Exclusive Search for Higgs Boson to Gamma-Gamma Decay via Vector Boson Fusion Production Mechanism

by

Dylan Sheldon Rankin

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Bachelor of Science in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Author ....
Department of Physics
May 21, 2012

Certified by ....
Markus Klute
Assistant Professor, Department of Physics
Thesis Supervisor

Accepted by ....
Nergis Mavalvala
Senior Thesis Coordinator, Department of Physics

Exclusive Search for Higgs Boson to Gamma-Gamma Decay via Vector Boson Fusion Production Mechanism

by

Dylan Sheldon Rankin

Submitted to the Department of Physics
on May 21, 2012, in partial fulfillment of the
requirements for the degree of
Bachelor of Science in Physics

Abstract

We perform an exclusive search for the Higgs boson to gamma-gamma decay via vector boson fusion. We utilize the characteristic features of vector boson fusion, such as the di-jet $\Delta\eta$ and mass, as well as the di-photon p_T, to search for the Higgs boson to gamma-gamma decay via the vector boson fusion process. The theoretical production cross section limit is analyzed over the accepted possible mass range for the Higgs boson, 120-130 GeV/c^2. We are able to reduce the theoretical production cross section limit to $\sim 6\sigma_{SM}$ in this range by using a boosted decision tree. Comparison to the cut based approach used by the CMS Collaboration shows no improvement in using a BDT as opposed to a cut based approach.

Thesis Supervisor: Markus Klute
Title: Assistant Professor, Department of Physics

Acknowledgments

First and foremost I would like to acknowledge my thesis supervisor, Professor Markus Klute, for guiding me and overseeing my work for the past year. I am very grateful for his advice and comments. The opportunities he has given me in high-energy physics, at the CMS Collaboration, have focused my interests in physics and I am confident that without him my interests in physics could be quite different.

I am also very grateful to Professor Christoph Paus who provided advice and support during my time working for Professor Klute. Max Goncharov has also been invaluable in my work, and also my development as a young particle physicist. The majority of my knowledge of the workings of the CMS detector and particle physics have come from his presentations and teachings.

Despite the arduous nature of the course, I am grateful for the time I spent in MIT Junior Physics Lab class. Under the guidance of my professors Nergis Mavalvala and Ulrich Becker I learned not only a great deal about experimental physics, and data and error analysis, but also about laboratory techniques and technical writing. I am also confident that the rigors of Junior Lab have helped me to become a much more focused and disciplined physicist.

Some of my most valuable experiences at MIT have come outside class or lab. My teammates on the varsity soccer team, Max Stein-Golenbock and Zach Kabelac, have been great friends to me since the first day we arrived on campus our freshman year. My brothers in the ΦΣK fraternity have been a constant source of support in my time here, without which MIT's curriculum would have been insurmountable.

I would like to personally thank Antony Speranza, who has been an integral part of my time at MIT. Antony and I braved the physics curriculum together, and without him I would have had a much more difficult time here. His help and friendship, especially in Junior Lab, are something I am very grateful for.

I would also like to thank my girlfriend, Tess Gannaway, who has been there for me throughout my time at MIT. Her support has been invaluable to me, and without her I would surely have faltered along my MIT journey. For this and so much more,

I am forever grateful.

Finally, I would like to dedicate my thesis to my parents. My mother, Janet Rankin, and father, Brian Sheldon, have done so much for me. Without their support and (sometimes unwanted) guidance I would not be where I am today. I hope that this symbolic gesture serves as a small display of how thankful for them I am and how much I love them.

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

The work presented in this paper is a report of an exclusive search for the Higgs boson created via the vector boson fusion process and decaying into two photons (gamma-gamma decay). We will detail the motivation for such a search, as well as the detector and tools used, and the specifics of the vector boson fusion process and the gamma gamma decay.

1.1 Motivation

In the mid-twentieth century the standard model of particle physics was developed. This theory, which explains the interactions of particles under the influence of the strong, weak, and electromagnetic interactions, has predicted, over the last four decades, the discovery of many different particles [1–3]. Although these particles have all been correctly predicted by the theory, it also predicts one last particle, which is yet to be discovered, called the Higgs boson. This particle is an excitation of a hypothesized field, called the Higgs field. This field's interactions with all particles are thought to give rise to the property of mass; as such, gluons and photons do not interact with the Higgs field, since they are massless.

However, despite the expectation that the particle should be detectable [4–9], the fact remains that the Higgs boson, a key part of the theory, has yet to be discovered. Although the mass of the particle is a free parameter of the standard model, it has been indirectly and directly ruled out of many different mass ranges over the past decade. A direct search at the Large Electron-Positron Collider (LEP) ruled out all masses below 114.4 GeV at 95% confidence level [10]. The Tevatron collider at Fermilab excluded the Higgs boson from the mass range 162-166 GeV at 95% confidence level [11]. Individual measurements also indirectly excluded all masses above 158 GeV at 95% confidence level [12]. Recently, the CMS and ATLAS groups at the LHC have reported excesses of events near 125 GeV [13, 14], and the CMS group has excluded all masses above 127 GeV at 95% confidence, further motivating the search for the Higgs boson in the vicinity of this excess. The range of available masses for the Higgs boson is narrowing, and now stands at ~115-127 GeV. It is one of the pressing issues in particle physics to resolve whether or not the Higgs boson exists. Using the highest-energy particle collider ever developed, the Large Hadron Collider (LHC) at CERN, physicists hope to be able to use new data to draw a conclusion.

1.2 The LHC and the compact muon solenoid detector

The Large Hadron Collider is the most advanced particle accelerator in the world, and is located in Geneva, Switzerland. The LHC is a proton-proton collider, and as it stands now is capable of producing collisions with a center of mass energy of 7 TeV. These collisions, being extremely energetic, produce a spray of different particles, which must be recorded and analyzed by detectors. There are four main detectors at the LHC: A Toroidal LHC Apparatus (ATLAS), Compact Muon Solenoid (CMS), LHC-beauty (LHCb), and A Large Ion Collider Experiment (ALICE). I will focus on the Compact Muon Solenoid (CMS) detector, shown in Fig. 1-1.

The CMS detector is comprised of four main sections: the particle tracker, the electromagnetic calorimeter, the hadronic calorimeter, and the muon chambers. The particle tracker is a silicon tracker, and serves to track particles paths through the

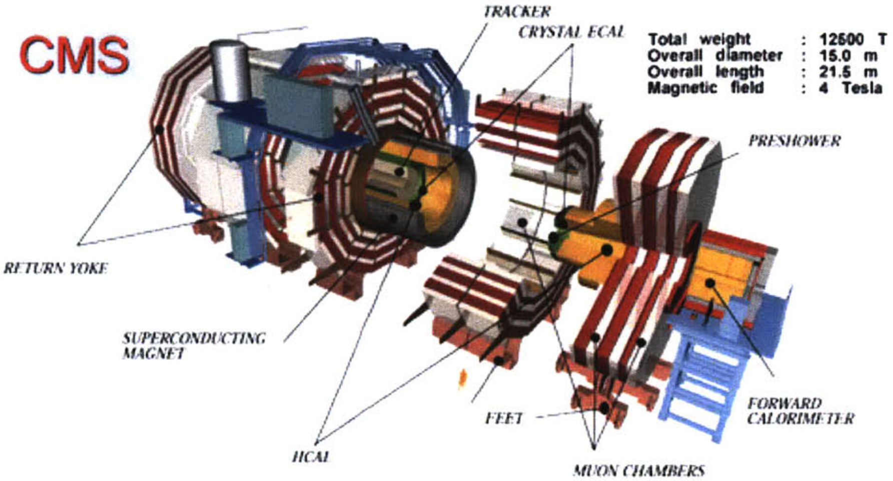

Figure 1-1: 3-dimensional view of the CMS detector and its different components. Certain statistics about the detector are also shown in the upper right corner of the diagram.

tracker's area. This allows one to distinguish between particles which may leave similar signatures in other areas of the detector. For example, both an electron and a photon may deposit energy in the electromagnetic calorimeter, but the photon will not leave a track, as opposed the the electrons track which, due to its charge, will be curved. The electromagnetic calorimeter is designed to capture particles with electromagnetic energy, and to record their energy. The hadronic calorimeter, similarly, is designed to capture and record the energy of any hadrons. Finally, the compact muon solenoid, and its accompanying muon chambers, is designed to confirm the existence of any muons, and to measure their momentum. Because muons are so massive (compared to electrons) they are not easily stopped by electromagnetic fields, and will penetrate through most materials. The CMS detector uses a superconducting solenoid to create a large magnetic field of 4 Tesla and the muon chambers allow the path of the muon to traced and the resulting momentum to be determined. A diagram of a slice of the detector is shown in Fig. 1-2.

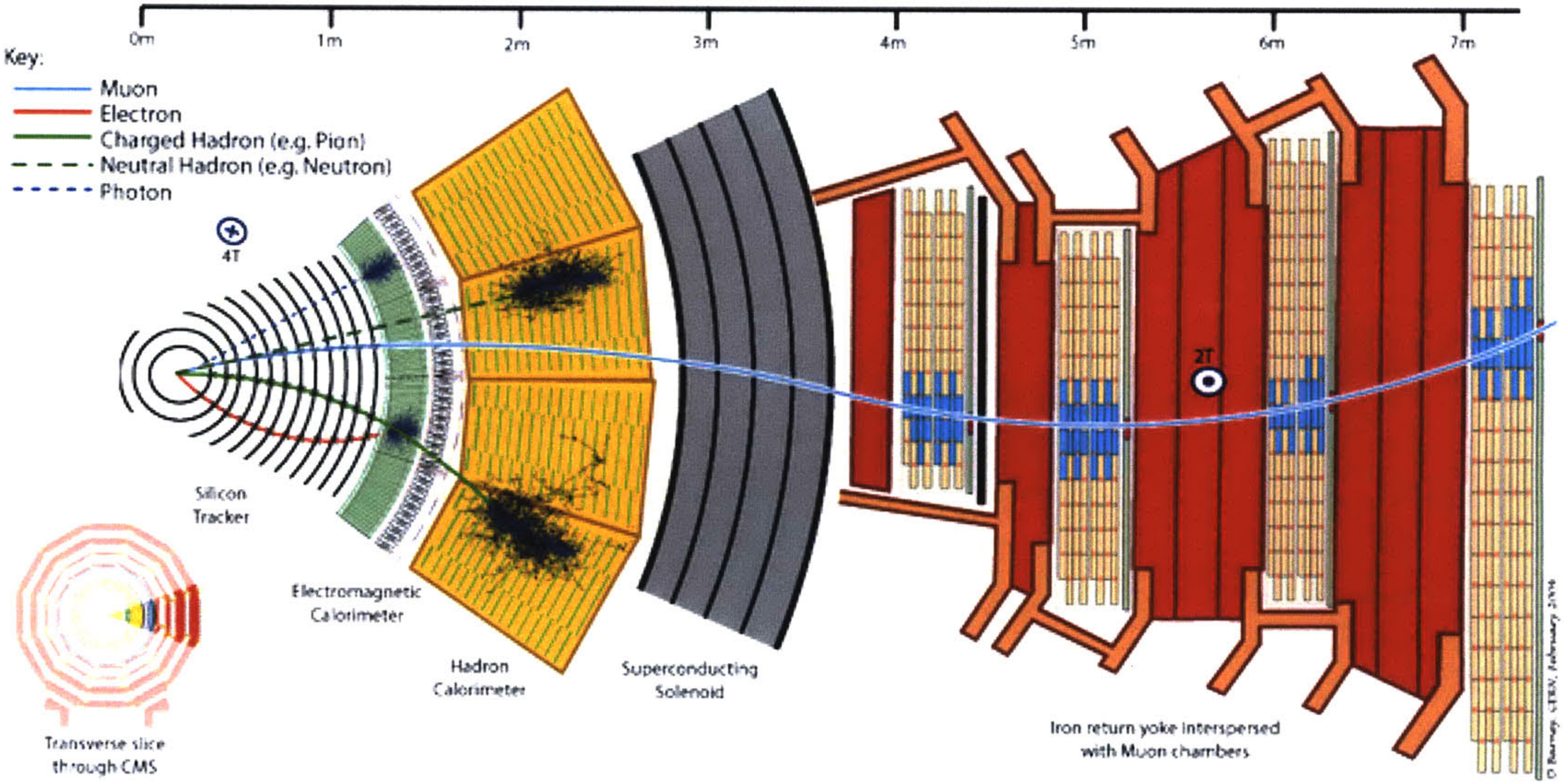
0m
1m
2m
3m
4m
5m
6m
7m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon
4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
Transverse slice through CMS
© Barney, CERN, February 2004

Figure 1-2: Diagram of a slice of the CMS detector, showing different particles paths through the detectors many layers. Of particular interest are the paths of the photon and hadrons.

1.3 Event reconstruction

In the LHC, with so many protons colliding every second, and the products of their collisions decaying and so on, a major difficulty of running the LHC is reconstructing what particles have been created and where. To determine what particles were involved in a collision, final products' paths and intersections must be calculated, and from this their parent particle's energy and momentum calculated. In this manner an event can be reconstructed, and the paths of each of the particles determined. Various algorithms are used to determine how to define a parent particle and its properties, allowing a comprehensive description of the interactions. Every component of the detector is therefore crucial to reconstructing an event and its cascade of particles.

1.4 Production cross section limits

Some particles are produced with such frequency and at such distinct mass ranges that their signals can be easily identifiable. Some particles, like the J/Ψ meson, have such strong signals that they are used for detector calibration. However, the Higgs boson is not such a particle; analysis has shown that if it exists, it is not produced with nearly enough frequency to easily identify its signal. Thus, in addition to attempting to refine the background and signal efficiencies to allow a discovery of the particle, we must also prepare ourselves for the possibility that the particle does not exist. To this end, a technique called limit setting is used. In limit setting, the amount of signal such that the particle would be detectable is determined. This amount is then compared to the predicted amount of signal, based on the standard model. If the required amount of signal drops below this predicted value, and analysis fails to discover the Higgs boson at this mass, then we conclude that the Higgs boson cannot exist at this mass. As the Higgs boson is excluded from certain mass ranges, the search for the Higgs boson can be narrowed to focus on the most promising remaining areas.

1.5 Vector boson fusion process

As an exclusive search, we will only concern our search with one specific production and decay mode of the Higgs boson. Here it is the vector boson fusion production mechanism, and the gamma-gamma decay mechanism that we will focus on. Vector boson fusion is a process by which two vector bosons, the W or Z bosons, fuse together to create the Higgs boson. The W or Z bosons mediate the weak interaction as two quarks pass by each other. Each quark radiates a W or Z boson, which come together to create a Higgs boson. The process is shown using a Feynman diagram in Fig. 1-3. [15]

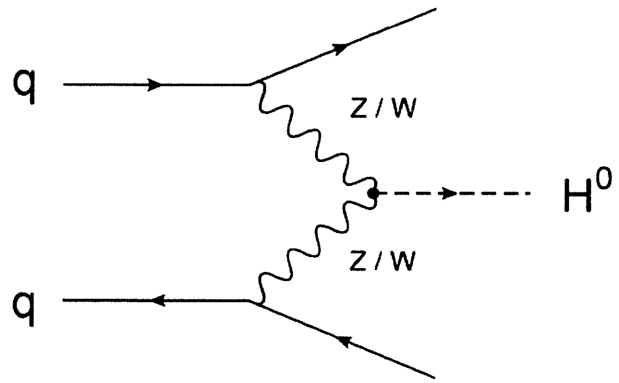

Figure 1-3: Feynman diagram of the vector boson fusion process creating a Higgs boson (H^0). Note that the two quarks are traveling in opposite directions, as they would be in a proton-proton collision at the LHC.

The most defining characteristic of the vector boson fusion process is the large angle between the two quarks involved in the process; this is known as a large rapidity gap of the jets. When protons collide, and two quarks interact, the quarks are deflected slightly as they radiate the vector bosons. However, they are not deflected much, as they are traveling with large momentum. Because they are deflected, however, the detector is able to detect them. The quarks leave a defining narrow cone of radiation due to a process known as hadronization. Since other particles, such as quark pairs or gluons may have caused this narrow cone, the detected particle is simply referred to as jet. The deflection angle is called η, but is not simply measured in degrees of radians with respect to the beam path. Instead, the variable pseudorapidity is used, which is zero if the particle travels perpendicular to the beam path, and infinity if the particle travels along the beam path, with a pseudorapidity of 1 being roughly equal to 40° from the beam path (see Fig. 1-4). [16]

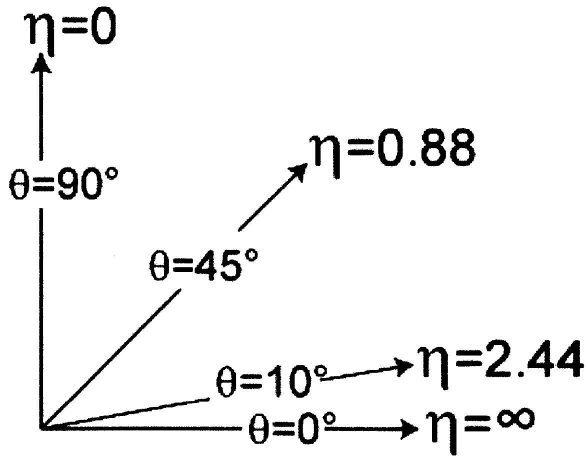

Figure 1-4: Diagram displaying the relation between pseudorapidity and angle from beam.

1.6 Higgs to gamma-gamma decay

Although the Higgs boson can decay in a variety of different processes, we will focus on one specific decay mechanism: gamma-gamma. As the name implies, in Higgs to gamma-gamma decay the Higgs boson decays into two photons. The photons' energies can be measured by the detector, and then reconstructed to calculate the mass of Higgs boson from which they were produced. If the Higgs boson was easily discoverable, a histogram of the reconstructed mass of all two photon events would produce a distinguishable bump at the actual mass of the Higgs boson.

The probability of the Higgs boson to decay into a given final state is called the branching fraction. The branching fraction is a function of the mass of the Higgs, and is shown in Fig. 1-5. Although choosing to focus on the gamma-gamma decay mechanism offers some advantages, a large disadvantage is the low branching fraction for Higgs to gamma-gamma, which is on the order of 0.002, and is close to 0 for many mass ranges. However, the branching fraction is only half of the picture; the background processes are of equal importance. In low mass ranges, the background for all processes producing two photons is quite low, and thus the small branching fraction is made up for by the lack of background (relative to other processes). This makes the Higgs to gamma-gamma channel an ideal channel for the discovery of a low-mass Higgs boson.

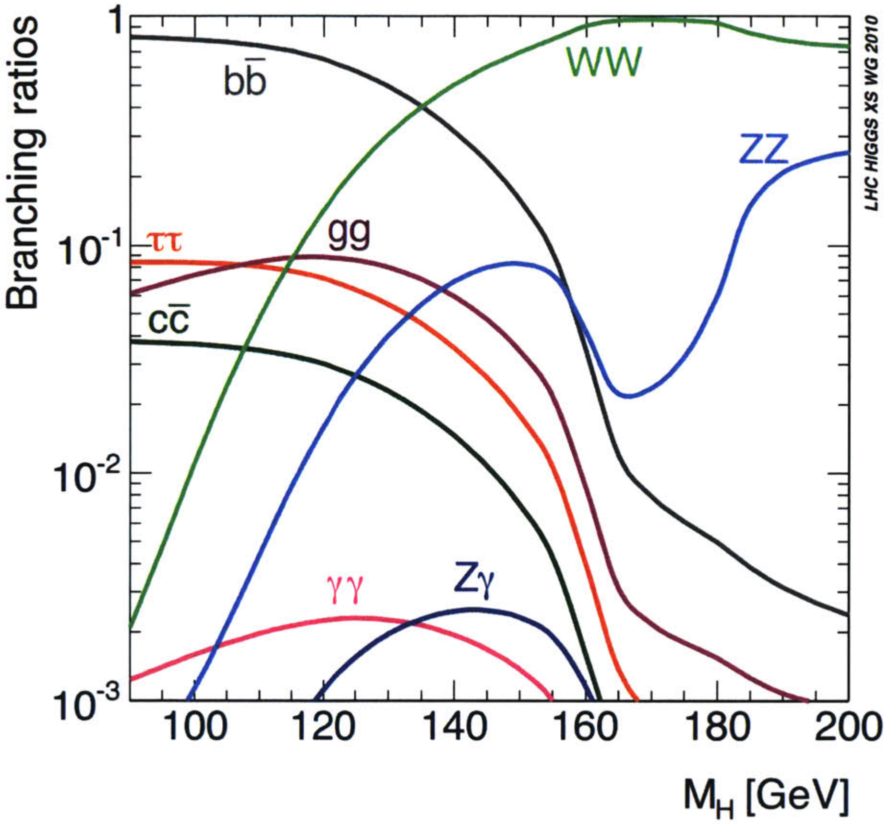

Figure 1-5: Diagram displaying the branching fraction of the Higgs boson for Higgs masses between 100 and 200 GeV. Note the gamma-gamma line in pink, which is maximized around 120 GeV.

Chapter 2

Methods

As mentioned above, if the Higgs boson were easily discoverable, it would be seen in a histogram of reconstructed mass of two photons (or reconstructed mass of other products of the Higgs boson). However, the lack of a distinguishable signal does not actually rule out the existence of the Higgs boson. In order to conclusively exclude the Higgs boson from a certain mass range, a technique called limit setting is used. This technique determines the largest possible signal that would be undetectable, and then compares that with the known amount of signal which should have been received. If it is discovered that the largest possible undetectable signal is larger than the observed amount of signal, then if no signal is observed we can conclude that the Higgs boson does not exist at that mass (or in that mass range).

2.1 Selection criteria

We have used the terms "signal" and "background", but have failed to define them explicitly. Loosely, a signal event is one with a Higgs boson, and a background event is one without a Higgs boson. However, this criteria alone would produce a huge number of background events. In order to perform some preliminary refinement of our overall dataset, we introduce various selection criteria. Of interest here are those selection criteria that concern photons or jets, since our analysis is most concerned with these two particles. The main selection criteria we establish for both particles utilizes mainly transverse momentum and pseudorapidity.

For photons, we restrict our analysis to photons with transverse momentum greater than 20 GeV/c. This restriction causes a much more smooth distribution of background events, since low p_T photons are likely to create much more variability in the reconstructed di-photon mass distribution. By forcing both photons to have a $p_T > 20$ GeV/c, the possible error in the reconstructed mass distribution is minimized. We also restrict our analysis to photons with a maximum pseudorapidity of 2.5. This restriction is a physical one, since the acceptance of the detector for photons has $|\eta| < 2.5$. The calculations of the properties of photons outside this range are not reliable enough, and so the error on all properties is minimized by this restriction. The variable R_9 is a measure of the spread of the deposition of a photon's electromagnetic energy. It is calculated by dividing the energy deposited in one central cell by the energy deposited in the 3x3 cell grid centered on the cell. A restriction of $R_9 > 0.93$ ensures that the photons selected are well-defined, and decreases the occurrences of falsely identified photons.

For jets, similar criteria are used. Restricting our analysis to jets with $p_T > 30$ GeV/c, we are able to remove variability from the reconstructed mass of the jets. We also require $|\eta| < 5.0$. Similar to the photon requirement, this criteria is imposed by physical restriction on our ability to obtain reliable data past this requirement. These selection criteria allow a clear determination of photons and jets and allow us to now develop clear definitions of signal and background. A background event is any event that passes all of the selection criteria, but is not the vector boson fusion process creating a Higgs boson that decays into two photons. A signal event is any event that passes all of the selection criteria and is the vector boson fusion process creating a Higgs boson that decays into two photons.

2.2 Production cross section limits

In order to set a limit on the production cross section for Higgs to gamma-gamma decay, we first use a combination of Monte Carlo simulations and real data to determine what a signal would look like, and what the background would look like in our mass region of interest. POWHEG and PYTHIA are used to model and generate the signal and background events, with real data being found to agree well with the background simulation. [17–19]

The method used to calculate limits is called the modified frequentist approach. [20, 21] We use fits to model the background and signal distributions of reconstructed di-photon mass using two functions: The background is modeled by a falling exponential (Eq. 2.1), and the signal is modeled by a Lorentzian (Eq. 2.2), where a and γ are constants.

$$M_{\text{back}} \approx e^{-ax} \tag{2.1}$$

$$M_{\text{sig}} \approx \frac{\gamma}{\pi} \frac{1}{x^2 + \gamma^2} \tag{2.2}$$

These forms are chosen for their simplicity and accuracy in describing the distributions; models such as polynomials and multiple exponential forms may yield technically better fits, but are more prone to overfitting. Additionally, the models in Eq. 2.1 and Eq. 2.2 are very distinct. Eq. 2.1 could never describe the signal distribution well, and visa versa.

Once the distributions of the reconstructed di-photon mass are adequately modeled, they can be used to generate pseudo-experiments. If we want to determine what the distribution of reconstructed di-photon mass would look like with a certain ratio of signal events to background events, then we turn the distributions into probability density functions by normalizing. Then they are no longer simply models; they provide us with the probability of a background event producing a certain value for the reconstructed di-photon mass.

Using the background distributions, we generate a histogram of reconstructed di-photon mass with no signal. We then need to quantify the success of the two possible models we have: signal and background. Although we know that there is no signal in the histogram, any method we generate will inevitably have false positives, and we need to account for this. We establish the form shown in Eq. 2.3 for the background model and the form shown in Eq. 2.4 for the signal model, where C_i are adjustable constants.

$$f_b(x) = C_1 e^{-C_2 x} \tag{2.3}$$

$$f_{s+b}(x) = f_b(x) + C_3 \frac{C_4}{\pi} \frac{1}{(x - C_5)^2 + C_4^2} \tag{2.4}$$

Proceeding to fit the histogram with both the functional forms, we generate a statistic called q, where $\mathcal{L}$ signifies the likelihood of the fit in relation to the histogram, defined by

$$q = -2 \log \left(\frac{\mathcal{L}(f_{s+b})}{\mathcal{L}(f_b)} \right) . \tag{2.5}$$

The likelihood of the fit in relation to the histogram is calculated by taking the value of the fit function at a bin to be the mean of a poisson distribution, and calculating the value of the distribution at the histogram bin value. This gives the likelihood that the bin was generated from the fit, or phrased differently, the likelihood that the fit function correctly describes the histogram. The likelihood for every bin is multiplied together to give the likelihood that the fit function describes the histogram. Mathematically,

$$\mathcal{L}(f) = \prod_{\text{all bins}} \text{Pois}(h; f(b)) = \prod_{\text{all bins}} \frac{f(b)^h e^{-f(b)}}{h!} , \tag{2.6}$$

where b is the bin center location, and h is the value of the bin.

The statistic q gives a rough idea of how likely the signal fit is to find an improvement over the background fit. Notice that if the signal and background fits are identical, then $q = 0$, and we do not expect $q < 0$ since the f_{s+b} has extra parameters to adjust, and part of it has the same form as the f_b. With no signal, we call the calculated statistic q_0, and generate a number of pseudo-experiments such that we can have a comprehensive distribution of the possible values of q_0. The previous procedure is repeated again, except that now a signal of strength μ is introduced into the previously background-only pseudo-experiments. The calculated statistic is now labelled q_μ, and again a number of pseudo-experiments are used to create a distribution. A typical result is shown in Fig. 2-1.

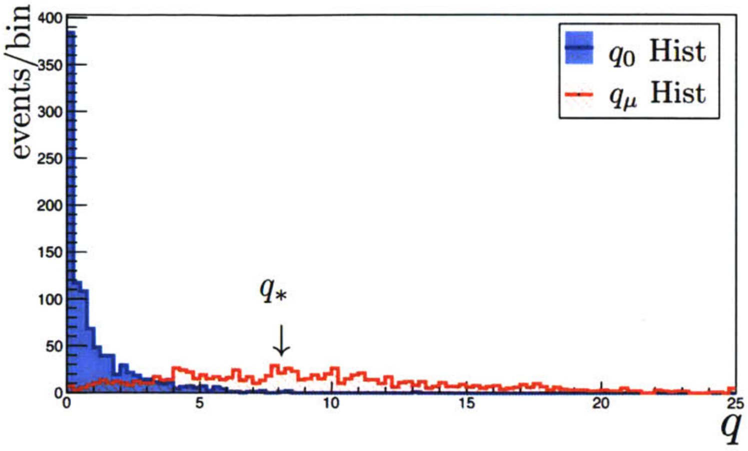

Figure 2-1: Histograms of q_0 and q_μ, with q_0 in blue and q_μ in red. The value of q_* is also shown graphically.

The distribution of q_0 is then used to calculate the value of q_*. This statistic is the value for which we are very certain (3σ) that the q calculated does not come from a pseudo-experiment with signal. The equation is shown in Eq. 2.7, where f_0 is the distribution of q_0.

$$0.997 = \int_0^{q_*} f_0(q)\, dq \tag{2.7}$$

Although we know that values of $q > q_*$ are unlikely to have come from an experiment without a signal, we would like to quantify how sure we are that an event chosen with a $q > q_*$ is generated from a pseudo-experiment with a signal. To accomplish this, we calculate the probabilities, based on the distribution generated, that a pseudo-experiment has $q > q_*$ given that it was generated with and without signal, as shown in Eq. 2.8 and Eq. 2.9.

$$p_\mu = P(q > q_* | f_\mu) \tag{2.8}$$

$$p_0 = P(q > q_* | f_0) \tag{2.9}$$

Notice that p_0 will be, by definition, 0.003. Defining the confidence limit statistic

CL_s as

$$CL_s = \frac{p_\mu}{p_0} \,, \tag{2.10}$$

we then see that CL_s gives a measure of how sure we are that an experiment selected at random with $q > q_*$ is from a pseudo-experiment with a signal. We are more interested, however, in the signal strength that would give a confidence limit of 0.95. We call this value $\mu^{95\%CL}$. Since CL_s is a function of the signal strength, μ must be varied in order to determine what value of μ will yield a CL_s of 0.95. By calculating $\mu^{95\%CL}$ many times, we can generate a distribution for $\mu^{95\%CL}$. This allows a range to be determined for the possible values of $\mu^{95\%CL}$. When quoting a production cross section limit, the value of $\mu^{95\%CL}$ is called the theoretical or expected limit. It is quoted in either units of Standard Model cross sections (σ_{SM}) (hence the phrase production cross section limit) or number of signal events. We calculate the standard model cross sections from the accepted cross section values and branching ratios in reference [22]. In the case of Standard Model cross sections, if the expected limit is $< 1\sigma_{SM}$, we can exclude the Higgs boson from this mass range.

2.3 Cut-based selection

Setting a production cross section limit allows us to establish how strong our signal must be to be detected or rule out the Higgs boson's existence from a certain mass region. Ideally the production cross section limit should be as low as possible, since we need the standard model signal production to be greater than the 2σ limit in order to rule out that mass region. There are multiple techniques that can be used to increase sensitivity. The first is the simplest, and is called variable cut analysis.

Up until this point we have only examined one of the properties about the particles available to us: reconstructed di-photon mass. However, there are many other properties, both about photons and about other particles such as jets, electrons, muons, and more. We can use these other variables, and take advantage of their specific distributions, in order to improve the amount of signal in relation to background.

Specifically, the metric used is

$$\frac{S}{\sqrt{S+B}}\,, \tag{2.11}$$

where S and B stand for the amount of signal and background events, respectively. We are looking to maximize this value, since the greater this value is, the better our signal will be in relation to the error on the total amount of signal and background. This will yield the most distinguishable signal, and thus will yield the lowest value of the theoretical limit. The task then is to determine what variables are most distinct between the signal and background events. Specifically we need to determine which variables are connected to whether or not the reconstructed di-photon mass came from a signal or a background event.

Once we establish a variable which we think is important, we then perform the cut analysis. Here, we scan over the range of the variable. At each point, we only look at events which have values of the variable which are greater or less than the value at the point (greater or less than is chosen for each scan, not each point). By removing some points, we see what effect that has on the ratio from Eq. 2.11. We then have a qualitative description of how our restriction of the dataset, or our cut, affects our signal and background. In particular, we know where to place our cut in order to get the best signal, according to our metric. Cut analysis can be performed on any number of variables, although most variables will be uncorrelated to the improvement of Eq. 2.11.

2.4 Multivariate selection

Cut-based analysis is not the only way to refine a production cross section limit and decrease the amount of data required to make a conclusion about the existence of the Higgs boson in a given mass range. There are many different, more complicated methods which are used to take multiple variables and their correlations into account at a time. These methods are grouped under the term "multivariate analysis", although they can be very different. The two types of methods we will focus on are called Boosted Decision Trees (BDT) and Multilayer Perceptrons (MLP). In each of these methods, the method is provided with a set of variables that we have decided are important in distinguishing signal from background. The methods are also provided with a sample dataset, called the training sample. Each method analyzes the training sample to determine how the variables are related, and how it can use them to distinguish between signal and background events. Unlike cut analysis, multivariate analyses will do more than make cuts. Each method has a different way of utilizing the information given.

In a boosted decision tree, each input variable is analyzed separately. A cut analysis is performed on each on in order to maximize $S/\sqrt{S+B}$. The variable cut that provides the greatest improvement is then chosen as the first branching point of a decision tree. The process is repeated multiple times to create a tree that allows event to be sorted into signal and background nodes. An example of a boosted decision tree is shown in Fig. 2-2. [23] In order to account for the variability present in the training sample, many different trees are created, to create a "forest" of decision trees. On an event by event basis, the decision as to whether an event is a signal or background event is then made by a majority vote of the trees.

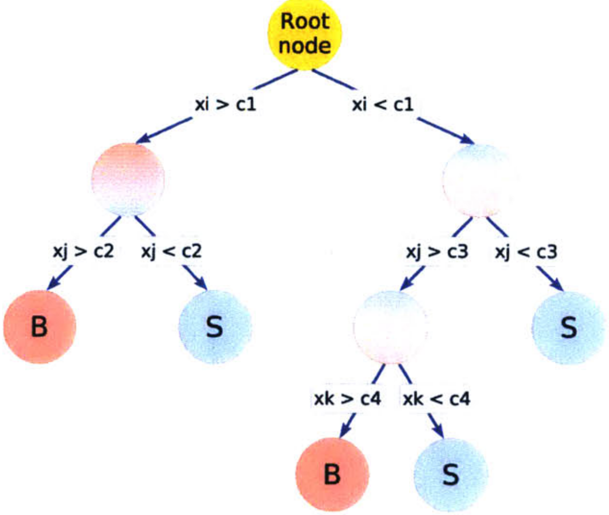

Figure 2-2: Schematic of boosted decision tree. The data begins at the root node and is sorted by successive cuts on the most significant variable until it is suitably separated. Here, xi, xj, and xk are different input variables.

In a multilayer perceptron, cuts on each input variable turn the initial input variables into groups containing one or more of the variables, with various weights associated with each variable-to-group relation. These groups are then cut and transformed in the same manner into more groups. The final output of the multilayer perceptron is a function of groups' values being combined at the end, and each step in the middle is called a hidden layer. An example of a multilayer perceptron is shown in Fig. 2-3. [23]

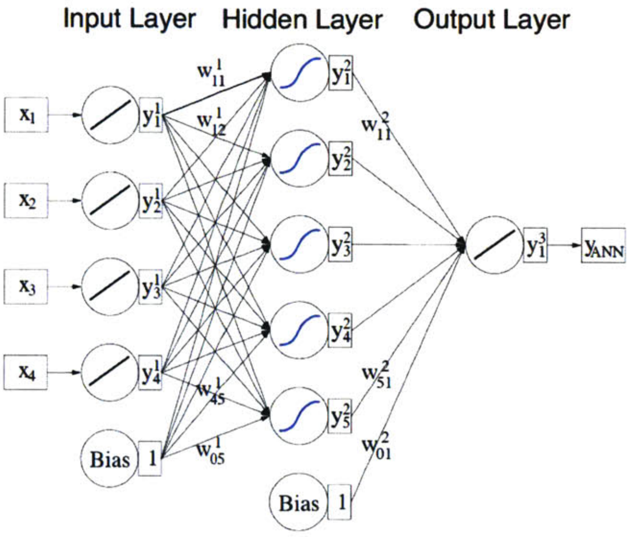

Figure 2-3: Schematic of a multilayer perceptron with one hidden layer. The data begins at the input layer, and is sorted into one or more hidden layers, and then sorted into the output layer. Here, x_1, x_2, x_3, and x_4 are different input variables.

The output of every method is called a response function, shown in Fig. 2-4. A response function, given the values for various selected variables, outputs a value. Finally we again perform a cut analysis; instead of the cut being on the value of a variable, it is on the value of the response function.

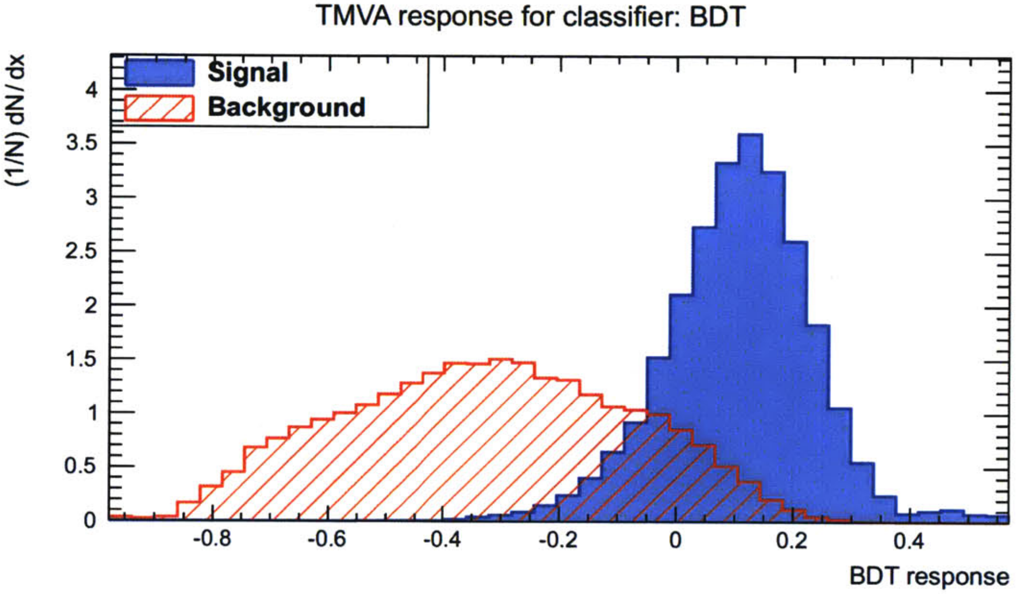

Figure 2-4: Example of a response function created using a boosted decision tree. Note how the signal and background distributions are quite distinct, allowing for a better distinction of signal for background.

Chapter 3

Results

The results section will focus on two main topics. The first will discuss the process of setting a production cross section limit for Higgs to gamma-gamma in the vector boson fusion production channel from 4.76 fb^{-1} of data. The second will focus on optimization, as well as display its advantages over other possible methods.

3.1 Initial expected limit

Using Monte Carlo simulations, background and signal events were generated for Higgs masses between 120 GeV and 130 GeV. The simulation data was modeled, and the fit functions were used to generate pseudo-experiments. The results of the limit setting procedure are shown in Fig. 3-1.

3.2 CMS cut analysis

In February 2012, the CMS Collaboration published an analysis of the data they had gathered up to that point. In the analysis, they performed a cut analysis to refine the vector boson fusion dataset, and obtained a signal efficiency of 0.3636 and background rejection of 0.9949. A diagram of their result is shown in Fig. 3-2. [24]

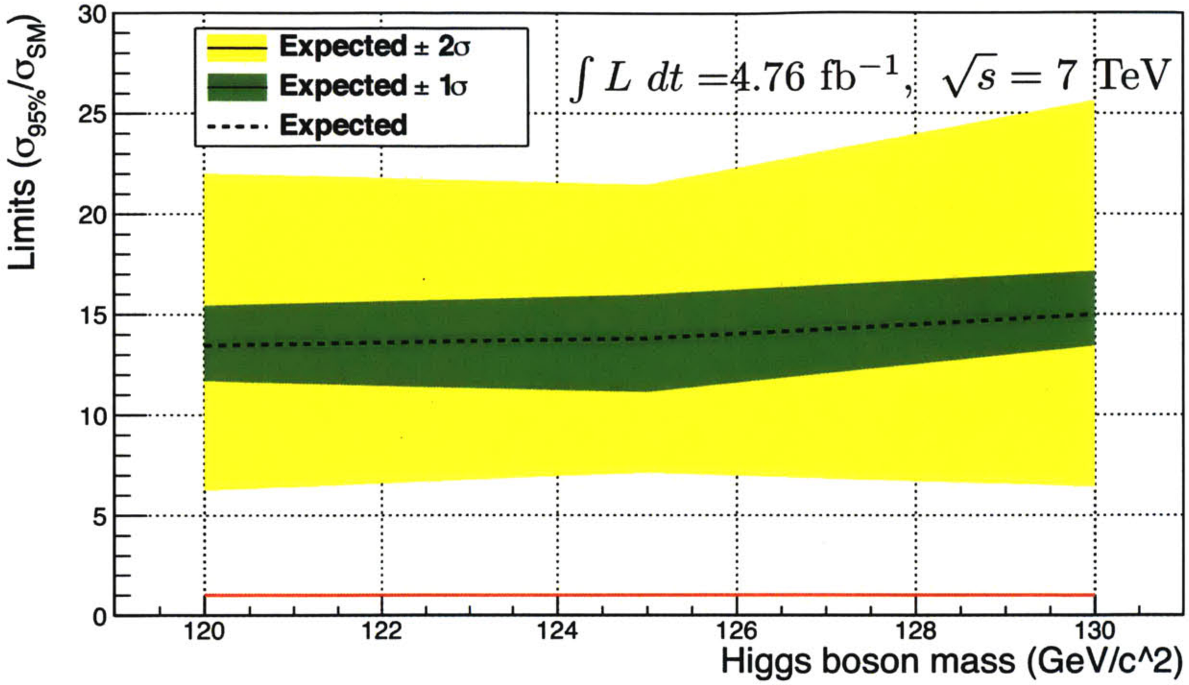

Figure 3-1: Initial limit on production cross section. There has been no refinement or restriction of data. The median, 1σ, and 2σ limits are shown in black, green, and yellow, respectively.

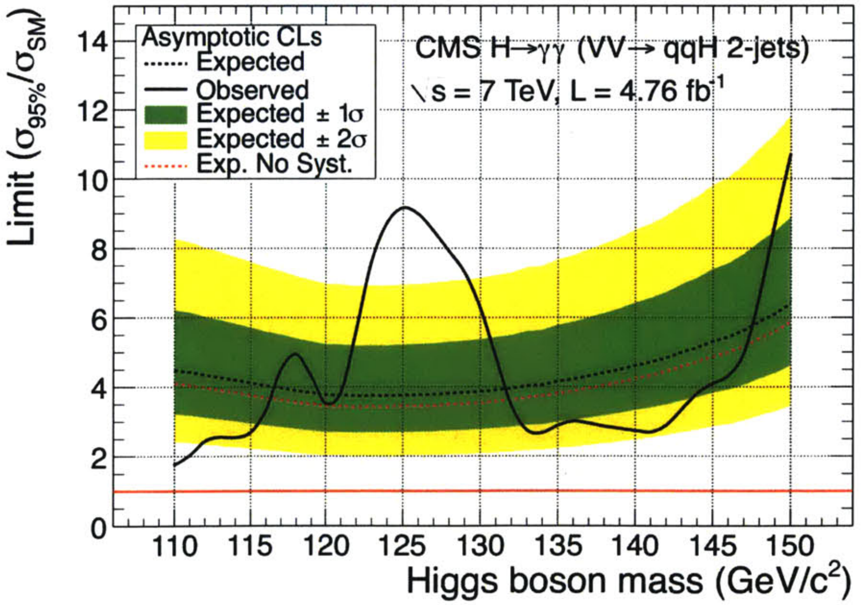

Figure 3-2: CMS production cross section limit with cut analysis. The median, 1σ, and 2σ limits are shown in black, green, and yellow, respectively.

3.3 Boosted decision tree theoretical limit

The first variable chosen was the difference in jet pseudorapidity, $\Delta\eta_j$. As previously discussed, one defining characteristic of vector boson fusion is the high pseudorapidity of the two jets created in the process. However, one of the jets will necessarily have positive pseudorapidity, while the other will have negative pseudorapidity. Thus, taking the difference in the two jet pseudorapidities will yield a large value for signal events. In contrast, the background events are not inclined to produce jets in any specific range of pseudorapidities. Further, the background events have no tendencies to yield jets with positive or negative pseudorapidities. Therefore, the difference in pseudorapidity for background events is clustered around 0, and then falls off to either side. The distributions of $\Delta\eta_j$ for signal and background are shown in Fig. 3-3.

The next variable chosen was the difference in jet azimuthal angle, $\Delta\phi_j$. In vector boson fusion the three products are a Higgs and two jets. Since the Higgs is much more massive than the jets, it causes the jets to be ejected at similar azimuthal angles in order to conserve momentum. Therefore, the signal distribution of $\Delta\phi_j$ is clustered around 0. For background events, there is not a massive particle created in general) to cause the clustering of the two jets. In fact, they are inclined to be ejected back-to-back, since most background comes from hard scattering events. Thus, the background distribution of $\Delta\phi_j$ is clustered around $\pm\pi$. The distributions of $\Delta\phi_j$ for signal and background are shown in Fig. 3-3.

The next variable chosen was the product of jet pseudorapidities, $\eta_j \cdot \eta_j$. As mentioned previously, the jets created in vector boson fusion commonly enter the front and back regions of the detector. One will have a positive pseudorapidity and the other a negative one. Thus, the signal distribution of $\eta_j \cdot \eta_j$ will be negative. Meanwhile, background events have no such bias towards the forward and backward regions of the detector. Additionally, they are not characteristically large, so their product is likely be small. The background distribution of $\eta_j \cdot \eta_j$ will therefore be tightly symmetric about 0. The distributions of $\eta_j \cdot \eta_j$ for signal and background are shown in Fig. 3-3.

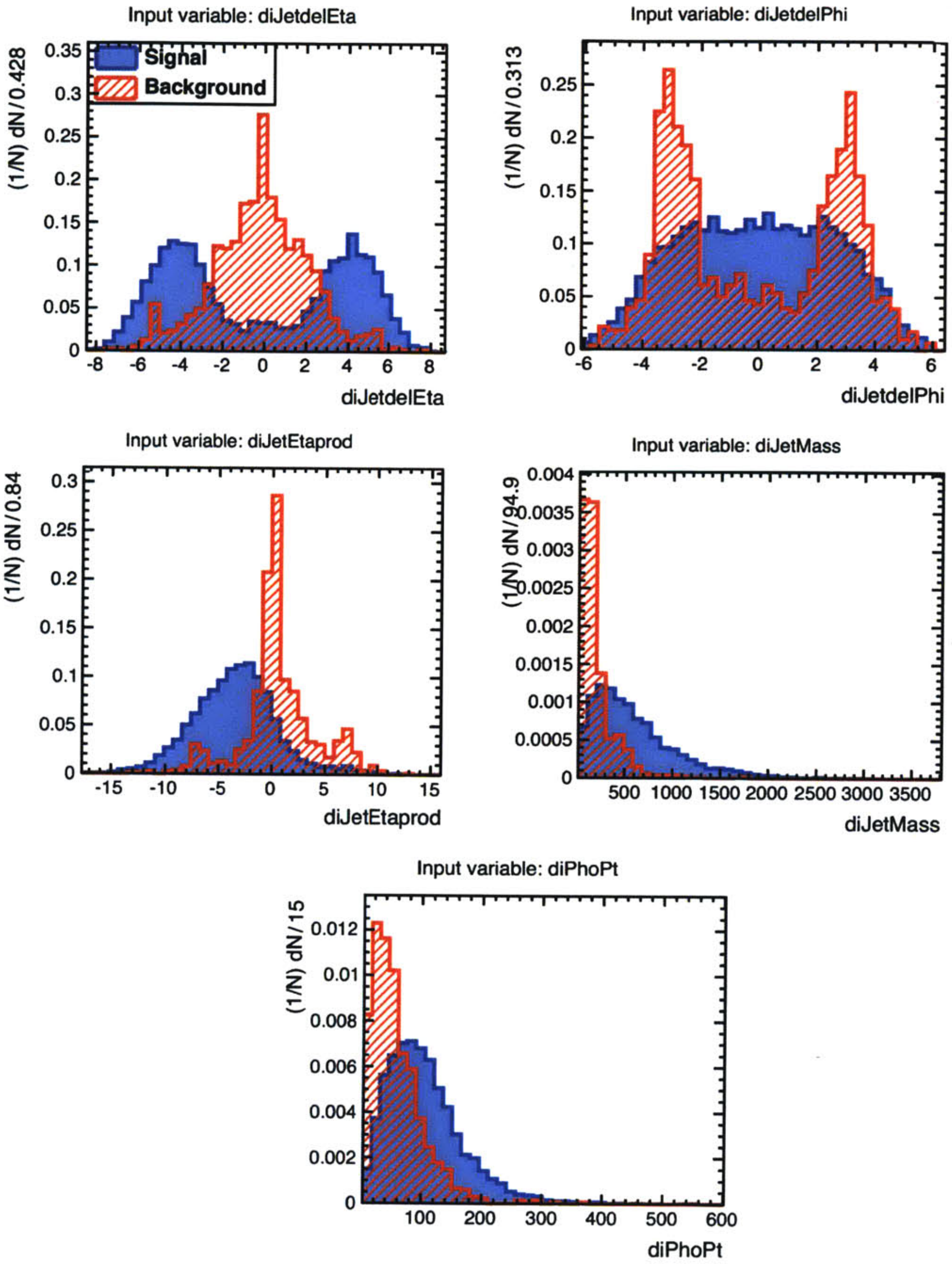

Figure 3-3: Distributions of various variables for signal and background. *Left to right, top to bottom*: difference in jet pseudorapidity ($\Delta\eta_j$), difference in jet azimuthal angle ($\Delta\phi_j$), product of jet pseudorapidity ($\eta_j \cdot \eta_j$), reconstructed di-photon transverse momentum ($p_{T\gamma\gamma}$), reconstructed di-jet mass (m_{jj}).

The next variable chosen was the reconstructed di-photon transverse momentum, $p_{T\gamma\gamma}$. The detector cannot detect momentum along the beam path, and thus the measured momentum is always in the transverse direction. In Higgs to gamma-gamma, the two photons are created by the decay of the Higgs boson. Since the Higgs boson has a large mass, its momentum is quite large, and so the reconstructed di-photon transverse momentum is correspondingly large. Unlike is signal events, in background events the two photons do not have to be created by any particle in particular. Thus, there is less constraint on their allowed momentum; specifically, their reconstructed transverse momentum is characteristically lower than that for the signal events. [25] The distributions of $p_{T\gamma\gamma}$ for signal and background are shown in Fig. 3-3.

The final variable chosen was the reconstructed di-jet mass, m_{jj}. As mentioned above, the Higgs boson's large mass gives it correspondingly large momentum. Given that the di-jet momentum and Higgs boson momentum must sum to 0 transverse momentum, the di-jet reconstructed mass must also be large. For the background events, there is no such restriction. Thus, in general the values of di-jet reconstructed mass are much smaller for background events than they are for signal events. The distributions of m_{jj} for signal and background are shown in Fig. 3-3.

With our important variables selected, we analyzed multiple MVAs to determine the most effective one. A plot of background rejection vs. signal efficiency, shown in Fig. 3-4, was was used to determine the most effective one. Given the success of the BDT method both at rejecting background and accepting signal, it was chosen as the most successful method. The BDT analysis produced a response function with the distributions shown in Fig. 3-5. A cut placed at 0.05 yielded the best separation of signal and background events: 94.0% of background events were rejected, while 71.5% of signal events were accepted. In contrast, using a MLP instead of a BDT yielded a background rejection of 83.84% and signal acceptance of 78.1%. Although slight improvement was shown in signal efficiency with a MLP over a BDT, the more than two-fold increase in background events made the BDT method the clearly better choice. With the BDT applied to the dataset, the production cross section limit was

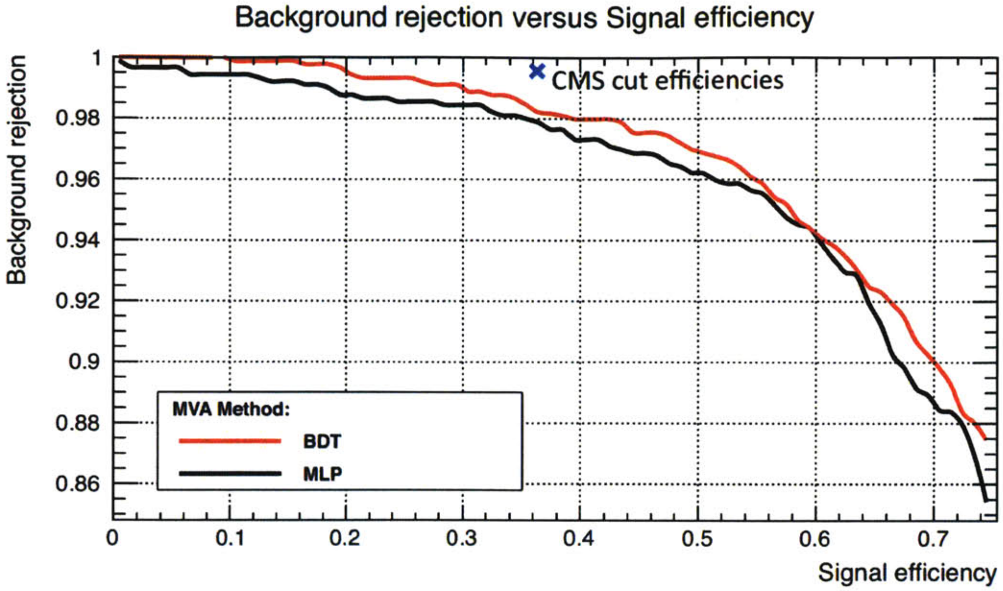

Figure 3-4: Plot of background rejection vs. signal efficiency for multiple MVA methods. The BDT method clearly provides the best background rejection and signal efficiency of the two. The CMS cut analysis efficiencies is also shown in blue.

recalculated, with a great improvement. The result is shown in Fig. 3-6.

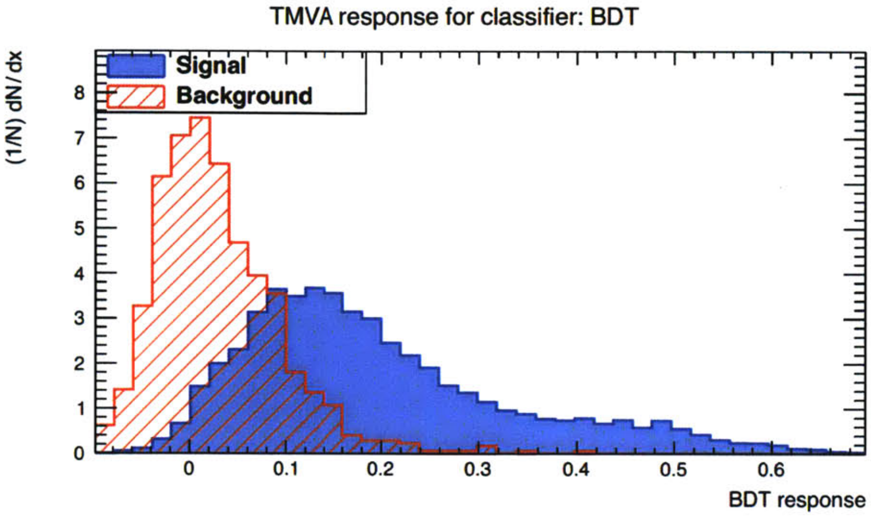

Figure 3-5: BDT response function for both signal and background. The background distribution is shown in red and the signal in blue.

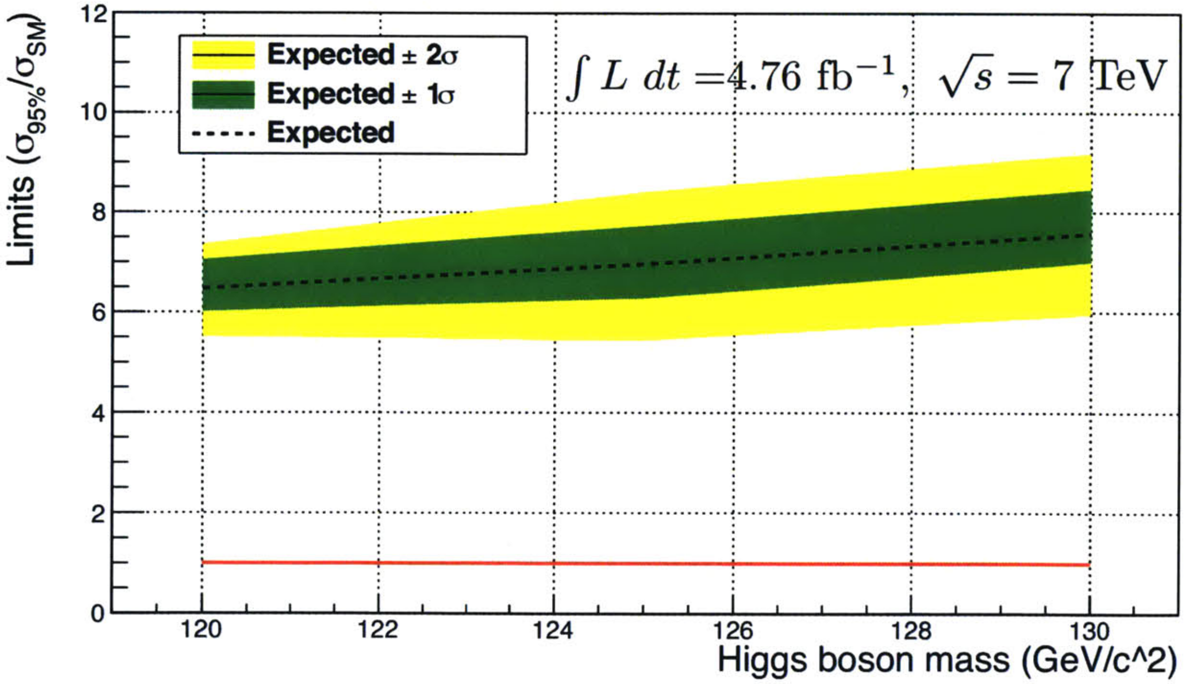

Figure 3-6: Production cross section limit after BDT refinement. The median, 1σ, and 2σ limits are shown in black, green, and yellow, respectively. Notice the improvement from the result in Fig. 3-1.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Conclusions

4.1 Discussion of results

The Higgs boson is the last remaining piece of the standard model yet to be discovered. Many physicists have come to take for granted that one day the particle will be discovered. However, it still eludes our searches; new techniques must developed and more data gathered in order to finally establish its existence. Our search for the Higgs boson aimed to search for the Higgs boson in the current mass range of interest, 120-130 GeV/c^2. A limit on production cross section was set to provide a benchmark, shown in Fig. 3-1. The simulated data was refined using a boosted decision tree to maximize $S/\sqrt{S+B}$, and the BDT was used to improve the production cross section limit. The result is shown in Fig. 3-6, and the lowest limit set was $\sim 6\sigma_{SM}$. This is not low enough to exclude the Higgs boson from any of the masses we analyzed.

This boosted decision tree analysis was also compared to the cut analysis done by the CMS Collaboration. Their results were compared to the results from the BDT, and found to be better. The cut analysis performed utilized more variables than the BDT analysis, and had better signal efficiency and background rejection. This allowed the development of a production cross section of $\sim 4\sigma_{SM}$, a clear improvement over our limit developed using a BDT.

Our inability to display an improvement over the cut analysis is unfortunate; there are two main reasons for this. The first reason is the larger set of variables used in the

CMS analysis. The larger number of variables used allowed an improved distinction of signal and background over the BDT analysis, as shown in Fig. 3-4. However, the more important reason is that most of the variables used were uncorrelated. The more correlated the variables used are, the worse a cut analysis will be in relation to more complex methods. Correlations only serve to decrease the effectiveness of a cut approach. Thus, the cut analysis was a very good method to refine the data, and the BDT could have only generated a small improvement, if any, over the cut analysis if the same variables had been used. It is also worth noting that the jet and photon selection used by the CMS Collaboration was more sophisticated than the jet and photon selection in our analysis. All these factors contributed to our inability to improve the production cross section limit calculated by the CMS Collaboration.

4.2 Future work

In the future, more variables could be utilized in the multivariate analysis. Variables like the Zeppenfeld variable

$$Z = \eta_{\gamma\gamma} - \frac{\eta_{j1} - \eta_{j2}}{2} \ , \tag{4.1}$$

and

$$\Delta_{\gamma\gamma,jj} = \eta_{\gamma\gamma} - \eta_{jj} \ , \tag{4.2}$$

as well as others, would only help to improve the background rejection and signal efficiency. This would improve the limit, hopefully enough to exclude the Higgs boson at various masses and help to further restrict our search to smaller and smaller mass ranges.

Bibliography

[1] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear Physics*, vol. 22, no. 4, pp. 579 – 588, 1961.

[2] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.

[3] A. Salam, "Weak and electromagnetic interactions," *Elementary Particle Theory: Relativistic Groups and Analyticity, Proceedings of the 8th Nobel Symposium*, p. 367, 1968.

[4] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964.

[5] P. Higgs, "Broken symmetries, massless particles and gauge fields," *Physics Letters*, vol. 12, no. 2, pp. 132 – 133, 1964.

[6] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.

[7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles," *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964.

[8] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons," *Phys. Rev.*, vol. 145, pp. 1156–1163, May 1966.

[9] T. W. B. Kibble, "Symmetry Breaking in Non-Abelian Gauge Theories," *Phys. Rev.*, vol. 155, pp. 1554–1561, Mar 1967.

[10] ALEPH, DELPHI, L3, OPAL Collaborations, and The LEP Working Group for Higgs Boson Searches, "Search for the Standard Model Higgs boson at LEP," *Physics Letters B*, vol. 565, no. 0, pp. 61 – 75, 2003.

[11] CDF and D0 Collaborations, "Combination of Tevatron Searches for the Standard Model Higgs Boson in the W^+W^- Decay Mode," *Phys. Rev. Lett.*, vol. 104, p. 061802, Feb 2010.

[12] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD Collaborations, LEP Electroweak Working Group, Tevatron Electroweak Working Group, SLD electroweak heavy flavour groups, "Precision Electroweak Measurements and Constraints on the Standard Model," 2010.

[13] CMS Collaboration, "Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV," *Physics Letters B*, vol. 710, no. 1, pp. 26 – 48, 2012.

[14] ATLAS Collaboration, "Combined search for the Standard Model Higgs boson using up to 4.9 fb^{-1} of pp collision data at with the ATLAS detector at the LHC," *Physics Letters B*, vol. 710, no. 1, pp. 49 – 66, 2012.

[15] UCSD CMS Group, "H → ZZ(*) → 4e Signal and Background Study at Generator Level."

[16] "http://en.wikipedia.org/wiki/File:Pseudorapidity2.png," May 2007.

[17] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 Physics and Manual," *Journal of High Energy Physics*, vol. 2006, no. 05, p. 026, 2006.

[18] P. Nason and C. Oleari, "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG," *Journal of High Energy Physics*, vol. 2010, pp. 1–18, 2010.

[19] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO Higgs boson production via gluon fusion matched with shower in POWHEG," *Journal of High Energy Physics*, vol. 2009, no. 04, p. 002, 2009.

[20] A. L. Read, "Modified frequentist analysis of search results (the CL_s method)," no. CERN-OPEN-2000-205, 2000.

[21] ATLAS and CMS Collaborations, LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011," Tech. Rep. ATL-PHYS-PUB/CMS-NOTE-2011-005, CERN, Geneva, Aug 2011.

[22] LHC Higgs Cross Section Working Group Collaboration, *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables.* No. CERN-2011-002, CERN, 2011.

[23] A. Hoecker, P. Speckmayer, J. Stelzer et al., "TMVA: Toolkit for Multivariate Data Analysis," *PoS ACAT*, 2007.

[24] CMS Collaboration, "Search for a Higgs boson decaying into two photons in *pp* collisions recorded by the CMS detector at the LHC." 2011/426, December 2011.

[25] A. Ballestrero, G. Bevilacqua, and E. Maina, "A complete parton level analysis of boson-boson scattering and electroweak symmetry breaking in $\ell\nu$ + four jets production at the LHC," *Journal of High Energy Physics*, vol. 2009, no. 05, p. 015, 2009.