threshold (MeV)	230	600

Table 4.1: PF ECAL Clustering parameters

calibrated to account for this energy loss. The calibration is derived using simulated photons across a large E_T range, and is applied to all clusters based on their uncorrected E_T. The calibrations are checked comparing the π_0 peak fit in 900GeV collision data and simulation. The mass measurement agrees within $\pm 2\%$ between simulation and data with resolution within 1 MeV. [24]. For the ECAL clusters in the endcap, there is a pre-shower detector of 2 layers of silicon strips with a lead absorber between them which corresponds to two radiation lengths. The measurement of MIPs in the strips can be converted to energy which can can be used to calibrate the ECAL cluster energy in front of the preshower clusters. The preshower scintillator strips are clustered in the same way as ECAL clusters. The ECAL clusters are associated with the preshower clusters by projecting the ECAL cluster envelope onto the preshower layers and collecting all the preshower clusters within the envelope. A calibration based on simulated photons is used to derive function that gives the total calibrated energy based on cluster energy and the collected energy deposited in the two preshower layers.

The HCAL clusters are also clustered from towers using the same calorimeter clustering procedure as above. HCAL towers are also "zero suppressed" if their energy is below 0.8 GeV. Likewise, the HCAL cluster enveloping is projected onto the ECAL to associate ECAL clusters within the HCAL cluster envelope. In this case, the ECAL and HCAL combination is used to give the energy of hadronic particles. The association between the track and the ECAL is made by extrapolating from the last hit of the track through the ECAL to a typical longitudinal shower depth.[24] If the track extrapolation is within the boundaries of a given cluster the track is associated to it. The envelope is enlarged a bit to account for the in-homogenous cracks and gaps in the ECAL. The track is also extrapolated to the HCAL at depth of approximately 1 interaction length of a hadron to match to HCAL clusters. The set of high purity tracks, ECAL clusters, and HCAL clusters and their associations, called PF Blocks, are the main input to the Particle Flow algorithm which then outputs final state particles.

The Particle Flow algorithm takes as input the blocks and creates final state particles: muons, electrons,photons, charged hadrons, neutral hadrons. The blocks make it easier to disambiguate which measurements correspond to the same particle. A muon is the least ambiguous since its track will have measurements in the muon chamber. As shown in Figure 4.5, the muon is the first to be reconstructed and passes through the full length of the detector and exits. At an early stage of the particle flow, the muon track and the energy deposits associated are combined into a Partcle Flow

muon. To prevent double counting the algorithm indices each sub-detector measurement and keeps a list of measurements that are already accounted for with a reconstructed Particle Flow particle.

Electrons are reconstructed next which are seeded from the Gaussian Sum Filter tracks. The GSF tracks are reconstructed in two main ways. The first uses a combination of ECAL information and tracking using a hypothesis track made from the helix given by the 3D information from pixel hits that extrapolate to a supercluster in the ECAL (described in Section 4.1). The second method involves only information from tracks that are reconstructed using the Kalman Filter algorithm. This approach is optimized for finding low p_T electrons below 10 GeV. These tracks are identified using a Boosted Decision tree to see if the Kalman filter track is compatible with an electron trajectory or another charged particle. [23] The discriminating variables used to identify electrons from other charged particles are:

- The KF track should be compatible with the electron path to the outer track layer. Electron path is given by the helical projection from the first few hits of the track under the hypothesis of an electron mass.
- The track momentum and its associated ECAL energy deposit is compared as E/p
- The large χ^2 value of the track corresponds to a very poor Kalman filter track, which does not account for bremsstrahlung energy loss
- The $(p_{out} - p_{in})/p_{in}$ indicates energy loss along the track corresponding to the electron hypothesis.

The curvature of these tracks are well determined enough to use tangents along the GSF track to find other tracks and clusters that come from the electron. The tangents are computed at each track layer which could be the hypothesis direction of a bremsstrahlung photon. The tangents that point to an ECAL cluster are collected as part of the electron. If the bremsstrahlung photon converts, the conversion track is identified and becomes part of the electron. The identification of the conversion track is done by combining discriminating variables into a single Boosted Decision tree discriminator. The signal sample is is made using tracks from simulated electrons that are found within proximity of the bremsstrahlung tangents along the GSF track: $|\Delta\phi| < 0.3$ and $|\Delta\eta| < 0.05$, and the background sample consists of tracks matched to charged pions from simulated jets. The discriminating variables for this BDT are used identify tracks from converted bremsstrahlung from KF tracks along the GSF track tangents:

- The number of track hits (small for converted bremsstrahlung tracks)
- The signed transverse impact parameter of the track

- $\Delta\eta$ between the tangent along the GSF track and the converted bremsstrahlung track is used, and should be small for converted bremsstrahlung tracks
- To prevent track double counting the number of shared track hits between the GSF track and the KF track are used.
- The ECAL energy associated to the Kalman Filter track is compared to the track momentum E/p
- The ratio of the track p_T of the GSF track and the KF track

The tracks are collected and are associated to the electron. The superclusters defined in Section 4.1 are used to collect additional ECAL deposits to improve the energy containment of electrons that radiate. The supercluster creates an envelope for ECAL clusters to be combined into a photon or an electron. If the GSF track extrapolates to one ECAL cluster in the supercluster, it is associated to all the ECAL clusters in the envelope. This technique is used to make photons, which are reconstructed after electrons in the Particle Flow reconstruction sequence. This will be described further in Section 4.3.1.

The set of particles that are reconstructed next in the particle flow sequence are additional photons, neutral hadrons and charged hadrons. The charged hadrons have a track always associated to them which can extrapolate to an ECAL cluster and also an HCAL cluster as described above. The four momentum of the charged hadrons is determined from a combination of the three measurements. If the energy found in the calorimeters is small compared to the track momentum or no energy is found in either the ECAL or the HCAL, the track is considered to be misidentified or mis-reconstructed and the track is cleaned away and not reconstructed as any particle. The identified charged hadrons combine the sub-detector measurements of the track momentum, the ECAL energy, and the HCAL energy to determine the 4-momentum of the charged hadrons:

$$E_{calib} = a + b\,(\eta, E)\,E_{ECAL} + c\,(\eta, E)\,E_{HCAL}$$

The coefficients are measured from simulated charged hadrons, and the energy E is taken from the track momentum or the measured calorimeter energy depending on which is greater. If the ECAL cluster energy is much greater than the track momentum, then the ECAL cluster is split to give a more compatible energy measurement with the track momentum. The above calibration makes the identification of tracks very important. Tracks misidentified as charged hadrons that come from converted photons or converted bremsstrahlung will be mis-calibrated as underestimated hadronic energy which will overestimate the energy of the event. The track classification for electrons is done

using a multivariate classifier as described above and the identification of conversions for photons is described in the next section. Neutral hadrons are made by combining the HCAL cluster energy with the ECAL cluster energy associated with it. For both the charged and the neutral hadrons the ECAL is used to calibrate the energy response of the HCAL. The remaining ECAL clusters that are not associated to a track or HCAL cluster are reconstructed as photons.

The photons not made from the supercluster footprint are assumed to come from additional photons from secondary decays like $\pi_0 \to \gamma\gamma$ within a jet. The set of reconstructed charged hadrons, neutral hadrons, and photons are input to jet clustering algorithms to form jets. Typically a jet consists of 20% of neutral hadrons, 65% of charged hadrons, and 15% of photons. Finally the τ can be reconstructed from leptons and also from the reconstructed jets. The sum of the 4-momenta of all final state particles is summed and the negative of the transverse component of the sum is the Missing E_T. In addition, the final state particles from pileup interactions can be clustered as jets as well using the FASTJET, and then the distribution of these "pileup jets" indicate the average amount of energy contamination. Also the same set of particles can be used to compute the isolation around isolated particles from signal processes like the photons $H \to \gamma\gamma$. The sum of the p_T of these final set particles in a cone around the signal photon which are used instead of the detector based isolation described in Section 4.1. The ECAL isolation is replaced with the sum of the E_T of the photons built in Particle Flow. The HCAL isolation is replaced with the sum of the E_T of the neutral hadrons built in Particle Flow. The track isolation is replaced with the sum of the p_T of the charged hadrons which can also be associated to a vertex.

For this thesis, the Higgs decay to photons, the above Particle Flow particles will be important to identify signal photons from jets based on the particles around a photon candidate. This experimental handle is known as Particle-based Isolation. Isolated photons produced from prompt decays like the Higg decay will not neighbor any high pt charged or neutral particles, but for photons reconstructed in a jet there will be a spray of particles measured close to the photon. The isolation is a measure of the activity around a given particle, and particle-based isolation is given by the p_T of the additional particles found around the photon. This will be an important handle to discriminate the large background from jets from QCD. Since jet clustering algorithms use Particle Flow particles as input, the isolation is a good representation of the 'fake' hypothesis of a photon. A map of the isolation around signal photons for the sum p_T of each of the Particle Flow particles:charged hadrons, neutral hadrons and photons will give softer energies relative to the photon energy. These isolation deposits mainly come from the pile-up, so these photons will have a good measure of isolation. For photons reconstructed inside of the jet from many neutral pion decays, the photon energy will comparable to neighboring particles in the isolation cone. These 'fake' photons will be observed

as less isolated. The isolation relies on well identified individual Particle Flow Particles. If a photon converts, the tracks from the conversion can be misidentified as charged pions. In this case, prompt signal photons appear as though they are within a jet and will be misidentified as backgrounds. This motivates the need for isolated photons to be reconstructed in Particle Flow (as described in Section 4.3.1) to clean up the isolation predictions.

The Particle Flow particles are used to measure the median background from particles originating from the Pile-Up denoted by ρ. The pile-up is from the softer overlapping pp collisions to a hard scattering process and will result in some 'fuzz' of energy across the calorimeter and tracks in the detector for every event. It is useful to have a prediction of the amount of this contamination in some average area. This idea came from studies to improve the jet energy resolution by determining an effective jet area and subtracting an average energy contamination to improve the jet momentum. The average number of particles is determined by a measurement of ρ which gives a particle energy density in a characteristic area. The characteristic area is determined by a rapid computation clustering particles in the event into jets, using the FASTJET package. Across all pile-up jets in a Higgs event one can compute the median of the jet p_T^j divided by the area of the jet $A_{eff}^j = \pi R_J^2$ to give $\rho = median\left\{\frac{p_T^j}{A_{eff}^j}\right\}$ to obtain the average amount of contamination from diffuse pile-up jets. [15] For Higgs events, this variable is useful to predict how many soft particles will be around the photon when computing the isolation, so it is input to the Photon identification BDT. For vector-boson fusion Higgs, there will be two jets from the scattered quarks. The ρ can be used to subtract the pileup contamination from the two jets: $p_T^{jcalib} = p_T^j - \rho A_{eff}^j$.

This section has described the Particle Flow algorithm, and the sequence by which it reconstructs a set of final state particles. These particles become important input for the identification of photons using particle-based isolation. They are used to reconstruct jets which are additional signatures for vector boson fusion Higgs, and also input to measuring the average energy density from pileup. The next section will describe the reconstruction of Particle Flow photons which remove the photon footprint from the isolation sum to improve identification and also allow for further conversion identification.

4.3.1 Particle Flow Photons

This section will describe how the reconstructed ECAL clusters assembled in the Particle Flow algorithm are aggregated into a supercluster. Compared to the superclustering approach used in Section 4.1, this approach will make use of already pre-assembled clusters instead of dominos of crystals. The assembly of Particle Flow ECAL clusters have particular advantages, the seeding

threshold makes it possible to reconstruct very soft energy clusters with low E_T, the iterative clustering allow neighboring clusters to share ECAL energy to split overlapping showers, and also the clustering gives a precise position of each subcluster. This information is useful for measuring the pattern of the electromagnetic shower on the ECAL and being able to separate the energy deposits from the pileup.

The first version of Particle Flow superclusters, also includes the standard superclustering used in 4.1. This is done by first reconstructing the standard superclusters. The total number standard superclusters will consist of a majority of photons from jets. These photons are rejected so that they do not harm the jet reconstruction. The detector isolation described in Section 4.1, is summed together and required to be less than 10GeV: $Tk_{iso}+ECAL_{iso}+HCAL_{iso} < 10GeV$. For the selected superclusters, each crystal is matched to one in a Particle Flow ECAL cluster. All matching clusters are grouped together to build a new supercluster from the Particle Flow ECAL clusters.

This prototype supercluster has two main functions for the analysis. The reconstruction of the photons from superclusters in Particle Flow removes the ambiguous candidates that would be reconstructed if a signal photon is misidentified as a jet. This improves the background rejection power of the particle-based isolation by accounting for prompt photons. This can be seen in the amount of predicted activity around signal photons as shown in Figure 4.8. Without removal of the EM footprint associated to the signal, signal photons appear to be surrounded by overlapping photons and charged particles, just as they would be for background photons in a jet. Cleaning the EM footprint shows much less activity as is is expected and accounts for only some activity from the pileup. This misidentification, however, cannot just be done with the above isolation cut and will rely on a combination of sub-detector information to interpret detector measurements as a photon or as a jet of particles. In particular, photon conversions need to be identified with the tracks associated to the ECAL clusters in the supercluster and properly disambiguated from charged pion tracks inside of the jet.

The conversion identification will play an important role in identifying prompt photons or jets. Section 4.4 described how conversion pairs could be built from tracks that can fit a conversion vertex. There it was also shown that the conversion pair reconstruction depends on where the conversion occurs and also whether tracks are reconstructed from the e^+e^- pairs. Tracks reconstructed as conversion pairs are matched to the high purity tracks to successfully identify them as coming from e^+e^- instead of charged pions. However, it was also found that many high purity tracks are still misidentified as charged pions. These additional tracks are the cases where the conversion pair is not reconstructed or sometimes only one track is present as was shown in Figure 4.4. A multivariate classifier is used to identify these tracks in the Particle Flow. As input discriminators it makes use of

some track information, and also combinations of sub-detector information like the energy associated to the track in the ECAL and the HCAL. The input discriminators are described as follows:

- The high purity tracks are fit with the Kalman Filter pattern recognition algorithm which was described in Section 3.2.1. This algorithm is designed for tracks with little p_T loss in the tracker and also a larger helical radius for heavy particles, so an electron trajectory will tend to have poorer track fit. So the χ^2 of the track fit is included as a discriminator.

- Since a conversion can happen anywhere in the track material, it tends to have shorter tracks compared to promptly produced particles from the collision point. The radius of the first track hit and the total number of tracks hits provide discrimination of conversion variables from promptly produced particles.

- The conversion vertex should give tracks that are parallel to the original direction of the photon. The direction of the photon can be estimated as the line form the primary collision point to the first hit of the track which is close to where the conversion takes place. From this line one can measure the angle $\delta\phi$ with the track momentum, for conversions it should be small while for charged particles coming from a decay it will be non-zero. This variable discriminates conversions from other displaced vertices that involve decays based on an opening angle.

- A missing hit along the trajectory of the track indicates that it could have converted in the track region where the hit is missing.

- The impact parameter is defined as the distance of closest approach between a track and a point, which can be a vertex. For displaced vertices, like those from conversion the impact parameter tends to be large when measured to the collision point (since that is not the origin of the track). The sign of the impact parameter is determined by the charge. For prompt tracks from pions in a jet, the impact parameter will be close to zero with symmetric positive and negative values. For the conversion tracks, they will tend have a positive impact parameter which is poorly measured from the track extrapolation. Dividing the impact parameter by the significance of the error exaggerates the difference between the prompt tracks and those from conversion.

- The Particle Flow algorithm allows for the track to be extrapolated to the ECAL and the HCAL and so the energy there can be associated to the track. Conversions should have a large amount of ECAL energy associated to the track while charged hadrons will have a large amount of HCAL energy associated to the track. Also for the conversion tracks the E_T measured of the sum of the ECAL clusters divided by the track pt should peak at one.

Figure 4.6 shows the signal and background discrimination for the input variables used to discriminate conversion tracks from those from promptly produced charged particles. These variables are combined into a single discriminating variable from the Boosted Decision tree procedure shown in Figure 4.7. The Particle Flow Photon reconstruction is done if there are not many tracks identified

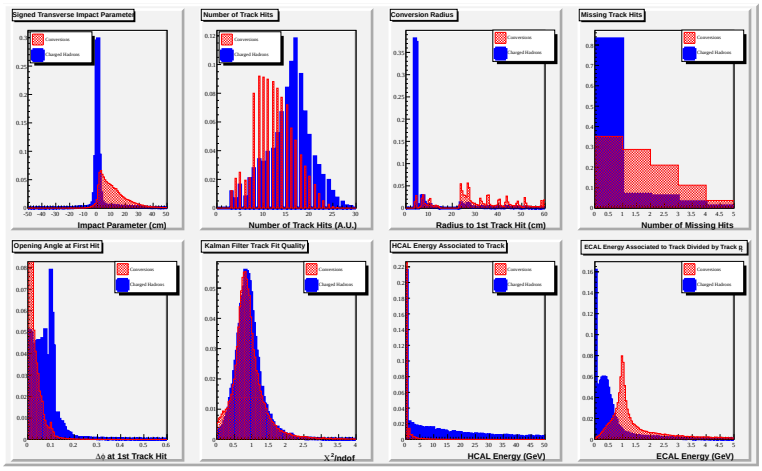


Figure 4.6: Input discriminating variables for the conversion identification Boosted Decision Tree.

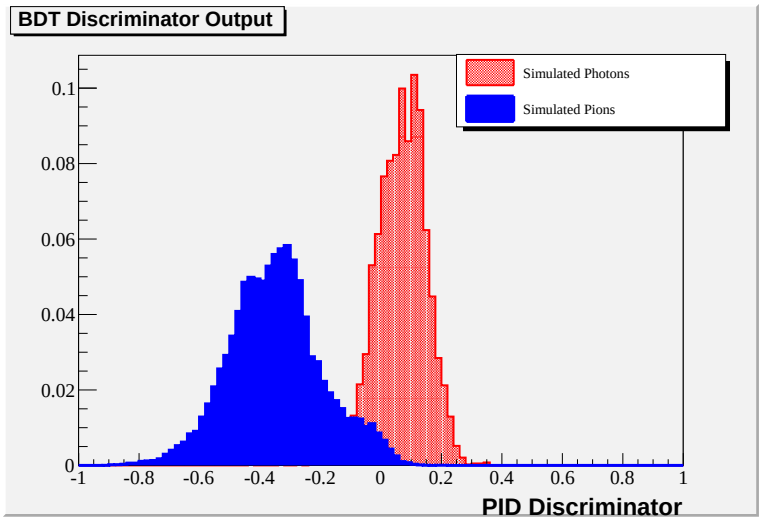
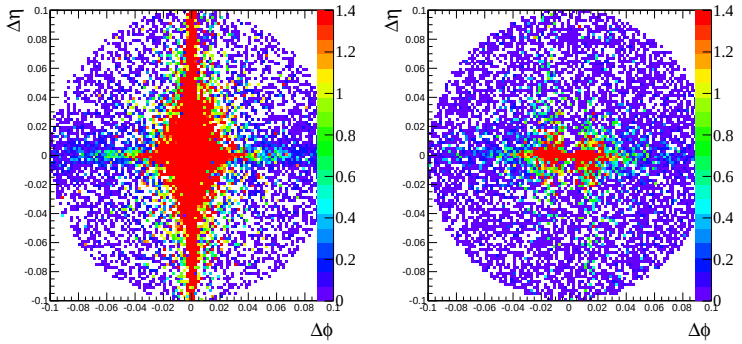
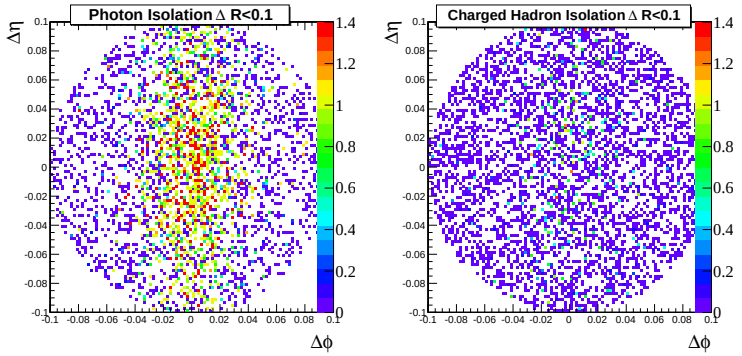


Figure 4.7: The Boosted Decision Tree output to classify conversion tracks.

as pions that extrapolate to the supercluster. This is done by taking the tracks below an a Boosted Decision tree discriminator value (shown in Figure 4.6) of $MVA < 0.0$ and applying a cut on the sum of their p_T as follows: $\sum p_T^{track} < 2.0 + 0.001 p_T^{pho}$. The $\sum p_T$ contribution of pions is required to be below a minimum of 2GeV and this cut is raised by the 0.1% of the photon energy. This cut very strictly rejects photons from jets based on the number of overlapping charged particles with the supercluster.



(a) Photon Isolation(left) and Charged Hadron Isolation (right) Before Photon Reconstruction



(b) Photon Isolation(left) and Charged Hadron Isolation (right) After Photon Reconstruction

Figure 4.8: The effect of introducing a Particle Flow photon to remove the EM footprint from the photon isolation. The photon isolation now does not count for the signal photon energy in the isolation sum when the Particle Flow identifes a photon. The identified conversions are removed from the Charged Hadron isolation.

The Particle Flow Photons contribute to the Higgs analysis in two main ways. Firstly, the photon footprint is removed from the particle based isolation. Instead of a single prompt photon being reconsidered as many Particle Flow Photons, one for each ECAL cluster, they are united in one photon object. This cleans the picture for signal photons since all the energy is included in one reconstructed photon object and the rest is summed as the isolation. Also the tracks are identified as charged pions or e^+e^- tracks associated with the photon. Figure 4.8 shows how the isolation around signal photons changes with this improved reconstruction. Thus the Particle Flow reconstruction to give a better description of signal photons and while preserving the description of jets. Also the single leg tracks add to the total number of reconstructed conversions which will further improve the vertex finding for the Higgs decays.

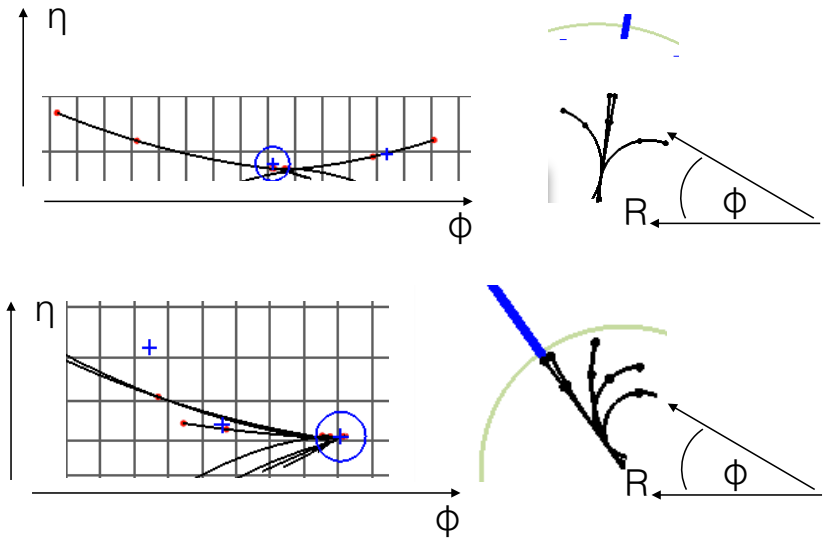


Figure 4.9: Helical Tracks projected on the ECAL

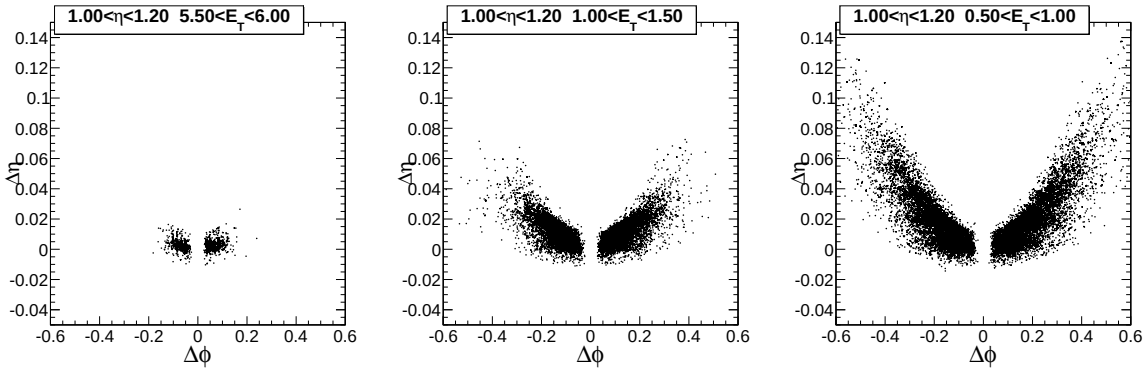
The latest version of Particle Flow photons aims at improving the photon reconstruction to include more ECAL clusters from the electromagnetic shower while minimizing contamination of ECAL clusters from pileup. The superclustering is again done dynamically to capture the photon fragments. But clusters are included not only based on $\Delta\eta$ and $\Delta\phi$ from the seed cluster, but also based on the E_T of the energy cluster. In Section 4.1, the hybrid supercluster was described which made the assumption that all the charged e^+e^- fragments of converted photons only bend only along the ϕ direction making a rectangular supercluster. However, charged particles can bend significantly in the field depending on their p_T. The deflection of a charged particle in the solenoidal field of CMS follows a helix, and the bend of the helix in $\Delta\eta$ and $\Delta\phi$ can be predicted from the E_T of the ECAL cluster. Figure 4.9 shows several tracks in the $R-\phi$ view, where two high p_T tracks point roughly straight along the direction of the original photon. The two low p_T tracks bend

away from the original direction. When the track is projected onto the ECAL, the helical trajectory makes roughly a parabola. At lower and lower p_T, the track bend of the track becomes tighter and tighter. When projected onto the ECAL the parabola becomes more and more curved. The pattern of ECAL clusters from a photon can be predicted using the E_T of the energy clusters. For softer low E_T clusters, one can measure the deflection along both $\Delta\eta$ and $\Delta\phi$, and as the clusters becomes harder the deflection in both directions becomes smaller. The correlation between $\Delta\eta$ and $\Delta\phi$ and the cluster energy does not exist for pileup clusters which are randomly distributed about a seed cluster. This parameterized envelope accounts for energy and also position, which makes it a robust EM footprint at high pileup.

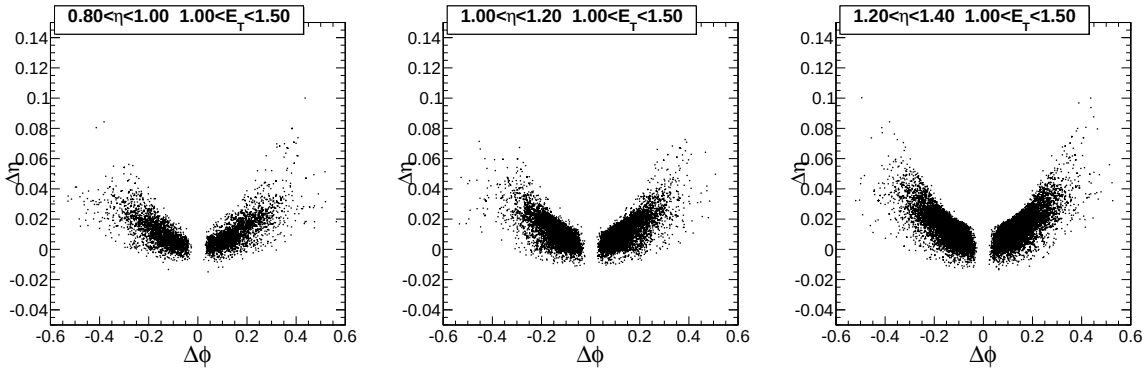
The pattern of ECAL clusters is studied using simulated photons to measure an EM shower shape region which depends on the E_T of the cluster, and how far it is η, ϕ from the seed cluster position which corresponds to the original direction of the photon. The bend in the direction will saturate for high E_T clusters so the region is parameterized in terms of $\log E_T$. Every region of the ECAL will have a different number of conversions which is shown in terms of η in Figure 3.9, this will give a larger population of bended tracks at large pseudorapidity. For simulated photons, the distribution of ECAL cluster positions about the seed position is finely divided into slices of $\log E_T - \delta < \log E_T < \log E_T + \delta$ with $\delta = 0.1$ and also slices of η. Each slice will have a distribution of cluster positions within a region bounded by two parabolas of the form: $\delta\eta = a\left(\delta\phi\right)^2 + b$. The curvature of the parabola is determined by the $\log E_T$ range and the spacing between them is determined by the η region. The envelopes trace out parabolic boundaries shaped like a mustache, so the supercluster algorithm is aptly named Mustache superclustering. Figure 4.10 shows the pattern ECAL clusters make in different η and E_T slice about the seed position. For the high E_T range the cluster positions lay along a straight line of $\delta\eta$ while it curves more and more for the low E_T slices. The "whiskers" of the Mustache region become more populated when the seed position is at larger η corresponding to more soft clusters where there is a larger material budget.

4.4 Photon Energy Multivariate Regression

The energy of reconstructed photons is corrected using multivariate regression trained on simulated photons measured at CMS. The aim of the regression is to map a set of inputs to the energy scale E_{reco}/E_{true}. The training is supervised since each event has a known energy for the simulated photon, E_{true}, and also the reconstructed energy from the superclustering procedure, E_{reco}. The Boosted Decision Tree technique is used for energy corrections due to its stability and fast-training time. A single decision tree is trained to give successive cuts on the input variables to optimally



(a) Mustache Shape for $0.5 < E_T < 1.0$ (b) Mustache Shape for $1.0 < E_T < 1.5$ (c) Mustache Shape for $5.5 < E_T < 6.0$



(d) Mustache Shape for $0.8 < \eta < 1.0$ (e) Mustache Shape for $1.0 < \eta < 1.2$ (f) Mustache Shape for $1.2 < \eta < 1.4$

Figure 4.10: Mustache shower shape for different slices of E_T on the top for an eta region and on the bottom for a different η regions with fixed E_T range.

group events with similar E_{reco}/E_{true}. The decision tree consists of nodes where cuts on the input variables are designed to split events which have a similar energy scale. This section will motivate the large improvements in using the multivariate technique. The MVA can be compared to simple corrections to demonstrate the effectiveness of multivariate techniques.

4.4.1 Boosted Decision Trees

Boosted Decision Trees, BDTs, were most prominently used for Particle Identification in High Energy Physics data analysis as a more robust technique compared to Artificial Neural Networks (MiniBoone Experiment [47]). The technique of boosting addresses the common problem of 'overtraining' in learning algorithms. Overtraining makes the output of the algorithm overly dependent on the inputs. Neural networks more closely model multivariate functions, so a small change in a input variable $x \to x + \delta x$ can lead to a large change in the output value $f(x + \delta x) - f(x) \gg \epsilon$. These jumps in output may not be properly modeled, and should be properly suppressed. The BDT technique trains many weak decision trees whose output is combined using a 'boosting algorithm'. The boosting algorithm is designed to up-weight or boost events that are misclassified when training a new tree leading to successively less mis-modeling. This section will describe the training algorithm and compare the output of the BDT regression routine to a simpler cut-based technique. The BDT will also be used to classify events and so the classification algorithm will also described. The effectiveness of this technique will be shown for simulated photons, which leads to improved signal model for the Higgs mass with a narrower width.

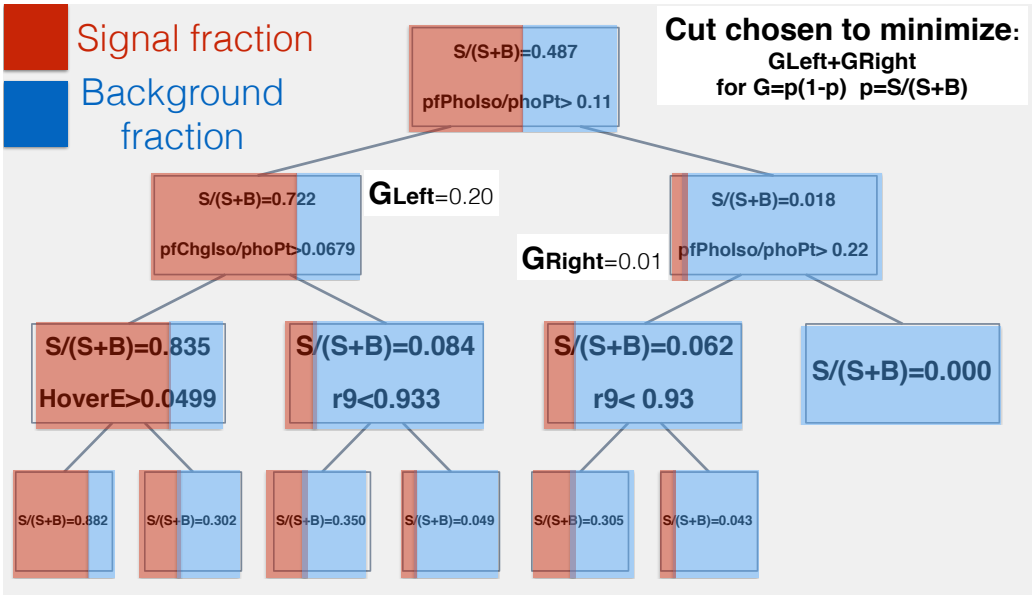


Figure 4.11: Example BDT Classifier Tree showing cut optimization procedure to separate signal and background events.

For a simple classification problem of events by signal or background, a decision tree consists of nodes split to give larger purity deeper into the tree. Each node is a cut on an input variable chosen based on its discriminating power, which can be quantified as $S/(S+B)$. The signal purity, p, can be defined at every node as the sum of signal events divided by the total number of events in the sample (Signal+Background). The best cut value is chosen such that if $G = P(1-P)$ for purity P, then the cut will minimize $Gini = G_{left} + G_{right}$ for the G computed on each side of the cut resulting in two nodes. For a cut that completely separates signal from background $Gini = G_{left} = G_{right} = 0$, where as for a cut with some impurity $Gini$ is pulled away from 0. The algorithm starts at a root node with the an admixture of signal and background in the training sample and each cut reduces the impurity by maximizing $Gini = G_{left} + G_{right}$ as shown in Figure 4.11. [48] The termination of the algorithm is defined based on the minimum number of training events on each node, a limit on the number of generations after the root node, or a requirement on the purity. The output value scores the signal-likeness of a particular event based on a number between 1(strongly signal-like) and -1(strongly background-like). Figure 4.11 gives an example of a BDT classifier tree where the colors represent p, signal fraction, and (1-p) for the background. The first cut is chosen to minimize$G_{left} + G_{right}$. The root node starts with equal fractions of signal and background, and the cut separates almost pure background to the right and node to the left has 70% signal. Successively cutting until reaching the base of the tree gives a much purer signal node with 80% signal. This procedure is also extended so that multiple variables can regress toward the mean of a target variable.

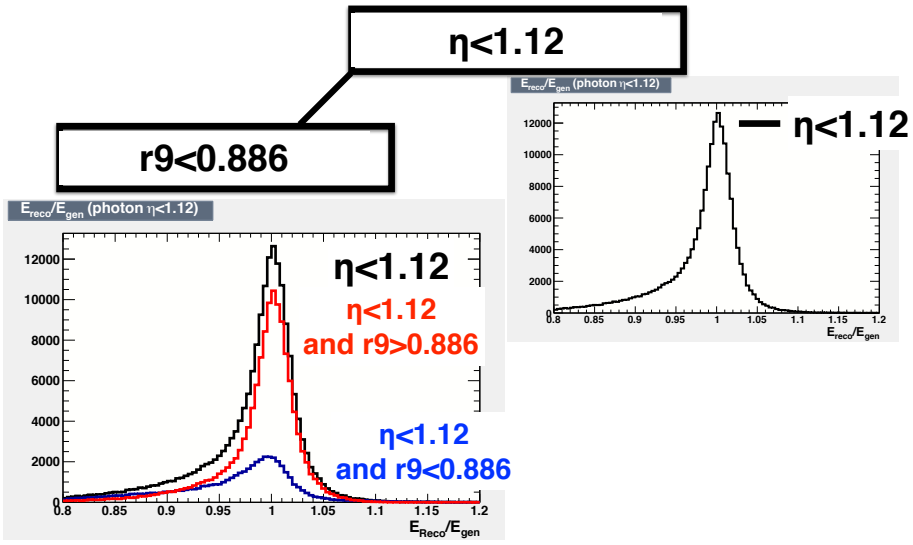


Figure 4.12: Boosted Decision tree for regression.

Any regression procedure attempts to determine the strength of the relation between a give set

of input variables and a set of responses. The splitting criterion for a regression tree is based on the average squared error of a target variable: $\overline{V} = 1/N \sum^{N} (y - \overline{y})^2$ where $\overline{y}$ is the average of the sample. Starting at the root node, $\overline{V}$ of the training sample is computed. For a given input variable x_1, the best chosen cut is the one that minimizes: $\overline{V}_{left} + \overline{V}_{right}$. Following the above procedure successive cuts result in a predicted value of the target y with a smaller and smaller average error. Figure 4.12 shows a regression tree with the target distribution of E_{reco}/E_{gen} shown at each node. The first node has a core distribution at one with a long low tail. The cut at $r9 < 0.0886$ splits the distribution into one that is narrow and peaked at one, while the other distribution is wide and will be further split into narrow distributions peaked at a value that corresponds to the energy correction.

Training many decision trees independently would not address the problem of overtraining, since each tree will have a different misclassification rate. Instead, the boosting procedure combines each decision tree to minimize the overall misclassification rate. The weighted sum of tree outputs can be written as:

$$F(x) = \sum_{m=0}^{M} \beta_m f(x, a_m)$$

The function $F(x)$ is the sum of weighted output of m trees denoted as base functions $f(x, a_m)$ where a well chosen set of values for a_m, β_m will minimize the overall error determined by a loss function. The boosting procedure is fully specified by the choice of loss function. The Gradient Boosting technique for classification makes use of the loss function: $L(F, y) = \ln\left(1 + e^{-2F(x)y}\right)$ which is the binomial log likelihood. The variable $y \in \{-1, 1\}$ and the term $F(x)\, y$ is the margin which gives the goodness of classification. The binomial log loss function penalizes $F(x)$ with opposite sign of the target y, which would correspond to signal misclassified as background or vice-versa. The loss function for regression is the Huber loss function:

$$L(F, y) = [y - F(x)]^2 \; for \; (y - F(x)) \leq \delta$$

otherwise

$$L(F, y) = \delta\left(|y - F(x)| - \delta/2\right)$$

The Huber loss function is quadratically increasing with the residual:$|y - F(x)|$ and at δ increases linearly along the tangent so that the loss function is differentiable. This form makes the loss function robust against outlying events and also allows for convergence with the gradient descent minimization technique. The minimum of the loss function is found numerically using the gradient, since incremental steps along the steepest descent of a function will lead to its minimum. The gradient boosting technique follows these steps to find the minimum of the loss function:

- Initial value set $F(x)_0 = \sum_i^n L(y_i, \beta_i)$ for the β_i that give the minimum loss for each target y_i, so $F(x)_0$ is a constant value independent of the inputs.
- In proceeding m iterations over the total number of trees, define a residual that defines the gradient step:
$$\left[\frac{\partial L}{\partial F}\right]_{F(x)=F(x)_{m-1}}$$
- A regression tree is grown with the inputs in the training sample to target the residuals in the previous step, the output nodes give $g_m(x)$
- Compute the one dimensional optimization of the weight: $\beta_m = \sum_i^n L(y_i, F(x)_{m-1} + \beta g_m(x))$ for the β that gives the minimum loss.
- Update $F(x) = F(x)_{m-1} + \beta_m g_m(x)$

Including a scaling term to the tree weights β_m improves the convergence of the above procedure by reducing the learning rate of the algorithm. The tree weight is shrunk by a factor $\beta_m \rightarrow \epsilon\beta_m$ which reduces the step size when a step is taken in the direction of the gradient. A "shrinkage" scale of 0.1 allows for 1000 trees trees to be trained in reasonable time which gives improved accuracy while further reducing the overtraining. [26] [47] [48]

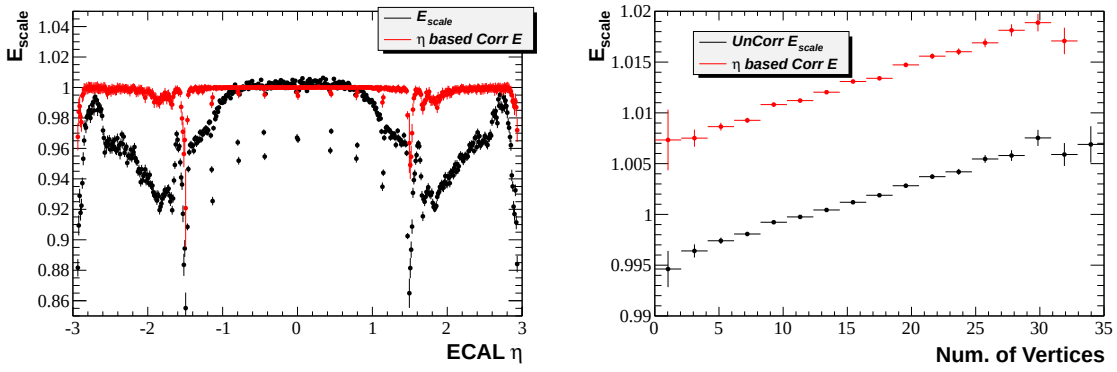
In general, multivariate techniques are used to algorithmically model the dependance between a set of inputs and a given set of dependant variables. For classification, each tree terminal node creates a multi-dimension boundary $t(x)$ and estimates the likelihood ratio: $\frac{\mathcal{L}(t|S)}{\mathcal{L}(t|B)}$ in the region. The regression technique creates a multi-dimension boundary to give the best estimate of a target dependent variable. The classification techniques have been commonly used in High Energy physics. It was first used as an improvement to Artificial Neural Nets for particle identification at MiniBooNE, and since then has also been used for particle identification in the LHC experiments. It has been used to identify electrons from converted photons from charged pions in the Particle Flow reconstruction described in Section 4.3. This technique will also be used to discriminate signal photons in Higgs events from those that come from meson decays. The next section will describe how the regression technique models various effects to give an accurate energy correction. The BDT robustness against overtraining reduces the sensitivity to certain weak variables while still preserving their effect on the energy correction. This allows many input variables to be included to improve the energy corrections as much as possible. The same inputs are also used to predict the energy resolution by targeting the variance about the reconstructed energy. Important gains in sensitivity are made by classifying Higgs events based on the energy resolution which will described in Section 5.6.

4.4.2 Energy Corrections with BDTG Regression

This section will motivate the gains made using a multi-dimensional energy correction as opposed to more standard corrections done by an analyzer one variable at a time. For demonstration purposes in this section, a BDTG regression is trained with only a small set of inputs to model the major features of the energy correction which can then be compared to a simple univariate correction. The simple set consists of the following inputs:

- η coordinate of the photon on the ECAL indicates the material budget in front of the tracker as well as the performance of the different super-modules. The low η region should have the best containment and the smallest energy correction.
- The r9 shower shape variable indicates the containment of the photon shower on the ECAL. The size of the energy correction should be inversely proportional to the r9 variable.
- The number of reconstructed vertices indicates the amount of additional interactions in the event. This gives the average energy contamination in the reconstructed photon energy.
- The reconstructed photon transverse momentum gives an indication of how well resolved the photon energy will be on the ECAL.

The above input variables will be used to estimate event-by-event the energy correction applied to reconstructed photons. A well trained BDTG correction should give an accurate energy correction that addresses the two major issues to the photon energy correction: the containment of converted photons and the contamination from the pileup interactions. The multi-dimension approach can be



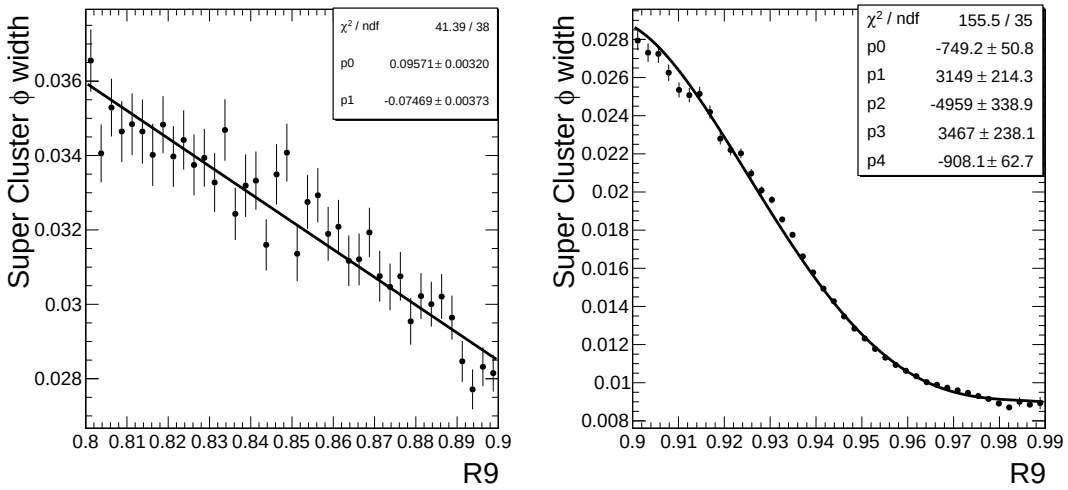
(a) a. Simple *eta* correction

(b) b. Simple Corr NVtx view

Figure 4.13: Corrections derived by hand based only on the η position in the detector.

compared to simple corrections done by hand with one variable. The η coordinate, pseudo-rapidity, can be used to derive a correction by computing the average energy scale: E_{reco}/E_{true} in small

ranges of η. This is shown in Figure 4.13 where each point is the average energy scale in small *eta* slices ($\Delta\eta = 0.012$) showing the energy response with respect to the position in the ECAL where it is measured. The energy scale in the central barrel region of the ECAL $|\Delta\eta| < 1.0$ is basically flat and close to 1.0. Outside of the central region the energy scale drops steadily at larger psuedo-rapidity where the photon will traverse larger and larger amounts of material in the tracker (as shown in Figure 3.9). The large discrete jumps in the energy scale show where the EM shower leaks into the support structure of the ECAL. In the endcap region of the ECAL, $|\delta\eta| > 1.5$ some of the energy is deposited in the Preshower lead strips while some is lost in early showers in the tracking material. To model an $f(\eta)$ single variable energy correction, one can take $1.0/\overline{E}_{scale}$ as the correction for every photon with an impact point within a particular η slice. This correction (in Red) makes the average energy scale rougly uniform and flat with respect to psuedo-rapidity. However, the energy scale with respect to the number of additional interactions is not flat. The average energy scale increases with the number of additional interactions as there is a larger and larger probability of including overlapping particles when reconstructing the photon. Events with photons at large pseudo-rapidity with contamination from pileup will be over-corrected with the η based correction. This motivates the need for a multi-variable correction that takes into account the correlation between different input variables and also their correlation with the energy scale.



(a) Linear Correlation $\phi - width$ and r9

(b) Linear Correlation $\phi - width$ and r9

Figure 4.14: Example linear and non-linear correlations between $\phi - width$ and r9

The input variables to the regression are not necessarily independent variables, but may have some linear and even non-linear correlations. One such example, is the correlation between the gaussian phi width of a supercluster and the shower shape variable R_9. As described in Section 4.1, the spread of the supercluster is allowed only in one direction. This spread should be correlated to

the the containment of the supercluster indicated by R_9. Figure 4.14 shows how the correlation looks for different values of containment. Very poor containment leads to basically a linear correlation between the gaussian phi-width and the containment. This is because dominoes are found further and further away from the seed with more radiation from the photon. A linear combination of the two variables can be used to make a single variable. Above $R_9 > 0.9$ the parameters of the algorithm give a higher order correlation. Figure 4.14 shows an inflection point $R_9 \sim 0.94$ where unconverted photons now cut off the superclustering close to the seed. The decision trees will sequentially split R_9 and the ϕ-width to model both the linear and non-linear behavior of the two variables. This example motivates the need to choose input variables that are not just strongly correlated with the target energy correction, but also have usable correlations amongst themselves to predict the energy correction.

Some optimization of the BDT performance comes from choosing the right parameters of the algorithm, but the main improvement is from choosing the best set of variables with appropriate correlations to each other and the target energy correction. Including additional variables allows for extending the depth of trees to finely split the target distribution as shown in Figure 4.12. Adding these variables:

- A measurement of the material budget in each of the three track layers (shown in Figure 3.9)
- The gaussian width of the supercluster in *eta* and also the gaussian width in ϕ
- The impact point of the shower on the face of the crystal
- The shower shape variable of the energy contained in a 5×5 crystal array to the total supercluster energy

Figure 4.15 shows the improvement in the energy correction which now has a smaller spread about one. The regression based energy corrections have been applied to the Partice Flow photons described in Section 4.3.1, following a strategy of factorizing the regressions into two steps. The first step is to correct the individual clusters within the photon supercluster for the inhomogenity in the ECAL. As described in Section 3.2.2, the ECAL is not entirely homogenous. The crystals are separated by small air gaps where small fractions of the shower can leak. The heavy modules and supermodules are held up by a support frame where a significant portion of the shower can leak. These effects are unrelated to the pileup or the photon showering in the tracker, and so can be trained using a separate regression routine. The energy correction is derived from mainly the position of the ECAL impact point:

- The cluster position on the crystal face

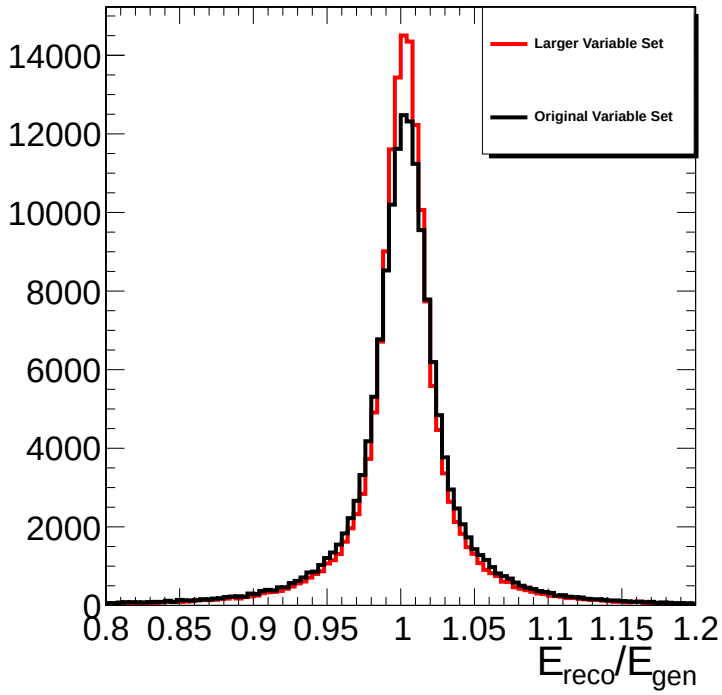
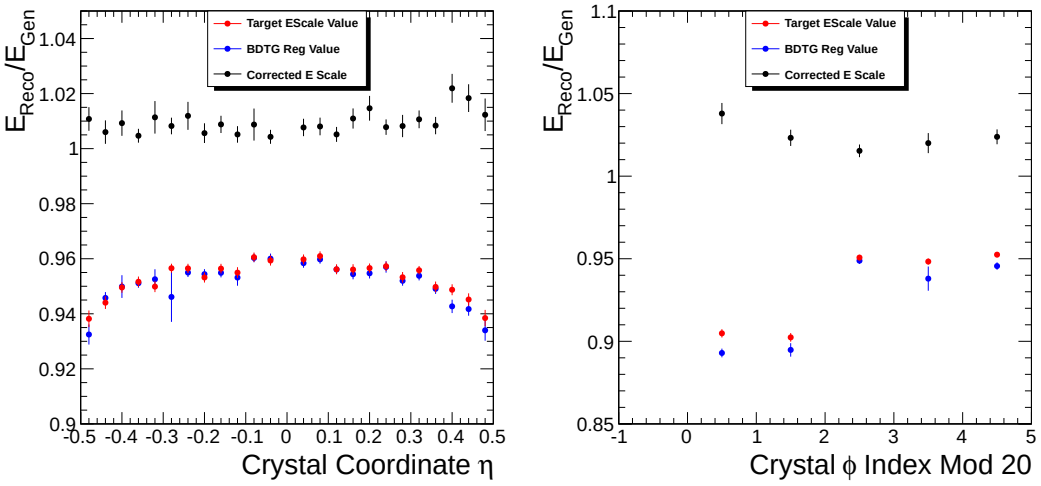


Figure 4.15: Improvement in the energy corrections adding additional variables.

- The crystal index is used to give the distance to the module and supermodule boundary.

The energy correction is trained with only unconverted photons with all the energy contained in a signal ECAL cluster. The corrections are then applied to every cluster within a supercluster for both converted and unconverted photons. These corrections give further improve the energy resolution to the most well measured photons in the ECAL. This can be seen for the first plot in Figure 4.18, for the E_{reco}/E_{gen} after applying these local containment corrections.

The two large corrections that are applied to the photon energy are due to the loss of the shower in the tracker and energy contamination from the pileup. These corrections are applied to the total energy collected in the superclustering. For the shower loss, r9 gives some indication of the loss along with the ϕ-width of the supercluster. The η coordinate of the photon position on the ECAL gives a coarse indication of the material budget as shown in Figure 3.9a. However, the measurement of the material budget more precisely describes the probability of conversion in terms of both coordinates η, ϕ which is shown in Figure 3.9b. The Particle Flow supercluster substructure, which was used to trace out the EM shower shape, can be used to predict the radiation loss as well. The $\Delta\eta$ from the seed position to the lowest energy cluster and the $\Delta\phi$ give the spread of energy. For photons that shower early and radiate in the detector, the smallest energy fragment will be deflected in ϕ and also in η from the seed, and indicates the loss of energy in the tracker (Figure 4.17). The information from



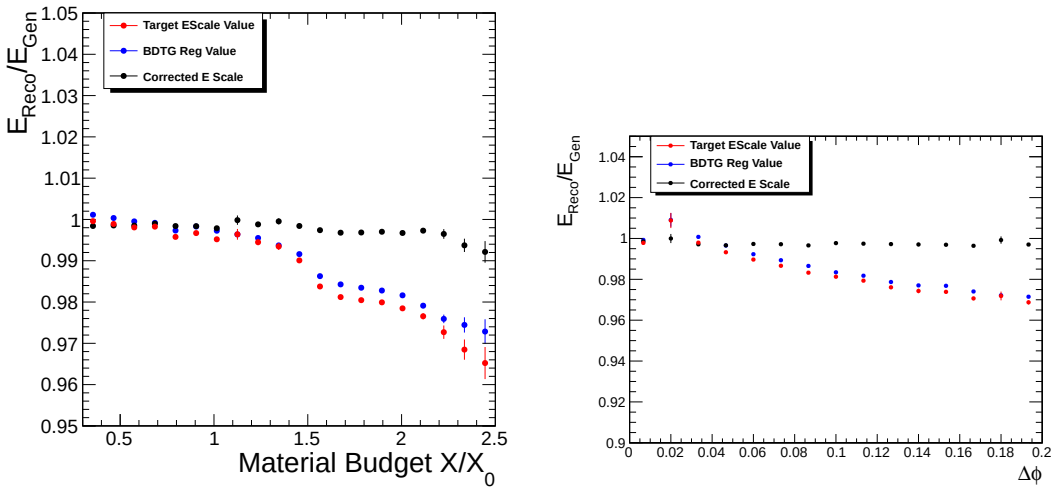
(a) Crystal η coordinate

(b) Crystal ϕ Module gaps

Figure 4.16: Two input variables fo the local containment corrections. The crystal coordinate η is the impact point of the photon on the face of the crystal, towards the edges of the crystal there is some leakage into air gaps. The second variable relates the crystal phi index to the module boundary. Every 20 crystals will border the boundary, so the modulus of 20 for the crystal index gives the proximity to the boundary.

the individual clusters in the supercluster can also be used to eliminate contamination event-by-event basis.

A major goal of the energy corrections is to give a stable energy resolution with increasing amounts of pileup. This was important in transitioning from the 7TeV collisions ramping up to the 8TeV with more data that has increasing pileup. The prototype Particle Flow photons made from standard superclusters together with the Mustache region are used to create variables to correct for the event by event pileup. The lowest energy cluster in the supercluster is most likely the contaminating cluster from pileup, so its energy is added as an input variable. When trained over many events the BDT will be able to learn a prediction of a threshold to include or exclude this cluster, so whether or not to subtract the cluster energy for the correction. The Mustache envelope is applied on top and the sum of the energy of clusters outside of the region give a prediction for the energy contamination for the photon. The energy within the Mustache envelope is the most like EM energy energy from the photon. The ratio of the lowest cluster energy to the energy in the Mustache also predicts how likely the lowest energy cluster is from pileup. Since the pileup cluster is typically soft compared to the photon energy, a small ratio would predict it is likely from pileup while a large ratio would dis-favour subtracting the cluster. The stability of the energy resolution of Particle Flow photons with increasing pileup is shown in Figure 4.19 compared to a standard regression without using any sub-cluster information for the standard superclusters.



(a) Material Budget

(b) $\Delta\phi$ to Lowest Energy Cluster

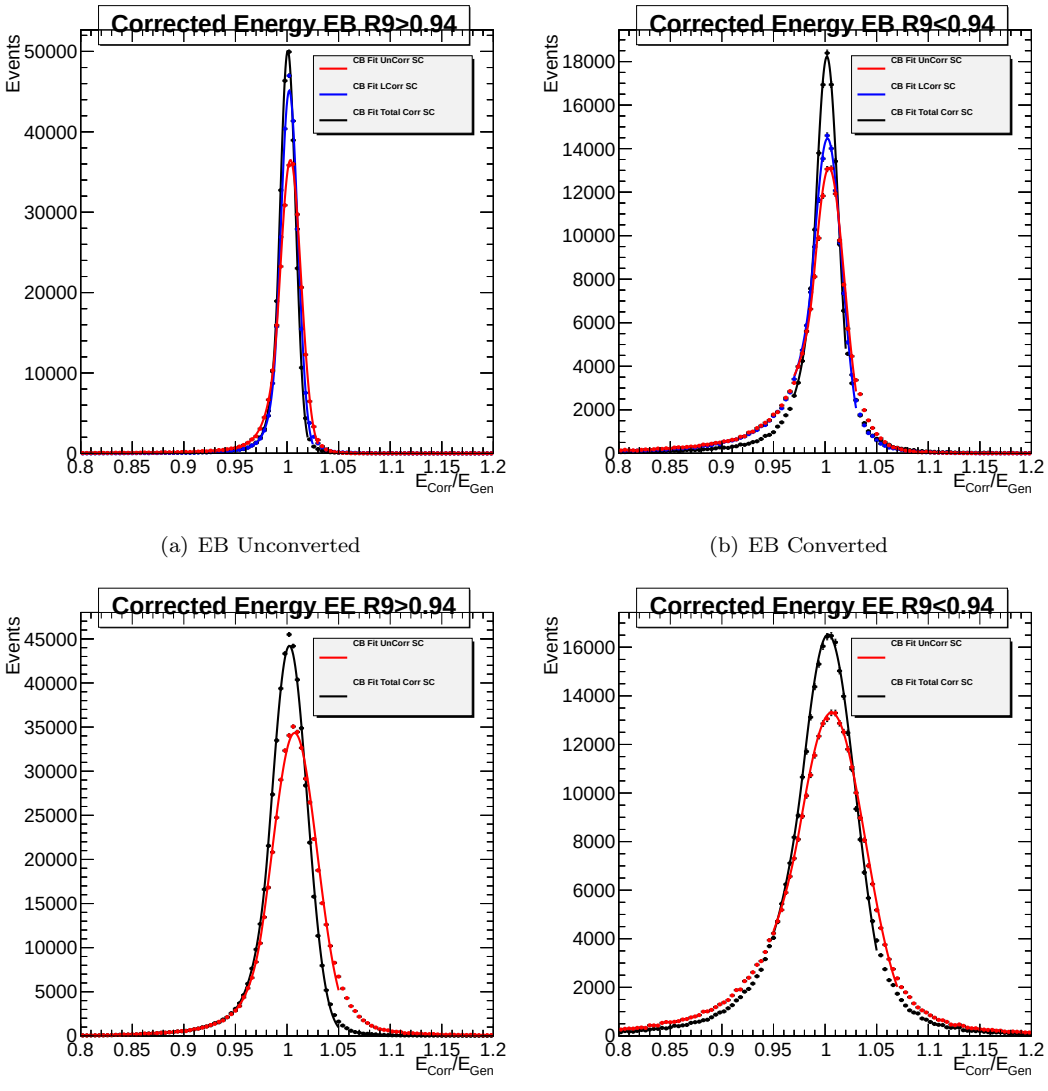
Figure 4.17: Two input variables for the global containment corrections that predict the loss of the EM shower in the tracker.

Figure 4.18a. shows the improvement for the total corrections in four resolution categories based on the photon position and whether it is converted based on the R9 variable. The local containment corrections for the first category further improve the energy for mainly well measured photons and make the energy response more uniform in the ECAL barrel. The local containment corrections are applied to the individual ECAL clusters which are then summed to give the most well measured supercluster energy. The shower loss and pileup contamination corrections, or total containment corrections, are applied to the total supercluster energy. Figure 4.18b. shows a large improvement from these corrections for the large set of converted photons which suffer from loss of the EM shower as well as contamination from pileup. The photons found in the endcap have only the total containment corrections. Figure 4.19 shows the stability of the corrections with increasing pileup comparing a regression to the standard superclusters with the superclusters from the Particle Flow.

4.5 Energy Corrections for the Higgs Analysis

This section will describe the input variables used for the Higgs analysis to correct the standard superclusters. These variables use the standard superclusters described in Section 4.1. A single energy correction is applied to a supercluster instead of the factorized approach applied in the previous section. A similar set of input variables is included to improve the containment of the shower on the ECAL based on:

- (η, ϕ) ECAL impact point on the crystal



(a) EB Unconverted

(b) EB Converted

(c) EE Unconverted

(d) EE Converted

Figure 4.18: Energy corrections for Particle Flow Photons shown in 4 resolution categories to show the effect on the corrections for photons measured in EB/EE and indicated as uncoverted/converted

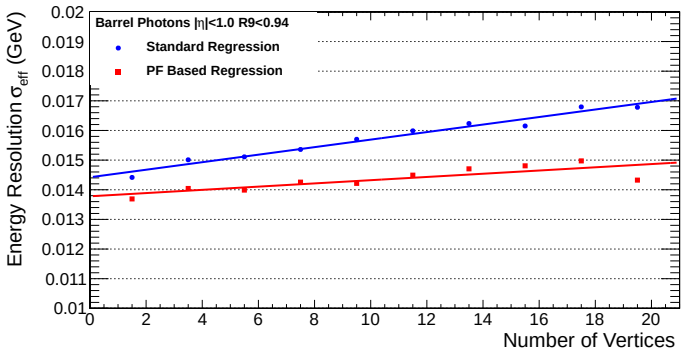


Figure 4.19: Comparison of the standard regression to the standard superclusters and the Particle Flow based regression that attempts to remove the pile-up contamination in each event to stabilize the energy resolution with increasing pileup.

- Crystal index (η, ϕ) which tells whether the crystal borders a module or supermodule boundary

These variables are used only for the photons in the barrel. The maximum energy containment of the supercluster for the standard supercluster is assumed to be in a 5x5 matrix of crystals. Several shower shape variables are used as input to predict the energy loss based on crystals within the 5x5 matrix in addition to the ones described previously: $R9$, η-width,ϕ-width.

- Energy contained in 3x3 matrix of crystals about the seed crystal divided by the 5x5 crystal energy
- The 4 nearest crystals to the seed are individually divided by the 5x5 crystal energy
- The maximum energy crystal, and also the 2nd largest energy crystal are divided by the 5x5 crystal energy
- The energy of crystals on each 2x5 edge of the 5x5 matrix divided by the total 5x5 energy

For the pileup stability, the total number of vertices and also the average density of particles, ρ are included as variables.

4.6 Scale Corrections for Transparency Loss

As described in section3.2.2, the laser monitoring system is designed to correct for the transparency loss in ECAL crystals. Imperfections in these corrections can result in time-dependance of the energy scale. A discrepancy between $Z \rightarrow ee$ data and simulation indicates a time-dependent energy response of the ECAL. The shift in the Z peak can in data compared to simulation can be used to derive a scale difference. The Z peak is fit with a Breit-Wigner convoluted with a Crystal Ball function. The Breit Wigner parameters are fixed to the particle data group values $M_Z = 91.188$

GeV and $\Gamma_Z = 2.495$ GeV. The Crystal Ball parameters give a description of the ECAL resolution as well as bremsstrahlung losses. The shift in the Crystal Ball peak in data is compared to shift in the simulation. The relative scale difference is:

$$\Delta P = \frac{\Delta m_{data} - \Delta m_{MC}}{m_Z}$$

This relative scale difference is found for different run ranges. The time dependence is not the same for all crystals since different ECAL regions get different radiation doses. So the relative scale difference is found in different psuedorapidity regions in each run range. The relative scale difference is applied to the data to mitigate the time-dependent behavior. [52]

4.7 Energy Smearing

The energy corrections trained on simulation as described in Section 4.4, give an overly optimistic correction based on the simulated response of the ECAL. However, in data the energy measurements of the ECAL can deviate from this ideal resolution. So the corrected energies derived in Section 4.4 need to also include a smear factor so that the final resonance signal peak resembles what will be observed in data. The residual simulation/data difference is measured using $Z \to ee$ where the energy corrections trained on photons are applied to electrons from the Z. The electron supercluster energy is modified by applying a Gaussian multiplicative factor centered in $1 + \Delta P$ and with a $\Delta\sigma$ resolution, where ΔP is the energy scale correction and $\Delta\sigma$ is the additional constant term in the energy resolution. After defining n exclusive electron categories, the $[n(n+1)/2]$ $Z \to ee$ invariant mass distributions are built in data and simulation. The method is based on the maximization of the likelihood between the data and the smeared MC in the $[n(n+1)/2]$ $Z \to ee$ invariant mass distributions. The smearing parameters $\Delta\sigma$ are applied in 8 categories: Barrel and Endcap are each split into two regions of psuedorapidty (Central and Non-Central). In each region the smear is computed for $R9 > 0.94$ and $R9 < 0.94$. The larger 8TeV dataset allows for the smear to be parameterized by two terms: $\Delta\sigma = \frac{\Delta S}{\sqrt{E_T}} \otimes \Delta C$ which closely relates to the stochastic and constant energy terms that give the ECAL resolution as described in Section 3.2.2. For 8TeV and 7TeV the ideal energy resolution for the expected Higgs signal peak will be smeared by these factors to give the resolution of the Higgs peak compatible with data. [52]

Chapter 5
Searching for the Higgs in Photon Data

The discovery of the Higgs has been the top priority of the LHC physics experiment [13], and its discovery has been a testament to the solid framework of the Standard model. The main observable that proves the existence of the Higgs field is the Higgs boson, which has a small branching ratio to photons of $\approx 0.3\%$. Chapter 2 has described the theory of the Higgs field with an associated Higgs boson which arises in the Standard Model to explain spontaneous-breaking of Electroweak symmetry. Chapter 3 described the Large Hadron Collider experiment of high energy proton-proton collisions where the Higgs is expected to be produced and can be observed to decay into two photons within CMS detector.This chapter will describe the analysis of the collected diphoton data at CMS to prove the existence of a Higgs boson. The $H \rightarrow \gamma\gamma$ channel has been studied since the initial planning of the LHC and the CMS detector. [18] Despite this small branching ratio, the Higgs decay to photons is the most promising decay channel for the discovery of a low mass ($< 130GeV$) Higgs boson. [18] The signal rate is much smaller compared to the background. The narrow Higgs resonance reconstructed from the diphoton mass is predicted atop a smoothly falling background of non-resonant diphoton invariant masses from QCD photons and also QCD jets that are misidentified as photons. The sensitivity of the analysis is driven by the detector resolution of the mass peak which drives the signal-to background ratio. The main goal of this thesis is to maximally improve the search sensitivity of $H \rightarrow \gamma\gamma$ and also measure its compatibility with Standard Model predictions.

This chapter will describe the bulk of the work done for this thesis in developing a multivariate analysis to give the most precise measurement of the Higgs signal and a measurement of its mass. The full Run 1 data taken with 7 TeV collision data at 5.1/fb and 8TeV collision data at 19.6/fb is used with the most current set of ECAL calibrations. The chain of multivariate analyzers will be described that gives the final set of signal categories to optimize the sensitivity to the Higgs peak. The measurement of resonances across these categories is combined to give a final measure of the Higgs signal strength: $\mu = \frac{\sigma_{obs}}{\sigma_{SM}}$ given as the ratio of the observed rate of Higgs production to the Standard Model prediction. Its statistical significance is denoted by a p-value and should correspond to 5σ for the threshold of discovery. The unknown parameter of the Standard Model, the mass of the Higgs boson is also measured.

5.1 Multivariate Analysis Strategy

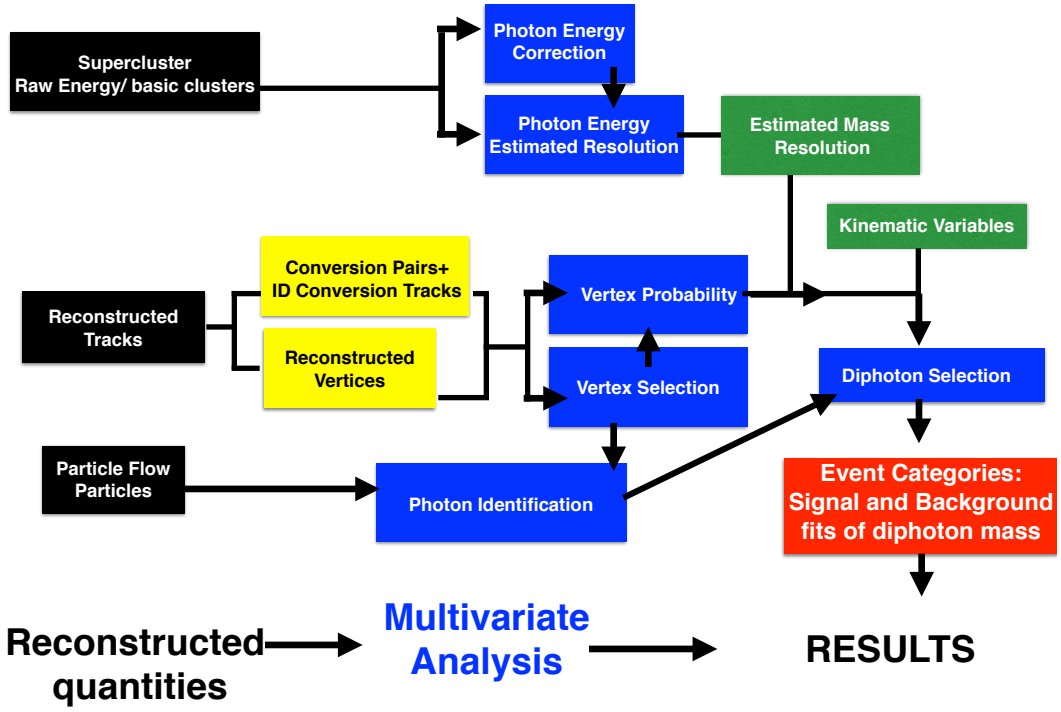
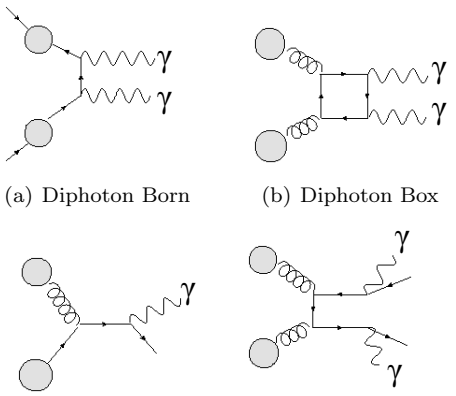


Figure 5.1: Multivariate Analysis Strategy Flow Diagram

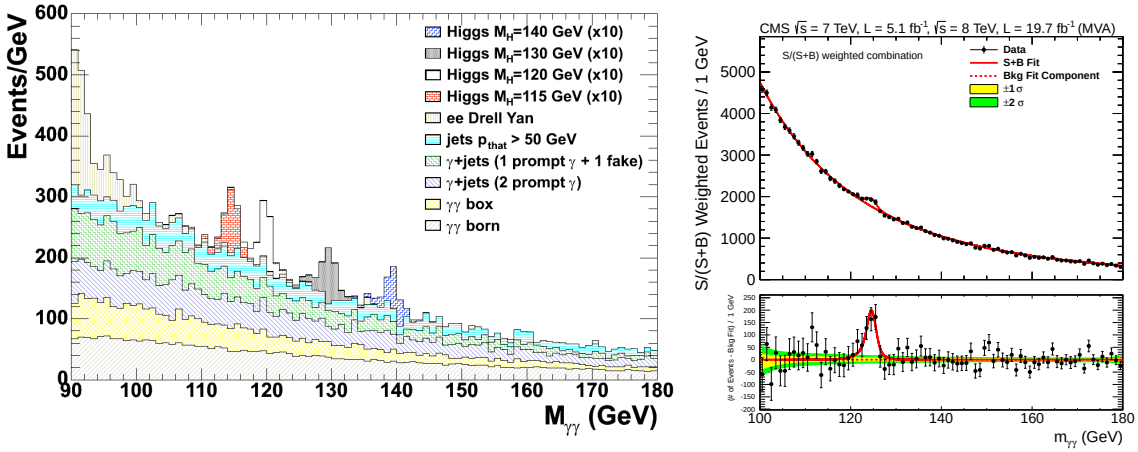
The Higgs to two photon decay occurs at Higgs masses below $\sim$ 130 GeV where the mass width of the Higgs is very narrow at $\sim$ 4 MeV, however the detector resolution for the two photons will spread out the mass peak so that it is observable. The energy resolution of the photons, as described in Chapter 4, is affected by many different factors, which will affect the expected width of the mass resonance. The advantage of a search of the Higgs decay to photons is the relatively large signal-to-background ratio which is orders of magnitude better than the dominant decays to heavy quarks. The main photon backgrounds are an irreducible background of non-resonant QCD photons produced in the proton collisions. Figure 5.2 shows QCD diphoton born and box diagrams that dominantly contribute to the irreducible background of non-resonant photons. The single and double fragmentation processes can produce one and two prompt photons respectively along with jets. The reducible background comes from photons reconstructed from meson decays inside of jets. There is also some background from electrons from the Drell Yann process that need to be disambiguated from photons that convert early in the tracker. Figure 5.3a. shows a stack plot of different reducible and irreducible background processes that bury the Higgs peak.



(a) Diphoton Born (b) Diphoton Box

(c) Single Fragmentation (d) Double Fragmentation

Figure 5.2: Four QCD processes that produce prompt photons and contribute to the irreducible background.



(a) Stacked Signal and Background [18] (b) Signal And Background Model

Figure 5.3: Potential Higgs Peaks atop the total simulated background

The mass reconstruction of a narrow peak from two photons is $m_{\gamma\gamma} = \sqrt{E_1 E_2 (1 - \cos\theta_{\gamma\gamma})}$ relies on a well-resolved energy of the two photons and their opening angle. The environment of the decay and the reconstruction of the photons determines how well the mass will be reconstructed. The opening angle is determined by the reconstructed decay vertex, but since photons are neutral there is no track from the decay vertex to the ECAL impact point as there is with electrons for example. As described in Section 3.1.1, the Higgs decay from a hard pp collision will have overlap with many softer pp collisions. There will be many vertices with which the two photons could potentially be associated, due to tracking inefficiencies for very low p_T tracks the primary interaction point with the largest $\sum |p_T^{Track}|^2$ is not the Higgs decay vertex. Assigning two photons to the wrong vertex within a 6cm gaussian smear of vertices along the beamline can degrade the Higgs resonance if the assignment is more than 1cm away from the true vertex.

Chapter 4 described the challenges of reconstructing photons at CMS as well as improvements made to the standard photon reconstruction. The energy correction made using Boosted Decision Trees improves both the loss of the EM shower in the tracker as well as the contamination from the overlapping soft collisions. Including a large set of variables could reduce the energy loss while keeping the energy resolution stable with respect to increasing contamination. With well resolved photon energies and also a vertex hypothesis, the main part of the analysis is to separate selected Higgs events by the resolution of the reconstructed mass. For well resolved diphoton masses, the signal peak will be very narrow to give high discrimination from the smoothly falling background (high signal-to-background ratio), while a very broad peak will give poor discrimination (lower signal-to-background ratio).

Each of the challenges described will be addressed in the analysis using Multivariate analyzers to take advantage of combinations of as many experimental handles to give optimal reconstruction and classification of Higgs events, thus give optimal sensitivity to events where a Higgs decays to photons. Figure 5.1 shows the flow of the analysis from the reconstructed quantities described in Chapter 4 to the final selected events to modeled by signal and background probability density functions. On the top of the flow chart, the collected energy in the supercluster window is corrected using a BDT regression. Trained in tandem with the energy correction is the energy resolution which is a strong indicator of the signal to background ratio as will be described in Section 5.6. The reconstructed tracks give some reconstructed conversion pairs and some that are identified as conversions using the BDT classifier described in Section 4.3.1. The information from the conversions and the reconstructed track information is input to a BDT classifiers that scores the most likely signal vertex. Also trained in tandem is a BDT that quantifies the probability of choosing the right vertex. The particle flow reconstructed particles are input to the photon identification BDT classifier, that

also uses the vertex information for the charged candidate isolation. The photon energy resolution estimate, the vertex probability, and the photon identification classifier values are input to the diphoton BDT classifier (the final blue box in the figure) along with additional kinematic variables. The final diphoton classifier value is used to divide selected events into categories of different signal-to-background ratio. The full flow of MVA analyzers finely filters the diphoton dataset based on the available information from the detector to give the most likely subset of Higgs events categorized by how likely they contain signal.

The chain of MVA analyzers is summarized as:

- Finding the Higgs Vertex: A Boosted Decision Tree classifier combines the kinematic information of a given vertex and the information from the conversions to rank the vertices based on how likely it is a Higgs decay vertex. A second Boosted Decision Tree is trained to quantify the probability of choosing the correct vertex in the event. This will be described in Section 5.4.

- Photon Identification: A photon identification Boosted Decision Tree is trained to discriminate prompt photons produced from the pp collisions from those that reconstructed inside of jets from meson decays. This reduces the bulk of the reducible background. This will be described in Section 5.5.

- Photon Energy: A Boosted Decision Tree regression routine is used to correct the energy of photons for energy loss in the tracker and also contamination from overlapping particles. A second Boosted Decision Tree regression is trained to predict the energy resolution of the photon, which can be used to compute a mass resolution: $\frac{\sigma_M}{M} = \sqrt{\frac{\sigma_E^1}{E_1} \oplus \frac{\sigma_E^2}{E_2}}$ This will be described in Section 5.6.

- Event Classification: From the above variables like the probability of the correct vertex, the mass resolution, and the photon identification classifier value an event classification Boosted Decision Tree is trained to classify events by their signal to background ratio. This will be described in Section 5.7.

The final set of selected events will consist of a majority of background events with an excess of signal events at a particular diphoton invariant mass. A probability distribution function is fit to the diphoton mass resonance of a simulated Higgs mass peak to create a Higgs signal model for a given Higgs mass. The non-resonant background decreases smoothly as a function of diphoton invariant mass and is modeled by a probability distribution function fit directly to the data which is mainly background. Selected events should be grouped according to how well resolved the Higgs mass would be, the narrowness of the Higgs signal clearly separates it from statistical fluctuations in

the background. Figure 5.3 shows the signal and background PDFs that model for all the selected Higgs events. Events are weighted by the estimated signal-to-background ratio in each category to emphasize the observed excess. This also shows the power of finely splitting events based on how likely they contain measurable signal. The final statistical treatment combines the fits across all event categories to give a final measure of the Higgs mass and also the rate of production $\mu = \frac{\sigma_{obs}}{\sigma_{SM}}$.

In addition to categories based on diphoton mass there are also categories built to tag the additional Higgs production mechanisms: vector-boson fusion and associated Higgs production where the Higgs couples to vector bosons. The rate of Higgs produced via vector bosons and the rate of Higgs produced via quarks can then be compared as μ_f vs. μ_v. This indicates the compatibility of the Higgs couplings to vector bosons and to fermions. The Higgs could potentially be 'fermiophobic' and couple only to vector bosons, which means that the top-quark interference term would not exist and the $H \rightarrow \gamma\gamma$ signal rate would be enhanced. The measurement of specifically Vector-boson fusion events will be described in Section 5.7.1 using a 2D PDFs to model the gluon-fusion mis-tagged events and the signal vector boson fusion events.

The multivariate analyzers intimately link the analysis to the simulated behavior of photons in Monte Carlo and also the simulation of the detector performance, which must be validated against the the observed performance in data. This is done by choosing control samples like the Z boson decaying to electrons and comparing simulation and data for MVA input variables, Z-mass resolution, electron efficiency applying the photon selection. The use of electrons from the Z to validate photons from the Higgs is one of the sources of systematics for this analysis. It was observed that the interaction of photons and electrons in the simulation have some differences and also the boost of the Z boson is different from the heavier Higgs boson. These uncertainties propagate to the final measure of the Higgs Mass.

The final results are found by modeling the signal and background in each event class by fitting a probability distribution function to the diphoton-mass and combining them. Depending on the amount of statistics in a given a category the probability distribution function can mismeasure the amount of background. The background function needs to be robust enough to give a consistent background prediction in the face of random statistical fluctuations. This ensures the signal is not artificially larger or smaller due to background fluctuations. The procedure for choosing a robust background probability distribution is described in Section 5.8. The full LHC Run 1 results for the Higgs decay to photons will be shown and described in Section 5.9.

5.2 Datasets and Monte Carlo Samples

The analysis presented here corresponds to the full LHC Run 1 photon dataset. The multivariate techniques used are trained with the latest simulation with the most up to date information on the detector performance and the pileup. The LHC photon dataset includes the most up-to-date ECAL calibrations, and the simulation includes the time dependent behavior of the detector. The photon dataset is divided into the 2011 runs where proton collisions occurred at a center of mass energy of 7TeV and give an integrated Luminosity of $5.1 fb^{-1}$. The bulk of the run 1 photon data is taken in 2012 with $8TeV$ proton-proton collisions which give an integrated Luminosity of $19.7 fb^{-1}$. The collision energy is ramped up through run periods, so the 8TeV data is divided into 4 run major run periods as shown in Table 5.1.

Run period	Int. lumi. (fb^{-1})
RunA+B	5.3
RunC	7.1
RunD	7.3
RunA+B+C+D	19.7

Table 5.1: Integrated Luminosity for the 2012 Data taken with $\sqrt{s} = 8TeV$ divided into four run periods

The simulation plays an important role in the analysis to train the decision trees. The simulation includes the pileup conditions for each run period which have a slightly different average number of interactions:

- **AB** $< PU >\sim 19$, $L = 5.3$ fb^{-1};
- **C** $< PU >\sim 20$, $L = 7.1$ fb^{-1};
- **D** $< PU >\sim 21$, $L = 7.3$ fb^{-1};

The conditions of the ECAL and the pile up, can change in different periods of the data taking. Each MC sample has been split in three periods, with event numbers proportional to the luminosity in the run periods AB, C, D and simulate the appropriate pileup and also the ECAL response for that run period. The simulation also includes an accurate distribution of the number of interactions taking place in each bunch crossing. This distribution is reweighted to match the observed distribution in data. The reweighting technique does not use the number of reconstructed vertices which has some inefficiency, but instead uses the simulated number of pileup interactions.

The signal samples for the gluon fusion and vector boson fusion processes were generated with the same generator POWHEG at NLO, whereas the associated production process was simulated with PYTHIA at LO. The cross-sections and branching ratios recommended by the LHC Cross-Section

Working-Group are used. In particular, the gluon fusion cross-section is computed at NNLO+NNLL for perturbative QCD and NLO for electroweak (EWK) contributions.The vector boson fusion cross-section is computed at NNLO for QCD and NLO for Electroweak contributions. The associated production cross-sections are computed at NLO QCD order. Signal samples were produced for Higgs mass values ranging from 90 to 150 GeV in steps of 5GeV.

The standard model background processes are also simulated and used to train the multivariate analyzers for identification and classification. The diphoton background with two real photons, prompt-prompt, includes the born processes with up to three jets at LO and the box process at LO. The photon plus jet process is used to train the identification MVA and the simulated jets are filtered to enhance the number of 'fake photons'. Simulated events are required to pass a filter which requires isolated ECAL activity with E_T above 15GeV within a small region of the ECAL. The simulation of QCD dijet provides a sample with two jets which is filtered as well to enhance the number of photon fakes. For data to simulation comparisons mainly $Z \to e^+e^-$ is used, but also $Z \to \mu\mu\gamma$ is used for comparisons with real photons.

5.3 Event Trigger and Preselection

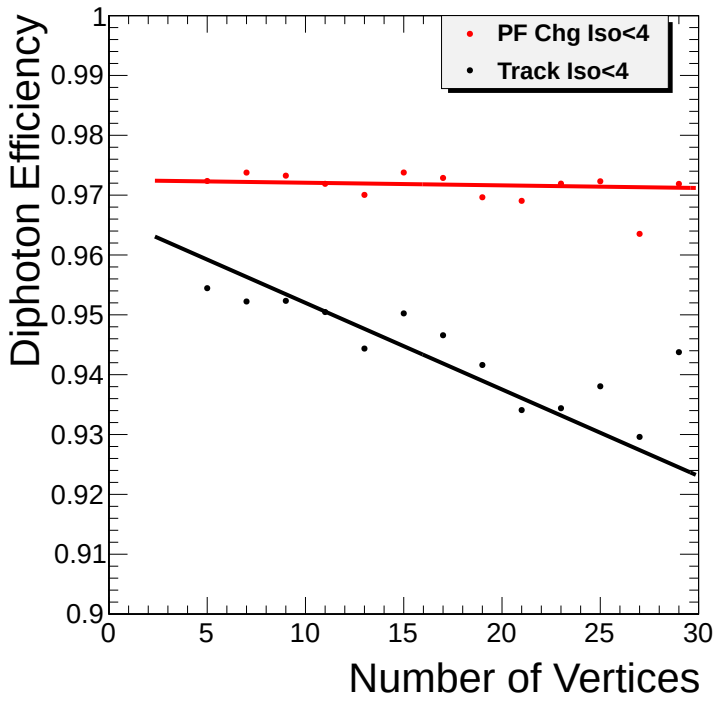


Figure 5.4: Preselection with Particle Flow Charged Hadron cut compared with the detector based Track isolation cut for gluon fusion Higgs signal at 125 GeV

Every LHC collision event at CMS is evaluated by the trigger system which provides electronic acceptance/rejection logic bits to filter events. The trigger paths have two levels. Level-1, which selects roughly reconstructed photons from square array of crystals, and a High Level Trigger. The events for this analysis must pass a diphoton trigger decision based on the E_T of the photon seed at Level-1 and also cuts on the detector isolation of each photon: $Track_{Iso}$, $ECal_{Iso}$, $HCal_{Iso}$ and also one shower shape variable $\sigma_{i\eta i\eta}$. [50]

All photons considered for the analysis are required to pass a preselection. The preselection basically rejects a large amount of the reducible background from jets that fake photons. This preselection is designed to be slightly tighter than the trigger requirements while keeping a match between the common phase space of data passing the trigger and the signal events in simulation where the trigger is not applied. The variables used are similar to those used in the trigger requirement. In addition to the detector isolations described in Section 4.1 and the preselection also includes the Particle Flow Charged Hadron isolation. The charged hadron isolation cut was shown to maintain the preselection signal efficiency with increasing amounts of pileup energy. The particle based variable is more powerful because the track can be associated to a vertex and the p_T of the hadron accounts for the track measure and the calorimeter measurements. A loose cut is applied on the charged hadron isolation which stabilizes the efficiency with increasing pileup. This is shown in Figure 5.3 where the track isolation cut has a lower signal efficiency at high pileup because off pointing tracks contaminate the isolation sum. In addition, the shower shape variables $sigma_{i\eta i\eta}$ and R9 are used to discriminate against photons from high p_T meson decays. [52] The loose selection is derived separately in 4 categories that separate the ECAL barrel and endcap, and also the converted and unconverted photons based on the variable R_9: Electrons faking photons are also

	Barrel		Endcap		Both Barrel and Endcap		
R9	HoE	CovIEtaIEta	HoE	CovIEtaIEta	HcalIso	TrkIso	ChargedPFIso
≤ 0.9	< 0.075	< 0.014	< 0.075	< 0.034	< 4 GeV	< 4 GeV	< 4 GeV
> 0.9	< 0.082	< 0.014	< 0.075	< 0.034	< 50 GeV	< 50 GeV	< 4 GeV

Table 5.2: Preselection photon ID cuts

rejected by applying the conversion safe electron veto. The conversion safe electron veto is designed to reject reconstructed photons with a well reconstructed Electron track. This veto is applied in two conditions. The first requires the GSF track to have all pixel layer hits along its trajectory. The layer with a missed hit would correspond to where the conversion occurred. The second condition is that the GSF track matches no GSF track used for reconstructing a conversion as described in Section 4.2. If the GSF track meets both of these conditions, it is considered an electron and the photon hypothesis is rejected. [52]

The signal efficiency in data and MC with the cuts is measured using tag and probe techniques on $Z \to ee$. The signal efficiency of the electron veto requires a photon sample so $Z \to \mu\mu\gamma$ events are used. The tag probe signal efficiency is measured with electrons in the 4 resolutions categories based on barrel/endcap and converted/unconverted. The efficiency is high for the unconverted photons $R_9 > 0.9$ at $\sim 98\%$ and slightly lower for the converted photons $R_9 < 0.9$ which look more like background with an efficiency of $\sim 96\%$. The ratio of the tag and probe data efficiency to the MC efficiency is used to scale the signal model by a small amount. [52]

5.4 Diphoton Vertex

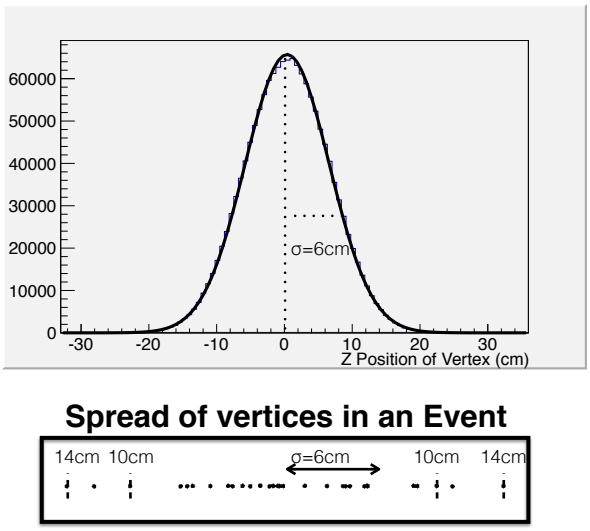
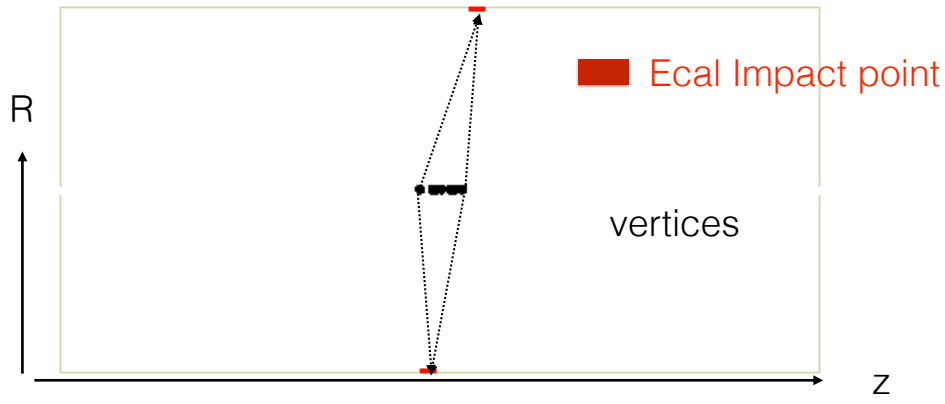
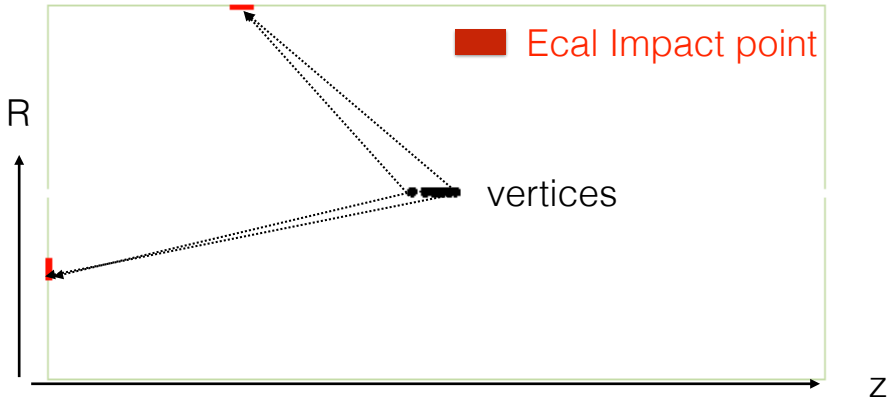


Figure 5.5: Average gaussian spread of vertices about the center of detector with gaussian $\sigma = 6cm$. (bottom) The spread of vertices in a typical high PU event with bulk of the vertices clumped within 6cm and some spread further to 14cm.

An error on the measurement of the opening angle between the photons can degrade the mass resolution if it is too large. The opening angle is determined by the ECAL impact points of the two photons and the Higgs Decay vertex which will lie along the beamline where the protons collide. The luminous region within the CMS detector has an RMS spread of about 7cm and the mean number of interactions per bunch crossing is about 19.9. Each interaction point is reconstructed from a vertex fit to all the tracks that can be interpolated to a common point of origin. By default, the vertex with the largest $\sum |p_T^{Track}|^2$ is considered the primary interaction point. However, inefficiencies in reconstructing very low p_T tracks can cause the highest $\sum ||p_T^{Track}|^2$ vertex to not be the signal vertex. A misidentified Higgs vertex will propagate to an error in the opening angle of photons. The opening angle can be written in terms of the polar angle θ and the azimuthal angle ϕ of each photon:



(a) Vertex finding error for back-to-back decays



(b) Vertex finding error for smaller angles

Figure 5.6: Vertex finding error when choosing the wrong vertex for almost back to back decays and also photons at narrower angles

$\theta_{\gamma\gamma} = \cos\theta_{\gamma 1}\cos\theta_{\gamma 2} + \sin\theta_{\gamma 1}\sin\theta_{\gamma 2}\cos\left(\phi_{\gamma_1} - \phi_{\gamma_2}\right)$ A small change in z corresponding to a different choice in the interaction point shifts the polar angle while keeping the the azimuth unchanged. The shift in the polar angle: $\delta\theta_\gamma = -\sin\theta_\gamma \dfrac{\delta z}{r_\gamma}$ where the r is the distance from the interaction point to the ECAL. So for a very large detector radius compared to the spread in the interaction points the vertex choice becomes irrelevant. However, the minimum value of r in the barrel would be 130 cm while for the endcap it is 310 cm. The typical spread in vertices has a gaussian sigma of about 6 cm the shift in the polar angle only becomes negligible for $\delta z < 1$cm. The vertex selection is necessary because the error, δz, of choosing the wrong vertex has an impact on the mass resolution.

This section will describe the vertex finding procedure and also the vertex finding uncertainty. Both techniques use a Boosted Decision tree. The vertex finding aims to choose the Higgs decay vertex based on the kinematics of the vertex as well as the information from conversion tracks from the photons. Since the vertex finding efficiency is not very high, the probability of finding the vertex in an event is also quantified. A second BDT classifier is trained to separate events where the Higgs vertex is likely to be chosen within $\delta z < 1$cm from those where there will be some degradation to the mass resolution. The classifier value for the vertex probability will be used when classifying signal events in Section 5.7.

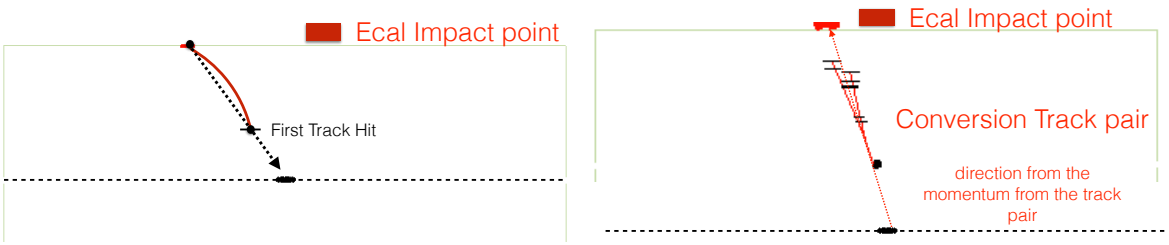
5.4.1 Vertex Identification

The vertex identification depends on whether the photon converts or not. For unconverted photons only the tracks associated to a vertex are used to discriminate the Higgs vertex from other reconstructed vertices. For converted photons, a track pair or even a single track can be used to determine the interaction point. The track-pair momentum can be used to determine the photon direction. Also the line from ECAL position to the first track first hit can be used to extrapolate to the vertex position.

For two unconverted photons, the Higgs decay vertex will have tracks that recoil against the diphoton system. The $\sum |\overrightarrow{p}_T|^2$ of the tracks can be combined with two other variables that are strongly correlated with this recoil:

- *ptbal*: $-\sum_i (\vec{p}_T^{\,i} \cdot \frac{\vec{p}_T^{\gamma\gamma}}{|\vec{p}_T^{\gamma\gamma}|})$.
- *ptasym*: $(|\sum_i \vec{p}_T^{\,i}| - p_T^{\gamma\gamma})\ /\ (|\sum_i \vec{p}_T^{\,i}| + p_T^{\gamma\gamma})$.

These variables are combined into an MVA classifier. The BDT discriminator value is used to rank the vertices by how likely they come from a Higgs decay vertex. The vertex is chosen by the highest rank given by the above variables.



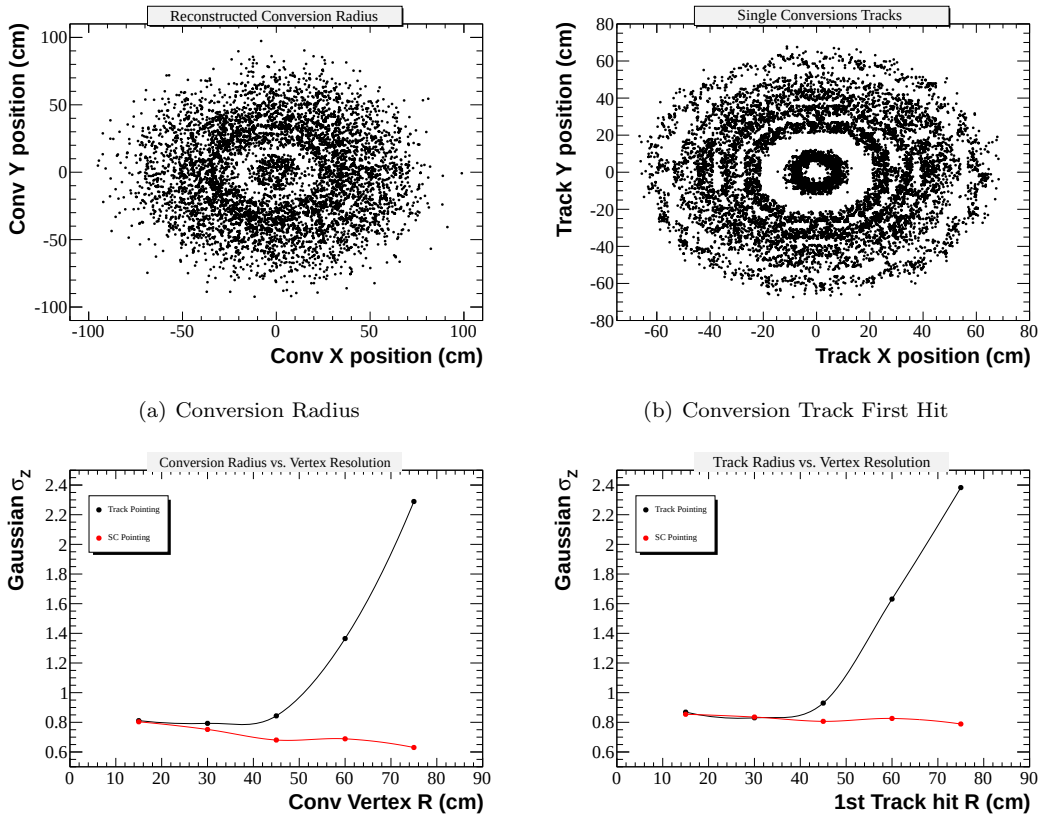
(a) ECAL Pointing

(b) Conversion Pointing

Figure 5.7: The ECAL plus track pointing method and the method using the conversion momentum

Photon conversion allows for the possibility of the photon to be associated to a reconstructed track from the e^+e^- . As described above, tracks can be grouped by a vertex algorithm based on the best estimate of a set of vertex parameters and a vertex fit is given by a minimizing strategy (often global least square minimization [17]). A conversion can be fit as well from reconstructed tracks of e^+e^- pairs. The conversion vertex is fit under the zero mass hypothesis which requires parallel tracks at the fitted vertex point as well as other criteria on the electron tracks to improve the conversion reconstruction quality. The conversion vertex will reconstruct the momentum of the e^+e^- pair which gives a measure of the photon direction. The momentum is extrapolated to the beamline, and the closest vertex to the intersection is identified as the Higgs decay vertex. Figure 5.4.1 shows the two types of conversion pointing techniques.

The conversion pointing technique relies on how well the trajectory of the e^+e^- is measured. Conversions in the pixel layers of the tracker will give a well determined photon direction from the well-tracked electron-positron pair. However, late conversions in the Tracker Outer Detector will have very few track hits (and many can be shared between the opposite charged tracks) so give a very poor measure of the photon direction. Figure 4.4 showed some of the cases for reconstructed conversions. In the case of only a few outer track hits that are not well separated, the track momentum can be poorly determined which can give poor vertex resolution when pointing. However, for later conversions, the impact point of the photon on the ECAL is well determined since the charged particles do not bend much in the field before striking the ECAL. The vector from the photon position on the ECAL to the the conversion vertex point can be used as the photon direction. This is propagated to the beamline to select a vertex. Figure 5.8 shows the conversion vertex radius and the first track hit radius for the conversions found in Higgs simulation. The vertex resolution can be computed as ΔZ to the true decay vertex in Monte Carlo. Just using the Conversion pointing or the projection of the track on the beamline depends on the track resolution. At larger radii, tracks have an increasing uncertainty on the polar angle which give poorer resolution on the vertex



(a) Conversion Radius

(b) Conversion Track First Hit

(c) Vertex Finding with Conversions

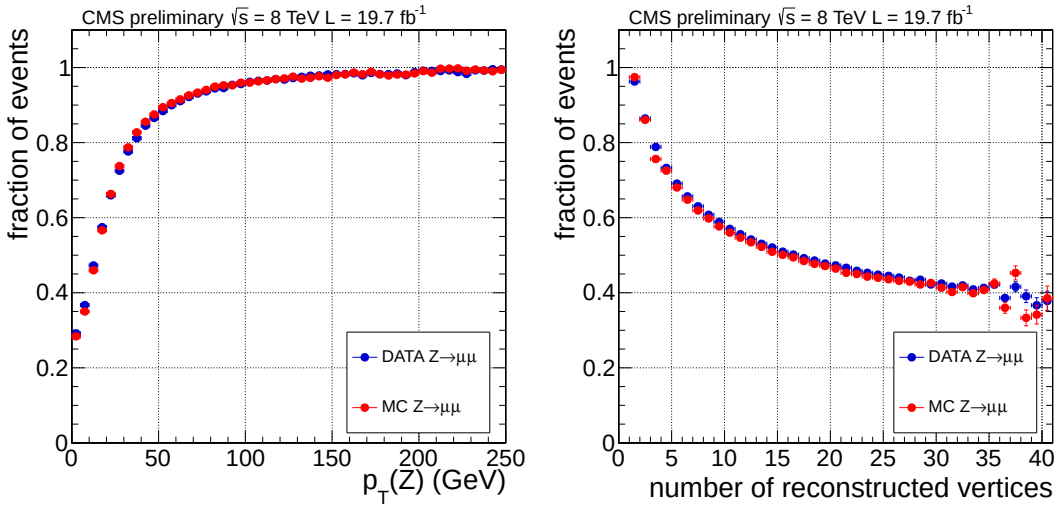
(d) Vertex Finding with Single Tracks

Figure 5.8: The radius of conversion and the radius of the first track hit. The Vertex finding resolution in simulation is measured as the gaussian sigma of the distribution of ΔZ to the true vertex for the conversion momentum pointin method and the ECAL plus track hit pointing.

resolution. The ECAL pointing does not depend on the polar angle of the track momentum, so the vertex resolution remains fairly constant at large and small radius of conversion as shown in Figure 5.8.

The stability of the conversion pointing at large radii using the ECAL position is key because the late conversions also have much better energy resolution. The well-resolved photons along with the well resolved vertex give events with the largest signal-to-background ratio.

The conversion pointing is done only for the inner regions of the tracker, while the ECAL position pointing is used for the outer regions. The vertex finding from the converted photons and the kinematic variables for the uncoverted photons are combined into a multivariate discriminate using the *Boosted Decision Tree* (BDT) algorithm. The inputs to the BDT are the *sumpt2*, *ptbal* and *ptasym* variables. In events with at least one photon conversion, a fourth variable $pull_{conv} = |z_{vertex} - z_{conv}|/\sigma_{conv}$ (where z_{conv} is the estimated primary vertex position and σ_{conv} is the resolution measured in data).



(a) Vertex ID Eff. vs. Z p_T

(b) Conversion Track First Hit

Figure 5.9: Vertex Finding Efficiency as a function of the p_T of the Z and increasing pileup [52]

The performance of the vertex finding is measured in data using $Z \to \mu\mu$ events and also $\gamma + jets$ for the case where a photon is converted. Events with the Z boson decaying into a muon pair are used to validate the vertex finding when there is no conversion. The muon tracks are ignored (which give the true vertex) and the vertex is refitting with the other tracks available, which mimics a vertex of a decay to neutral particles. The BDT veretex finding is evaluated on $Z \to \mu\mu$ data and simulation where a vertex is considered found if it is within $\Delta Z < 1$ cm of the true value given from the muon tracks. In the $\gamma + jets$ sample, the conversion pointing is validated by checking to see if the conversion points to the primary vertex of the jet. Figure 5.9 shows the $Z \to \mu\mu$ vertex identification efficiency. At low Z p_T the variables correlated with the recoil become more alike for signal and background so the vertex finding efficiency begins to drop. Also with more pileup there is a larger probability of choosing a vertex with $\Delta Z >$ 1cm from the true value.

5.4.2 Probability of Correct Vertex ID

For the final event classification, it is necessary to have a measure of how probable the right vertex has been selected. As shown in Section 5.4 the performance depends on the number of vertices and the best vertex efficiency is 75% in the current 2012 pile-up conditions. The vertex probability is computed as a function of a second multivariate discriminant. This discriminant is trained to seperate events with a high probability of choosing the correct vertex from those with lower probabilities. The BDT inputs use the diphoton p_T. The larger the diphoton p_T the larger the discriminating power of the kinematic variables for vertex finding. The inputs of the BDT are the following:

- p_T of the diphoton system.
- number of vertices in each event.
- value of the per-vertex BDT discriminant for the best three vertices in each event.
- Δz between the best vertex and the second and third choices.
- number of photons with associated conversion tracks (0, 1, or 2).

The MVA discriminator folds the above variables into a single BDT output which is a linear function of the vertex finding probability (the linear function is found in simulation where the true vertex finding probability is known).

5.5 Photon Identification

The background of the $H \to \gamma\gamma$ signal is a composition of prompt photons from QCD processes and an irreducible component from $pp \to \gamma$+jet and also dijet processes. The jets can fake photons when the jet consists of one or more mesons that decay to photons. In the photon p_T range for the Higgs decay, mainly $\pi_0 \to \gamma\gamma$ can have collimated photons that can fake a signal photon. To reject a bulk of the reducible background a combination of isolation and shower shape variables is input to an MVA discriminator.

The particle flow isolation was described in Section 4.3. The photon shower is properly removed from the isolation sum, the photon isolation and charged hadron isolation more precisely gives the sum of the p_T of overlapping particles. Since the charged hadrons can be associated to a vertex, the isolation is computed both for charged hadrons that come from the selected Higgs and from a different vertex that gives the largest isolation sum.

Shower Topology Variables

1. $\sigma_{i\eta i\eta}$: The energy weighted standard deviation of single crystal eta (in crystal index) within the 5x5 crystals centered at the crystal with maximum energy.
2. $cov_{i\eta i\phi}$: The covariance of the single crystal η and ϕ (in crystal index) values within the 5x5 crystals centered at the crystal with maximum energy.
3. E_{2x2}/E_{5x5}: Ratio of the energy in the 2x2 crystals containing the crystal with maximum energy which have the maximum energy sum to the energy in the 5x5 crystals centered at the crystal.
4. R9 (R9): The ratio of the energy in the 3x3 crystals centered at the crystal with maximum energy to the Supercluster energy.

5. σ_η: The energy weighted standard deviation of single crystal eta within Supercluster.

6. σ_ϕ: The energy weighted standard deviation of single crystal phi within Supercluster.

7. Preshower σ_{RR} (for Endcap): The standard deviation of the shower spread in the x and y planes of the preshower.

The above variables are designed to discriminate the photon EM shower shape from the shower shape of photons that originate from one or more meson decays like $\pi_0 \rightarrow \gamma\gamma$. The first variable was described earlier in Section 4.1 which gives the deviation of the pion shower in η which is wider than the EM shower. The $cov_{i\eta i\phi}$ indicates the deviation in both η and ϕ. The three other shower shape variables $\mathrm{E_{2x2}/E_{5x5}}$, σ_η, σ_ϕ discriminate the overall spread in energy for signal photons and background fake photons inside of jet, which tend to give larger shapes. The additional information of the preshower hits can be used to discriminate the hit pattern from fake photons which have a radial spread of hits in the preshower from those that come from converted photons where the two electrons give a different hit pattern through the preshower.

Isolation Variables:

1. PF Photon ISO: Particle flow photon isolation sum ($\Delta R < 0.3$).

2. PF Charged ISO (selected vertex): Particle flow charged hadron isolation sum ($\Delta R < 0.3$) with respect to the selected vertex.

3. PF Charged ISO (worst vertex): Particle flow charged hadron isolation sum ($\Delta R < 0.3$) with respect to the vertex for which the isolation sum is largest.

The particle-based isolation is computed from the four-vector of the reconstructed particles found in Section 4.3. The sum of the E_T of the different types of particles in a cone around the photon gives a measure of the additional activity in the region of the photon and the sum is divided by the E_T of the photon. The charged particles can be associated to a vertex so the isolation can be computed under two hypotheses: that all the charged particles from the same vertex as the potential Higgs photon or the charged particles come from the vertex that gives the largest amount of activity. The isolation variables combined with the shower shape provide a powerful discriminator between signal photons and photons inside of jets.

1. ρ (rho): The median energy density per solid angle.

2. Supercluster η: the eta of the supercluster corresponding to the reconstructed photon.

3. Supercluster E_{RAW}: the raw energy of the supercluster corresponding to the reconstructed photon.

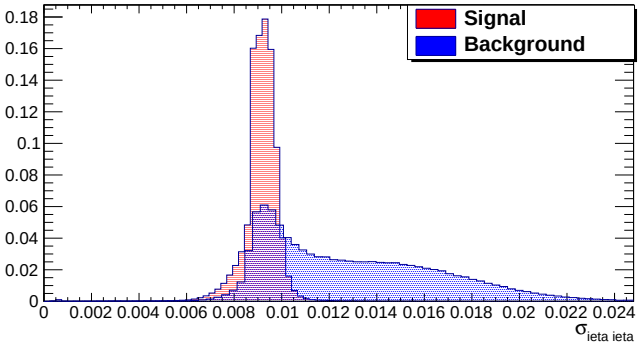
The median energy density was described in in Section 4.3, and accounts for characteristic that signal photons start to look more like fake photons at high pileup. The η coordinate and the energy of the supercluster is used to give the same discriminating power across the full pseudorapidity range and also give the same discrimination for different photon p_T ranges.

The advantage of combining shower shape variables in an MVA classifier can be seen in Figure 5.10. A one dimensional cut on the $\sigma_{i\eta i\eta}$ can discriminate some of the background but there is still significant overlap between the signal and background values of $\sigma_{i\eta i\eta}$. However, a 2D scatter plot of $Pho_{iso} + Chg_{iso}$vs $\sigma_{i\eta i\eta}$ shows a clear two dimensional boundary between signal and background. An MVA cut of $MVA > -0.2$ carves out the signal region in the 2D scatter plot. In the larger multidimensional parameter space of all the variables islands of signal can be carved out of the reducible background. A cut on the photon ID MVA output of $MVA > -0.2$ retains more than 99%of the signal events passing the other preselection requirements (in Section 5.3), while removing 23.5% of the data events in the mass range $100 < m_{\gamma\gamma} < 180GeV$.

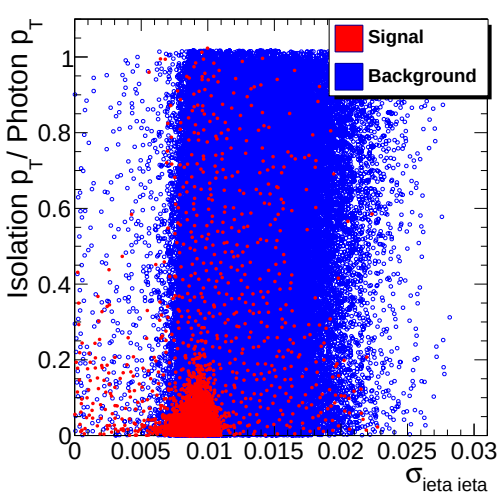
Each of the input variables to the BDT discriminator is compared in Data/Monte Carlo Simulation for $Z \to ee$. Though the shower shape variables for electrons can be different for electrons and photons, this comparison indicates how well the electron shower is simulated in Monte Carlo compared to data. For large discrepencies, the shower shape variables are scaled to better match the data. The input variable distributions between data and Monte Carlo agree to within the systematic uncertainty quoted in Section 5.8.4. Also a photon sample of $Z \to \mu\mu\gamma$ is also used to further cross-check the comparison of the input variables in Data and Monte Carlo. This study gives comparable results with the $Z \to ee$ control sample.

5.6 Photon Energy Resolution

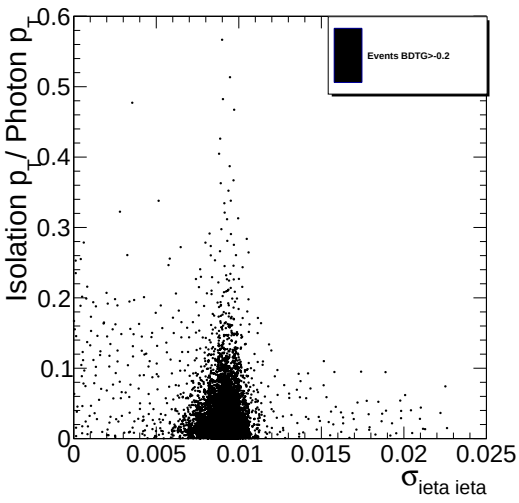
The detector resolution of a Higgs decay to diphoton event is determined by the decay vertex resolution and the energy resolution of the two photons. For the classification of events as signal-like or background-like it is useful to have a single variable which can be used to categorize events. The category boundaries will be described in Section 5.7. Major discriminating power can be given by the energy resolution of each photon which drives the Higgs signal peak. Section 4.4.2 described how the photon energy correction is used to improve the photon energy measurement on the ECAL. The same technique can be used to predict how well-resolved the energy of a reconstructed photon is. The BDT regression is trained with the same variables used to derive the energy correction but the target is now: $\frac{|E_{corr} - E_{sim}|}{E_{sim}}$. The energy scale plots shown in Section 4.4.2 have a characteristic spread about the mean value, and this BDT target gives a measure of how likely the photon energy



(a) $\sigma_{i\eta i\eta}$ for signal and background



(b) PF Isolation vs. $\sigma_{i\eta i\eta}$



(c) PF Isolation vs. $\sigma_{i\eta i\eta}$

Figure 5.10: Comparision of the photon Identification BDT with a one dimensional cut on the shower shape variable $\sigma_{i\eta i\eta}$. The 1D plot of the shower shape variable shows the overlap between signal and background. The 2D scatter plot with the shower shape variable and the isolation carves out a cleaner boundary. The final plot shows the scatter plot for signal events with a cut on the multivariate discriminator.

lies from mean value of the energy scale. Figure 5.11 shows the total corrected energy distribution which is then split into BDT energy resolution ranges. The BDT values correspond to photons narrowly spread about the mean value of one, and in increasing ranges of BDT the spread becomes larger and larger. Figure 5.11 shows the spread about the energy scale for different ranges of the BDT.

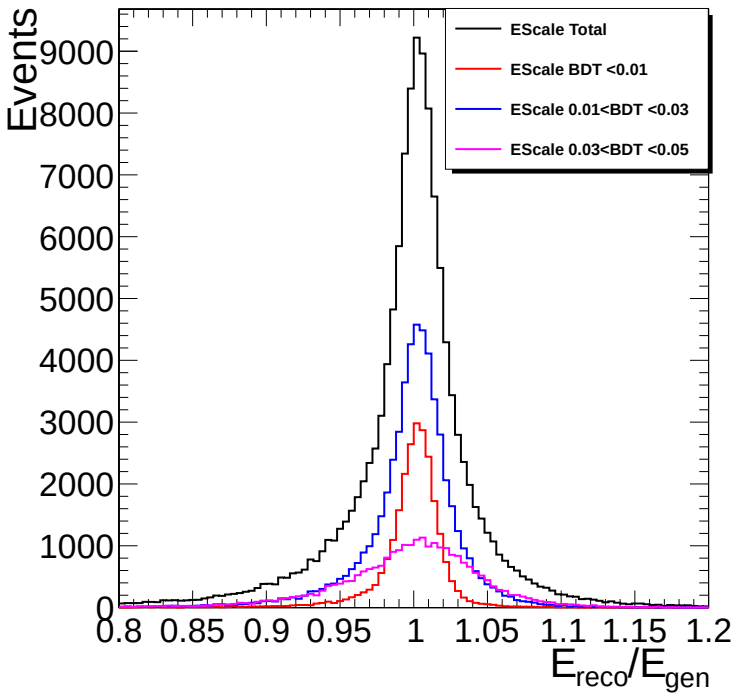
The BDT target $\frac{|E_{corr} - E_{sim}|}{E_{sim}}$ will be only half of a distribution because of the absolute value, so $\frac{E_{corr}}{E_{sim}}$ is binned finely in the BDT regression values and fit to a gaussian. This gives the linear correspondence between the gaussian width of the energy resolution and the BDT values. Figure 5.11 shows the linear function that gives the proportionality between the BDT predicted energy resolution and the gaussian energy resolution. The value of the proportionality is measured for the ECAL Barrel and the ECAL Endcap to give: $\sigma_E^{EB} = 1.07BDT$, $\sigma_E^{EE} = 1.045BDT$.

Figure 5.12 shows how the BDT resolution values for photons and overlays the cut-based categories based on EB/EE and converted/unconverted given by the R9 value. The first category in red shows the best resolved photons measured in the ECAL Barrel and are indicated as unconverted. Comparing this category with the BDT resolution bins shows that it can be split further with more variables to reduce the overlap with the next best category of barrel converted photons. Some portion of the next best photon category in blue, contains some really well resolved photons at the BDT resolution values, which overlap with the first distribution in red. The majority of this category has very poorly resolved photons with energy resolution comparable to photons measured in the endcap. The BDT resolution with additional variables more finely splits photons by their energy resolution prediction than the more coarse categories based on the two variables.

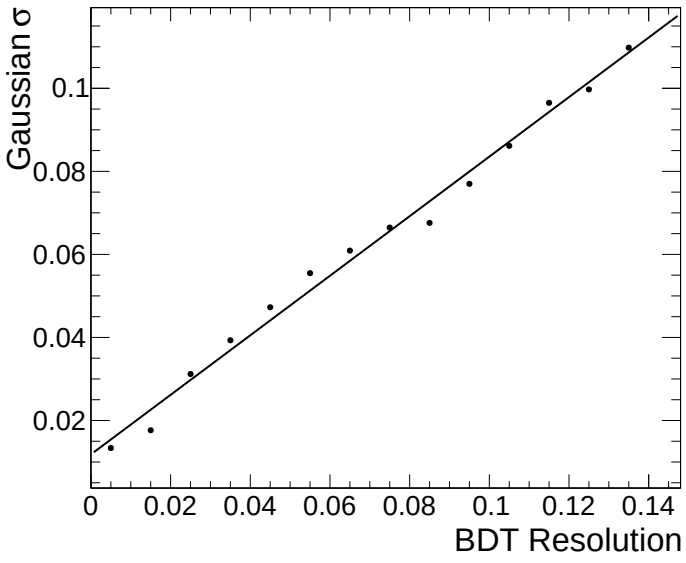
The mass resolution can be computed from the gaussian energy resolution of each photon:

$$\left(\frac{\sigma_M}{M}\right)_{Eres} = \frac{1}{2}\sqrt{\frac{\sigma E_{\gamma 1}}{E\gamma 1} \oplus \frac{\sigma E_{\gamma 2}}{E_{\gamma 2}}}$$

The mass resolution becomes a very strong indicator of the signal to background ratio. As described at the beginning of the chapter, the natural Higgs width is very narrow and the number of signal events is small compared to the background. For all photon resolutions combined, the signal peak is very wide as shown in Figure 5.13 for a simulated gluon fusion Higgs boson with mass of 125 GeV. The signal is modeled by a gaussian probability density function of $m_{\gamma\gamma}$ and the signal strength is measured as the normalization times the signal PDF. The background is modeled by an exponential probability density function where the background is estimated from the normalization of the PDF. The sum of the two functions and their normalizations is fit to an admixture of background simulated QCD diphotons from the diphoton box process and Higgs signal events to give an estimate of the signal strength and background from the fit normalization. For a very broad signal peak, the

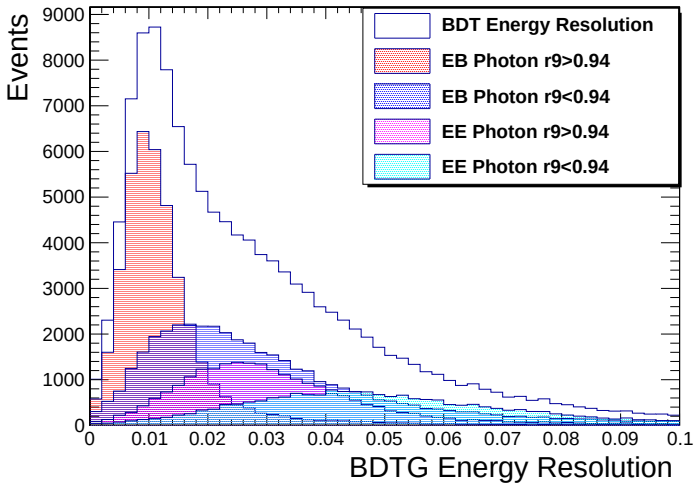


(a) E-scale in ranges of BDT resolution



(b) Sigma of the Energy Resolution vs. BDT resolution

Figure 5.11: Energy Scale split into ranges of BDT. Each range corresponds to a different energy resolution. Finely binning in the BDTG and fitting the energy scale to a gaussian gives a linear function.



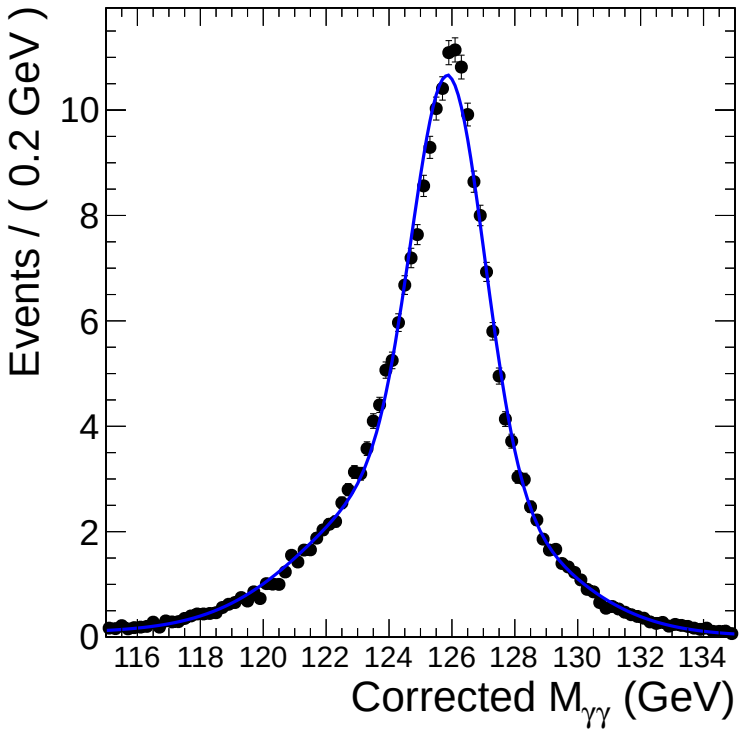
[htb]

Figure 5.12: Energy Scale split into ranges of BDT. Each range corresponds to a different energy resolution.

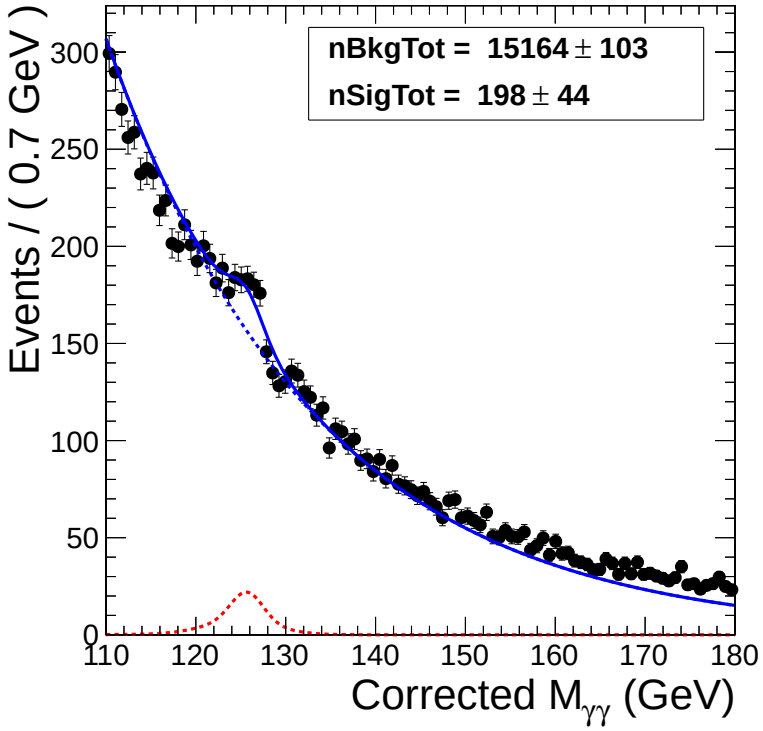
gaussian is spread over a large range of $m_{\gamma\gamma}$ and the integral within the large signal range includes large fractions of background. The result of the fit is shown in Figure 5.13, where the Higgs events are measured as a broad bump on top of the background predicting 200 signal events.

Dividing the events into categories gives two signal models based on $\frac{\sigma_M}{M}$: $\frac{\sigma_M}{M} < 0.1$ and $\frac{\sigma_M}{M} > 0.1$. The first category corresponds to very well resolved photons and gives a very narrow signal window as shown in Figure 5.14. For the high resolution category, the signal and background model consist of the most well-measured photons to reconstruct $m_{\gamma\gamma}$, so the fit to the combined signal and background PDF gives a very sharp peak which lies atop a smaller number of background events. This category, however, gives a smaller yield of 85 events because there are a minority of signal events with very excellent energy resolution. These are the most clearly identifiable Higgs events from the well measured background events. The second category expects a much broader signal from less well measured photons compared to the less well measured background. The second category has a bulk of the yield of 185 Higgs signal events modeled by a broader Higgs signal peak. The split into two categories by $\frac{\sigma_M}{M}$ gives a high purity signal category with 30% of the total signal yield while the other category carries the bulk of signal events. The second category has a smaller signal-to-background ratio since the signal window is larger, so includes a large fraction of background.

Naively summing the yields from the two categories gives 270 events, which is larger than the measured signal yield without categorizing. This gives an indication of the increase in sensitivity by classifying events. Splitting the second category further can boost the signal yield even further creating one category with a larger signal to background ratio and another category where it is smaller but still measures signal events. Compared to the uncategorized signal model, the two

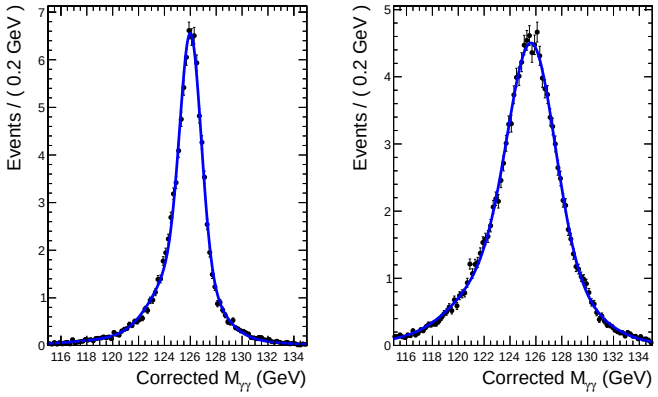


(a) Total Signal Model

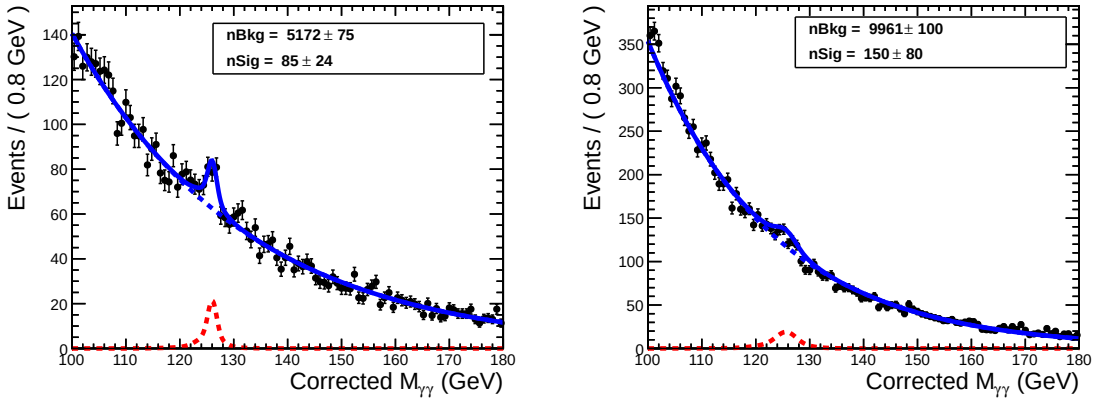


(b) Signal plus Background Combined Model

Figure 5.13: Expected Higgs signal for gluon-gluon fusion Higgs and the Signal plus Background model to measure the Higgs signal in an admixture of signal and background events.



(a) Total Signal Model



(b) Signal plus Background Combined Model

Figure 5.14: Splitting events by the $\frac{\sigma_M}{M}$ into resolution categories gives a narrow signal peak in the high resolution category with a larger signal-to-background ratio. The second category has a broader peak but measured in a category with a smaller signal-to-background ratio.

categories both also give a more significant signal observed above the background model. The next section will describe the final event classification to give the final set of categories which uses the mass resolution as input to develop the final set of categories.

5.7 Event Classification

A major gain in the expected sensitivity to the Higgs decay to photons is achieved in categorizing events according to sensitivity: $\frac{signal}{\sqrt{signal+bkg}}$. After passing the preselection in Section 5.3 and the cut on the photon identification MVA discriminator in Section 5.5, the diphoton events are divided into classes that group events according to their signal-to-background ratio. This is done by training an MVA discriminator on the mass resolution described in the previous section, and kinematic variables that are correlated with $\frac{signal}{\sqrt{signal+bkg}}$.

The mass resolution can be computed solely from the photon energy resolution:

$$\left(\frac{\sigma_M}{M}\right)_{Eres} = \frac{1}{2}\sqrt{\frac{\sigma E_{\gamma 1}}{E\gamma 1} \oplus \frac{\sigma E_{\gamma 2}}{E_{\gamma 2}}}$$

Or it can include the error from the vertex finding:

$$\left(\frac{\sigma_M}{M}\right)_{Eres+vtx} = \frac{1}{2}\sqrt{\frac{\sigma E_{\gamma 1}}{E\gamma 1} \oplus \frac{\sigma E_{\gamma 2}}{E_{\gamma 2}} + \oplus\frac{\sigma_{vtx}}{M}}$$

The first case will be used under the hypothesis that the correct vertex is chosen, so there is negligible error on the opening angle of the photons. If the error on the opening angle is negligible this will group the well measured Higgs events with a well measured prompt photon background. For the non-negligible effects from a poorly chosen vertex, the $\left(\frac{\sigma_M}{M}\right)_{Eres+vtx}$ term becomes important. The luminous region of the LHC at CMS has an RMS spread of about 5 cm so the distance between the true vertex and the selected vertex is distributed as a Gaussian with width $\sigma_Z = \sqrt{2}5cm$ and the contribution of σ_{vtx} is computed from the photon positions and their distance to the vertex (as in Section 5.4).

In the training of the BDT, information needs to be provided that $\frac{signal}{\sqrt{signal+bkg}}$ is inversely proportional to mass resolution. This is achieved by weighting the signal events used to train the BDT by the probability of the vertex choice and the mass resolutions:

$$w_{sig} = \frac{p_{vtx}}{\left(\frac{\sigma_M}{M}\right)_{Eres}} + \frac{1-p_{vtx}}{\left(\frac{\sigma_M}{M}\right)_{Eres+vtx}}$$

The MVA discriminator is designed to fulfill the following criteria:

1. The variable should classify with a high score events with:

   (a) signal-like kinematic characteristics ,

(b) good diphoton mass resolution events,

(c) High purity signal photons given by the photon identification discriminator values,

2. The variable should be mass independent; it should not select events according to the invariant mass.

[52] The following variables are used for the diphoton event classification:

1. the relative transverse momenta of both photons, $p_{\mathrm{T}}^{1(2)}/m_{\gamma\gamma}$,
2. the pseudo-rapidities of both photons, $\eta^{1(2)}$,
3. the cosine of the angle between the two photons in the transverse plane, $\cos(\phi_1 - \phi_2)$,
4. the photon identification BDT output value for both photons,
5. the per-event relative mass resolution estimate assuming the mass has been constructed using the correct primary vertex
6. the per-event relative mass resolution estimate assuming the mass has been constructed using the incorrect primary vertex
7. The per-event probability that the correct primary vertex has been used to reconstruct the diphoton mass, computed from a linear fit to the event-level vertex selection MVA as described in Section 5.4.2.

[52]

The BDT discriminator has been used to define 5 signal categories. The procedure is to successively split the events into classes by introducing a boundary value. The first boundary results in two classes, and then these classes are further split. Each split is introduced using the boundary value that gives rise to the best expected exclusion limit.

As was shown in the previous section, a narrow signal window squeezes the sensitivity to the signal events among the majority of background events. The category of most signal-like events predicts a Higgs mass peak of an estimated Gaussian width of 1GeV and a full-width at half-maximum of 2.01 GeV. The signal width in the categories corresponding to smaller MVA values grows slightly larger. Figure 5.17 shows the expected signal events in all the defined categories for the 8TeV dataset, the signal width, and the estimated $S/(S+B)$. The signal is computed from the expected number of signal events and the background estimate is given by integrating the background model in the signal window region. The signal width for the top four categories is inversely proportional to the

effective signal width. As stated in the previous section, the mass resolution is a good indicator for the signal to background ratio for this analysis.

The diphoton MVA used for classification also includes the photon identification MVA values for each photon and also some kinematic information about the diphoton system. For the background model each category also includes a different composition of backgrounds as shown in Figure 5.16. The main backgrounds of the $H \to \gamma\gamma$ analysis are an irreducible background of photons produced via QCD from the pp collision (denoted as prompt) and also a reducible background from photons that come from meson decays inside of jets (denoted as fake). The figure shows stacked distributions of the Monte Carlo background used for training and also the 8TeV data. The first two categories contains mainly the prompt photons in the high p_T range corresponding to well identified photons with signal like kinematics, and the lower two categories shows a larger admixture of the fake background. A cut on the BDT value close to zero removes a bulk of the background. Figure 5.15 shows a comparison of the MVA categories with simple cut based resolution categories. A fraction of the most well-resolved photon pairs (in red) in the $EBEB$ $R9 > 0.94$ category are included in the highest BDT category. The poorly resolved photon pairs, which have at least one photon in the endcap and one photon to be converted, fall below an MVA value of zero so would be classified as mainly background-like events.

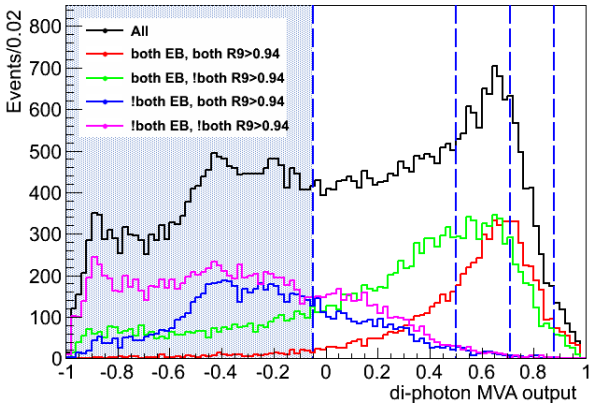
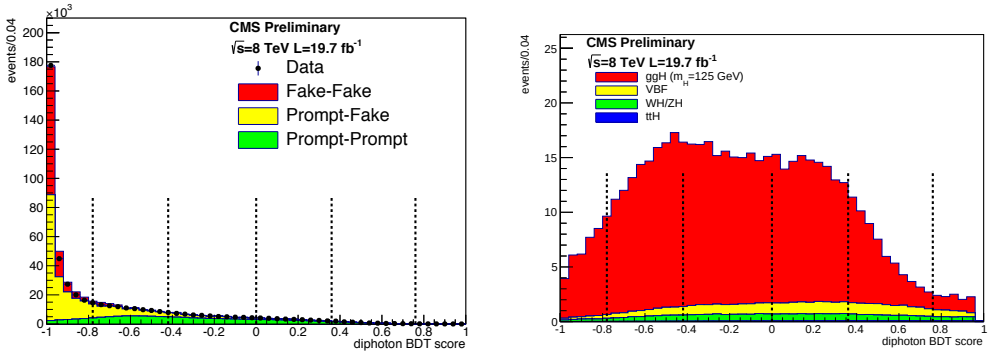


Figure 5.15: Cut based categories overlapped with Diphoton BDT values

In addition to the diphoton resolution-kinematic categories defined by the previously defined MVA, events can also be split by the different final states that arise from different Higgs production mechanisms. Higgs bosons produced by VBF are accompanied by a pair of jets separated by a large rapidity gap. Those produced by VH may be accompanied by a charged lepton, large E_T^{miss} , or jets from the decay of the W or Z boson. Those produced by ttH are accompanied by $b-$quarks, and may be accompanied by charged leptons or additional jets. Unlike the gluon-fusion events, the



(a) MVA categories 8TeV Backgrounds

(b) MVA categories 8TeV Signal

Figure 5.16: Signal processes contained in diphothon BDT categories and also the Background composition.

additional object-tagging allows for very pure signal categories. This can be seen by looking at the estimated $S/(S+B)$ for the tagged categories in Figure 5.17. Though these categories have much fewer expected signal events, they also have smaller backgrounds because they require a tag on the additional object. However, some tagged categories can have significant leakage from the dominant gluon fusion production mechanism like the dijet tag. The tagging of dijet events not only significantly increases the overall sensitivity of the analysis and precision on the measured signal strength, but is aimed to measure the coupling of the Higgs to vector bosons from the VBF production mechanism. The procedure to try to disambiguate dominant gluon fusion process from the vector-boson fusion process will be described in the next section. The tagging aimed at the associated production processes, VH and ttH, increases the sensitivity of the measurement of the coupling to vector bosons and also coupling to the top quark, and further probes the compatibility of the observed signal with a SM Higgs boson.The p_T spectrum of Higgs bosons produced by the VBF, VH and ttH processes is significantly harder than that of Higgs bosons produced by ggH, or of background diphotons. This results in a harder leading photon p_T spectrum. This advantage is used in the tagged class selections by raising the p_T requirement on the leading photons to reject background.

These categories were first used in a Higgs analysis searching for a fermiophobic Higgs boson targetting only production of the Higgs boson where is couples only to vector bosons. The sensitivity of the search is maximized by tagging the additional objects in the final state. For the associated production of Higgs with a vector boson, leptons from the leptonic decays of the vector boson are tagged in two categories: One with a final state muon and one with a final state electron based on selection cuts. The dominant mode of fermiophobic Higgs production is the vector boson fusion which results in final state jets. The dominant mode of fermiophobic Higgs production was discriminated

from other background processes using a two dimensional PDF with $m_{\gamma\gamma}$ and $\pi_T^{\gamma\gamma} = p_T^{\gamma\gamma}/m_{\gamma\gamma}$. This used the fact that a single kinematic variable, $\pi_T^{\gamma\gamma}$ discriminates the Higgs signal which has larger transverse momentum from the background processes. For the Standard Model Higgs search, the $\pi_T^{\gamma\gamma}$ is used in the diphoton MVA, and the dijet-tag also makes use of a single kinematic variable m_{jj} in a 2D PDF of $m_{\gamma\gamma}$ and m_{jj} to discriminate the VBF signal from gluon-fusion and also the diphoton background. The same categories can be used for the search for a Standard Model Higgs boson in order to target the vector boson modes of production. The tagging of additional objects gives categories with a lower background rate than the diphoton categories defined by the BDT. The categories based on production mode boost the sensitivity of the Standard model Higgs analysis. It also allows for a measure of the Higgs coupling to fermions and also the Higgs coupling to vector bosons, since the events are split by the Higgs production mechanism. [14]

For the 2013 analysis in [51] four categories gave optimal sensitivity, these are shown with dashed lines in Figure 5.15, where the minimum BDT value allowed into the analysis was for $BDT > -0.05$, and the 7TeV also was analyzed with four BDT categories. The first 8TeV analysis, included two cut-based categories for the dijet tag targeting vector boson fusion Higgs. Higgs produced associated with a vector boson was targeted in three categories based on the lepton decays: muon tag, electron tag, and also targeting Missing Energy from the neutrinos that come from the W decay. For the analysis in [52] the Diphoton MVA is retrained for the 2011 and 2012 data. Also the energy corrections and energy resolution are simultaneously trained using a modified version of Boosted Decision Tree. The 7TeV dataset, is still divided in four new categories while for the 8 TeV dataset an additional category was added for lower diphoton BDT scores in order to include more events from vector boson fusion. Additional tagged categories are added in [52] to include hadronic decays of vector bosons produced in association with the Higgs. Also additional categories are included to target the ttH process. This new categorization for the 8TeV is shown in Figure 5.16, and the order of the tagging is chosen to minimize the overlap of the four production processes.

To encapsulate all production mechanisms even those with smaller production rates are included like VH and ttH. The associated production, VH, targets Higgs events with additional leptons or with large E_T^{miss} from the W-decays, while ttH targets events with leptonic or hadronic top decays. In the most recent analysis, even VH Higgs are tagged where the vector boson decays hadronically. In [51], the 7TeV dataset included only dijet tagged events in a single category from vector-boson fusion, while the 8TeV dataset contained two dijet tagged categories targeting VBF and three categories to target VH production. The three Vector boson associated to Higgs production categories each target one additional signature: either an electron, a muon, or E_T^{miss}. In [52], the exclusive categories are further split and also include categories that target ttH production. The order of the selection is of

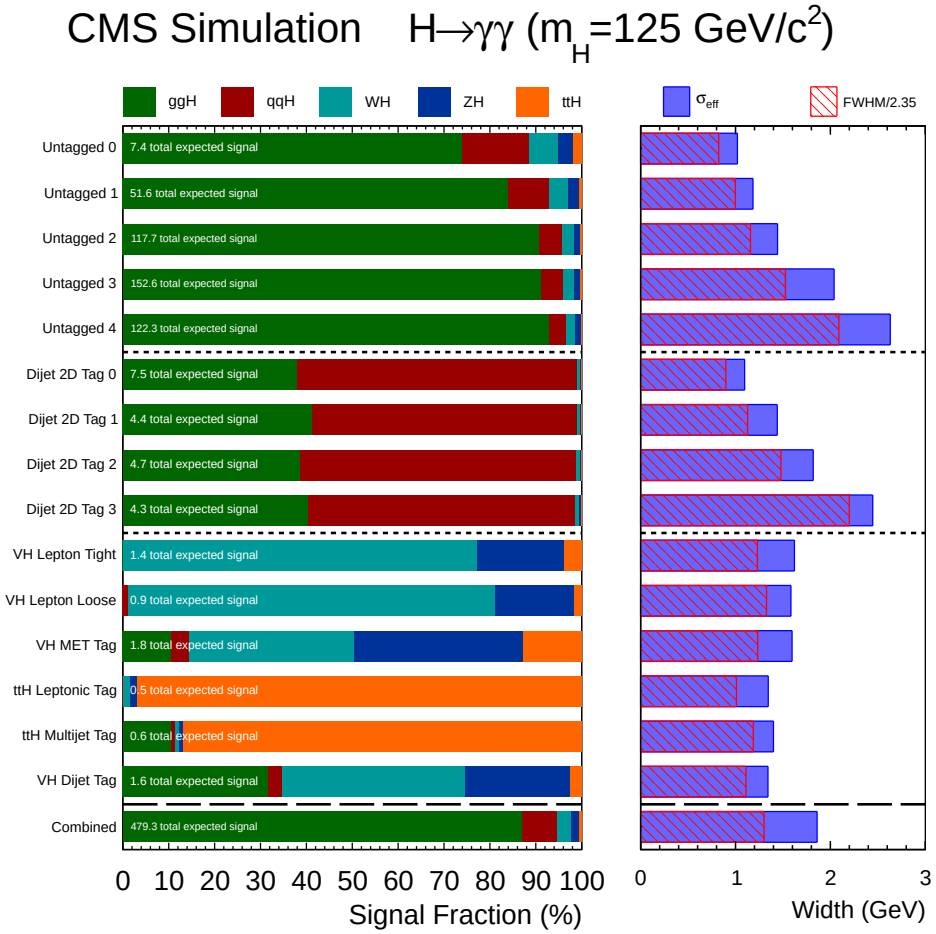


Figure 5.17: Signal composition based on Higgs production mechanism of each event class and the expected signal width (measured as the Full-width at Half Max and the effective gaussian width)

prime importance to disambiguate the overlap between different production mechanisms:

- First events are selected with leptons from the leptonic or semi-leptonic top decays.
- Events with leptons from the VH process are selected in two E_T^{miss} categories.
- Events with two jets are selected and divided into categories to target the VBF process. Events from the gluon-gluon production mechanism are also separated. This selection is described in Section 5.7.1.
- Events with high E_T^{miss} are selected to target the VH process.
- Jets from hadronic top decays are selected.
- Jets from a hadronically decaying vector-boson are selected.

[52]

To summarize: events with leptons are selected first for the top decays and then for the VH process. The VBF selection is then done to tag events with jets. Then jets from top decays are considered and finally jets from the VH process where the vector boson decays hadronically are considered. The signal purity of each process in each category is shown in Figure 5.17. Section 5.8 describes how the signal measurements over all categories is combined taking into account the uncertainty measured in each category.

5.7.1 Dijet Tag for Vector Boson Fusion Higgs

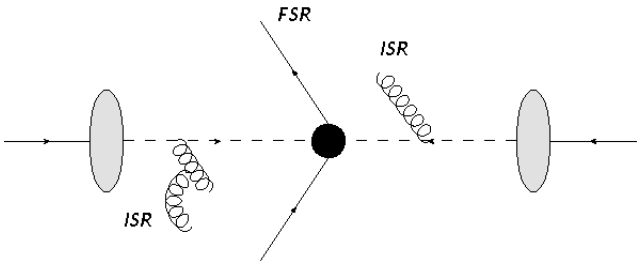


Figure 5.18: Initial State Radiation vs. Final State Radiation for incoming protons in a collision.

This section will describe the classification of Higgs decays produced through vector boson fusion. This production mode has two jets in the final state from scattered quarks. The four body system of two jets plus the two photons allow for combinations of dijet and diphoton kinematic variables to be used to discriminate the vector boson fusion process from the background as well as from the dominant process of gluon-fusion. However, each kinematic variable needs to be theoretically well understood so that the simulation describes a well predicted VBF Higgs event. There are two

strategies for classifying Vector Boson fusion events. One is an MVA approach that uses a set of kinematic variables from the four body system, as well as the diphoton classification described in the previous section. The second approach uses only one discriminating variable from the dijet system, the dijet mass, together with the diphoton mass and uses it to model 2D probability density functions for VBF Higgs events, gluon-fusion Higgs events, and the diphoton background.

The tagging of Vector boson fusion events involves identifying the event topology which corresponds to the most likely Vector boson fusion events while rejecting gluon-fusion events as well as background. The gluon-fusion Higgs events will be accompanied by a mess of radiated gluons from the deep inelastic scattering of protons as shown in Figure 5.18. These emissions occur before the hard process, so it is called the initial state radiation. The gluons will cascade into jets which can mimic the vector boson fusion jets. This process is not theoretically well understood and so could be mis-modeled in the parton shower simulation. The vector boson fusion signature has two quarks in the final state, final state radiation, that are emitted along with the Higgs boson from the initially scattered quarks. One process that is very similar to the vector boson fusion production is the higher order QCD process of Drell Yann production where $qq \rightarrow qqZ$. These events are used to study background rejection of the large amount of QCD background and validate the kinematic variables for the vector boson fusion process in data. The background for vector boson fusion from gluon fusion is more difficult because the emission of initial state radiation is not theoretically well predicted for the loop-induced gluon-fusion process. However, $t\bar{t}$ production is a process which proceeds dominantly by gluon fusion (85% of the cross section). Top quark production can then be used to verify the jet kinematics for initial state radiation accompanying the gluon fusion process.

In [51], a set of cuts on kinematic variables defined the two dijet categories. The loose categories corresponds to a low dijet mass category of $250GeV < m_{jj} < 500GeV$ which also contains gluon-fusion events and some diphoton background events. The tight category with $m_{jj} > 500GeV$ contains a larger fraction of VBF compared to gluon-fusion events and background events. The 2D fit approach instead models the m_{jj} of the signal VBF, the gluon fusion, and the diphoton background with PDF which acts as a continuum of m_jj categories. The kinematic variables used to tag Vector boson fusion are as follows:

- Photon Kinematics: $\frac{p_T}{m_{\gamma\gamma}}$ of each photon
- Jet Kinematics: Jet p_T, dijet mass, difference in pseudorapidity between jets
- Photon-Jet system kinematics: $\eta|_{obs} = \eta\left(\gamma_1 + \gamma_2\right)$ is used to shift the average pseudorapidity of the two jets $Z = \eta|_{obs} - \frac{\eta_1^{j1} + \eta_2^{j2}}{2}$. The difference in azimuthal angle between the dijet and diphoton system is also used:$\Delta\phi_{jj,\gamma\gamma}$.

Many jets are reconstructed from the soft emissions from pileup interactions being clustered together to form a jet. Jets from the VBF signal are clustered using the anti-kT algorithm described in Section 3.2.3 to give narrow circular cone jets. The jets from soft emissions are estimated by very quickly clustering jets using the FASTJET algorithm described in Section 4.3 which can result in very wide cones with a large jet pT. These pileup jets are removed using a selection criteria based on the compatibility of jets coming from the primary vertex, which has the largest $\sum p_T$ of tracks as well as the width of the jet. The variables used to pileup jets are :

- $\beta^* = \frac{\sum_{jet PU vtx} p_T^{tk}}{\sum_{jet} p_T^{tk}}$
- $RMS = \frac{\sum_{constituents} p_T^2 \Delta R^2}{\sum_{constituents} p_T^2}$

The sum of constituents is the sum of particle flow charged hadrons that make up about 2/3 of the substructure of real jets. A key advantage to the Particle Flow jet algorithm is the ability to associate each charged hadron to a vertex, which is useful to create the above variables to separate particles from the hard process from the overlapping particles that come from the pileup. All jets identified as coming from the pileup are removed. Cuts on the variables are chosen to keep a steady background rejection with respect to the number of overlapping interactions, and maintain a high signal efficiency above 95%. This pileup cleanup of jets is done for both VBF tagging techniques.

The first technique uses a Multivariate analyzer to identify VBF events and it is built in two stages. The first stage trains a Boosted Decision Tree algorithm trained on the kinematic variables above to identify VBF events from the background and also from the gluon fusion. The input variables combine dijet variables with variables of the full dijet-diphoton system. The output of this MVA is input to another MVA which combines it with the diphoton MVA output described in the previous section and also the diphoton kinematic variable: $p_T^{\gamma\gamma}/m_{\gamma\gamma}$. This combined MVA is designed to maximally discriminate the background by using all the information available from the photons, which is combined in the diphoton MVA value, and all the information from the jets which is combined in the first MVA. The diphoton variable $p_T^{\gamma\gamma}/m_{\gamma\gamma}$ is also included because it correlates to both MVA values. The final combined MVA value is cut on to categorize VBF events. The combined MVA uses VBF events as signal but is only trained with the diphoton background and not with the gluon fusion. This technique suffers because it relies on a sensitive combination of many input variables, some of which are not theoretically well predicted and cannot be easily validated in data. Only the first stage discriminates the gluon-fusion from the vector boson fusion, which mainly contaminates the vector boson fusion events. The MVA makes use of variables of the full 4 body final state :$\Delta\phi_{jj,\gamma\gamma}$ which can be verified in data using $qq \to qqZ$ events for the signal VBF process, but no such control sample exists for the gluon fusion process.

Instead of using all possible dijet and diphoton kinematics for discrimination, the second approach uses the two main observables $m_{\gamma\gamma}$ and m_{jj} to build a 2D probability density function for signal VBF events, and background gluon-fusion events and also the diphoton background. Only the m_{jj} variable needs to be validated for each process. As before $qq \to qqZ$ can be used to validate the dijet mass for the VBF signal process. The gluon fusion m_{jj} variable can be studied using top production in data which is produced dominantly by gluon fusion.

The 2D approach can be summarized in Figure 5.19. The background jets from gluon fusion and the diphoton background come from soft emissions mainly from the initial state radiation which have a smaller dijet mass. The signal VBF process in contrast has a large m_{jj} high tail from the p_T spectrum of the scattered quarks from a hard collision. The three shapes can be modeled by a probability density function of m_{jj}. The background events and the gluon fusion are modeled by a Landau function, which is characterized by two parameters (μ, σ) where μ is the most probable value for m_{jj} and σ is the width. The function was first used to describe energy loss by ionization so it has a steep turn on on the low side and then a long radiative tail at high m_{jj}. Since the signal VBF process has a less steeply falling radiative tail than the background processes, it is modeled by simply adding a Gaussian function and a Landau function together with appropriate weights as $W_g G\left(m_{jj}|\mu_g, \sigma_g\right) + W_L L\left(m_{jj}|\mu_L, \sigma_L\right)$ with $W_g + W_L = 1$. The dijet mass is modeled by these functions and becomes part of the model for signal and background to discriminate the VBF process from the diphoton background and also the gluon-fusion Higgs signal. The probability density functions for m_{jj} are combined with the probability density function for $m_{\gamma\gamma}$ to form a 2D probability function for both functions. If m_{jj} and $m_{\gamma\gamma}$ are very nearly uncorrelated than the combined PDF factorizes as a product: $P\left(m_{jj}, m_{\gamma\gamma}\right) = p\left(m_{jj}\right) \cdot p\left(m_{\gamma\gamma}\right)$. However, the two observables can have some measurable correlation. Any correlation must be introduced as a conditional PDF:$P\left(m_{jj}, m_{\gamma\gamma}\right) = p\left(m_{jj}\right) \cdot p\left(m_{\gamma\gamma}|m_{jj}\right)$. The combined fit to $P\left(m_{jj}, m_{\gamma\gamma}\right)$ now allows the measured signal in the dijet category to be split into the two signal processes: gluon fusion and vector boson fusion as shown in Figure 5.19. Instead of applying a flat uncertainty for the maximum fraction of processes contaminating the dijet tag, the gluon fusion is measured together with the VBF Higgs signal. The error in determining the contamination now comes from the uncertainty on each of the shapes and its normalization.

The events selected use only simple cuts on the kinematic variables above and the cuts are shown in Table 5.3. The kinematic cuts are the same as used in the dijet categories in [51]. In addition, cuts on the photon ID MVA value described in Section 5.5 is applied to reject events from the diphoton background.

As described in Section 5.6, the dijet tag is split into four resolution categories for 8TeV and

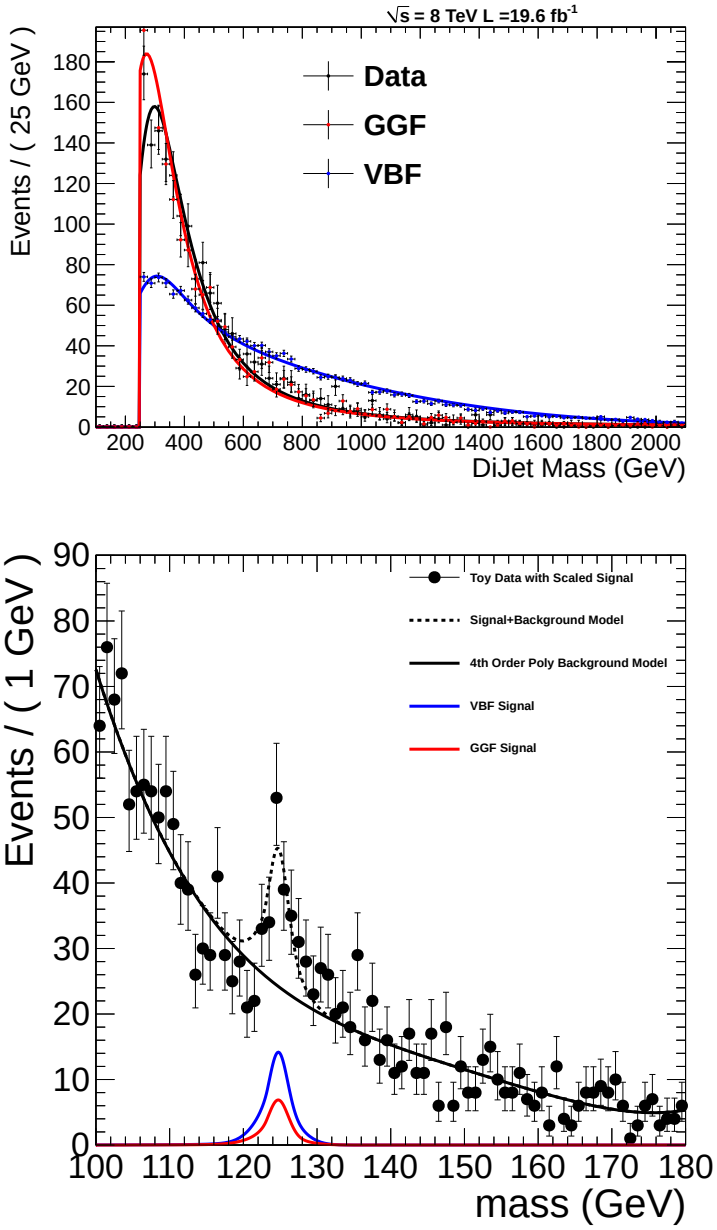


Figure 5.19: 2D modeling of the vector boson fusion signal and the gluon-fusion signal. The PDF component for M_{jj} separates the gluon fusion process in red from the vector boson fusion in blue. The signal peak (scaled by a factor of 3) in $m_{\gamma\gamma}$ becomes split between the two production mechanisms.

Variable	Cut Value
$p_T^{\gamma_1}/m_{\gamma\gamma}$	> 0.5
$p_T^{\gamma_1}/m_{\gamma\gamma}$	$> 25/120$
$p_T^{j_1}$	> 30 GeV
$p_T^{j_2}$	> 20 GeV
$\|\Delta\eta_{j_1 j_2}\|$	> 3.0
$\|Z\|$	< 2.5
$M_{j_1 j_2}$	> 250 GeV
$\|\Delta\phi(\mathrm{jj}, \gamma\gamma)\|$	> 2.6
$IDMVA\gamma_1$	> 0.03
$IDMVA\gamma_2$	> 0.03

Table 5.3: Cut-based Dijet tag

2 resolution categories for the smaller 7TeV dataset. This improves the sensitivity to VBF events which can have high p_T photons, so many well resolved events. Due to the presence of jets, the primary vertex becomes is the signal vertex so the mass resolution only depends on the energy.

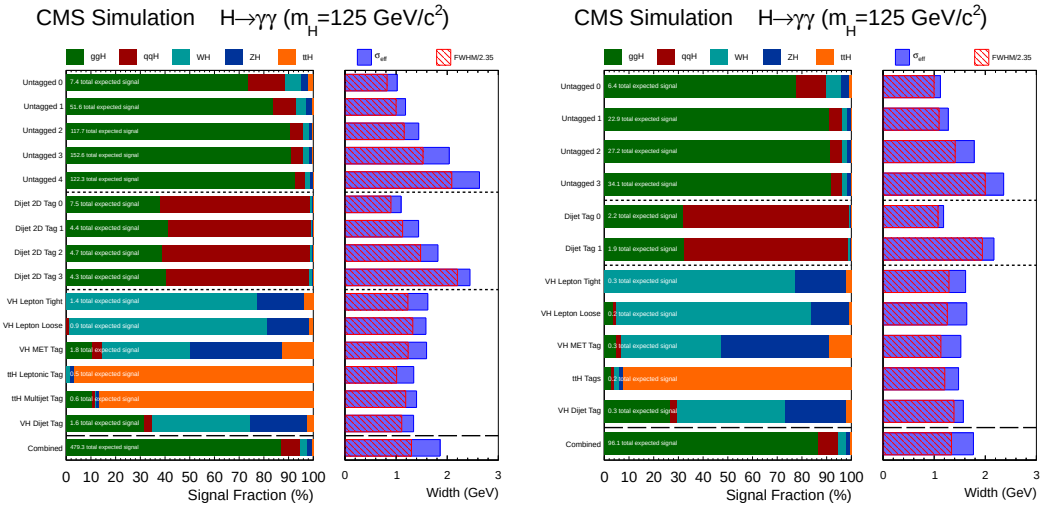
Category	$\frac{\sigma_m}{m}$ range (%)
0	$[0.00 - 1.00]$
1	$[1.00 - 2.00]$

Table 5.4: Event categorization in 2D VBF analysis for 7TeV data

Category	$\frac{sigma_m}{m}$ range (%)
0	$[0.00 - 0.80]$
1	$[0.80 - 1.05]$
2	$[1.05 - 1.40]$
3	$[1.40 - 2.00]$

Table 5.5: Event categorization in 2D VBF analysis for 8TeV data

The signal models are modeled by 2D PDFs that factorize m_{jj} and $m_{\gamma\gamma}$, while correlations are included between the two variables for the data. This will be described in the final statistical treatment for the 2D dijet categories. The 2D fit method simultaneously measures both the gluon-fusion and the vector-boson fusions from the complete signal model. Both signals can be resolved with some uncertainty depending on the shape uncertainty for m_{jj} of each signal model. The MVA approach attempts to separate both VBF-like events with GGF-like events in the first stage where gluon-fusion is included as a background and relies on all the diphoton+dijet for the four body kinematics. Some theoretical uncertainty is associated with the full four body kinematics and it is not clear how these variables are simulated for ISR jets in the gluon fusion background simulation and FSR jets in the VBF signal simulation. The $qq \to Zqq$ process where the Z decays to electrons can be used to compare discrepancies between data and simulation for the VBF signal. However, here the scattered quarks may not have the same kinematics as the FSR from VBF Higgs. Also



(a) 2D VBF categories for 8TeV data

(b) 2D VBF categories for 7TeV data

Figure 5.20: Signal Composition for the 2D PDF model categories for 7TeV and 8TeV. The overlap of ggH (in Green) and VBF (in Red) will be simultaneously measured using the 2D PDF composite signal model.

the boost of the Z is not the same as that of VBF Higgs, so the electrons from the Z are not exactly comparable to the photons from the Higgs. For the gluon fusion process there is no such control sample with a final state of ISR jets. One could try to mix gluon fusion Higgs events with a control sample that features ISR jets but this will also have some unpredictable uncertainties. The $t\bar{t}$ production process is produced mainly via gluon-fusion ($\approx 85\%$) of the total production cross-section), so this sample can be used to compare the ISR from gluon-fusion in data and simulation. For leptonically decays of $t\bar{t}$, the overlap from the final state jets with the initial state radiation can be more easily disambiguated and a selection criteria on the leptons gives a pure conrol sample of $t\bar{t}$ events in data.

The full set of 2D signal and background models for each dijet category will be described in Section 5.8.5 and the final results will be shown in Section 5.9.

5.8 Statistical Analysis

The previous sections described how events with two photons (and additional objects) are selected and also described how events are classified based on their signal-to-background ratio. This section describes the statistical treatment of the data to claim the discovery of a Higgs boson as well as measure some its properties. The two main types of results presented in particle physics experiments are searches where a claim can be made about a particular theory and measurements where the best numerical value and uncertainty are assigned to a physical observable. The Higgs decay to photons

analysis began as a search for the Higgs boson within the mass range of $100 < m_{\gamma\gamma} < 180$. Once the claim was made that a Higgs-like boson was discovered, the emphasis was switched to measuring physical observables like the Higgs mass and also the compatibility with the Standard Model Higgs production rate.

The major advantage for looking for the Higgs decaying to photons is the mass resonance that can be observed across the smoothly falling background of non-resonant photon pairs. Each event class described in Section 5.7 predicts a resonance at a particular mass with a characteristic width. Ideally the background of non-resonant photons would be a smoothly falling function, but realistically there will be both upward and downward fluctuations. The model of the background of diphoton mass pairs must be robust enough to minimize a signal bias due to statistical fluctuations which can enhance or degrade the signal. The procedure for gauging the robustness of the background model will be described in Section 5.8.3. An important measure for claiming discovery of the Higgs boson is the p-value and its statistical significance. The p-value relates to the agreement between the data and a signal hypothesis of a significant resonance above the background fluctuations. The significance of this p-value is usually required to be above 5σ for a discovery so that the peak is highly probable above any fluctuations of the background.

Section 5.9 will show the exclusion limit and the p-value that confirm the existence of a Higgs boson with a mass of approximately 125 GeV as well as excluding other hypothesis Higgs masses. The rate of Higgs decays to photons is normalized by the predicted rate from the standard model as $\mu = \frac{\sigma\left(H \to \gamma\gamma\right)_{obs}}{\sigma\left(H \to \gamma\gamma\right)_{SM}}$. This quantity indicates the compatibility of the observed data with the Standard Model prediction, and it will be shown that the ratio is compatible with 1 within the measured uncertainty. The production rate is also measured based on the different production mechanisms described in Section 2.2. The most compatible value for μ_F for Higgs production from fermions is measured as well as μ_V for Higgs production from vector bosons. The narrow resonance of the Higgs decaying to photons is used to measure a major missing parameter in the Standard Model: the Higgs Mass. The mass measurement will be described as well as the uncertainty in the measurement. The results show strong compatibility of the observed mass resonance with the Higgs boson predicted by the Standard Model.

5.8.1 Signal Hypothesis Tests

In order to take great care in preventing false-positive results, much care was given to statistically combining Higgs measurements across different decay channels and also across categories in a single decay channel. This section will briefly summarize the statistical treatment which will give the results described in the next section.

For setting exclusion limits, a frequentist statistical test is used. The frequentist interpretation defines a probability by its limiting frequency. A simple frequentist test is to roll a dice and count the rolls with 3 in N trials. N does not need to be infinite to establish that the probability is $\frac{1}{6}$ within a given statistical confidence interval. For a resonance a simple frequentist game can be played by generating 100 Monte Carlo $m_{\gamma\gamma}$ distributions with some statistical fluctuations and then adding one distribution that is actual data. Then ask a graduate student with a good eye to choose the 5 most likely distributions with a resonance from the 101 $m_\gamma\gamma$ distributions. If 1 of the 5 that she chooses is the data distribution, then one could say at 95% confidence the data has a real resonance. This relies on the fact the graduate student's choice involves negligible uncertainty. Quantifying or modeling the goodness of the generated models, introducing uncertainty, will change the criteria for a given confidence interval. [42]

The exclusion limit is a standard representation of a particle physics result that tests a hypothesis of signal plus background processes against a background only hypothesis. The diphoton mass spectrum for signal and background is modeled by a composite of probability density functions. The sum of the signal and background PDFs gives a signal plus background hypothesis and the normalization in front of the signal PDF gives the signal yield while the normalization in front of the background PDF gives the estimate of the background. The normalization and the PDF shapes can be subject to uncertainties which can be modeled by a set of nuisance parameters θ. The signal and background PDFs can be denoted as $s\,(\theta)$ for signal and $b\,(\theta)$ for background.

The first step of hypothesis testing is to define a test statistic. It is common to define this test statistic in terms of a log likelihood ratio. The likelihood function gives the likelihood of particular dataset being compatible with a given hypothesis. For the hypothesis of a signal peak on top of a smoothly falling background the Likelihood is written in terms of the signal and background probability density functions:

$$\mathcal{L}(x|\vec{\theta}) = \frac{e^{-\mu_s n_s - n_b}}{N!} \prod_{i=1}^{N} (\mu_s n_s \mathcal{P}_s(x^i|\vec{\theta_s}) + n_b \mathcal{P}_b(x^i|\vec{\theta_b})) \tag{5.1}$$

The product is over the Poisson probabilities: $(\mu_s n_s \mathcal{P}_s(x^i|\vec{\theta_s}) + n_b \mathcal{P}_b(x^i|\vec{\theta_b}))$ for the dataset x^i with n_s expected signal and n_b estimated background. N is the total number of signal categories. The dataset can consist of 1-dimension information like $m_{\gamma\gamma}$ or for 2-dimensional $(m_{\gamma\gamma}, m_{jj})$ used for the 2D likelihood for Vector boson fusion. The probability density functions $\mathcal{P}_s$, $\mathcal{P}_b$ are the signal PDF and the normalized background PDFs. The PDF shape parameters are specified by θ. The parameter of interest in the signal hypothesis test is μ_s which scales the expected number of signal events measured in Monte Carlo n_s. Maximizing the likelihood function gives the values of θ and μ_s that are most compatible with a given dataset. Common practice is to take the negative log

of the likelihood function. Near the maximum of the likelihood function, the function can become gaussian if it is fit to a real measurement with a gaussian uncertainty. For $\mu^{obs}_{bestfit}$ the negative log then gives a parabola near the minimum since $\mathcal{L}(\mu|\theta) \approx \exp\left(-\left(\mu - \mu^{obs}_{bestfit}\right)^2/2\sigma\right)$. The width of $\mathcal{L}$, σ, corresponds to a range of likely values centered on best fit value. For the results the two ranges to consider are the 1σ and 2σ ranges of $\mathcal{L}$ corresponding to the 68% and 95% probabilities. The compatibility of data with the background only and signal plus background hypotheses is done by building a test statistic in terms of the signal strength modifier μ_s and the parameters θ:

$$q_\mu = -2\ln\frac{\mathcal{L}(\mu_{test},\theta)}{\mathcal{L}(\mu_{fit},\theta_{fit})}$$

In the numerator, μ is fixed to a particular test value of 1, while only the θ are varied. The denominator is maximized at (μ_{fit},θ_{fit}) for the data. The above test statistic can be computed for two values of μ, $q_{\mu=0}$ which is the background only hypothesis and $q_{\mu=1}$ where there is signal plus the standard model predicted number of signal events. This will give the values for θ under different hypotheses. The test statistics are now used to generate pdfs under the two hypotheses. Denote these as $f_\mu(q_\mu)$ and $f_0(q_0)$. These functions can be integrated from $q_{obs\mu}$ to infinity to obtain the confidence limits. The ratio of the integrals gives the confidence limit of the observation: $CL = \frac{p_\mu}{p_0}$. The interpretation of this number is that for $CL < \alpha$ the Higgs boson produced with signal strength $\mu = 1$ is excluded with the confidence level of $1-\alpha$. For an upper limit on μ at 95% confidence the value of μ would be adjusted to give CL=0.05. [12]

The expected limit is defined by the median upper limit and the band of 1σ and 2σ probabilities as function of μ and the hypothesis Higgs mass. It may appear that to determine the expected limit requires a huge number of test-statistics to determine $\mu^{95\%}$ for each Higgs mass. This process becomes very CPU intensive with many categories in a large dataset. Also it is very useful to be able to relate the p_μ to a gaussian significance $Z\sigma$ which expresses the significance of an excess above fluctuations of the background. The background p-value written as:

$$p_0 = \int_{q_{obs}}^{\infty} f_0(q_0)$$

The integral relies on generating many background only test-statistics to well model the tail of this hypothesis. A value of Z=5 for discovery corresponds to $p_0 = 2.8\times10^{-7}$. [12] However, one can rely on the asymptotic behavior of the test-statistics from the likelihood ratio to give the p-value in terms of a one sided Gaussian tail :$p = \frac{1}{2}\left[1 - erf\left(\sqrt{q_{obs}/2}\right)\right]$. The test statistic in this asymptotic approximation is a well defined formula. The formula is determined in terms of the most nominal test-statistic, the Asimov dataset, which corresponds to the expected background without any fluctuations. [12] This asymptotic formula for the $f(q_\mu)$ removes the need to throw many toys

to determine f, instead the most typical dataset gives both the median expected limit and the error bands from a smooth functional form. The p-value, the upper limit and the significance now all relate simply to the test-statistic from the Asimov data. [25] However, for categories with a small number of events, the asymptotic formula is not guaranteed and can result in some bias. [12] The procedure for testing potential signal bias (overoptimistic results) is described in Section 5.8.3.

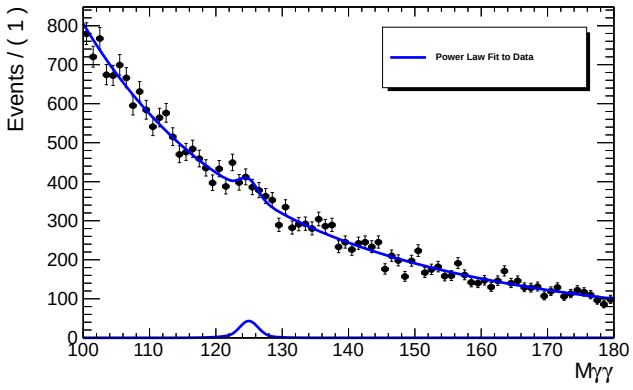
5.8.2 Signal Models

In order to do the statistical analysis of the data, it is necessary to describe the signal specifying the overall efficiency×acceptance and also the shape for the diphoton mass distribution that accounts for how well $m_{\gamma\gamma}$ is reconstructed. The efficiency times acceptance accounts for all the efficiency corrections and scale factors applied to the normalization of the signal PDF for a particular signal category. The final signal models give the expected parametric model of the signal expected in data.

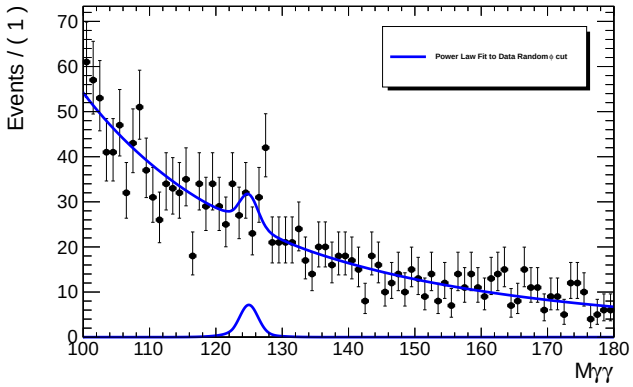
The signal PDF itself is modeled by the sum of 2 gaussians or the the sum of 3 gaussians depending on the signal width of the category. The signal PDF has to be smeared to widen the total gaussian so that it reflects the energy resolution observed in data as described in Section 4.7. The uncertainty on the energy scale is incorporated as a shift in the mean of each Gaussian, and uncertainty on the resolution is incorporated by convoluting the signal model, *sigma*, for each gaussian with an additional gaussian width. The signal PDF are fit to Monte Carlo simulation which is generated for each Higgs production mechanism in steps of 5GeV from 100 to 150 GeV. Within the 5GeV mass points the signal shape is linearly interpolated to provide a continuous signal model across the full mass range. The normalization is first scaled to account for the predicted cross section at the luminosity $N_{MC}/\mathcal{L}\sigma\left(H \to \gamma\gamma\right)$.

The normalization of the signal PDF accounts for the efficiency corrections and scale factors. Each signal event is weighted so that the signal model accounts for the pileup in data. The signal pileup weighting is designed to account for the difference in the simulated pileup and what is observed in data, so the signal PDF most accurately accounts for what would be observed in data. As described in Section 5.2, each run period for the 8TeV dataset corresponds to a different distribution of pileup events and the 7TeV has a smaller amount of pileup. For the most current results, the simulation is run dependent to account for these run-dependent effects, so events are weighted according to the particular data run period with which they correspond. Small corrections are also applied to account for efficiency loss when applying the selection criteria, these corrections are measured using $Z \to ee$ events. The preselection efficiency is described in Section 5.3 as well as the trigger efficiency and also there is an efficiency for passing the Photon ID cut described in Section 5.5. Each of the efficiency scale factors is applied to the signal model.

5.8.3 Background Models



(a) Total Power Law fit to all statistics



(b) Power Law fit to statistics in Random category

Figure 5.21: An example to show the effect of a bias when categorizing. The top plot shows a reasonable fit of the diphoton background with a power law and the resulting measured signal peak. Applying a random cut to creates a random category with still a measured signal, but now it is not clear if this signal is due entirely to statistical fluctuations.

The background seems deceptively simple since it parameterizes the smoothly falling non-resonant diphoton mass spectrum. However, the choice of a background function that is not robust against variations and fluctuations in the data could result in enhancements or degradations of the signal. Figure 5.21 shows how some biases can even be seen by eye to compare how some functions can mis-model what is observed in data. A power law function can be used for the background model of an event class which results in the given signal yield. Randomly cutting up the total dataset in fine ranges of the ϕ of each photon (which is uniform so corresponds to randomly dividing the data) gives categories with much larger statistical variation in data points as shown in Figure 5.21. Now it is not clear by eye whether the function could be mis-modeling the data, but there is still clear signal in the random category which is entirely a false positive fitting to fluctuations in the background.

Combining all categories for the final signal measure relies on having a robust background model for each category. This motivates the need to quantify the bias and develop a procedure to test a background model that minimizes the bias in the signal for each category.

To quantify the signal bias, the standard pull definition of the signal: $(\mu_{obs} - \mu_{true})/\sigma_{\mu}$ is used. If μ_{obs} is a gaussian random variable distributed with the same mean value as μ_{true} and the same σ, then the distribution of $(\mu_{obs} - \mu_{true})/\sigma_{\mu}$ for large statistics is a gaussian centered at zero with unit width. If the mean value of μ_{obs} deviates significantly from the μ_{true} the pull distribution shifts away from zero in the direction of the deviation. The mean of the pull distribution is the measure of the bias as shown in Figure 5.22b.

A statistical test is devised to generate test statistics to generate a pull distribution. A set of smoothly falling background PDFs is chosen to make these statistical tests which are shown in the equation below. Each function is fit directly to the data, since it is dominantly background, so the background model does not rely on any predictions from simulation. The order of each function is chosen by requiring the fit to the data in a given category to be within a given fit criteria. The family of exponential functions can be added to create higher orders thus creating a fit with more degrees of freedom. The Bernstein family of polynomials increase order with a combinatorial rule for their coefficients. The final family of functions is a function with a fixed power determined by a Laurent series. The ROOFIT algorithm fits each PDF to the data using a negative Log-Likelihood fit where this quantity is minimized. To choose the order of a function, the improvement in fit quality is measured as the difference in negative-log likelihood value of a fit from the nominal order N and the larger value $N+1$ which has more degrees of freedom: $\chi^2_{N\to(N+1)} := 2(NLL_N - NLL_{N+1})$. The $\chi^2_{N\to(N+1)}$ follows a χ^2 distribution which can be used to compute the significance of the change in the fit when increasing the degrees of freedom:$p\left(\chi^2 \geq \chi^2_{N\to(N+1)}\right)$. This significance is required to be below 0.05 for the change to be significant, and indicates the higher order function fits the data much better than the lower order.

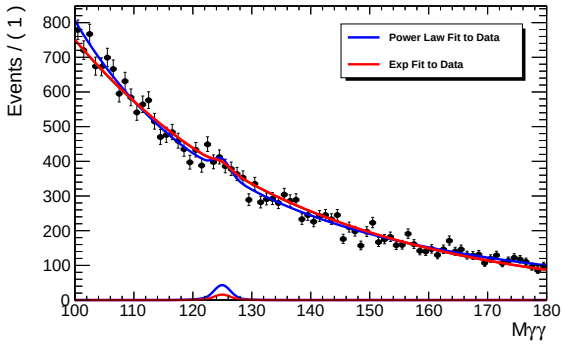
$$
\begin{aligned}
N\text{Exp}(x) &:= \sum_{i=1}^{N} \beta_i e^{\alpha_i x}, \\
N\text{Pow}(x) &:= \sum_{i=1}^{N} \beta_i x^{-\alpha_i}, \\
N\text{Ber}(x) &:= \sum_{i=0}^{N} \beta_i b_{(i,N)}, \text{ with } b_{(i,N)} := \begin{pmatrix} N \\ i \end{pmatrix} x^i (1-x)^{N-i} \text{ and} \\
N\text{Lau}(x) &:= \sum_{i=1}^{N} \beta_i x^{\left(-4+\sum_{j=1}^{i}(-1)^j(j-1)\right)},
\end{aligned}
\tag{5.2}
$$

The truth function is a PDF that can now be used to generate test datasets of $m_{\gamma\gamma}$ that represents a background model with a known functional form. Each test dataset has a Poisson variation of the $m_{\gamma\gamma}$ points about the values specified by the PDF. Only test datasets for the background model are generated. They are fit to a test function which is a composite of a background model from the above set of functions and the signal model for the particular category. For the background model that corresponds to the truth the pull distribution: $(\mu_{obs} - \mu_{true}/\sigma_{\mu})$ will have a peak at zero, since the number of signal events in each dataset is zero by construction. The pull will also have unit width. For a background model that is very different from the truth model the pull will indicate the bias. The most robust background model is required to be $\mu/\sigma_{\mu} < 14\%$. Increasing the degrees of freedom of the fit reduces the bias for the Bernstein polynomial functions as shown in Figure 5.22d. The polynomial set of functions is the one not monotonically decreasing so the inflections points in the fit can model many different variations in test datasets for a large enough order of the Bernstein polynomial.

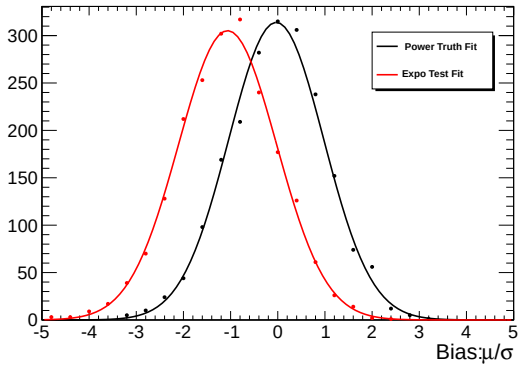
5.8.4 Systematic Uncertainties for MVA Analysis

Table 5.7 lists the full table of the systematics associated with the diphoton selection. As described in Section 5.7, the event classification is done using a diphoton BDT classifier that scores events as signal-like or background-like. The background is modeled in a fully data-driven manner. The systematics due to the background mis-modeling become negligible because the background that gives minimal signal bias is chosen as described in Section 5.8.3. The signal modeling is however described in terms of the Monte Carlo simulation and relevant data$\rightarrow$ Monte-Carlo scale factors are measured and applied. Any uncertainties on these scale factors are propagated to an uncertainty on the diphoton classifier, which then propagates to an uncertainty on the final extracted signal. Most of the efficiency and scale factors are measuring using a $Z \rightarrow ee$ control sample. As described in Section 5.3, the efficiency of the pre-selection applied on top of the trigger is measured using tag and probe with an electron, the uncertainty on this measurement is quite small. However, the difference between electrons and photons becomes more important for other systematics. The two major sources of systematics for the diphoton classification come from two input variables the Photon ID classification BDT and the energy resolution predicted from BDT regression. Also the Z boson differs from a Higgs signal in that the mass peak has a natural width similar to the contribution to the total width due to resolution. The spin-1 Z has different kinematics to the scalar Higgs boson, and some kinematic variables are used to train the diphoton event classifier.

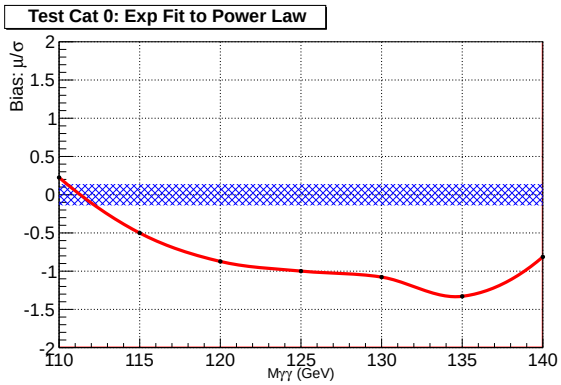
The photon ID shower shape variables have been compared between data and Monte Carlo for electrons in $Z \rightarrow ee$ but photon shower shape variables like R9 need not be the same between



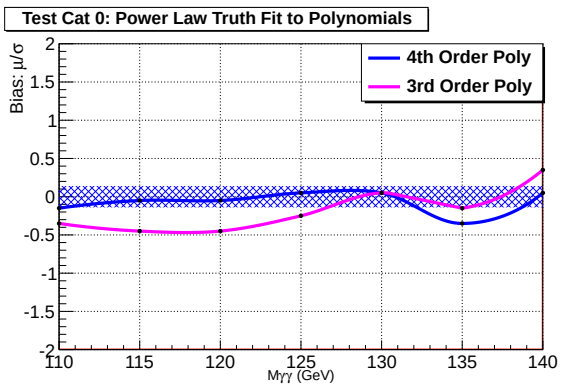
(a) Power Law and Exponential Fit to Data



(b) Exponential Fit Bias at $M_H = 125$ GeV for Power Law data



(c) Fit Bias at Different Mass points



(d) Fit Bias at Different Mass points

Figure 5.22: Example bias test for the exponential function fit to power law truth, and the resulting bias. The Bernstein polynomial functions give tolerable bias in the range of the signal for a large enough order.

electrons and photons, which can have different interactions with the tracker material. The imperfect modeling of the shower shape variables and also imperfect modeling of any correlations between variables leads to discrepancies which must be covered by shape uncertainties. The photon ID discriminator value is continuous in $[-1.0, 1.0]$ with a bulk of the distribution in $[-0.2, 0.4]$. The procedure to propagate the resulting systematic to the diphoton classifier is to take every photon in the signal Monte Carlo and translate it's Photon ID output by ± 0.01 fully correlated across all photons in the sample. Since the diphoton MVA is in general monotonic with the Photon ID MVA output, this monotonic transformation of the ID MVA leads to a maximal possible migration of events in the diphoton MVA output, which is propagated as a migration of the signal yield among the final event classes. The resulting error to due events migrating across diphoton event categories is shown in Table 5.7.

As was described in Section 5.7, the diphoton classification depends on the predicted mass resolution computed from the energy resolution. The energy resolution is compared to data using $Z \to ee$ as described in Section 4.7. The smearing term is summed in quadrature to the energy resolution in order to dilute the predicted energy resolution. This dilutes the expected energy resolution, which is overoptimistic, and scales it so that it better matches it to what is observed with electrons in data. The additional systematic uncertainty from the modeling of σ_E is estimated to be $\pm 10\%$, which gives an 8.1% migration of events across diphoton event classes. Also additional systematics are assigned to mis-modeling of the amount of tracking material in the simulation and an uncertainty in determining the smear (parameterized as a stochastic and constant term).

In addition, there are uncertainties on the predicted cross sections for each Higgs production process and also the uncertainty of the branching fraction. Both are shown in Table **??**.

The mass measurement from the two photons from the Higgs decay also exploits the copious production of the Z boson peak, and relies on reconstruction between electrons and photons being identical. The uncertainty of the mass measurement originates from the uncertainties defined earlier to cover the imperfect Monte Carlo simulation of the differences between electrons and photons, and also the uncertainty on the energy scale set at m_Z to the higher mass of m_H. The systematic errors are evaluated by making comparisons between data and MC of the $Z \to ee$ and $H \to \gamma\gamma$ simulation at a mass of 125GeV. The main effect is the linearity of the ECAL response for the Z peak at 90GeV up to the Higgs mass of 125 GeV. The linearity of the energy response is studied by examining the dependance of the ECAL energy to track momentum ratio for electrons as a function of the E_T of the isolated electrons from the Z. The stability of the ECAL energy response relates to the dielectron invariant mass as a function of $H_T = E_T^1 + E_T^2$, so the E/P is measured in bins of H_T corresponding to different boosts of the Z-boson. The linearity scale is determined by fitting E/P,

Table 5.6: Sources of systematic uncertainty on the signal accounted for in the first analysis of the 8TeV data in [51], and applicable to events in all classes.

Sources of systematic uncertainty		**Uncertainty**	
Per photon – numbers given are for 8TeV dataset (numbers for 7TeV dataset are similar)		**Barrel**	**Endcap**
Energy resolution	$R9 > 0.94$ (low η, high η)	0.05%, 0.10%	0.07%, 0.03%
	$R9 < 0.94$ (low η, high η)	0.05%, 0.09%	0.09%, 0.06%
Energy scale	$R9 > 0.94$ (low η, high η)	0.05%, 0.10%	0.10%, 0.05%
	$R9 < 0.94$ (low η, high η)	0.05%, 0.05%	0.10%, 0.10%
Photon energy scale from	$R9 > 0.94$	0.03%, 0.12%, 0.34%	0.34%
material mismodelling	$R9 < 0.94$	0.04%, 0.06%, 0.22%	0.22%
Photon preselection efficiency		1.0%	2.6%
Photon identification BDT (Effect of up to 4.3% event class migration.)		±0.01 (shape shift)	
Energy resolution estimation (Effect of up to 8.1% event class migration.)		±10% (scaling)	
		7TeV dataset	**8TeV dataset**
Stochastic/constant mixing angle (Barrel only)	$R9 > 0.94$ (low η, high η)	50%, 50%	10%, 14%
	$R9 < 0.94$ (low η, high η)	50%, 50%	10%, 4%
Per event		**7TeV dataset**	**8TeV dataset**
Integrated luminosity		2.2%	2.6%
Vertex finding efficiency		0.2%	0.2%
Trigger efficiency		1.0%	1.0%
Energy scale non-linearity		0.1% (0.2% for Untagged 0)	
Production cross sections		**PDF**	**Scale**
7TeV:	Gluon-gluon fusion	+7.6% -7.1%	+7.1% -7.8%
	Vector boson fusion	+2.5% -2.1%	0.3%
	WH production	2.6%	0.9%
	ZH production	2.7%	2.9%
	ttH production	8.1%	+3.2% -9.3%
8TeV:	Gluon-gluon fusion	+7.5% -6.9%	+7.2% -7.8%
	Vector boson fusion	+2.6% -2.8%	0.2%
	WH production	2.3%	1.0%
	ZH production	2.5%	3.1%
	ttH production	8.1%	+3.8% -9.3%
Branching fraction		+5.0% -4.9%	

Table 5.7: Sources of systematic uncertainty on the signal accounted for in the final analysis of the 7TeV and 8TeV data in [52], and applicable to events in all classes.

Sources of systematic uncertainty		**Uncertainty**	
Per photon		Barrel	Endcap
Photon selection efficiency		0.8%	2.2%
Energy resolution ($\Delta\sigma/E_{MC}$)	$R9 > 0.94$ (low η, high η)	0.23%, 0.72%	0.93%, 0.36%
	$R9 < 0.94$ (low η, high η)	0.25%, 0.60%	0.33%, 0.54%
Energy scale ($(E_{data} - E_{MC})/E_{MC}$)	$R9 > 0.94$ (low η, high η)	0.20%, 0.71%	0.88%, 0.12%
	$R9 < 0.94$ (low η, high η)	0.20%, 0.51%	0.18%, 0.12%
Photon identification MVA		$\pm$0.01 (shape shift)	
(Effect of up to 4.3% event class migration.)			
Photon energy resolution MVA		$\pm$10% (shape scaling)	
(Effect of up to 8.1% event class migration.)			
Per event			
Integrated luminosity		4.4%	
Vertex finding efficiency		0.2%	
Trigger efficiency		1.0%	
Global energy scale		0.47%	
Dijet selection			
Dijet-tagging efficiency	VBF process	10%	
	Gluon-gluon fusion process	30%	
(Effect of up to 15% event migration among dijet classes.)			
Muon selection			
Muon identification efficiency		1.0%	
Electron selection			
Electron identification efficiency		1.0%	
MET selection			
MET cut efficiency	Gluon-gluon fusion	15%	
	Vector boson fusion	15%	
	Associated production with W/Z	4%	
	Associated production with $t\bar{t}$	4%	
Production cross sections		Scale	PDF
Gluon-gluon fusion		+7.6% -8.2%	+7.6% -7.0%
Vector boson fusion		+0.3% -0.8%	+2.6% -2.8%
Associated production with W/Z		+2.1% -1.8%	4.2%
Associated production with $t\bar{t}$		+4.1% -9.4%	8.0%
Scale and PDF uncertainties		(y, p_T)-differential	
(Effect of up to 12.5% event class migration.)			

the relative response, as a linear function of H_T and the deviation of the line from unity gives the residual non-linearity. The uncertainty on the measurement of the Higgs boson mass is extracted by applying the observed non-linearity is applied to the Monte Carlo. This gives the uncertainty on the final Higgs boson mass measurement.

5.8.5 2D Models for Vector Boson Fusion

The previous sections described the one-dimensional PDFs which model a dataset of $m_{\gamma\gamma}$, this section will describe the parameterization of the 2D PDFs for the dijet datasets: $(m_{\gamma\gamma}, m_{jj})$ found in the dijet tag described in Section 5.7.1. There are two dijet categories for the 7TeV dataset, and four dijet categories for the 8TeV dataset. The PDF now can consist of a conditional PDF: $\mathcal{P} = \mathcal{C}\left(m_{\gamma\gamma}|m_{jj}\right) \times \mathcal{D}\left(|m_{jj}\right)$. The correlations between the two variables is included in the first PDF while second only includes the dijet model. The four categories are again combined as a product of probabilities just as in the previous section. For the signal models of gluon fusion and vector boson fusion, the two observables have factorized PDFs: $\mathcal{P} = \mathcal{C}\left(m_{\gamma\gamma}\right) \times \mathcal{D}\left(|m_{jj}\right)$. The correlations for the signal are negligible because the $m_{\gamma\gamma}$ peak is localized across a small range of just a few GeV at most. However, for the background model the correlations will be important across the full range of $m_{\gamma\gamma}$. The one dimensional signal models as described in Section 5.8.2 along with the systematic uncertainties incorporated into the parameterization. The second one dimensional PDF models only the m_{jj} observable which discriminates the gluon-fusion signal process from the vector boson signal process. The signal PDF for vector boson fusion is modeled by a Landau function ($\mathcal{L}$) plus a gaussian function ($\mathcal{G}$) : $V_s^{VBF}\left(m_{jj}|\mu_L, \sigma_L, \mu_g, \sigma_g, f\right) = f\mathcal{G}\left(m_{jj}|\mu_g, \sigma_g\right) + (1-f)\,\mathcal{L}\left(m_{jj}|\mu_L, \sigma_L\right)$This function describes the longer radiative tail of the m_jj for FSR jets from the Vector boson fusion process. For the gluon fusion the signal is modeled with only the Landau so this tail drops more steeply for ISR jets: $\mathcal{I} = \mathcal{L}\left(m_{jj}|\mu_L, \sigma_L\right)$.

The diphoton background fits as for the 1-dimensional case, is fit directly to the data. The data could consist of an admixture of many different processes so it is not clear if the background model will have correlations between the two observables. However, the same statistical tests applied for choosing the 1-Dimensional background models can be applied to the 2-dimensional case. The family of truth functions include the Power law, Polynomial, Exponential family of functions just as for the 1-dimensional case for $m_{\gamma\gamma}$. The family of functions can also be used to make conditional PDFs which explicitly include correlation terms between $m_{\gamma\gamma}$ and m_{jj}. The 0th, 1st, and 2nd order correlations are also included in the $m_{\gamma\gamma}$ PDFs so that the full set of statistical tests not only chooses the most robust background model accounting for correlations between the two observables as well. The correlations by parameterizing in terms of a_i for the ith order correlation . For negligible correlation

the fit will predominantly yield $a_0 \sim 1$ and $a_1 \sim 0$. This simple parameterization works really well for single parameter functions like the Power law where $m_{\gamma\gamma}^{-\alpha} \to m_{\gamma\gamma}^{-\alpha(a_0+a_1 m_{jj})}$. The correlation can be included in polynomial function by multiplying the whole function by the correlation factor e.g: for the 2nd order polynomial with first order correlation: $\left(\beta_0 + \beta_1 m_{\gamma\gamma} + \beta_2 m_{\gamma\gamma}^2\right)(a_0 + a_1 m_{jj})$.

The $m_{\gamma\gamma}$ background shapes are 1st order Power law functions, 1st oder exponential functions, and 2nd, 3rd and 4th order polynomials. Each of these functions is also considered with 1st and 2nd order correlations. The m_{jj} functions considered that form the other part of the product of the total PDF are the sum of a Landau function plus a gaussian (similar form to the signal VBF) and also a Landau function convoluted with a 3rd order polynomial. The two m_{jj} models and together with the fifteen models for $m_{\gamma\gamma}$ gives 30 background tests. These tests are done using the same procedure as defined in Section 5.8.3 with the pull bias. The models chosen are those where the bias is within the 14% criteria. The 7TeV categories do not show any significant bias without including the correlations. Both 7TeV categories give minimal bias using the 2nd order polynomial for the $m_{\gamma\gamma}$ shape and the sum of the Landau and Gaussian functions for m_{jj} for all 7TeV categories. The tests for the first category of the 8TeV data for the highest resolution photon events show that the correlation term becomes necessary. Not including the term can result in signal bias above the required value, so for this category a power law with the first order correlation term is used for the $m_{\gamma\gamma}$. For the three other 8TeV categories the polynomial functions are used with no correlation. The 2nd category is modeled with a 4th order polynomial and the remaining two categories are modeled with 3rd order polynomials. The m_{jj} observable is modeled with the sum of a Landau and a Gaussian for all 8TeV categories. These are the most robust background models based on the bias tests for the dijet tagged events for the 8TeV and 7TeV data.

With this approach instead of including an estimated uncertainty on the VBF measurement due to the contamination from gluon fusion, both signals are modeled and extracted in the dijet tag. The extraction of the signal now depends on the known systematics from the uncertainty of the shape. The background models are fit to data and tested rigorously to minimize the signal bias. The signal shapes are fit to simulation and rely on accurate models to simulate ISR and FSR jets in the gluon-fusion and the vector-boson fusion process respectively.

5.8.6 Systematics for Vector Boson Fusion

This section will describe the specific systematics for the dijet tag for the two methods described in Section 5.7.1. For the MVA approach the uncertainty on the jet and jet plus photon kinematics need to be propagated to the uncertainty in the classifier value to quantify the resulting migration of events across the categories. The categories for the 2D models need to account for the migration

of events across the four categories due to the uncertainty in the predicted energy resolution, but some of the jet uncertainties become uncertainties on the shape of the m_{jj}. The 2D models also apply straight cuts on the kinematic variables described in Section 5.7.1 where the uncertainty on the kinematics gives a systematic for the dijet tagging. There are theoretical uncertainties for the MVA kinematic variables due from the mis-modeling of the underlying event and also the impact of different parton distribution functions on the signal efficiency. The MVA relies on the simulation to also give the prediction of the kinematic variables for the background. The 2D modeling of m_{jj} is fit directly to the data to predict the background so the main theoretical uncertainties are associated to the mis-modelling of the gluon-fusion ISR parton shower. The modeling of m_{jj} in a $t\bar{t}$ control sample can be used to obtain the systematic due to mis-modeling of the ISR jets in the gluon fusion 2D model.

The MVA categorization combines the photon+jet kinematics in one stage and then the diphoton and the dijet discriminator in a combined MVA. Since it relies on the diphoton MVA as input the same uncertainties mentioned for the event classification in Table **??** for the photon identification, photon energy scale, and resolution. The uncertainty on the jet energy scale and the jet energy resolution cause events to migrate between the two VBF MVA categories and the untagged categories. The identification of PU jets also has an uncertainty which is evaluated using a $Z \to \mu\mu$+jets sample with a resulting uncertainty of 2%. The theoretical uncertainty of the kinematic variables is accounted for by evaluating the systematics related to the underlying event. Different simulations of the parton shower from the underlying event are generated for the VBF and gluon-fusion Higgs signal at 125 GeV and the resulting category migration is evaluated. Also different parton distribution functions are used to evaluate the change in the efficiency of the selection.

The 2D model of the m_{jj} and $m_{\gamma\gamma}$ relies on tagging cuts for the signal VBF photons using the VBF kinematic variables and also cuts on the Photon ID variables, which results in a Jet ID efficiency. The four categories based on the mass resolution will have some event migration due to the uncertainty on the energy resolution. The main source of the jet energy scale uncertainty is that the signal models are determined from the run dependent Monte-Carlo but the jet energy scale derived using the non-run dependent. This uncertainty will propagate to an uncertainty on each of the signal shapes for gluon-fusion and vector boson fusion, as additional 'nuisance parameters'. These nuisance parameters introduce the uncertainty directly into the likelihood model of the PDF. If a fit parameter θ is affected by the jet energy scale uncertainty in an interval $[\theta_{min}, \theta_{max}]$ which is modeled by $g(\theta_s)$, then the PDF modifies as $f \to f'(x,\theta) = f(x,\theta) \cdot g(\theta_s)$ for uncorrelated uncertainty. This then becomes a penalty term in the negative log likelihood: $-\log \mathcal{L}_{total} = -\log \mathcal{L} - \log(g_s) = NLL_0 + NLL_{pen}$. For many of the shape systematics the uncertainty is modeled as a gaussian so

the measured uncertainty σ_θ and the penalty term becomes $NLL_P = \log(\sigma_\theta 2\pi) + \frac{1}{2}\left(\frac{\theta-\mu}{\sigma_\theta}\right)^2$. The jet energy scale systematic is included as a correlated systematic so depends on multiple variables and their correlations in a matrix V so the penalty term has a more complicated form: $NLL_{pen} = \log\left(|V|2\pi^{N/2}\right) + \frac{1}{2}\left(\overrightarrow{\theta} - \overrightarrow{\mu}\right)^T V^{-1}\left(\overrightarrow{\theta} - \overrightarrow{\mu}\right)$. The jet energy scale corrections becomes correlated with the underlying event simulations so the two uncertainties give a correlated penalty term.

The uncorrelated uncertainty depends mainly on the Monte Carlo mis-modeling of the parton shower for the signal processes. As described above one strategy is to estimate this systematic from different simulations of the underlying event and evaluating the effect. This then becomes correlated to the Jet Energy scale uncertainty. However, this again relies on simulation to characterize the mis-modeling. A data driven-measure gives the most accurate measure of the penalty term, and can reduce the need for the effect of the above penalty term. The $t\bar{t}$ production proceeds dominantly through gluon fusion at the LHC (85% of the total fraction produced). The simulation of the $t\bar{t}$ leptonic decay process uses the same parton shower modeling as the Higgs gluon fusion and relies on some prediction of the initial state parton shower. This simulation, which is modeled in the same way for the gluon-fusion Higgs, can be compared to $t\bar{t}$ data. The data with the same selection gives the observed spectrum of m_{jj} from ISR jets, and the ratio to the Monte Carlo gives an accurate systematic for the uncertainty in modeling the ISR parton shower.

A pure control sample of $t\bar{t}$ is obtained by requiring fully leptonic decays. The selection of events for the $t\bar{t}$ control sample is:

- The muon is reconstructed as a global muon and also a Particle Flow muon
- $\chi^2/ndof < 10$ for the muon track and there is at least one muon chamber hit in the global muon track fit. Also the muon segments are in at least two muon stations.
- To suppress cosmic muons the track signed impact parameter is less than 2 mm and the ΔZ to the primary vertex is less than 5mm
- There is at least one pixel hit for the muon track
- Cut on number of track layers with more than 5 hits

The above cuts give comparable efficiency of muons to the electron efficiency with the loose Identification below

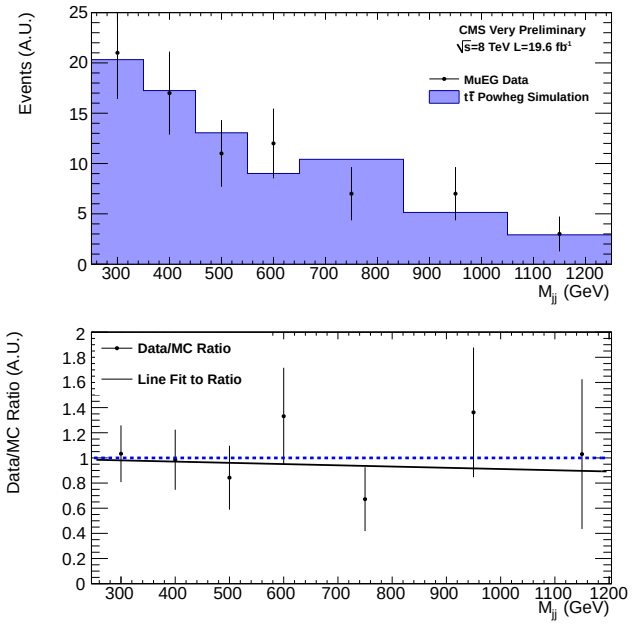
- For EB: $\Delta\eta < 0.007$ and $\Delta\phi < 0.15$ between the supercluster and electron track
- For EB: $\sigma_{i\eta i\eta}$¡0.01 and $H/E < 0.12$

- For EE: $\Delta\eta < 0.009$ and $\Delta\phi < 0.1$ between the supercluster and electron track
- For EE: $\sigma_{i\eta i\eta}$¡0.03 and $H/E < 0.1$
- For EB and EE: $d0 < 0.02$ and $dz < 0.2$ and fabs$1/E - 1/p) < 0.05$
- For EB and EE: PF Isolation in cone of 0.3 divided by Electron pt is less than 0.15
- For EB and EE: Conversion Vertex probability less than $1x10^{-6}$ and no missing hits in the pixel layer

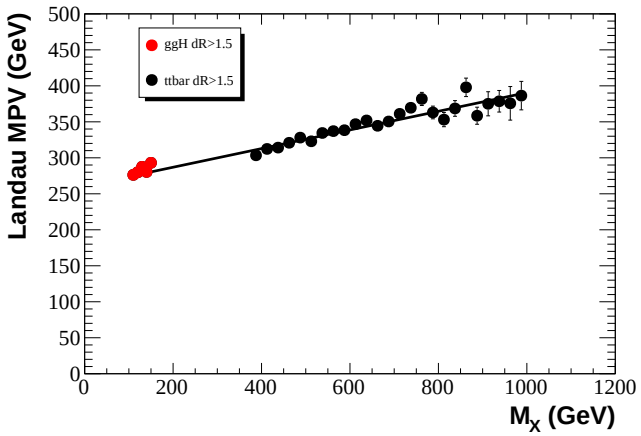
In addition, the ISR jet is required to be isolated from the b-jet to prevent overlap between the ISR and FSR jet (which is tagged as coming from a b-jet) with $\Delta R > 1.5$ between the candidate ISR jet and the bjet. Finally, jet selection requirements are applied to obtain the same ISR jets that pass the dijet tag. The data yield is smaller than the MC but gives roughly linear relation for the data/MC ratio as a function of m_{jj} as shown in Figure 5.23. The line gives the uncertainty of the Monte Carlo mis-modeling as function of m_{jj}:$1.01 + \left(-9.821 \times 10^{-5}\right) m_{jj}$. This can be applied as a correction factor with the shape uncertainty given by the uncertainty of the linear fit. However, the $t\bar{t}$ production still needs to be compared to the gluon fusion Higgs simulation. Though both final states are produced via gluon fusion, the initial state radiation can depend on the center of mass energy of the incoming gluons. The initial state radiation for the Higgs boson can be different than the heavier top anti-top system, since the center of mass energy of the incoming quarks is greater for the top anti-top final state. To gauge this effect, the dijet mass shape of the ISR jets selected in the gluon fusion Higgs is fit with the Landau function for different Higgs masses and the same is done for ranges of the $t\bar{t}$ invariant mass. Figure 5.23 shows the result for the gluon fusion Higgs masses and the $t\bar{t}$ invariant mass. The most probable value of the landau shifts at high invariant mass of the $t\bar{t}$ which is proportional to the center of mass of the incoming gluons. This has to be accounted for when computing the uncertainty for the gluon fusion process for Higgs which matches events with smaller center-of mass energy of the incoming gluons.

5.9 Results

This section will show two sets of results corresponding to two analyses. Section 5.9.1 corresponds to the first results from 2013, following the discovery period where a new "Higgs-like" boson was announced at CERN on July 4th 2012. Many of the analysis techniques described were applied and studying using the updated 8TeV dataset with the best available ECAL calibrations at the time and combined with the 7TeV results analyzed in 2011. Section 5.9.2 shows the latest results for the full



(a) Data/MC ratio



(b) MPV of the Landau fit versus invariant Mass of the final state

Figure 5.23: The data monte-carlo comparison of ISR jets in $t\bar{t}$ leptonic decays. Also the comparison of the MPV of the Landau shape for gluon fusion Higgs (red) at different masses and the $t\bar{t}$ invariant mass (black)

LHC Run 1 data with all the data calibrated with the latest ECAL calibrations and the 7TeV and 8TeV datasets analyzed using the same procedure.

5.9.1 Results 2013

These set of results from [51] describe the first major analysis of the bulk of 8TeV data taken in 2012. The results consist of data taken in the Run: A,B,C,D periods as described in Section 5.2, but the calibration of the first two run periods is done differently than the last two run periods which contain the majority of the data. These results include a dijet-tag using only explicit cuts on the kinematic variables described in Section 5.7.1, and the exclusive tags for VH are included for the first time. The diphoton MVA is trained using Monte Carlo with no run dependent information, and the energy reconstruction and energy resolution estimation is trained in tandem as described in Section 5.6, which is the same procedure also used for the July 2012 Higgs discovery. The conversion pointing described in Section 5.4 was used for the first time analyzing the 8TeV datasets with the updated single-leg conversion reconstruction.

Figure 5.24 shows the 95% confidence limit in [51] which excludes at 95% CL Higgs masses within the range of 110 to 123 GeV and 128 to 149 GeV, leaving a narrow mass range for observed signal within 123 to 128GeV. Figure 5.25 shows the combination of signal models and background models weighting the data in every category by the signal to background ratio. The signal is given by the expected number of signal events in each category and the background is estimated from the normalization of the likelihood fit to the data under the hypothesis that there is no signal. The p-value of the excess at 125 GeV is 3.2σ.

Figure 5.26 shows the best fit value of $\mu_{obs} = 0.78^{+28}_{-26}$times the Standard Model rate for the signal hypothesis of $M_H = 125GeV$. Figure 5.27 also shows the compatibility of the μ_{obs} for all combined categories with the measured μ in each of the 14 categories. For the 7TeV dataset, mainly the best untagged (gluon-fusion) category deviates from the combined μ toward larger values. The same is true for the dijet tagged category using just a cut-based tag. For the 8TeV, all categories are compatible with the μ_{obs} value, except the best untagged (gluon-fusion) category which is pulled toward larger μ values.

Figure 5.28 shows the one-dimensional negative log likelihood scan of Higgs mass with the best fit value of $M_H = 125.4 \pm 0.5 \pm 0.6$GeV. The scan is done for only the statistical uncertainty and the combination of statistical and systematic uncertainties described in Section 5.8.4. The horizontal lines show the region of 1σ and 68% probable range of M_H around the best fit value and also the 2σ of the 95% probable range. The 2D likelihood scan of hypothesis μ and M_H gives a contour around the best fit value of both μ_{obs} and M_H and shows the 2D region of 68% and 95% probability of

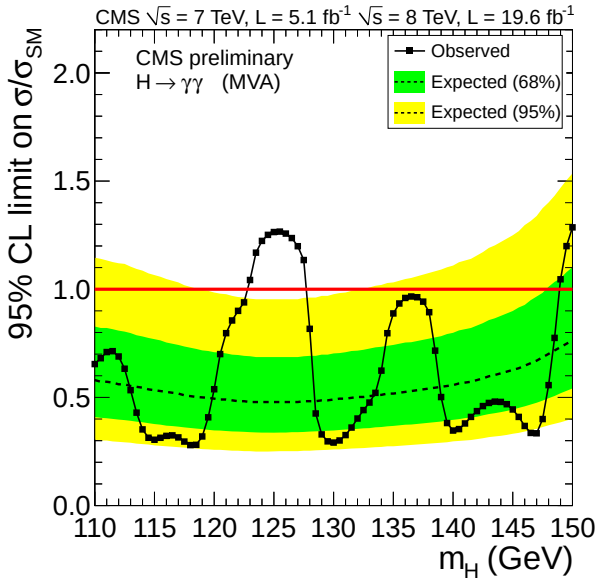
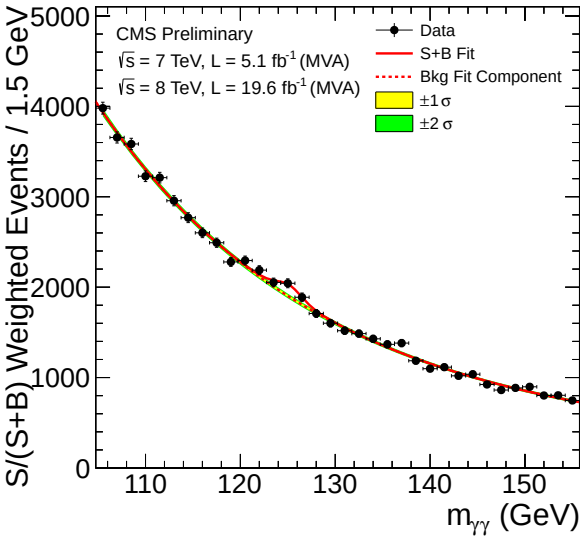
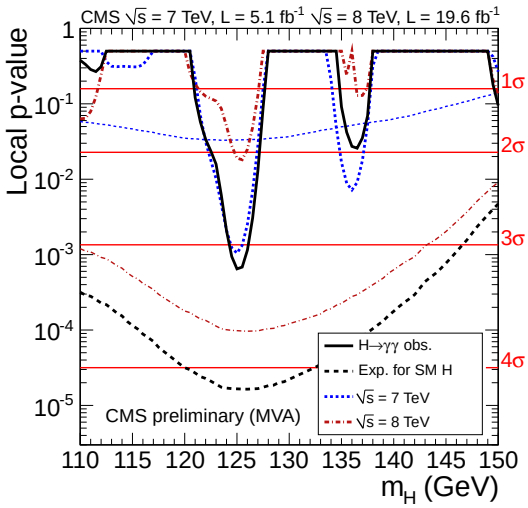


Figure 5.24: Observed and median expected (95% CL) limits on the cross-section ratio to the expected Standard model Higgs cross-section (μ) for hypothesis Higgs Masses using the Asymptotic CLs Approximation



(a) Combined Signal And Background fits weighted by $S/(S+B)$



(b) Observed p-value

Figure 5.25: The combined diphoton mass spectrum across all categories (7TeV and 8TeV) combined and weighted by $S/(S+B)$. The observed p-value of the excess at $M_H = 125GeV$

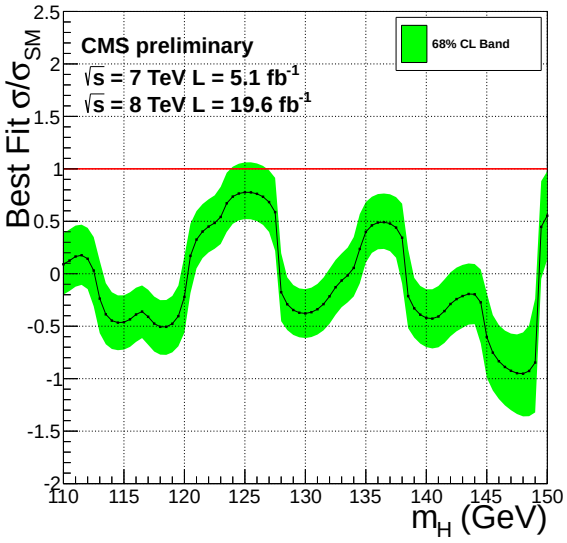


Figure 5.26: Best likelihood fit values for μ_{obs} for different hypotheses Higgs Masses.

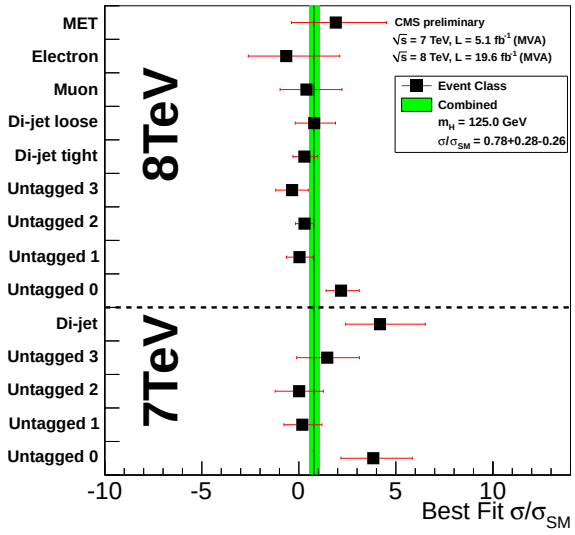
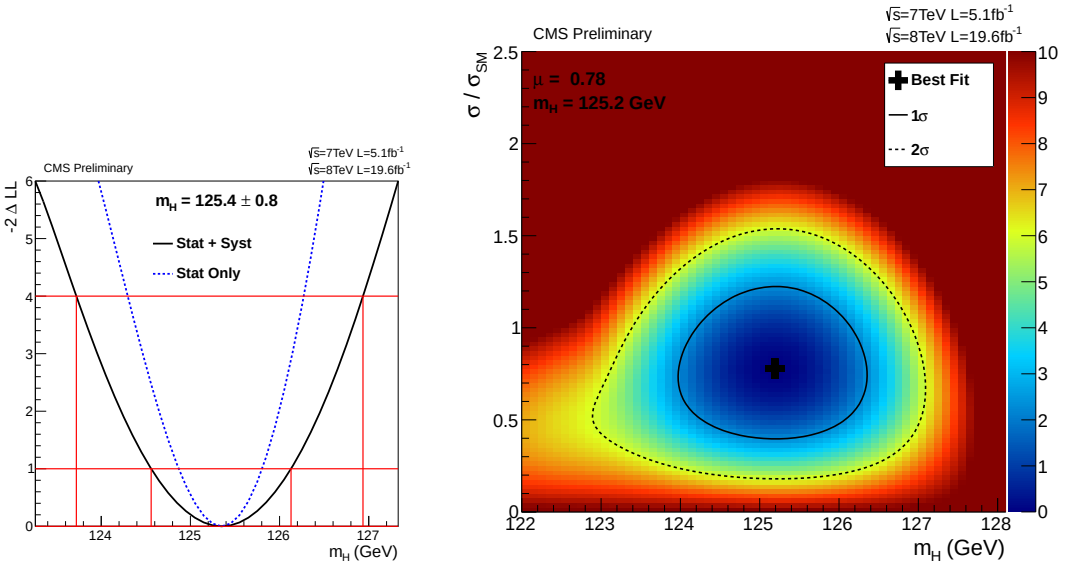


Figure 5.27: Compatibility of best-fit value of μ in each category compared to the combined μ_{obs}



(a) Best fit Higgs Mass

(b) 2D Negative Log Likelihood of M_H and μ

Figure 5.28: 1D Likelihood scan of hypothesis Higgs masses and the 2D Likelihood scan of both μ and the Higgs masses

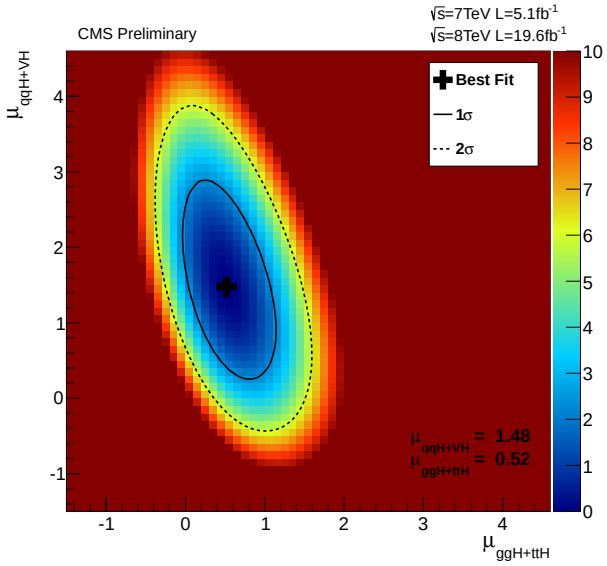


Figure 5.29: 2D negative log likelihood scan of the measure of the Higgs coupling to fermions μ_F and the Higgs coupling to vector bosons μ_V

signal strength and Higgs mass.

The four main Higgs production mechanisms are associated to production via top-quarks or via vector bosons. The Higgs coupling to fermions can be given from the gluon fusion and ttH production categories as μ_F and the categories targeting associated vector boson production and vector boson fusion gives μ_V. The two values are scanned in a 2D likelihood scan to find the value of (μ_F, μ_V) most compatible with the observed data. The best fit value of the measure of the Higgs coupling to fermions, which is predicted by the Standard Model to be proportional to the top mass, is $\mu_F = 0.52$ times the Standard model cross-section. The best fit value of the Higgs coupling to vector bosons, which is predicted by the Standard Model to be proportional to M_V^2, is $\mu_V = 1.48$. Figure 5.29 shows the 2D likelihood scan regions around the best fit values. The standard Model value of $(\mu_F, \mu_V) = (1, 1)$ is within the 68% confidence region of the scan.

This section summarizes the results in [51] for the first analysis of the total 8TeV and 7TeV datasets. The next section will give the most current results of $H \to \gamma\gamma$ with the most current ECAL calibrations and the most current optimizations as described in [52].

5.9.2 Legacy Run 1 Results

The results presented here correspond to the full LHC Run 1 photon dataset. The Run 1 dataset consists of the majority of data $19.7 fb^{-1}$ taken at 8TeV collected in 2012 and 5.1 fb^{-1} at 7TeV collected 2011. These results correspond to the first consistent treatment of the 2 datasets of 2011 and 2012. The results are for the latest re-training of all the MVAs described. A modified energy regression procedure is used to simultaneously give the energy correction and also the energy resolution. They are trained with the latest simulation of the 7TeV and 8TeV taking into account run dependent effects. The background samples for training the event classifier diphoton MVA have an improved description of the photon kinematics which are input to the classifier. The latest energy calibrations are applied for the 2011 and 2012 datasets to give a more precise reconstructed energy which is input to the training of the energy corrections and energy resolution estimates. The set of exclusive categories which target production mechanisms with additional objects in the final state is greatly expanded. The production of $t\bar{t}$H is included and disambiguated from the associated vector boson production. The vector boson fusion tag is modified from the cut-based approach on kinematic variables to and MVA approach or alternatively 2D modeling of $m_{\gamma\gamma}$ and m_{jj}. The final set of fits from all 25 categories in the 2011 and 2012 datasets are statistically combined to give the final results.

Figure 5.30 shows the 95% confidence limit for the standard model Higgs boson decaying to photons for different values of the Higgs boson mass. A majority of the mass range of possible Higgs

boson masses is excluded at 95% confidence with a clear excess of events in a window around 125 GeV: $[122, 126.5]$. To examine the the full signal excess in the selected data the diphoton invariant mass distribution for all event classes is combined together with the sum of the background model fits for each category as shown in Figure 5.31. In order to account for the signal-to-background ratio of each category, the categories are combined by weighting the signal model, the background model and the data based on the number of signal events in the expected signal window. To quantify, how probable the observed excess of events is above the background fluctuations the combined p-value is computed for the best-fit Higgs mass of 124.7GeV with a statistical significance of 5.65σ shown in Figure 5.31. Figure 5.32 shows the best fit negative log likelihood value of $M_H = 124.726 \pm 0.3 \pm 0.15$GeV and compared to the previous section the 68% confidence interval is within a very narrow interval.

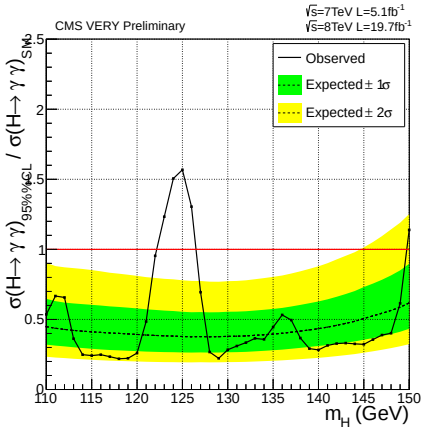
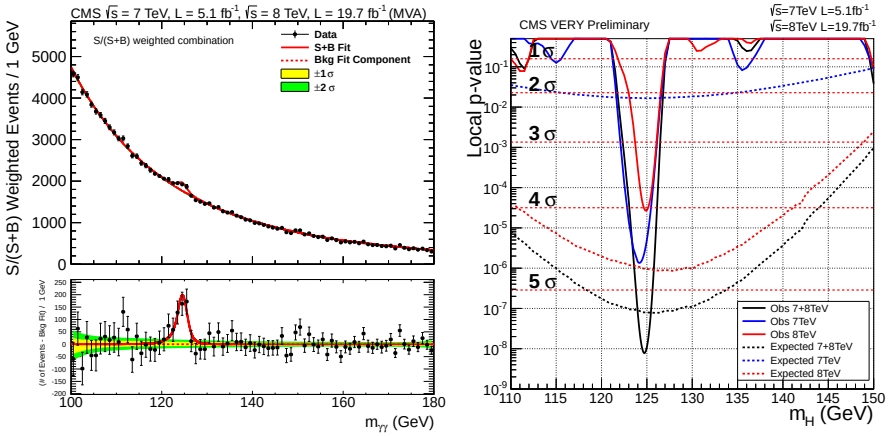


Figure 5.30: Observed and median expected (95% CL) limits on the cross-section ratio to the expected Standard model Higgs cross-section (μ) for hypothesis Higgs Masses using the Asymptotic CLs Approximation for the most current Higgs analysis.

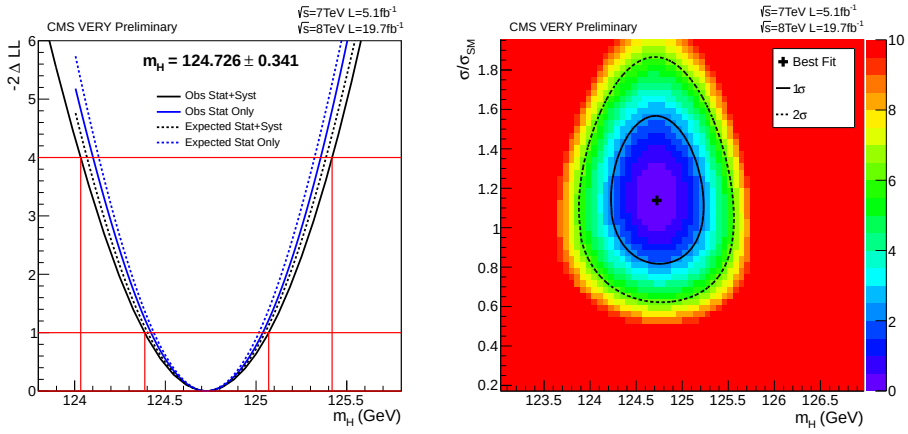
The production of the Higgs is split into two based on the production via top quarks with mainly gluon fusion and production from vector bosons mainly via vector boson fusion. Using the 2D statistical models of m_{jj} and $m_{\gamma\gamma}$, Figure 5.33 shows the 1D likelihood scans of the fermionic signal strength while profiling the bosonic one to give $\mu_F = 0.98^{+0.37}_{-0.34}$. Figure 5.33 also shows the 1D likelihood scans of the vector boson signal strength while profiling the fermionic one to give $\mu_V = 1.18^{+0.70}_{-0.63}$.



(a) Legacy: Weighted Signal+Background Fit

(b) Legacy: Observed p-value

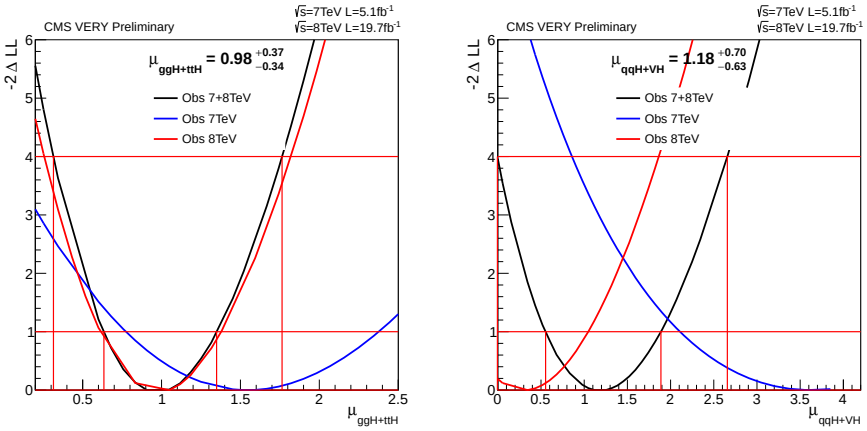
Figure 5.31: The combined diphoton mass spectrum across all categories (7TeV and 8TeV) combined and weighted by $S/(S+B)$. The observed p-value of the excess at $M_H = 124.7GeV$



(a) Legacy: Best fit Higgs Mass

(b) Legacy: 2D Negative Log Likelihood of M_H and μ

Figure 5.32: Most current 1D Likelihood scan of hypothesis Higgs masses and the 2D Likelihood scan of both μ and the Higgs masses



(a) Measure of the fermionic coupling using the 2D Fit

(b) Measure of the vector boson coupling using the 2D Fit

Figure 5.33: 1D negative log likelihood scan of the measure of the Higgs coupling to fermions μ_F and also the Higgs coupling to vector bosons μ_V using the 2D PDF models

Chapter 6
Conclusion

Presented in this thesis, is the observation of a new boson which is confirmed to be the Standard Model Higgs boson. A major contribution to this discovery came from the high signal-to-background ratio of the $H \to \gamma\gamma$ decay channel, where the sensitivity is optimized by including multivariate techniques. Firstly, the reconstructed photon energy is greatly improved by using multivariate regression energy correction based on the simulation. This gives optimized energy measurements of reconstructed photons to correct for energy loss in the tracker, energy loss in gaps and cracks in the ECAL, and also energy contamination from pileup. The same technique is used in the final classification to divide diphotons according to their reconstructed mass resolution. The MVA classifier technique was used to improve the vertex selection, and also quantify the probability of choosing the wrong vertex. The Particle Flow reconstruction and the photon shower shape variables are combined into an MVA classifier to remove the bulk of the reducible background from jets. Finally, all available indicators of signal-like events are combined into a diphoton BDT. The signal and background are modeled using one-dimensional signal models in $m_{\gamma\gamma}$ and also two-dimensional models of $m_{\gamma\gamma}$, m_{jj} to seperate the gluon-fusion and vector boson signal processes in the dijet category. The Higgs boson is observed as an excess of events in the diphoton invariant mass spectrum at $124.73^{+0.34}_{-0.34}$ GeV with a statistical significance of 5.65σ. The observed cross section is consistant with the one predicted by the Standard Model prediction: $\mu = \sigma_{obs}/\sigma_{SM} = 1.14^{+27}_{-0.23}$. Also measuring the Higgs production rate seperating based on different production mechanisms, gives the Higgs production via fermions $\mu_F = \sigma_{ggH+ttH}/\sigma_{SM} = 0.98^{+0.37}_{-0.34}$ and $\mu_V = \sigma_{VBF+VH}/\sigma_{SM} = 1.18^{+0.70}_{-0.63}$ via the vector bosons. Both are consistent with the Standard model predications of the cross-section.

The measurements in the $H \to \gamma\gamma$ decay channel are combined with four other Higgs decay channels at CMS:$H \to ZZ \to 4\ell$, $H \to WW \to 2\ell 2\nu$, $H \to \tau\tau$, $H \to b\bar{b}$. This gives the largest selected dataset give the combined $\mu = \sigma_{obs}/\sigma_{SM}$ of the Standard Model compatibility of the total Higgs production (fixing the Standard Model branching fractions for each decay). The current combined measure of the Standard Model compatibility of the cross-section is $\mu_{comb} = 0.80 \pm 0.14$ The two channels with the largest mass resolution $H \to \gamma\gamma$ and $H \to ZZ$ are combined to give the

final measure of the Higgs mass. Each decay channel is also categorized to target the four production mechanisms: ggH, $t\bar{t}H$, VH, VBF. This gives a full array of production and decay processes where the event yield for a given production and a particular decay is given by: $(\sigma \cdot BR)(x \to H \to ff) = \frac{\sigma_x \cdot \Gamma_{ff}}{\Gamma_{total}}$. [21] This allows for the coupling scale to be measured κ which gives the deviation from the Standard Model prediction of a Higgs coupling to a particle which can be measured across decay modes and production modes simultaneously. A 2D likelihood Scan of κ_F, a common scale factor for the Higgs coupling to fermions, with κ_V a common scale factor for the Higgs coupling to vector bosons, excludes the pure fermiophobic Higgs hypothesis at 95% confidence.

Another key confirmation that the new boson is the Higgs boson is to test its compatibility with the spin 0^+ hypothesis for a scalar particle (instead of a pseudo-scalar with negative parity 0^-). Since the potential Higgs boson is observed decaying to photons, the spin 1 hypothesis is excluded. The kinematics of the four body decay: $H \to ZZ \to 4\ell$ provides a large number of discriminating variables to show that the pseudo-scalar hypothesis is highly disfavored. A major alternative hypothesis to test is the spin 2^+_m hypothesis of a gravition-like particle produced from either gluon fusion or $q\bar{q}$ annihilation. Using $H \to \gamma\gamma$ with a neutral final state is possible to create a kinematic variable based on the diphoton scattering angle in a particular rest frame that discriminates a scalar particle with spin 0^+ and a graviton-like particle with spin 2^+_m. This scattering angle denoted $\cos\theta^*$ is defined as $\cos\theta^* = 2 \times \frac{E_2 p_{z1} - E_1 p_{z2}}{m_{\gamma\gamma}\sqrt{m^2_{\gamma\gamma} + p^2_{T\gamma\gamma}}}$ where the energy E and the logitidunal momentum p_Z is given for the leading and trailing photon denoted 1 and 2 respectively. This variable is isotropic for a spin 0^+ scaler particle and this is not the case for spin 2^+_m particles. The discrimination is made by throwing test statistics from a log likelihood ratio of the two spin hypotheses. The 2^+_m final state is added with an admixutre of gluon fusion production and $q\bar{q}$ annihilation production. This can be combined with the angular measurements of $H \to ZZ \to 4\ell$ and $H \to WW \to 2\ell 2\nu$ to show that the 2^+_m is disfavored for all admixtures. The $H \to \gamma\gamma$ spin hypothesis test disfavors the 2^+_m model produced purely from gluon fusion and also the 2^+_m model purely from $q\bar{q}$.

The discovery of the Higgs boson, as with the discovery of any new particle, changes the landscape of particle physics. [16] The year 2012 is historical as the year the hypothesized particle more than 40 years earlier was discovered in one of the world's largest particle physics experiments. Within the current experimental evidence and also theoretical uncertainty, all measurements of the scale of the coupling of the discovered boson to W and Z bosons, tau leptons, top quarks and bottom quarks are consistent with the expected couplings of the Standard Model Higgs boson. Also the hypothesis that the new boson is a scalar has been tested against an alternative spin-parity hypotheses with the available decay channels. [16] The hypothesis of the fermiophobic Higgs has also been excluded, and

will become conclusive with a significant observation of the Higgs decaying to fermions. This gives a picture of the current landscape for the Higgs. One could have considered alternate symmetry breaking mechanisms to explain the mass of the W and Z bosons, like a technicolor theory with a new gauge interaction similar to QCD. The coupling of techniquarks is hypothesized to become strong at an energy scale of $\Lambda_{TC} \approx 250GeV$ at the Electroweak Symmetry breaking scale. When the chiral symmetry of the techniquarks is spontaneously broken the longitudinal components of the Goldstone bosons become the observed W and Z massive bosons. Instead, the Higgs field, which is the only fundamental scalar field in the Standard Model gives the mechanism for symmetry breaking with its non-zero vacuum expectation value. The future LHC goals will require high precision measurements of the Standard Model Higgs boson to answer several open questions. The most important measurement after the discovery of the Higgs boson is the measurement of the Higgs potential itself. [16] The mass of the Higgs and the vacuum expectation value $v = \left(G_F\sqrt{2}\right)^{-1/2} = 246GeV$ define the Higgs potential and will be measured with greater precision. Multiple Higgs boson production is possible with larger datasets of $3000fb^{-1}$ at the LHC with promising final states that give a precise measure of M_H or have a large branching fraction like $bb\gamma\gamma$ with smaller experimental backgrounds. The production of multiple Higgs bosons can be enhanced in Beyond Standard model scenarios as well. Another key open question is the mass of the Higgs itself and its self-couplings, which would give a much more massive Higgs boson than observed. Quantum corrections from loops of virtual particles must give large corrections to bring the Higgs mass to the observable value of 124.7 GeV. These virtual particles could contain new particles that have been unobserved at present that come from a theory of supersymmetry. [16]

The discovery of the Higgs boson also provides the opportunity to look for physics beyond the Standard Model. Any deviations of the Higgs couplings to particles can indicate physics beyond the Standard Model. The observed Higgs boson can be a fundamental scalar or composite of scalar particles leading to an extended Electroweak symmetry breaking sector. The mixing between the fundamental scalar particles forming the Standard model-like Higgs would give deviations from the predicted Standard model couplings. New states can also be investigated in the Standard Model decay channels or in new channels where a heavy Standard-Model-like Higgs boson decays to lighter scalar particles. The search for an extended Electroweak symmetry breaking sector is constrained by the currently measured coupling scales which is dominated by the most sensitive decay channels. With additional integrated luminosity, the more poorly constrained couplings will further constrain these models and guide the search for an extended Electroweak symmetry breaking sector. This gives an example of a post-discovery phase of the Higgs search, where the Standard Model Higgs measurements and potential deviations can guide the search for beyond Standard Model phenomena

like additional scalar particles. [27]

As was discussed in the introduction of this thesis, the universe is still the answer and is still full of questions. When particle theorist John Ellis was asked about his work by former Prime Minister Margaret Thatcher, he explained that he told experimentalists what to look for and hoped they didn't find it. He explained that not finding something resulted in learning and guided future theories. In this thesis, quite the opposite was described a theory proposed nearly 50 years ago was proven through the existence of a Higgs boson and has resulted in knowledge that will guide the exploration of future frontiers along the landscape of particle physics.

Bibliography

[1] Rohini Godbole Abdelhak Djouadi. *Physics at the Large Hadron Collider*, chapter Electroweak Symmetry Breaking at the LHC. Indian National Science Academy, 2009.

[2] A.N. Annenkov, E. Auffray, A.E. Borisevich, G.Yu. Drobychev, A.A. Fedorov, A. Inyakin, M.V. Korzhik, A.R. Lopatik, P. Lecoq, J.-P. Mendiburu, P. Nedelec, J.P. Peigneux, P. Rebecchi, and D. Sillou. On the mechanism of radiation damage of optical transmission in lead tungstate crystal. *physica status solidi (a)*, 191(1):277–290, 2002.

[3] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Microscopic theory of superconductivity. *Phys. Rev.*, 106:162–164, Apr 1957.

[4] Vernon D. Barger and Roger J.N. Phillips. *Collider Physics*. Addison Wesley Publishing, 1997.

[5] J.D. Bjorken. Asymptotic Sum Rules at Infinite Momentum. *Physical Review*, 179:1547–1553, March 1969.

[6] Oliver Bruning, Vinod Chohan, and Swapan Chattopadhyay. *Physics at the Large Hadron Collider*, chapter Accelerator Physics and Technical Challenges of the LHC. Indian National Science Academy, 2009.

[7] Oliver Sim Bruning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN, Geneva, 2004.

[8] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 0804:063, 2008.

[9] Robert N. Cahn, Stephen D. Ellis, Ronald Kleiss, and W. J. Stirling. Transverse-momentum signatures for heavy higgs bosons. *Phys. Rev. D*, 35:1626–1632, Mar 1987.

[10] Nicolo Cartiglia. Measurement of the proton-proton total, elastic, inelastic and diffractive cross sections at 2, 7, 8 and 57 TeV. *arXiv*, 2013.

[11] CMS. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[12] The CMS and ATLAS Collaborations. Procedure for the LHC Higgs boson search combination in summer 2011. Technical Report ATL-PHYS-PUB-2011-011, CERN, Geneva, Aug 2011.

[13] CMS Collaboration. CMS at the High-Energy Frontier. Contribution to the Update of the European Strategy for Particle Physics. Technical Report CMS-NOTE-2012-006. CERN-CMS-NOTE-2012-006, CERN, Geneva, Oct 2012.

[14] CMS Collaboration. Search for the fermiophobic model Higgs boson decaying into two photons. Technical Report CMS-PAS-HIG-12-002, CERN, Geneva, 2012.

[15] CMS Collaboration. Pileup Jet Identification. Technical Report CMS-PAS-JME-13-005, CERN, Geneva, 2013.

[16] CMS Collaboration. Projected Performance of an Upgraded CMS Detector at the LHC and HL-LHC: Contribution to the Snowmass Process. *arXiv*, 2013.

[17] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006.

[18] CMS Collaboration. CMS Physics: Technical Design Report Volume 2: Physics Performance. *J. Phys. G*, 34(CERN-LHCC-2006-021. CMS-TDR-8-2):995–1579. 669 p, 2007. revised version submitted on 2006-09-22 17:44:47.

[19] CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009.

[20] CMS Collaboration. Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. *JINST*, 6:P11002, 2011.

[21] CMS Collaboration. Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV. Technical Report CMS-PAS-HIG-13-005, CERN, Geneva, 2013.

[22] CMS Collaboration. Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s}$= 7 TeV. *J. Instrum.*, 8(arXiv:1306.2016. CMS-EGM-11-001. CERN-PH-EP-2013-097):P09009. 51 p, Jun 2013.

[23] CMS Collabroration. Commissioning of the Particle-flow Event Reconstruction with leptons from J/Psi and W decays at 7TeV. Technical Report CMS-PAS-PFT-10-003, CERN, Geneva, 2010.

[24] CMS Collabroration. Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector. Technical Report CMS-PAS-PFT-10-001, CERN, Geneva, 2010.

[25] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *European Physical Journal C*, 71:1554, February 2011.

[26] Glen Cowan. *"LHC Physics*, chapter Topics in Statistical Data Analysis for HEP, pages 179–231. CRC Press, 2012.

[27] Nathaniel Craig, Jamison Galloway, and Scott Thomas. Searching for Signs of the Second Higgs Doublet. *arXiv*, 2013.

[28] Gunther Dissertori. *Physics at the Large Hadron Collider*, chapter LHC Detectors and Early Physics. Indian National Science Academy, 2009.

[29] Abdelhak Djouadi. The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model. *Phys.Rept.*, 457:1–216, 2008.

[30] ECALDPG2010. Electromagnetic calorimeter calibration with 7 TeV data. Technical Report CMS-PAS-EGM-10-003, CERN, Geneva, 2010.

[31] John Ellis, Mary K. Gaillard, and Dimitri V. Nanopoulos. A Historical Profile of the Higgs Boson. *arXiv*, 2012.

[32] W Erdmann. Vertex reconstruction at the cms experiment. *Journal of Physics: Conference Series*, 110(9):092009, 2008.

[33] Rick Field. Min-Bias and the Underlying Event at the LHC. *arXiv*, 2012.

[34] H. M. Georgi, S. L. Glashow, M. E. Machacek, and D. V. Nanopoulos. Higgs bosons from two-gluon annihilation in proton-proton collisions. *Phys. Rev. Lett.*, 40:692–694, Mar 1978.

[35] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. Broken symmetries. *Phys. Rev.*, 127:965–970, Aug 1962.

[36] Dan Green. Vector boson fusion and quartic boson couplings. *arXiv*, 2003.

[37] W. Heisenberg. Uber den bau der atomkerne. *Zeitschrift fur Physik*, 77(1-2):1–11, 1932.

[38] Konrad Kleinknecht. *Detectors for Particle Radiation*. Cambridge University Press, 1986.

[39] Philippe Lebrun. *"LHC Physics*, chapter The LHC Accelerator: Performance and Technology, pages 179–231. CRC Press, 2012.

[40] Leon Lederman. *The God Particle: If the Universe is the Answer, what is the Question?* Dell Publishing, 1993.

[41] LHC Higgs Cross Section Working Group and Dittmaier, S. and Mariotti, C. and Passarino, G. and Tanaka (Eds.), R. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. *CERN-2013-*, CERN, Geneva, 2013.

[42] Lous Lyons. *Statistics for Nuclear and Particle Physics*. Cambridge Univeristy Press, 1986.

[43] Nancy Marinelli. Track finding and identification of converted photons. Technical Report CMS-NOTE-2006-005, CERN, Geneva, 2006.

[44] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.J.*, C63:189–285, 2009.

[45] Bruce Mellado. *Physics at the Large Hadron Collider*, chapter Prospects of Searches for the Higgs Boson at the LHC. Indian National Science Academy, 2009.

[46] Donald H. Perkins. *Introduction to High Energy Physics*. Addison-Wesley Publishing, 3rd edition, 1987.

[47] Byron P. Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 543, 2005.

[48] P Speckmayer, A Hocker, J Stelzer, and H Voss. The toolkit for multivariate data analysis, tmva 4. *Journal of Physics: Conference Series*, 219(3):032057, 2010.

[49] G. 't Hooft and M. Veltman. Regularization and renormalization of gauge fields. *Nuclear Physics B*, 44(1):189 – 213, 1972.

[50] The Higgs to Gamma $\gamma\gamma$ working group. Search for a standard model higgs boson decaying into two photons employing multivariate methods. *CMS Analysis Note*, CMS-AN-2012/160, 2012.

[51] The $H \to \gamma\gamma$ working group. Updated measurements of the new higgs-like boson at 125 gev in the two photon decay channel. *CMS Analysis Note*, CMS-AN-2013/008, 2013.

[52] The $H \to \gamma\gamma$ working group. Measurements of the higgs boson at 125 gev in the two photon decay channel in the full lhc run 1 data at cms. *CMS Analysis Note*, CMS-AN-2013/253, Expected 2014.

[53] Hermann Weyl. *Symmetry*. Princeton University Press, 1952.

[54] C.N. Yang and R.L. Mills. Conservation of Isotopic Spin and Isotopic Gauge Invariance. *Physical Review*, 96:191–195, October 1954.