Search for the Vector Boson Fusion Production of the Higgs Boson in the $H \rightarrow WW^* \rightarrow l\nu l\nu$ Channel using Support Vector Machines

A dissertation submitted by

Jeffrey Berwyn Wetter

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Phyiscs

Tufts University

May 2015

Advisor: Krzysztof Sliwa

Abstract

This thesis presents a search for vector boson fusion production of Higgs bosons in the $H \rightarrow WW^* \rightarrow l\nu l\nu$ channel using the $20\ fb^{-1}$ 2012 ATLAS data set. The analysis uses support vector machines in order to distinguish signal and background events. First, an analysis is performed using a 2-class support vector machine, designed to separate signal events from all backgrounds. In addition, a 3-class support vector machine analysis is performed in which background events are separated into two classes. The results of both the 2 and 3-class analyses are consistent with the findings of other ATLAS analyses and with Standard Model predictions to within 1σ.

Acknowledgements

This thesis would not have been possible without the love and support of my wonderful parents, Berwyn and Delores Wetter. I would also like to thank Tracy McAskill for her constant support and companionship. Finally, I am grateful to my advisor, Krysztof Sliwa, for his guidance throughout my years of graduate school.

Contents

List of Figures

List of Tables

Search for the Vector Boson Fusion Production of the Higgs Boson in the $H \rightarrow WW^* \rightarrow l\nu l\nu$ Channel using Support Vector Machines

Chapter 1

Introduction

In 2012, both the ATLAS and CMS experiments at the Large Hadron Collider in Geneva, Switzerland, announced the discovery of a new boson with a mass of $125 GeV$ [1][2]. This new boson is believed to be the Higgs boson, predicted by Peter Higgs and his collaborators 50 years ago. The Higgs boson is crucially important to our understanding of the universe as it gives mass to all particles within the Standard Model of particle physics. The joint discovery of this new boson both validates the Standard Model and provides new opportunities to test its predictions.

Since its discovery, both the CMS and ATLAS experiments have been working to measure the spin, couplings and mass of the Higgs boson. These measurements are important in order to verify that the newly discovered boson is in fact the Higgs boson predicted by the Standard Model. Any differences between the Standard Model predictions and the observed properties of the

newly discovered boson might give insight into new physics or possible flaws in the Standard Model.

The mass of the Higgs boson is not predicted by the Standard Model and therefore the properties of most interest are the spin and couplings of the newly discovered boson. The Standard Model predicts that the Higgs boson will have zero spin and will couple to particles in proportion to the particles' mass. Studies of the spin of the newly discovered boson favor the spin-zero hypothesis [3]. Determining the coupling of the Higgs boson to fundamental particles is a much larger task, as the Higgs' coupling to each particle must be determined independently.

The ability to measure the coupling of the Higgs boson to various particles is highly dependent on the particle of interest, the final state of the system, the mass of the Higgs boson, and the limits of the LHC luminosity. All of these factors combine to make certain measurements significantly easier than others. Due to these factors, the discovery of the Higgs boson in 2012 was based primarily on studies of $H \to ZZ$ and $H \to \gamma\gamma$ decays.

The $H \to ZZ \to llll$ channel is ideal for Higgs boson studies as the final state contains four leptons, zero jets and no missing energy. The ATLAS detector is well-equipped to detect this final state, having multiple lepton detection systems. In addition, many background processes contain high-p_T jets, facilitating the removal of background events. Finally, at a mass of 125 GeV, the Higgs boson has a large branching ratio into two Z bosons, allowing for the

creation of large numbers of $H \to ZZ$ events with relatively limited luminosity.

Like the $H \to ZZ$ measurement, there are several factors which facilitate the study of $H \to \gamma\gamma$ events within the ATLAS detector. $H \to \gamma\gamma$ events are characterized by two photons in the final state with zero jets and no missing energy. This relatively simple final state eliminates a large number of background processes, allowing for high signal-to-background ratios with limited luminosity. As with leptons, the ATLAS detector is well-equipped to detect high energy photons, allowing for proper identification of final state particles. Studies conducted in the $H \to ZZ$ and $H \to \gamma\gamma$ channels indicate that the Higgs boson branching ratios in these channels agree with Standard Model predictions [4].

The final prominent Higgs boson decay channel studied at the LHC is $H \to WW^*$. At a mass of 125 GeV, the $H \to WW^*$ channel has the second highest predicted branching ratio after $H \to bb$. Despite this relatively large cross section, large backgrounds and final state neutrinos make $H \to WW^*$ studies difficult at the LHC. In order to reduce hadronic backgrounds, most $H \to WW^*$ searches are performed in the $H \to WW^* \to l\nu l\nu$ channel. Due to the relatively large cross section, the primary $H \to WW^*$ searches look for Higgs bosons produced via gluon-gluon fusion. The gluon-gluon fusion process is advantageous due to its relatively large cross section and because it has no initial-state particles in the final state. The choice of a fully leptonic final state greatly reduces potential backgrounds, but it also reduces the number of

potential Higgs events due to the relatively small $W \to l\nu$ branching ratio. In addition, since τ leptons have a distinct signature within the ATLAS detector, only $W \to e\nu$ and $W \to \mu\nu$ decays are considered, further reducing the number of potential signal events. In total, only 4% of all $H \to WW^*$ decays result in an $H \to WW^* \to l\nu l\nu$ final state, resulting in a relatively small number of expected signal events.

In addition to the limited number signal events, $H \to WW^* \to l\nu l\nu$ analyses also struggle with missing energy due to the presence of two neutrinos in the final state. The presence of neutrinos in the final state has several detrimental effects: First, missing energy in the final state forbids an accurate reconstruction of the Higgs boson mass. Second, elimination of background events is significantly more difficult due to inaccuracy in the reconstruction of missing energy. Despite these difficulties, studies of gluon-gluon fusion Higgs production in the $H \to WW^*$ channel have found evidence for the Higgs boson consistent with Standard Model predictions [4].

Beyond the well-studied decay channels, there are several decay channels in which the search for definitive evidence of the Higgs is ongoing. These channels include the $H \to WW^*$ decays in which the Higgs boson is created through vector boson fusion (VBF). The process of vector boson fusion involves the emission of two vector bosons by colliding quarks. In VBF $H \to WW^*$ decays, the vector bosons fuse to form a Higgs boson, which then decays via two W bosons. Studies of VBF Higgs production are important in that they provide

independent measurements of Higgs boson couplings to vector bosons. Any variation in VBF Higgs production from Standard Model predictions could indicate new physics beyond the Standard Model.

The production cross section of Higgs bosons via vector boson fusion is more than 10 times smaller than the cross section for the dominant gluon-gluon fusion process. At the current LHC luminosity, this small cross section results in a very small number of expected VBF $H \rightarrow WW^*$ events. In addition, the large cross sections of many background processes, including $t\bar{t}$ events, makes the VBF $H \rightarrow WW$ channel very difficult to study using current LHC data.

This thesis will present a search for the Higgs boson produced via vector boson fusion decaying through two W bosons using the 2012 ATLAS data set. The final state of this decay includes two high-p_T leptons, missing energy and two high energy forward jets. Due to the complicated nature of this final state, it is beneficial to use multivariate learning algorithms in order to separate signal and background events. These algorithms are commonly referred to as multivariate analysis techniques (MVA), and are capable of distinguishing between different types of events by learning event properties. In this thesis we use a type of multivariate analysis technique known as support vector machines (SVM). SVMs use topological event information in order to separate events in n-dimensional parameter space. Support vector machines can be trained to distinguish between different types of events (i.e. signal and background events) by using training samples in which the true type of each event is

known. Once trained, the SVM is capable of distinguishing events of unknown type. We are therefore able to train support vector machines using Monte Carlo events of known type, and then used these trained machines to search for $H \rightarrow WW^*$ events in the 2012 ATLAS data set.

The analysis presented in this thesis is performed within the context of an ongoing ATLAS search for evidence of VBF $H \rightarrow WW^*$ decays. The primary ATLAS subgroup devoted to VBF $H \rightarrow WW^*$ searches uses a multivariate analysis technique known as boosted decision trees (BDT). The goal of our analysis is to validate the findings of the ATLAS VBF $H \rightarrow WW^*$ subgroup group as well as to investigate the possible benefits of using a system of support vector machines to enhance signal-background resolution. We begin by conducting an analysis parallel to the BDT analysis using a single support vector machine to separate signal and background events. The goal of this analysis is to validate the SVM methodology and establish a baseline result. We then conduct an analysis using a system of three support vector machines with the goal of improving on this baseline result. This analysis is referred to as a 3-class SVM analysis and involves splitting background samples into heavy and light classes in order to increase signal-to-background separation.

Chapter 2

The Standard Model of Particle Physics

The Standard Model of particle physics is a field theory describing three fundamental forces: electromagnetism, weak interactions and strong interactions. Within the Standard Model particles are separated into two categories: fermions, which are spin $\frac{1}{2}$ particles, constitute the fundamental particles of the Standard Model, and integer-spin bosons, which mediate interactions between fermions and other bosons. The electromagnetic force describes the interactions of electrically charged particles such as electrons through mediating photons. The photon is a massless boson, giving the electromagnetic force infinite range. In contrast, the weak force interacts with particles with weak isospin and is mediated by three heavy gauge bosons, W^+, W^- and Z, giving the weak force a very short range. Due to its short range, we do not notice the

weak force within our daily lives. However, the weak force is responsible for sustaining life on earth by mediating nuclear fusion within the sun. Within the Standard Model, the electomagnetic and weak interactions are combined into one force known as the electroweak force. The strong force, as its name implies, is the strongest fundamental force and is responsible for holding together atomic nuclei. The strong force is mediated by eight massless bosons known as gluons. Only quarks and gluons are able to interact via the strong force, which requires that the quarks are bound in pairs or triplets to make up more complex particles known as hadrons.

2.1   Standard Model Particles

The Standard Model separates fermions into two separate groups: leptons and quarks. These groups are further subdivided into three generations. The majority of visible matter in the universe is made up of first generation particles. Particularly, up and down quarks bond into protons and neutrons, which form stable nuclei and together with electrons create atoms. Higher-generation fermions tend to be more massive and less stable, decaying too quickly to form stable atoms.

Lepton and quark generations are described as SU(2) doublets of weak isospin (I_3), with each generation containing one particle with $I_3 = +\frac{1}{2}$ and one particle with $I_3 = -\frac{1}{2}$. For leptons, each generation contains a massive charged particle with negative isospin and a neutral particle with positive

isospin known as a neutrino. The electron and electron neutrino make up the first generation of leptons, with the muon and muon neutrino composing the second generation and the tauon and tau neutrino in the third generation. For quarks, each generation contains an up-type quark with positive electric charge of $Q = +\frac{2}{3}$ and positive weak isospin of $I_3 = +\frac{1}{2}$ and a down-type quark with negative charge, $Q = -\frac{1}{3}$ and negative weak isospin, $I_3 = -\frac{1}{2}$. Additionally, quarks come in three colors (red, green, blue) and are only found as a part of color-neutral bound particles. The first generation of quarks is composed of up and down quarks, the second generation contains the charm and strange quarks and the third generation is made up of the top and bottom quarks.

The Standard Model includes 12 gauge bosons: 8 gluons mediating the strong force and 4 gauge bosons responsible for mediating electroweak interactions. In addition, the Higgs boson, which allows for massive particles, is introduced to the Standard Model through spontaneous symmetry breaking as discussed in section 2.5. Figure 2.1 lists the properties of all fundamental particles within the Standard Model. In addition to the particles described in Figure 2.1, each fermion in the Standard Model has an associated anti-particle. Anti-particles have identical mass to their associated particle, but opposite quantum numbers such as charge and electroweak isospin. Interactions are symmetric with respect to particles and anti-particles within the Standard Model, and the matter/anti-matter discrepancy in our universe is an ongoing field of research.
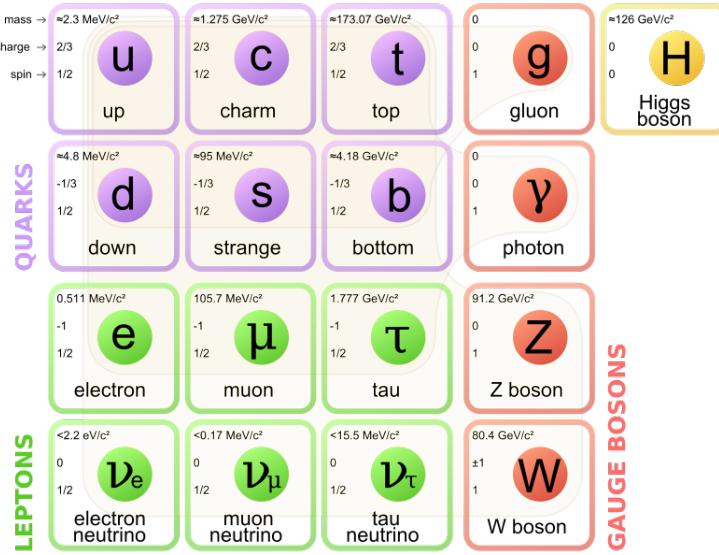
Figure 2.1: Properties of Standard Model particles [6].

2.2 Lagrangian Formulation of the Standard Model

The Standard Model describes fundamental particles as spin $\frac{1}{2}$ fields within the context of Lagrangian dynamics. The Lagrangian is defined to be

$$L = T - V, \tag{2.1}$$

where T is the kinetic energy of the system and V is the potential energy. In place of a Lagrangian, the Standard Model uses the Lagrangian density, $\mathcal{L}$, defined as

$$L = \int \mathcal{L} d^3 x. \tag{2.2}$$

For the remainder of this thesis we will follow common practice and refer to

$\mathcal{L}$ as the Lagrangian, though this is technically inaccurate. The equations of motion of a system can be derived by requiring the Lagrangian of that system satisfy the Euler-Lagrange equation,

$$\partial_\mu\Big(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\Big) - \frac{\partial\mathcal{L}}{\partial\phi} = 0. \tag{2.3}$$

In this equation ϕ is the field and ∂_μ is a common shorthand for $\frac{\partial}{\partial x_\mu}$.

Within field theories, such as the Standard Model, each term in a Lagrangian represents a particle interaction or energy. For example, a term within a Lagrangian of the form $\lambda\phi^4$, represents the interaction of four ϕ scalar fields. Of particular importance to this thesis are mass terms of the form $\mu^2\phi^2$, where the coefficient μ^2 is related to the mass of the particle ϕ by $m = \sqrt{\mu^2/2}$.

We will begin our discussion of the Standard Model by writing down a Lagrangian which includes only fermions. We will then add mediating gauge bosons into our Lagrangian through the requirement of gauge invariance. Gauge invariance simply means that a given Lagrangian is invariant with respect to gauge transformations of a given type. The simple example of local $U(1)$ gauge invariance is discussed in the following section.

2.3 Local Gauge Invariance

In order to illustrate the technique of generating gauge bosons through the requirement of local gauge invariance, we will look at the example of the Lagrangian of a massive scalar field,

$$\mathcal{L} = \partial^\mu \phi^* \partial_\mu \phi - \mu^2 \phi^* \phi - \lambda (\phi^* \phi)^2. \tag{2.4}$$

As discussed in section 2.2, our Lagrangian (2.4) contains only a massive scalar field with no mediating gauge bosons. The $\partial^\mu \phi^* \partial_\mu \phi$ term in (2.4) represents the kinetic energy of the scalar particle, while the $\mu^2 \phi^* \phi$ term indicates that ϕ is a massive particle. Finally, $\lambda (\phi^* \phi)^2$ represents the four-particle vertex of ϕ with coupling constant λ.

We now require that $\mathcal{L}$ be invariant under a local gauge transformation of the form

$$\phi \to \phi' = e^{i\alpha(x)} \phi. \tag{2.5}$$

Phase transformations of the form $U(\alpha) = e^{i\alpha}$, $(\alpha \in \Re)$ make up the unitary Abelian group $U(1)$, where the Abelian label indicates that the group multiplication is commutative. Equation (2.5) is defined as a local gauge transformation because the phase, α, is dependent on the position, x. Substituting equation (2.5) into the first term of equation (2.4) we find that $\partial_\mu \phi$ transforms as

$$\partial_\mu \phi \rightarrow e^{i\alpha(x)}\partial_\mu \phi + ie^{i\alpha(x)}\phi\partial_\mu \alpha. \tag{2.6}$$

The presence of the term $ie^{i\alpha(x)}\phi\partial_\mu \alpha$ indicates that $\partial_\mu \phi$ is not invariant under local $U(1)$ transformations. In order to account for this additional term in the transformed derivative, we introduce a covariant derivative of the form

$$\partial_\mu \rightarrow \mathcal{D}_\mu = \partial_\mu - ieA_\mu, \tag{2.7}$$

where A_μ transforms as

$$A_\mu \rightarrow A_\mu + \frac{1}{e}\partial_\mu \alpha. \tag{2.8}$$

We note that the covariant derivative is invariant under $U(1)$ gauge transformations up to a phase,

$$D_\mu \phi \rightarrow \mathcal{D}_\mu \phi' = e^{i\alpha(x)}\partial_\mu \phi + ie^{i\alpha(x)}\phi\partial_\mu \alpha -$$
$$ieA_\mu e^{i\alpha(x)}\phi - ie^{i\alpha(x)}\phi\partial_\mu \alpha = e^{i\alpha(x)}\mathcal{D}_\mu \phi. \tag{2.9}$$

We can easily see that the second term in our Lagrangian is invariant under $U(1)$ transformations by noting that

$$\mu^2 \phi^* \phi \rightarrow \mu^2 e^{-i\alpha(x)}\phi^* e^{i\alpha(x)}\phi = \mu^2 \phi^* \phi. \tag{2.10}$$

The third term in (2.4) can be shown to be invariant in the same way. We can therefore write the $U(1)$ invariant Lagrangian of a massive scalar field as

14

$$\mathcal{L} = (\partial^\mu + ieA^\mu)\phi^*(\partial_\mu - ieA_\mu)\phi - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2. \qquad (2.11)$$

The Lagrangian, (2.11), contains a new vector gauge field, A_μ, introduced due to the requirement of $U(1)$ invariance. This new field is massless, since there is no corresponding mass term of the form $\frac{1}{2}m^2 A_\mu A^\mu$. In fact, such terms are forbidden as they are not gauge invariant. Adding an additional gauge invariant term to account for the kinetic energy of the gauge boson associated with the field A_μ, we write our final Lagrangian as

$$\mathcal{L} = (\mathcal{D}_\mu\phi)^*\mathcal{D}_\mu\phi - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \qquad (2.12)$$

where $F_{\mu\nu}$ is constructed from derivatives of A_μ,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \qquad (2.13)$$

We have now shown that by starting with a Lagrangian describing massive scalar particles (e.g. electrons), we are able to generate massless vector bosons through the requirement of $U(1)$ gauge invariance.

This simple formulation is sufficient for describing electromagnetism when the scalar field ϕ is replaced by a Dirac spinor, ψ. In this case we can associate A_μ with the photon and $F_{\mu\nu}$ with the electromagnetic field strength tensor. We are able to describe electromagnetism using this simple Lagrangian because photons are massless. However, in order to create a theory which encompasses the electromagnetic, weak and strong forces, we will need to introduce massive

gauge bosons in a gauge invariant manner. To do this, we utilize spontaneous symmetry breaking and the Higgs mechanism.

2.4 Spontaneous Symmetry Breaking in $U(1)$

In section 2.3 we generated a massless gauge boson through the requirement of local $U(1)$ gauge invariance. The requirement of gauge invariance precludes the addition of a massive gauge boson of the form $\frac{1}{2}m^2 A_\mu A^\mu$ to the electromagnetic Lagrangian (2.11). However, the Standard Model, particularly the weak interactions, require massive gauge bosons. We must therefore find a means of introducing massive particles to our theory while maintaining the gauge invariance of the initial Lagrangian. In order to accomplish this we introduce the Higgs mechanism. For simplicity, we shall introduce the Higgs mechanism within the confines of the $U(1)$ Lagrangian developed in 2.3 and then expand on this formalism to create a more complete theory including the weak and strong interactions.

Remembering that the Lagrangian is defined as $\mathcal{L} = T - V$, we can write the potential energy of (2.11) as

$$V(\phi) = \mu^2 \phi^* \phi + \lambda(\phi^* \phi)^2. \tag{2.14}$$

If we assume that both μ^2 and λ are positive then $V(\phi)$ has a single minimum at $\phi = 0$ as shown in Figure 2.2. However, if we assume that $\mu^2 < 0$ and $\lambda > 0$ then the potential has a maximum at $\phi = 0$ and minima satisfying the
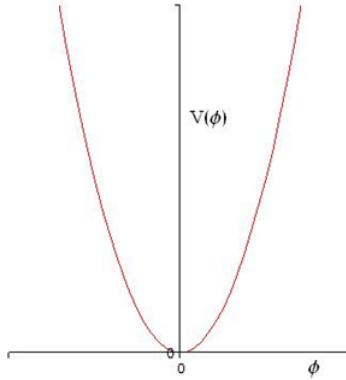
Figure 2.2: The potential energy, (2.12), assuming that both μ^2 and λ are positive.

requirement

$$\frac{\partial V}{\partial \phi} = \phi(2\mu^2 + 4\lambda\phi^2) = 0. \tag{2.15}$$

This requirement is satisfied by two degenerate minima at

$$\phi = \pm v, \; v = \sqrt{\mu^2/2\lambda} \tag{2.16}$$

as shown in Figure 2.3.

We now have two possible stable minima for our potential at $\phi = \pm v$. For small perturbations, we can describe ϕ as the constant v, representing the value of ϕ which minimizes the potential, plus a small perturbative term $\eta(x)$,

$$\phi = \sqrt{\frac{1}{2}}(v + \eta(x)). \tag{2.17}$$

Here we have chosen to expand around the positive value of v, however, this is an arbitrary choice and does not result in a loss of generality since $\phi = -v$ can be reached through reflection symmetry. Inserting equation (2.17) into
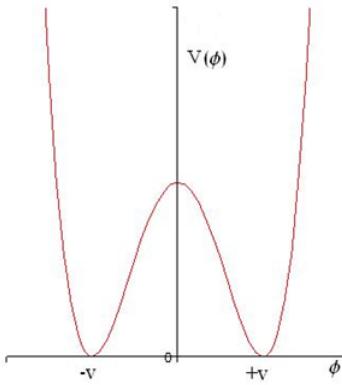
Figure 2.3: Two degenerate minima of the potential energy (2.12) satisfying (2.16).

our invariant Lagrangian (2.12) we find

$$
\begin{aligned}
\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)^2 - v^2 \lambda \eta^2 - \frac{1}{2}e^2 v^2 A_\mu A^\mu - \lambda v \eta^3 - \frac{1}{4}\lambda \eta^4 \\
- \frac{1}{2}e^2 A_\mu A^\mu \eta^2 + v e^2 A_\mu A^\mu \eta - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}.
\end{aligned}
\tag{2.18}
$$

The term $\frac{1}{2}e^2 v^2 A_\mu A^\mu$ in equation (2.18) indicates that A_μ is the field of a massive gauge boson. As discussed in section 2.3, the A_μ field is associated with the photon and the presence of the photon mass term makes this theory unphysical. However, the procedure used to generate this Lagrangian is analogous to the method used to generate massive gauge bosons in electroweak interactions within the Standard Model. This method of generating massive particles is known as spontaneous symmetry breaking. The term spontaneous symmetry breaking refers to the fact that the field ϕ spontaneously breaks the symmetry of the system when it chooses one of the two possible minima shown in Figure 2.3. By utilizing a simple $U(1)$ invariant Lagrangian we were able to explore the formalism of spontaneous symmetry breaking within the confines

18

of scalar fields while avoiding the complexities of more physical systems.

In addition to the massive photon, A_μ, a new massive boson, η, with mass term, $v^2\lambda\eta^2$, has been added to the Lagrangian. Though this Lagrangian is unphysical, η is analogous to the Higgs Boson found in the Standard Model Lagrangian. η is an unavoidable result of the spontaneous symmetry breaking of the field ϕ and the generation of the massive field A_μ. Along with the two boson mass terms, the Lagrangian (2.18) contains kinetic energy terms for both A_μ and η as well as interaction terms such as $\frac{1}{2}e^2 A_\mu A^\mu \eta^2$.

As we have seen in section 2.3, electromagnetic interactions are associated with the $U(1)$ symmetry group. Likewise, weak interactions are associated with the $SU(2)$ symmetry group and strong interactions with the $SU(3)$ symmetry group. In order to incorporate the electromagnetic, weak and strong interactions into one theory, the Standard Model is required to be $SU(3) \times SU(2) \otimes U(1)$ invariant. In the following sections we will discuss the creation of a Lagrangian containing this three-fold symmetry along with the use of spontaneous symmetry breaking to generate massive particles.

2.5 ElectroWeak Interactions and the Higgs Mechanism

We have so far discussed the $U(1)$ structure of electromagnetism. We will now combine our understanding of electromagnetism with the $SU(2)$ structure of
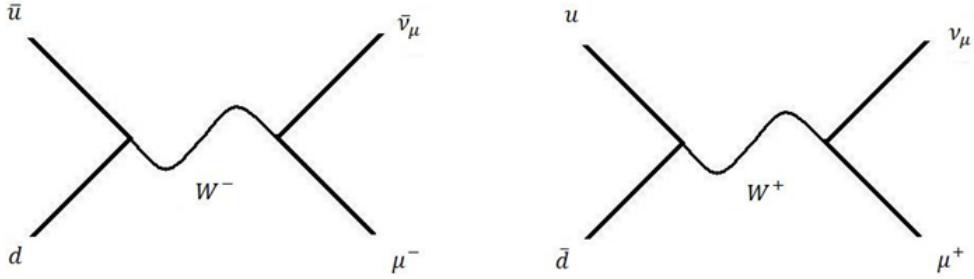
Figure 2.4: Feynman diagrams depicting the primary decay modes of the π^+ and π^- particles via the weak interaction.

weak interactions to form a combined electroweak theory. Before beginning with the formalism of electroweak interactions, it is useful to discuss the properties of weak interactions.

Unlike electromagnetic interactions, which have infinite range, weak interactions occur over very small distances. This short range indicates that the mediating bosons of weak interactions are massive and unstable. Experimental results, such as pion decay shown in Figure 2.4, reveal the need for at least two weak gauge bosons, one with positive and one with negative charge. In addition, electron-neutrino scattering, as shown in Figure 2.5, indicates the need for a neutral weak boson. Combined, these results indicate that any theory describing electroweak interactions must contain four gauge bosons: three massive bosons mediating weak interactions, one neutral (Z) and two with opposite charge (W^+,W^-), and one massless photon meditating electromagnetic interactions.

Experimental results indicate that only chiral left handed particles interact
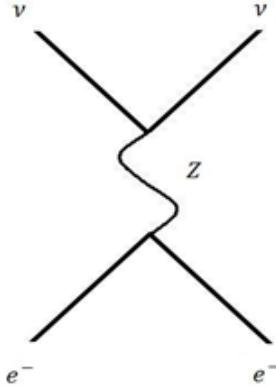
Figure 2.5: Feynman diagram of electron-neutrino scattering via the weak interaction

weakly, while chiral right handed particles are excluded from weak interactions. This property of weak interactions is known as V-A as a result of the vector minus axial-vector couplings found within the electroweak Lagrangian. The chirality of a particle is determined by it's transformation under a Poincare group of transformations. However, chirality is more intuitively understood in relation to helicity. Helicity is the relative orientation of a particle's spin with respect to its direction of motion as shown in Figure 2.6. Particles with spin oriented in the same direction as the particle's momentum are referred to as right handed, while particles with spin and momentum anti-aligned are referred to as left handed. Since the direction of a particle's momentum depends on the frame of the observer for massive particles, it is always possible to Lorentz boost into a frame where a particle's helicity is flipped. For massless particles, such a Lorentz boost is impossible and chirality and helicity are equivalent.
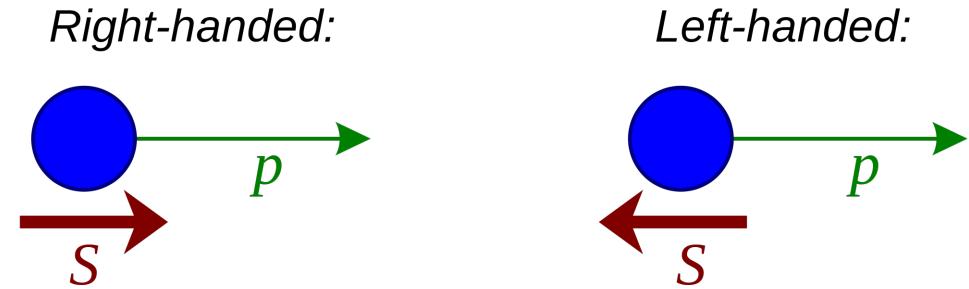
Figure 2.6: The relative orientation of particle spin and momentum in left and right handed particles [7].

Just as electromagnetic interactions can occur only between particles with electric charge, weak interactions can only occur between particles with weak isospin. Left handed particles are described as doublets containing particles with the same weak isospin, T, but different projections of weak isospin onto an arbitrary axis, $T_3 = \pm\frac{1}{2}$. Equation (2.19) is an example of one such doublet involving up and down quarks. This doublet formation allows left handed particles to interact weakly within the electroweak model. In contrast, right handed particles have zero weak isospin and are described as singlets, able to interact through the electromagnetic force, but not through the weak force.

$$\begin{pmatrix} u \\ d \end{pmatrix} = \begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \tag{2.19}$$

In addition to weak isospin, electroweak particles are describe by the weak hypercharge quantum number Y. Weak hypercharge, electric charge and weak isospin are all conserved within weak interactions and are related by the equation

$$Q = T_3 + \frac{Y}{2}. \tag{2.20}$$

Table 2.1 lists the values of weak isospin and weak hypercharge for left and right handed particles. Right handed neutrinos have never been observed and are omitted from Table 2.1.

Particle	Handedness	Q	Y	T_3
e^-, μ^-, τ^-	Left	-1	-1	$-\frac{1}{2}$
ν_e, ν_μ, ν_τ	Left	0	-1	$+\frac{1}{2}$
u, c, t	Left	$+\frac{2}{3}$	$+\frac{1}{3}$	$+\frac{1}{2}$
d, s, b	Left	$-\frac{1}{3}$	$+\frac{1}{3}$	$-\frac{1}{2}$
e^-, μ^-, τ^-	Right	-1	-2	0
u, c, t	Right	$+\frac{2}{3}$	$+\frac{4}{3}$	0
d, s, b	Right	$-\frac{1}{3}$	$-\frac{2}{3}$	0

Table 2.1: Table of electroweak quantum numbers for chiral left and right handed particles.

Weak hypercharge is associated with the $U(1)$ symmetry group while weak isospin is associated with the $SU(2)$ symmetry group, making the symmetry group of electroweak interactions $SU(2) \otimes U(1)$. $SU(2)$ transformations take the form

$$\phi \rightarrow \phi' = e^{i\alpha_a \sigma_a/2}\phi, \tag{2.21}$$

where $a = 1, 2, 3$, α_a are coupling constants and σ_a are the Pauli matrices,

$$\sigma_1 = \begin{pmatrix} 0 & +1 \\ +1 & 0 \end{pmatrix} \tag{2.22}$$

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ +i & 0 \end{pmatrix}$$

$$\sigma_3 = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Using the same technique as we used in the $U(1)$ case, we introduce a covariant derivative in order to maintain the $SU(2) \otimes U(1)$ symmetry of the Lagrangian. The appropriate covariant derivative is of the form

$$\mathcal{D}_\mu = \partial_\mu - ig\frac{\sigma_i}{2}W_\mu^i - \frac{1}{2}ig'YB_\mu. \tag{2.23}$$

Y is simply the identity matrix while g and g' are coupling constants. W_μ^i represents a set of three gauge vector fields associated with the $SU(2)$ symmetry and B_μ is the vector gauge field generated by the $U(1)$ symmetry. Using (2.23) we can write down a simple $SU(2) \otimes U(1)$ invariant Lagrangian of the form

$$\mathcal{L} = -\frac{1}{4}W_{\mu\nu}^i W_i^{\mu\mu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \left|\left(\partial_\mu - \frac{1}{2}ig\sigma_i W_\mu^i - \frac{1}{2}ig'YB_\mu\right)\phi\right|^2 - V(\phi) \tag{2.24}$$

Our goal is to introduce mass terms for the three $W_{\mu\nu}$ gauge fields. These mass terms will be generated by the degrees of freedom in the ϕ field after spontaneous symmetry breaking. As such, we must choose a representation for ϕ which provides at least three degrees of freedom and which is able to interact with $W_{\mu\nu}$. The simplest representation of ϕ that meets these criteria is an SU(2) doublet containing four real scalar fields of the form

$$\phi = \sqrt{\frac{1}{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.25}$$

The first two terms in equation (2.24) represent the kinetic energies of the $W_{\mu\nu}$ and $B_{\mu\nu}$ fields. The third term is the most interesting to our discussion as it holds interactions between the field ϕ and the gauge bosons, and the fourth term represents the potential energy of the field ϕ and takes the familiar form

$$V(\phi) = \mu^2 \phi^* \phi + \lambda (\phi^* \phi)^2. \tag{2.26}$$

Following the same procedure as we used in the $U(1)$ case, we will use spontaneous symmetry breaking to generate massive gauge bosons. In this case, by forcing $\mu^2 < 0$ we will generate an infinitely degenerate vacuum. The shape of $V(\phi)$ is described as a Mexican hat, as can be seen in Figure 2.7, with the possible vacuum states lying in the circular valley within the potential. After generating our infinitely degenerate potential we then force the field, ϕ, to chose a specific vacuum state, v, and then introduce small perturbations to
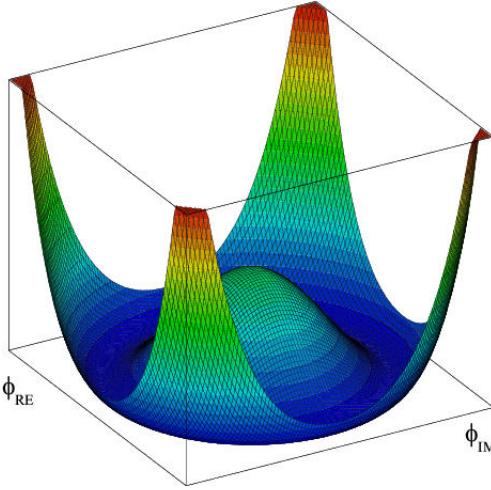
Figure 2.7: Potential energy of the field ϕ under spontaneous symmetry breaking in the electroweak theory [8].

the vacuum state, $h(x)$. The constant vacuum state will generate mass terms for our gauge bosons and the additional perturbation term will introduce a new scalar boson referred to as the Higgs boson.

Reducing our scalar field to a single real scalar we can write the vacuum expectation of ϕ as

$$\phi_0 = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.27}$$

Perturbing this field around the vacuum expectation value we write

$$\phi = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{2.28}$$

where $h(x)$ is the Higgs field. For the moment we will only consider the vacuum expectation of ϕ in order to generate gauge boson mass terms in our

Lagrangian. By inserting the vacuum value of ϕ into the third term in our Lagrangian, (2.24), we find

$$\left| \left(\partial_\mu - \frac{1}{2} i g \sigma_i W^i_\mu - \frac{1}{2} i g' Y B_\mu \right) \begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2. \tag{2.29}$$

Removing the ∂_μ term we are left with only terms relating to W^i_μ and B_μ,

$$\left| \left(-\frac{1}{2} i g \sigma_i W^i_\mu - \frac{1}{2} i g' Y B_\mu \right) \begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2$$

$$= \frac{1}{8} v^2 g^2 (W_\mu)^+ (W^\mu)^- + \frac{1}{8} v^2 \begin{pmatrix} W^3_\mu & B_\mu \end{pmatrix} \begin{pmatrix} g^2 & -gg' \\ -gg' & g^2 \end{pmatrix} \begin{pmatrix} W^{3\,\mu} \\ B^\mu \end{pmatrix}, \tag{2.30}$$

where $W^\pm$ fields are superpositions of the W^1 and W^2 fields,

$$W^\pm = \sqrt{\frac{1}{2}} (W^1 \mp i W^2). \tag{2.31}$$

We can see that the first term in (2.30) is a mass term for the $W^\pm$ bosons with $M_W = \frac{1}{2} v g$. The Second term in (2.30) can be simplified to

$$\frac{1}{8} v^2 \left(g W^3_\mu - g' B_\mu \right)^2, \tag{2.32}$$

which is the mass term of the Z boson, defined to be a superposition of the W^3_μ and B_μ fields,

$$Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2 + g'^2}}. \tag{2.33}$$

The mass of the Z boson is found to be $M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}$. The photon, A_μ, can be shown to have zero mass and is also represented by a superposition of the W_μ^3 and B_μ fields as

$$A_\mu = \frac{gW_\mu^3 + g'B_\mu}{\sqrt{g^2 + g'^2}}. \tag{2.34}$$

We have now created a theory describing electroweak interactions containing three massive gauge bosons and one massless gauge boson, fitting our initial predictions. To accomplish this we introduced an $SU(2) \otimes U(1)$ invariant Lagrangian and then used spontaneous symmetry breaking to generate mass terms for our gauge boson fields. The masses of the weak gauge bosons are not predicted by the Standard Model, however experimental results give the W and Z masses as: $M_W \approx 80 GeV$, $M_Z \approx 91 GeV$. The boson mass terms in our electroweak Lagrangian did not come for free, as we have still to investigate the perturbation term $h(x)$ in our Lagrangian.

The field $h(x)$ represents a neutral, massive scalar boson, known as the Higgs boson, generated by the spontaneous symmetry breaking of the potential, $V(\phi)$. The mass term of the Higgs boson within our electroweak Lagrangian is found by inserting our perturbed field (2.28) into the first term in our Lagrangian potential (2.26). Taking only the term of interest we find

28

$$\mu^2 h^2 = -2\lambda v^2 h^2, \tag{2.35}$$

where we have used the fact that $v^2 = -\mu^2/2\lambda$ as found in equation (2.16). The value of λ is unknown, meaning that the Standard Model does not predict the mass of the Higgs boson. However, recent discoveries at the Large Hadron Collider have shown the mass of the Higgs to be $M_H = 125.36 GeV$ [9].

Since this thesis specifically deals with the Higgs boson decaying via two W bosons as shown in Figure 2.8, let us look at the Higgs coupling to the W bosons. The term of interest in our Lagrangian contains $h(x)$ multiplied by $W_\mu W^\mu$ and is found by inserting the perturbed ϕ field, (2.28), into the third term in our electroweak Lagrangian, (2.24). Doing so yields the Higgs-W interaction term of the form

$$\frac{1}{4} v g^2 h W_\mu^+ W^{-\mu} \tag{2.36}$$

Remembering that the mass of the W boson is given by $M_W = \frac{1}{2} v g$, we note that the coupling of the Higgs boson to the W is given by M_W^2/v. This means that the strength of the Higgs coupling to the W boson is directly dependent on the mass of the W boson. This relation is not unique to the W boson. In fact, the coupling of the Higgs to all particles is dependent on the particle's mass. Using this knowledge, along with the experimentally-determined masses of Standard Model particles, it is possible to predict the branching ratios of the Higgs Boson. This property of the Higgs boson, along with the relative

29
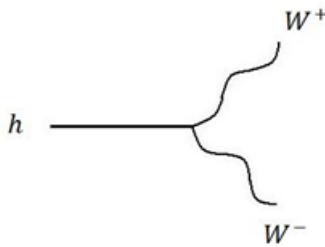
Figure 2.8: Feynman diagram showing the decay of a Higgs boson into a pair of W bosons

coupling of the Higgs to various Standard Model particles will be discussed further in section 2.7.

2.6 Quantum Chromodynamics

The strong interaction is described by a theory known as Quantum Chromodynamics (QCD). Only quarks are affected by the strong force, which is mediated by eight massless gauge bosons known as gluons. Similar to charge in electrodynamics and weak isospin in weak interactions, only particles with color charge interact with the strong force. There are three possible color charges, (red, green, blue), with anti-quarks having anti-color (anti-red, anti-green, anti-blue). Individual quarks are described in terms of a single color and quarks are only observed in color-neutral two or three quark combinations, known as mesons and barions. For barions, this means that they contain one quark of each color (RGB), while for mesons, color neutrality is only achievable

by combining a color and it's anti-color $(R\bar{R})$.

In order to generate eight massless gauge bosons, the Lagrangian of QCD is required to be invariant under $SU(3)$ gauge transformations. The generators of $SU(3)$ transformations are given by

$$T_a = \frac{1}{2}\lambda_a \qquad (2.37)$$

where λ_a are the eight Gell-Mann matrices. The gauge-invariant QCD Lagrangian can be written as

$$\mathcal{L} = \bar{q}(i\gamma^\mu\partial_\mu - m)q - g(\bar{q}\gamma^\mu T_a q)G^a_\mu - \frac{1}{4}G_{\mu\nu}{}^a G^\mu_a\nu. \qquad (2.38)$$

$G_{\mu\nu}$ is the field strength tensor of QCD given by

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - gf_{abc}G^b_\mu G^c_\nu \qquad (2.39)$$

and f_{abc} is determined by the commutation relations of the Gell-Mann matrices,

$$[T_a, T_b] = if_{abc}T_c. \qquad (2.40)$$

The constant g in equation (2.38) is related to the strong coupling constant α_s by

$$\alpha_s = \frac{g^2}{4\pi}. \qquad (2.41)$$

The strong coupling constant is in fact not constant. Rather, the strong coupling depends inversely on the energy of the interaction. This means that at short distances the strong force is relatively weak, while at large distances the strong force becomes stronger. At large enough distances, it becomes energetically favorable for quarks to form bound states with quarks spontaneously generated from the vacuum in a process called hadronization.

2.7 Higgs Boson Creation at the LHC

One of the primary goals in the design of the Large Hadron Collider was the discovery of the Higgs boson. Since the LHC is a proton-proton collider, the primary mode of Higgs creation is through gluon-gluon fusion (ggf). Gluons, the mediating gauge bosons of the strong force, are found within protons and often interact within pp collisions. Since gluons do not interact weakly, they interact with the Higgs boson through a top quark loop as shown in Figure 5.3. Along with being the primary mode of Higgs creation at the LHC, the gluon-gluon fusion process is a convenient creation mode for experimental study as there are no tree-level particles from the initial state found in the final state.

The second most prevalent Higgs production mode at the LHC is vector boson fusion (VBF), which is the focus of this thesis. Vector boson fusion involves the emission of a pair of vector bosons, either W^+W^- or two Z bosons, by initial state quarks. The two vector bosons then fuse to form a Higgs boson as shown in Figure 2.10. This production mechanism results in the two initial
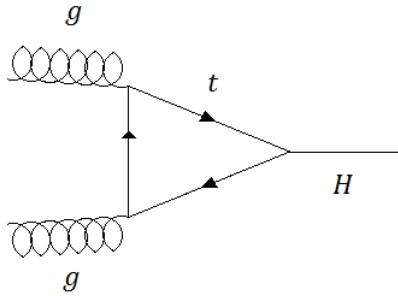
Figure 2.9: Feynman diagram of Higgs production via gluon-gluon fusion.

state quarks being present in the final state, creating a unique Higgs via VBF signature.

In addition to ggf and VBF, co-creation of the Higgs boson with a vector boson or with two top quarks is also possible as shown in Figures 2.11 and 2.12. The cross sections for these Higgs production mechanisms are smaller than the cross sections for ggf and VBF, making them difficult (though not impossible) to study using the current LHC luminosity. Figure 2.13 shows the cross sections for various Higgs production mechanisms as a function of Higgs mass at the LHC. As one can see, ggf is by far the most dominant Higgs production mode.

2.8   Higgs Decay Modes

As discussed in section 2.5 the coupling of the Higgs boson to Standard Model particles is dependent on the mass of the particle. We can use this information, along with the experimentally discovered masses of particles in order to
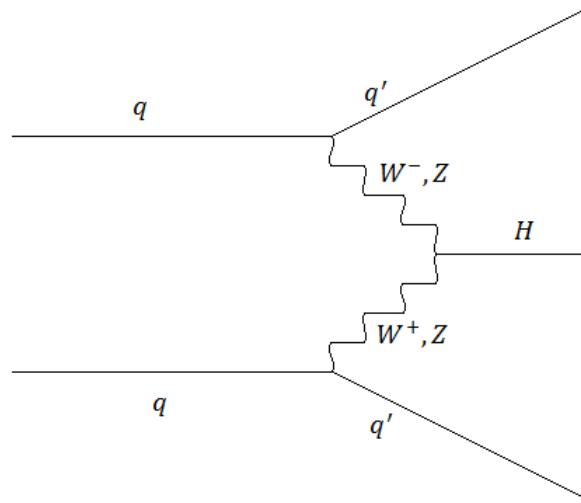
Figure 2.10: Feynman diagram of Higgs production via vector boson fusion.
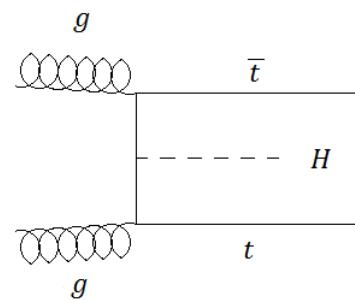
Figure 2.11: Feynman diagram of Higgs production in association with two top quarks.
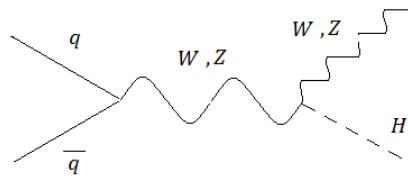
Figure 2.12: Feynman diagram of Higgs production in association with a vector boson.
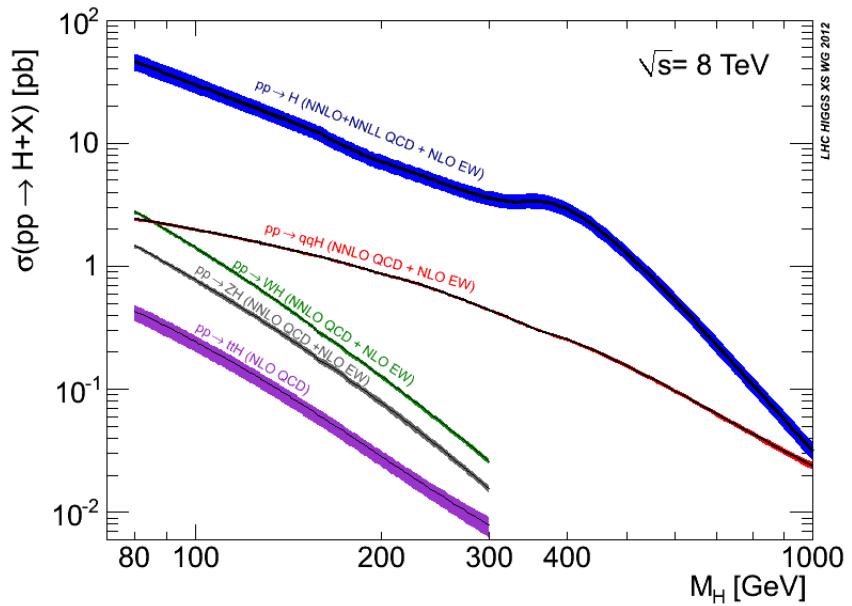
Figure 2.13: Cross sections for various Higgs boson production modes in pp collisions with center of mass energy $\sqrt{s} = 8\,TeV$ at the LHC.
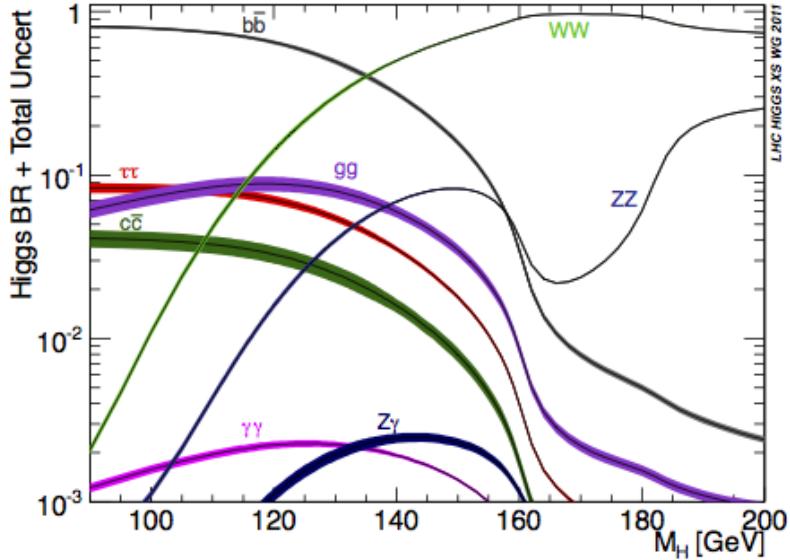
Figure 2.14: Branching ratios of the Higgs boson into Standard Model particles [13].

determine the relative branching ratios of the Higgs boson for various M_H values. A plot of the Higgs branching ratios based on the Higgs boson mass is shown in Figure 2.14. Noting the experimentally discovered Higgs mass of $M_H = 125.36 GeV$, we can see that the Higgs decays primarily into b quarks and W bosons, with significant contributions from Z boson, gluon and τ lepton decays.

In terms of experimental searches, not all Higgs channels are equally viable. From Figure 2.14, one would assume that it would be most fruitful to search for the Higgs boson in the $H \rightarrow bb$ channel. However, the high branching ratio of $H \rightarrow bb$ events is overshadowed by the very large hadronic backgrounds associated with this channel. The most significant measurements of the Higgs

boson have actually come from channles with much smaller branching ratios.

The discovery of the Higgs boson was based primarily on findings in the $H \rightarrow ZZ$ and $H \rightarrow \gamma\gamma$ decay modes. Though these channels have lower branching ratios, they have very distinct signals with no tree level hadronic activity. The $H \rightarrow ZZ \rightarrow llll$ channel in particular has a final state containing four high energy leptons and no hadrons. This signature allows for the removal of most backgrounds, allowing for a high signal-to-background ratio with a limited number of signal events. Likewise, the $H \rightarrow \gamma\gamma$ decay mode results in only two high energy photons in the final state, allowing for the removal of most backgrounds. In addition, both channels allow for a measurement of the Higgs boson mass, which is not possible in the $H \rightarrow WW^*$ and $H \rightarrow \tau\tau$ channels due to the presence of final state neutrinos.

In this thesis, we focus on the $H \rightarrow WW^*$ decay mode in which both W bosons decay leptonically. Specifically, we are interested in Higgs bosons created through the vector boson fusion process decaying into two W bosons, which then decay into two first or second generation leptons plus neutrinos as shown in Figure 2.15. This process is characterized in the final state by 2 quarks from the initial state, two leptons and missing energy. The 2 final state quarks originate within the two colliding high energy protons, meaning that their initial momenta are oriented along the beam line. These quarks are deflected by the emission of a vector boson. However, they maintain much of their forward momentum and are therefore typically found in opposite forward
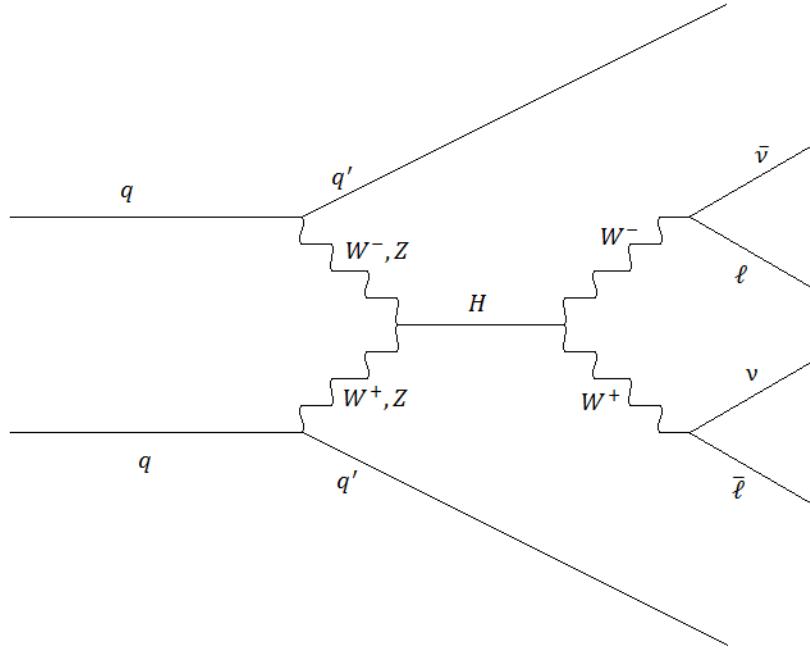
Figure 2.15: Feynman Diagram of VBF Higgs production with the Higgs decaying via two W bosons.

regions of the detector.

The unique jet topology of the VBF $H \rightarrow WW^* \rightarrow l\nu l\nu$ final state is further distinguished by the topology of the final state leptons and neutrinos. First, the two leptons tend to be clumped together due to the spin zero nature of the Higgs boson. The two W bosons produced by the Higgs decay each have spin 1. Therefore, in order to conserve angular momentum, the spins of the two W bosons must be anti-aligned. Since the Higgs boson has zero electric charge, one W boson must be a W^+, which decays leptonically into a neutrino and an e^+ or μ^+ and the other W boson must be a W^-, which decays leptonically into an anti-neutrino and an e^- or μ^-. W decays into

38

τ leptons are considered background within this analysis due to their short lifetimes and distinct signature. Due to the V-A nature of weak interactions, the W^+ boson preferentially emits an e^+ or μ^+ along the direction of its spin. In order to conserve momentum, the corresponding neutrino is emitted in the opposite direction. Likewise, the W^- boson preferentially emits an e^- or μ^- in the direction opposite to its spin and an anti-neutrino along the direction of its spin. The result of these two W boson decays is that the two leptons are emitted preferentially in the same direction, as are the neutrinos. This leptonic decay process is depicted in Figure 2.16. The combination of two clumped leptons, missing energy and two high energy jets close to the beam line create a very distinct signature for VBF $H \rightarrow WW^* \rightarrow l\nu l\nu$ events.
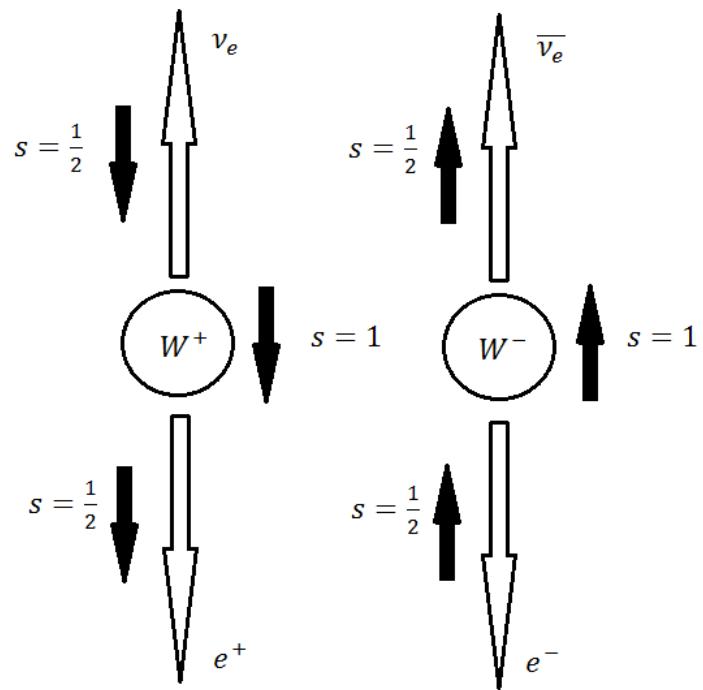
Figure 2.16: The decay of two W bosons with opposite spin. The spin of particles is represented by the solid arrows while the outlined arrows represent momentum.

Chapter 3

Experimental Apparatus

The analysis described in this thesis is based on data collected using the AT-LAS detector, which is one of four major detectors located around the ring of the Large Hadron Collider (LHC). The LHC is the worlds largest particle collider, located in Geneva, Switzerland and operated by the European Organization for Nuclear Research (CERN). The Large Hadron Collider is primarily a proton-proton collider, designed to work at center of mass energies up to $\sqrt{s} = 14$ TeV. The collider lies in a ring, 27 km in circumference and approximately 100 m underground. The ring is located primarily in France, but a section of the ring does cross the border into Switzerland.

The Large Hadron Collider provides particle collisions to seven different experiments, with each experiment having its own detector. The two largest are the ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solendoid) experiments. Both the ATLAS and CMS experiments use multi-purpose

detectors to investigate a variety of physical processes. In addition to the two multi-purpose experiments, there are two other major experiments at the LHC: LHCb (Large Hadron Collider Beauty), which explores specifically b-physics and ALICE (A Large Ion Collider Experiment), which investigates quark-gluon plasma using heavy Ion collisions.

Along with the four major experiments, the LHC also houses three smaller experiments. LHCf (Large Hadron Collider Forward) investigates particles specifically in the forward region along the beam line in order to study cosmic rays. TOTEM (TOTal Elastic and diffractive cross-section Measurement) also studies particles close to the beam line as a supplement to ATLAS and CMS, which can not detect particles in this region. MoEDAL (Monopole and Exotics Detector At the LHC) searches for highly ionizing stable massive particles (SMP) and magnetic monopoles.

3.1 Creation and Acceleration of the Proton Beams

The majority of experiments at the LHC rely on two colliding beams of protons. This requires the creation and acceleration of a beam of pure protons, which can then be split into two separate beams. The two beams move in opposite directions around the LHC ring while they are accelerated to their final collision energy. Once this energy has been reached, the two beams are

collided approximately head-on within the experiment detectors.

In order to create a proton beam, hydrogen gas is passed through an electric field to remove the electrons. Once the electrons have been removed, the protons are injected into a linear accelerator known as Linac2, which is designed to accelerate the protons to an energy of 50 MeV. Linac2 uses a series of radio frequency cavities to charge cylindrical conductors which accelerate protons in a straight line. The conductors ahead of the protons are negatively charged such that they attract the approaching protons. The conductors behind the protons are positively charged such that they push the protons down the beam pipe. This procedure requires that the conductors along the beam pipe are constantly oscillating between positive and negative charge and is ineffective for a continuous stream of protons. Thus the protons are broken into bunches of up to 10^{11} protons.

Once the protons have been accelerated to 50 MeV, they are inserted into a series of increasingly large synchrotrons. Synchrotrons use magnetic fields to bend particles into a circular ring and electromagnetic fields to accelerate particles within the ring. The fields are time-dependent such that the rate at which the fields oscillate increases as the speed of the particles increases. The fields also vary in strength as the protons increase in energy such that the radius of curvature of the protons remains constant. Using three consecutive synchrotrons known as the Proton Synchrotron Booster, the Proton Synchrotron and Super Proton Synchrotron, the protons are accelerated to an

energy of 450 GeV.

After reaching the final output energy of the Super Proton Synchrotron, the protons are injected into the Large Hadron Collider through two beam lines. One beam line circulates the the protons counter-clockwise, while the other circulates clockwise. The two beams travel in completely independent lines except for inside of the four major LHC detectors, ATLAS, CMS, LHCb and ALICE, where the beams intersect. Before colliding, the two beams are accelerated inside the LHC to an energy of 4 TeV. In order to accelerate the proton beams to their final energies, the LHC uses a synchrotron design.

In order to bend the highly energetic protons into a circular path, 1232 superconducting dipole magnets are used. Each dipole magnet is $15m$ long and generates a magnetic field of $8.4T$. Since the LHC houses two separate beams of protons moving in opposite directions, the dipole magnets have a two-in-one design. This design allows for the magnetic field in the two beam pipes to be oriented in opposite directions, causing both beams to bend in the same direction.

In addition to the bending dipole magnets, quadrupole magnets are used to tighten the beam and sextuple and higher order magnets are used to correct for imperfections in the dipole magnetic fields. Near the 4 LHC detectors, the proton beam is narrowed from a width of $2mm$ to $16\mu m$ in order to increase the proton-proton interaction probability and the instantaneous luminosity. This narrowing of the beam is accomplished with a set of three quadrupole magnets

known as the triplet. In total there are eight triplets within the LHC, one on either side of the four detectors (ATLAS, LHCb, ALICE, CMS). The use of superconducting electromagnets is required to achieve the strong magnetic fields necessary to bend the relativistic protons along the circular beam path. The superconducting magnets require extremely cold temperatures to operate and liquid helium is used to keep the magnets at a temperature of $-271.3C$.

Throughout most of the LHC the two proton beams are kept in separate beam pipes in order to prevent unwanted pp collisions. For $70m$ on either side of each LHC detector, the two beams travel in parallel within the same pipe. The beams are then allowed to cross at the center of each detector.

3.2 Luminosity

In any particle physics study using the ATLAS detector, it is important to know the rate at which a process of interest, Higgs boson production for example, occurs in relation to energy and frequency of pp collisions. To put this more simply, we need to be able to estimate the expected number of Higgs bosons in the 2012 ATLAS data using information about the number of recorded collisions. In order to accomplish this, we use the concept of luminosity. Luminosity describes the number of pp events per unit time per unit area and is generally written in units of $[\text{cm}^{-2}\ \text{s}^{-1}]$. We can use information about the number of particles in a bunch, the bunch size and the interaction frequency to calculate the instantaneous luminosity of the LHC. The instan-

taneous luminosity of head-on particle bunches, assuming the bunches are Gaussian-distributed in the transverse direction, is given by

$$\mathcal{L} = \frac{n_b N_b^2 f}{4\pi\sigma_x\sigma_y}. \tag{3.1}$$

N_b is the number of particles per bunch, n_b is the number of bunches per beam, f is the revolution frequency and σ_x and σ_y are the root mean square width of the beam in the transverse directions. The LHC does not collide bunches exactly head-on and therefore a scale factor is used to compensate for the beams' crossing angle. The total luminosity of an LHC run is determined by integrating the instantaneous luminosity over the duration of the run. Using the total luminosity, the expected number of events for a given process is:

$$N = \sigma \int \mathcal{L}dt. \tag{3.2}$$

N is the expected number of events for a given process, σ is the process cross section in units of $[\text{cm}^2]$ and $\int \mathcal{L}dt$ is the integrated luminosity in units of $[\text{cm}^{-2}]$. At the LHC, luminosity is measured in terms of inverse femtobarns, $1 \text{ fb}^{-1} = 10^{39} \text{ cm}^{-2}$, with one inverse femtobarn equaling approximately one trillion proton-proton collisions. Figure 3.1 shows the integrated luminosity for 2010, 2011 and 2012 run periods. This thesis includes analysis of data taken during the 2012 run period.
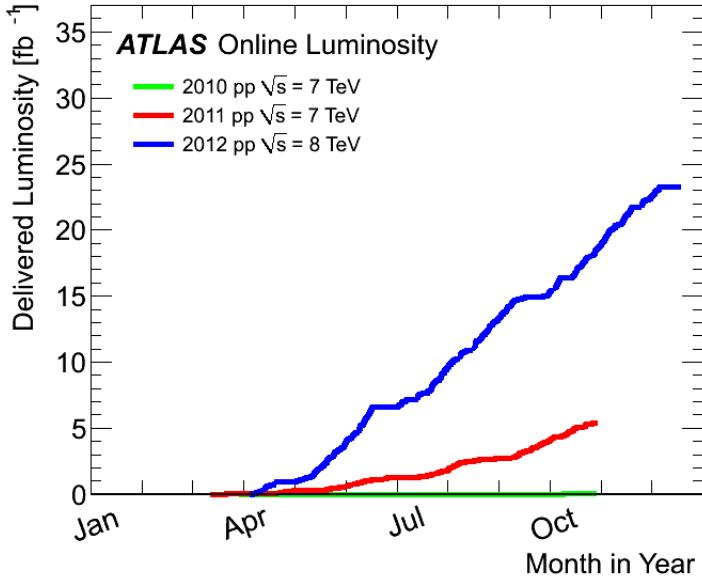
Figure 3.1: Delivered luminosity for 2010, 2011 and 2012 LHC run periods [13].

3.3 The ATLAS Experiment

The ATLAS experiment is a collaboration of over 3000 scientists from 38 countries analyzing data gathered from proton-proton collisions in the ATLAS detector. The ATLAS detector is a general purpose detector, designed to gather information on a wide variety of physics objects at many different energies. The ATLAS detector is made up of a series of smaller detectors, each designed to detect specific types of particles. The detector is laid out as a cylinder with two end caps as shown in Figure 3.3. Each of the subdetectors serves a unique purpose and is designed specifically to detect a certain class of particles.

The inner most set of detectors, known as the inner detector, is designed for

high precision measurements of charged particles. Since the inner detector is closest to the interaction point, it is able to provide crucial information about the location of the interaction as well as detailed measurements of particle momenta. In addition, the inner detector is designed to interact minimally with all particles in order to preserve information about particle momentum.

Outside of the inner detector lie two calorimeters. Calorimeters are designed to serve two purposes: first, they measure the energy and location of incident particles; and second, they absorb the energy of particles to prevent high energy particles besides muons from reaching the muon spectrometer. The final layer of the ATLAS detector is the muon spectrometer, designed specifically to detect high energy muons. Neutrinos, which interact only through the weak interaction, are not detected by the ATLAS detector. The interactions of various particles with the different components of the ATLAS detector are shown in Figure 3.2. Note that various types of particles interact with the detector in distinct ways, making it easier to identify individual particles such as jets, electrons and muons. In the following sections the design and purpose of each of the ATLAS subdetectors will be discussed.

3.3.1 Detector Coordinates

The ATLAS experiment uses a right-handed Cartesian coordinate system in which the x-axis points towards the center of the LHC ring, the y-axis points straight up and the z-axis points along the beam line. The point of interac-
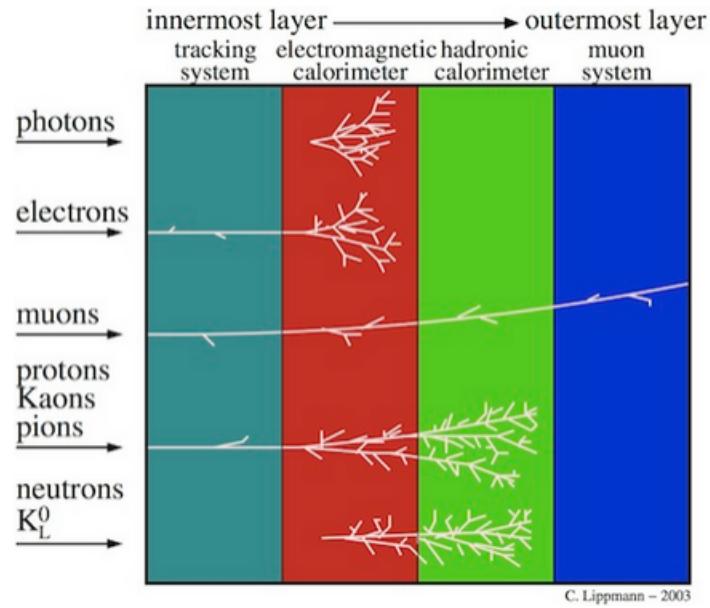
Figure 3.2: A diagram of the interactions of various particles with each of the

ATLAS subdetector systems [12].
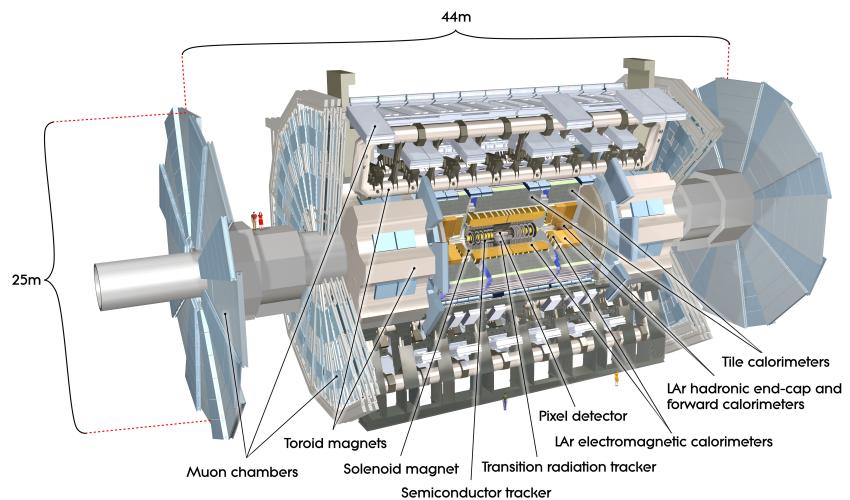
Figure 3.3: Computer generated schematic of the ATLAS Detector [13].

tion at the center of the detector is chosen as the origin for this coordinate system. To describe the location and direction of particles within an event, spherical coordinates are generally used. The transverse direction of a particle is described by the variable ϕ, which can take values from $[-\pi, \pi]$ with $\phi = 0$ corresponding to the x-axis. ϕ increases by moving counter-clockwise around the positive z-axis. The forward-backward direction of the particle is described by the polar angle θ, which can take on values from $[0, \pi]$ and is zero along the positive z-axis.

3.3.2 Detector Variables

When describing objects in high-energy particle collisions, the use of Lorentz-invariant variables is essential since the longitudinal momentum of the interacting partons is unknown. Due to this requirement, final state particles are generally described in terms of transverse momentum, p_T, or transverse energy, E_T and pseudo-rapidity η. The transverse momentum is defined as

$$p_T = |\mathbf{p}| sin\theta \tag{3.3}$$

and transverse energy is defined as

$$E_T = Esin\theta. \tag{3.4}$$

Both variables are Lorentz invariant with respect to boosts along the beam line and are therefore good variables for physics analysis.

In order to preserve Lorentz invariance, the angle of a particle with respect to the beam line is measured using pseudo-rapidity rather than the angular variable θ. Pseudo-rapidity, defined as

$$\eta = \frac{1}{2}\ln\left(\frac{|\mathbf{p}| + p_L}{|\mathbf{p}| - p_L}\right) = -\ln\left(\tan\frac{\theta}{2}\right), \tag{3.5}$$

where p_L is defined as momentum along the beam line, is not in itself Lorentz invariant. However, for particles with high energy and low mass such that $E \approx |\mathbf{p}|$, pseudo-rapidity is a very close approximation to rapidity. Rapidity is defined as

$$y = \frac{1}{2}\ln\frac{E + p_L}{E - p_L}. \tag{3.6}$$

In the limit of massless particles, pseudo-rapidity is equivalent to rapidity. The difference in rapidity between two objects , Δy, is invariant under Lorentz transformations and therefore, for high energy particles with low mass, $\Delta\eta$ is also Lorentz invariant. The value of η varies from 0 to ∞ with $\eta = 0$ corresponding to objects with momentum perpendicular to the beam line and $\eta = \infty$ corresponding to particles moving along the beam line. Due to the Lorentz invariant properties of the pseudo-rapidity, η is used in place of θ do describe angular distance from the beam line. The direction of a particle within the ATLAS detector can therefore be written in an approximately Lorentz invariant form using η and ϕ and the angular distance between two particles can be written as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{3.7}$$

3.3.3 Magnets

A charged particle moving within a magnetic field follows a curved path. The radius of curvature of this path can be used, along with the particle's mass, to calculate the particle's momentum. ATLAS tracking subdetectors are situated within magnetic fields in order to exploit this property. A $2T$ superconducting solenoid located between the inner detector and the electromagnetic calorimeter provides a constant magnetic field in the $+z$-direction to the inner detector. In addition, three superconducting toroid magnets are used to provide magnetic fields within the muon spectrometer. The barrel toroid is constructed using eight separate coils arranged symmetrically in ϕ. The end-cap toroids are also made of eight coils arranged symmetrically in ϕ, but rotated $\pi/8$ radians with respect to the barrel toroids. The magnetic field of the muon spectrometer is designed to be perpendicular to the direction of motion for most muons within the muon spectrometer. The layout of the toroid and solenoid magnet systems is shown in Figure 3.4.

3.3.4 Inner Detector

The inner detector is located at the center of the ATLAS detector and is closest to the proton-proton interaction point with an inner radius of $50mm$.
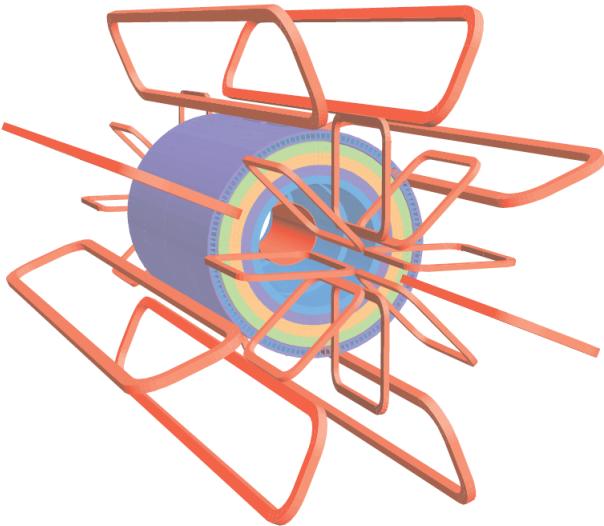
Figure 3.4: Schematic of the toroid and solenoid magnets in the ATLAS detector. Orange areas represent the magnets while the mulit-colored concentric cylinders represent the hadronic and electromagnetic calorimeters.

The inner detector lies within a $2T$ magnetic field and consists of three separate subdetectors: the pixel detector, semiconductor tracker (SCT) and transition radiation tracker (TRT). The three systems are laid out in concentric cylinders along the barrel and as discs perpendicular to the beam pipe on the end caps. The inner detector covers the region $|\eta| < 2.5$.

The goal of the inner detector is to provide precision measurements of the position and momentum of charged particles as they are bent by the magnetic field. In addition, the inner detector is capable of reconstructing the position of interaction vertices. The proper reconstruction of vertices is essential since there are an average of 20 particle collisions per bunch crossing. In order to study an event of interest, such as a Higgs boson decay, the ATLAS detector

needs to be able to filter out all particles not associated with the primary interaction vertex. In addition, the high-precision vertex reconstruction of the ATLAS inner detector helps to identify longer-lived particles such as b-hardons and τ leptons, which can travel a significant distance before decaying. Components of the inner detector must also be robust against radiation damage due to the high rate of incident particles. Finally, the inner detector is designed to absorb as little energy as possible from incident particles in order to allow for precision energy measurements within the calorimeter.

3.3.4.1   Pixel Detector

The pixel detector is the inner most layer of the inner detector and consists of three concentric cylindrical detectors in the barrel region and three consecutive disc layers in the end caps as shown in Figure 3.5. The pixel detector is specifically designed to have very high resolution in order to accurately measure the position, momentum and impact parameter of incident charged particles. For most particle tracks the pixel detector provides 3 high precision measurements of η and ϕ, one for each pixel layer.

The inner-most layer of the pixel detector is positioned as close as possible to the interaction point at a radius of 50.5 mm with the outer two layers at a radius of 90 mm and 120 mm respectively. When high energy particles interact with the silicon chips, electrons are freed from the silicon. These electrons generate an electric current which is then processed by readout chips. The
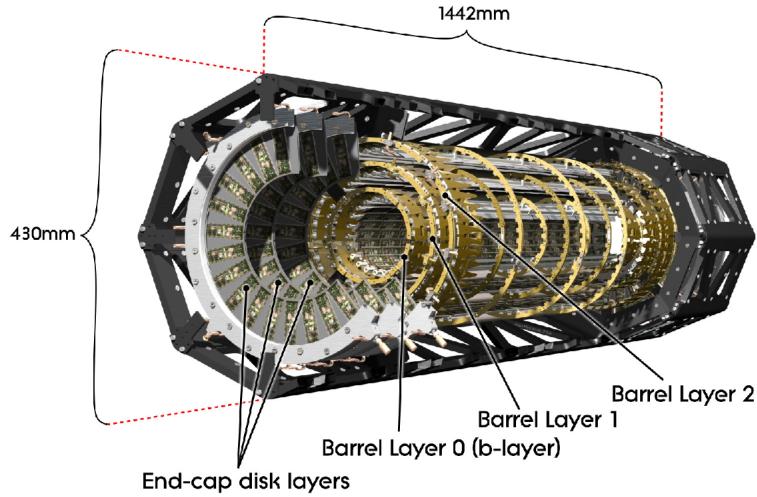
Figure 3.5: Schematic diagram of the ATLAS pixel detector [13].

pixel chips are arranged into 1456 modules in the barrel region and 288 in the end cap with each module holding 46080 pixels. The output from the pixel sensors is interpreted by 16 read-out chips. The pixel detector has a resolution of $10\mu m$ in ϕ and $115\mu m$ in the z direction along the barrel and in the R direction in the end caps.

3.3.4.2 Semiconductor Tracker

Once particles have traversed the pixel detector, they enter the Semiconductor Tracker (SCT). The semiconductor tracker uses silicon technology similar to the pixel detector in order to provide additional high-precision tracking of charged particles. The SCT consists of 4088 modules, 2112 in the barrel and 1976 in the end caps. Each module contains silicon microstrips, 12 cm in length and 80 μm pitch. There are four layers of microstrips in the barrel region and

nine layers in each endcap, providing an average of 8 hits per track. In the end cap the microstrips are mounted back-to-back in order to increase resolution. The SCT has a resolution of $7\mu m$ in ϕ and $580\mu m$ in the z direction along the barrel and in the R direction in the end cap.

3.3.4.3   Transition Radiation Tracker

The Transition Radiation Tracker (TRT) is the final piece of the inner detector and is positioned outside of the SCT. The TRT detector is composed of approximately 350,000 straw tubes aligned parallel to the beam along the barrel and radially in the end caps and provides approximately 36 additional hits for each track. The drift tubes are 4 mm in diameter with a 31 μm diameter wire at the center. The tubes are filled with a gas mixture of 70% Xe, 27% CO_2 and 3% O_2. This gas is ionized by charged particles passing through the detector causing electrons and ions to drift toward the anode wire. These electrons and ions result in a current which is read out by the detector.

The TRT detector has lower granularity than the SCT and pixel detectors, but is able to measure more points per particle and therefore contributes significantly to the measurement of track momentum. In addition, the TRT detector is capable of electron identification by detecting transition radiation created when relativistic charged particles pass between different materials. In order to induce transition radiation in electrons passing through the detector, the space between the straws is filled with an array of polymer fibers.

Electrons generate more intense transition radiation than other particles since the intensity of transition radiation is proportional to the Lorentz factor of a particle, $\gamma = E/m$. In order to help distinguish between electrons and heavier hadronic particles the TRT operates at two different thresholds: a lower threshold used to detect radiation from hadrons and minimum ionizing particles, and a higher setting used to detect intense transition radiation from electrons. Due to the long length of the tubes, the TRT has highly limited resolution in the z direction along the barrel and in the radial direction in the end caps. In the ϕ plane the TRT has a resolution of $130\mu m$.

3.3.5   Calorimeters

The inner detector is designed specifically to detect charged particles which ionize as they move through the detector, while neutral particles are able to pass through the inner detector without being measured. In order to measure the momentum of neutral particles, such as photons and neutral π^0 mesons, as well as to assist in the measurement of charged particles, the ATLAS experiment employs a calorimeter system.

The ATLAS detector contains two concentric calorimeters: the EM calorimeter designed to detect electrons and photons; and the hadronic calorimeter designed to measure hadronic showers escaping the EM calorimeter. Both systems employ a series of alternating layers of dense absorbers and active material. The absorbing layers are designed to interact with particles passing
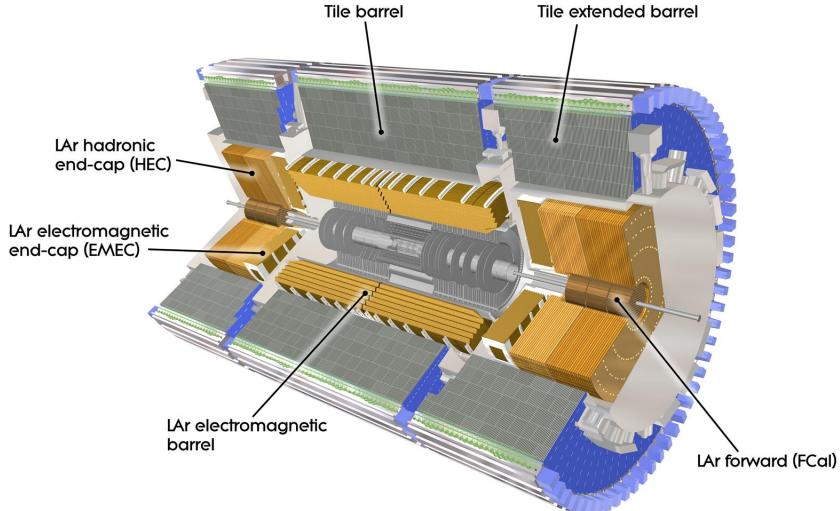
Figure 3.6: Diagram of the ATLAS calorimeter system [13]

through the calorimeter, creating a shower of particles. These showered parti-
cles are then measured by the active layer. The number of particles created in
the shower is proportional to the energy of the incident particle, allowing for
a measurement of the incident particle's energy. It is important that all par-
ticles except muons and neutrinos are completely absorbed by the calorimeter
systems. In order to accomplish this goal, very dense absorbing materials are
used in order to shorten the interaction lengths and many layers of material
are used to ensure total absorption.

3.3.5.1 Electromagnetic Calorimeter

The EM calorimeter is designed to measure electrons and photons using an
accordion design. To ensure complete absorption of electrons and photons, the
EM calorimeter contains 24 radiation lengths, X_0, in the barrel region and 26

in the end cap. A radiation length is the distance a charged particle must travel in order to have $1/e$ of its initial energy. The radiation length of a charged particle varies based on the material through which the particle is traveling, with more dense materials resulting in shorter radiation lengths. In order to reduce the radiation length and minimize the size of the EM calorimeter, the absorbing layers are made of lead. Placed between the lead absorbing layers are active layers containing liquid argon.

The EM calorimeter is composed of two separate sections, a barrel-region covering the range of $|\eta| < 1.475$ and an end cap with two coaxial wheels covering $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$ respectively. The EM calorimeter contains three layers in the region $|\eta| < 2.5$, which are used for precision measurements and only contains two layers in the region $2.5 < |\eta| < 3.2$, resulting in coarser granularity. To account for energy loss of particles in the inner detector, a pre-sampling liquid argon layer is employed in the region $|\eta| < 1.8$. The EM calorimeter provides an energy resolution of $\sigma_E/E = 10\%/\sqrt{E} \oplus 7\%$.

3.3.5.2 Hadronic Calorimeter

Hadronic showers escaping the EM calorimeter are detected and absorbed by the hadronic calorimeter. The hadronic calorimeter is composed of four separate components: the tile barrel, tile extended barrel, the hadronic end cap calorimeter (HEC) and the forward calorimeter (FCal).

The tile barrel covers the central interaction region of $|\eta| < 1.0$ with the

two extended tile barrels covering the regions $0.8 < |\eta| < 1.7$ on either side of the central barrel. These calorimeters use steel absorbers and active layers composed of polystyrene scintillator tiles. Showers within the active layers produce photons, which are amplified and converted to an electric current via photo-multiplier tubes.

The hadronic end-caps cover the region of $1.5 < |\eta| < 3.2$ and are composed of copper absorbing layers and liquid argon active material. The HEC surrounds the forward calorimeter, which covers the region $3.1 < |\eta| < 4.9$ and is made up of three separate layers. The inner most layer of the FCal uses copper as the absorber, with the goal of measuring electromagnetic showers. The outer two layers of the FCal use tungsten absorbers in order to measure hadronic showers. All three layers use liquid argon as the active material.

Hadronic showers produce fewer particles than EM showers, leading to smaller statistics, larger statistical fluctuations and lower energy resolution in the hadronic calorimeter. The HEC and tile calorimeters have a combined resolution of $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$ while the forward calorimeter has a resolution of $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$.

3.3.6   Muon Spectrometer

The outermost system of the ATLAS detector is the muon spectrometer, which is designed to provide standalone triggering on high-p_T muons while also measuring the properties of incident muons with high efficiency and resolution.

To accomplish this goal, the muon spectrometer is made up of four separate components: monitored drift tubes (MDT), cathode strip chambers (CSC), resistive plate chambers (RPC) and thin gap chambers (TGC). The thin gap chambers and resistive plate chambers provide triggering signals in the region $|\eta| < 2.4$. High resolution tracking is performed by the monitored drift tubes with the exception of regions of high $|\eta|$. For regions of $2.0 < |\eta| < 2.7$ cathode strip chambers are used as they provide higher granularity and are better suited to the high interaction rate in these areas. The two high resolution systems cover the region $|\eta| < 2.7$.

A magnetic field is generated throughout the muon spectrometer using three superconducting toroid magnets. The barrel toroid consists of eight coils spaced symmetrically within the muon spectrometer and provides a magnetic field in the region $|\eta| < 1.0$. Two end cap toroid magnets provide magnetic fields in the region $1.4 < |\eta| < 2.7$. The magnetic field in the transition region, $1.0 < |\eta| < 1.4$, is provided by a combination of the barrel and end-cap magnets. The layout of the various systems within the muon spectrometer can be seen in Figure 3.7.

3.3.6.1 Monitored Drift Tubes

The monitored drift tubes, designed for high-resolution tracking of incident muons, are made of extruded aluminum tubes, 30 mm in diameter, containing a mixture of 93% Ar and 7% CO_2 gasses. At the center of each tube is a
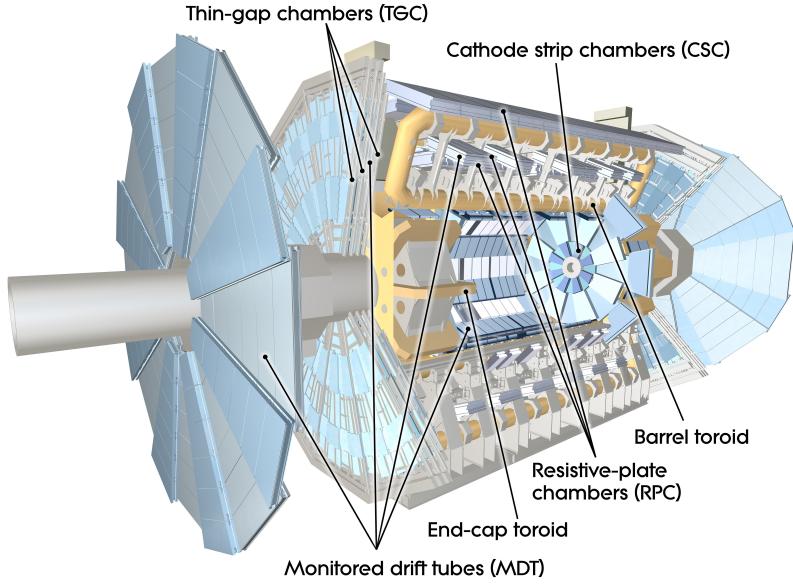
Figure 3.7: Diagram of the ATLAS Muon Spectrometer [13].

tungsten-rhenium anode wire with a diameter of $50\mu m$. The gas is ionized by passing charged particles, causing electrons to drift toward the anode, creating a current as shown in Figure 3.8. As a muon transverses an MDT many electrons are freed. By measuring the time of arrival of these electrons, the distance of closest approach of the incident muon to the anode wire can be estimated. The monitored drift tubes are split into a barrel region covering $|\eta| < 1.1$ and an end-cap covering the range $1.1 < |\eta| < 2.7$. Each MDT has an resolution of $80\mu m$.

3.3.6.2 Cathode Strip Chambers

Cathode strip chambers have a higher granularity than monitored drift tubes and are therefore better suited to the high rates of incident particles found
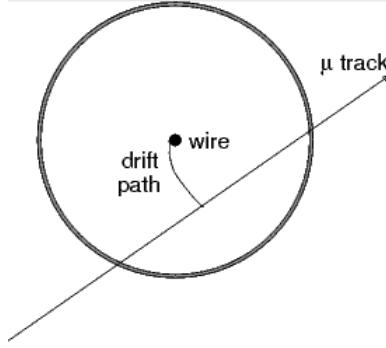
Figure 3.8: Schematic diagram of a muon ionizing gas within a monitored drift tube in the muon spectrometer. The resulting electron drifts towards the anode, resulting in a current [15].
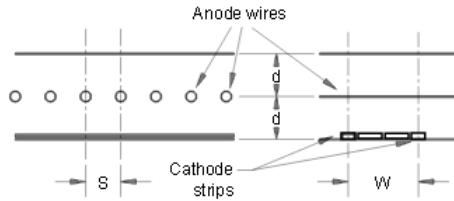
Figure 3.9: Schematic diagram of a cathode strip chamber [15].

in the inner-most layer of the end-cap covering the range $2.0 < |\eta| < 2.7$. Cathode strip chambers are multiwire proportional chambers with cathode strip readout. The spacing of the anode wires is equal to the radial distance between the anode wires and cathode strips as shown in Figure 3.9. A precision measurement of the muon coordinates is made by measuring the charge induced on the cathode due to the avalanche formed on the anode wire. Cathode strip chambers have a resolution of $5mm$ in ϕ and $60\mu m$ in z.

3.3.6.3   Resistive Plate Chambers

Resistive plate chambers are capable of fast and coarse tracking of muons within the muon spectrometer and are used in the ATLAS trigger systems as well as for azimuthal coordinate measurements. Resistive plate chambers consist of two insulating plates separated by a gas layer and held within an electric field. Resistive plate chambers are used within the barrel of the muon spectrometer in the range $|\eta| < 1.05$. Resistive plate chambers have a resolution of $10mm$ in ϕ and no appreciable resolution in z.

3.3.6.4   Thin Gap Chambers

Thin gap chambers are used for triggering and to assist particle tracking in the range $1.05 < |\eta| < 2.4$. Thin gap chambers are very similar to cathode strip chambers except that the anode-to-anode spacing is larger than the anode-to-cathode spacing. This spacing allows for faster signals with good resolution, making thin gap chambers useful for triggering. In addition, thin gap chambers have high granularity and rate-tolerance, making them ideal for the high particle flux found at large values of η. Thin gap chambers have a resolution of $3mm$ in ϕ and $2mm$ in z.

3.3.7   Trigger System

Within the ATLAS detector, proton-proton collisions occur at a rate of approximately $8GHz$. In contrast, the ATLAS computing systems can only store

events at a rate of approximately $200Hz$. This discrepancy requires that only a small percentage of pp collisions are stored for analysis, while the vast majority of pp collisions are discarded. In order to preferentially store events containing interesting physics, a 3-tier trigger system is implemented. The trigger system looks for characteristics typical of events containing interesting phyiscs, such as large missing transverse energy and high p_T jets or leptons.

The first step in the trigger system is the hardware-based level 1 trigger (L1), which is used to pre-select events for further investigation by the level 2 trigger (L2). The goal of the L1 trigger is to eliminate events which do not contain well-defined physics objects and to determine regions of interest for further study by the L2 trigger. The L1 trigger is composed of two separate pieces: the L1 muon trigger and the L1 calo trigger. The L1 muon trigger uses information from the RPC and TGC components of the muon spectrometer to test for low and high p_T muons. By comparing hit patterns in the RPC and TGC to pre-defined look-up tables, the L1 muon trigger is able find events with likely muon candidates. The L1 calo trigger uses coarse granularity calorimeter information to select events containing objects such as leptons, photons, jets and missing transverse energy. Together the two L1 trigger systems reduce the rate of events moving on to the L2 trigger system to $75kHz$. In order to keep pace with this high rate, the L1 trigger system is able to analyze events within $2.5\mu s$.

After passing the L1 trigger, events are processed by the software-based

level 2 trigger (L2). The L2 trigger uses reconstruction algorithms in order to analyze information from the subdetectors in regions of interest of the detector defined by the L1 trigger. The L2 trigger is able to analyze an event within $40ms$ and reduces the event rate to approximately 3 kHz. Events passing the L2 trigger are passed on to the final event filter trigger.

The event filter (EF) trigger is the most thorough and sophisticated of the three triggers. The EF trigger performs a full reconstruction of the event, a process which takes approximately 4 seconds. Due to the large amount of time required to process events in the EF trigger, parallel processors are employed to analyze multiple events simultaneously. The EF trigger selects events based on object p_T and object quality requirements in order to reduce the event rate to approximately $200Hz$.

Chapter 4

Object and Event Selection

In order to eliminate the majority of background events, a cut-based approach is used. The full set of cuts are known as preselection. These cuts are optimized using Monte Carlo samples in order to retain the maximum number of signal events while removing as many background events as possible. Once all relevant cuts have been applied, a support vector machine analysis is conducted using relevant event parameters as input.

In order to properly distinguish signal and background events, the identification and selection of objects such as leptons and jets is crucial. In this section, the identification criteria for event objects will be discussed. In addition, this section will discuss the parameters used for preselection cuts and as SVM inputs.

4.1 Objects

4.1.1 Jets

Hadronic particles, such as pions, protons and b-quarks, are detected as showers of particles within the ATLAS calorimeter systems and are referred to as jets. As discussed in Chapter 3, particles incident on the ATLAS calorimeter systems interact with the calorimeter's absorbing layers, resulting in the decay of the incident particle into many lower-energy particles. These lower energy particles then also decay and are further absorbed and measured by the calorimeter. This process is repeated until all of the original energy of the incident particle has been measured and absorbed.

The result of this decay process is a cascade of particles emitted from the initial point of interaction. Because of the showering aspect of these hadronic decays, the energy of the incident particle is not measured by a single cell in the calorimeter, rather it is measured by combining measurements of topologically connected calorimeter cells in order to determine the energy of the initial particle. Figure 4.1 shows a plot of the energy measured by cells within the hadronic calorimeter. As you can see, there are several cells which contain a large amount of energy over the background. In order to determine the energy of the incident jet, the energy deposited in each of these cells must be summed.

In order to determine the energy of the incident particle, jets are reconstructed using the anti-k_T algorithm [25] with a distance parameter of
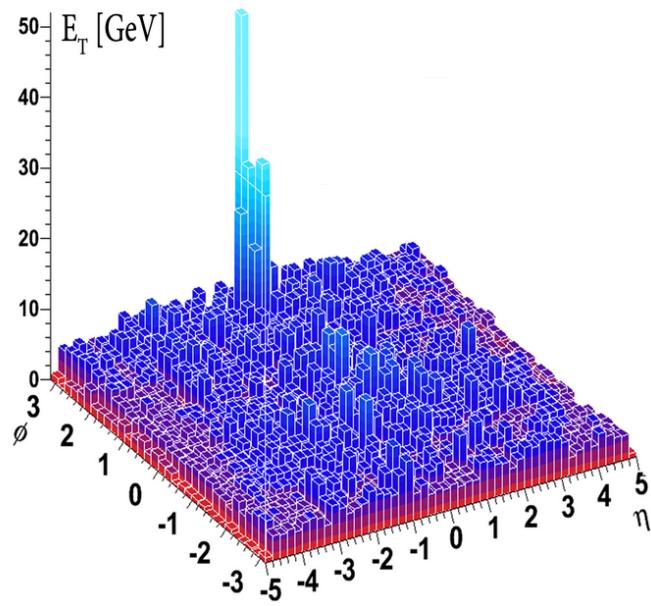
Figure 4.1: Plot showing the energy distribution of cells within the ATLAS calorimeter systems. The set of clusters containing large amounts of E_T represent a jet [13].

$R = 0.4$. The algorithm uses topological energy clusters created from connected calorimeter cells with a significant signal-to-noise ratio. The jet energies are initially reconstructed at the EM scale by measuring the energy of electromagnetic shower particles deposited in the calorimeter. In order to increase resolution, the topological clusters are then scaled using the Local Cluster Weight (LCW) method. Finally, after the LCW method has been applied, the jets are calibrated using the jet energy scale (JES) to compensate for pile-up, Monte Carlo and other in-situ effects [24]. The jet collection used for this analysis is therefore referred to as Anti-k_T4 LCW+JES.

After reconstruction, jets are required to pass LOOSER cleaning criteria, designed to remove jets due to cosmic rays, beam induced background and calorimeter noise [26]. Finally, jets are required to pass the following series of cuts on η, p_T and jet vertex fraction (JVF):

- $|\eta| < 4.5$

- $p_T > 25(30)GeV$ for $|\eta| < (\geq)2.4$

- $|JVF| > 0.5$ for $p_T < 50GeV$ and $|\eta| < 2.4$.

The jet vertex fraction (JVF) measures how well the jet points back to the primary vertex and a cut on JVF is effective at eliminating pileup jets. The other cuts are optimized through MC studies to create the highest possible statistical significance for H^* analyses and to reduce pileup effects [27].

70

Due to the major top quark background in the HWW VBF channel, proper tagging of b-jets is essential. Jets within the range $|\eta| < 2.4$ are tagged using the MV1 tagging algorithm with an efficiency threshold of 85% [28]. Events containing a b-tagged jet with $p_T > 25 GeV$ are vetoed.

4.1.2 Electrons

Electrons are reconstructed using a combination of information from the EM calorimeter and the inner detector. Both multivariate and cut-based methods are used to select electrons depending on the electron's transverse momentum. Low energy electrons, with $E_T < 25 GeV$, are selected using the multivariate electron likelihood method [29] and are required to meet the requirements of the "Very Tight Likelihood" criteria. For electrons with $E_T > 25 GeV$ the cut-based "Medium++" criteria is applied, in addition to two additional cuts imposed in order to improve conversion rejection. In addition to the normal "Medium++" requirements, electrons must have at least one hit in the innermost pixel layer if a hit is expected and their conversion flag is required to be false.

In addition to the above selection criteria, electrons must also pass isolation requirements to remove jets misidentified as electrons. For the calorimeter, the total E_T deposited in topological clusters surrounding the electron candidate is summed, with corrections applied to account for the energy of the electron, underlying event and pile-up effects. For this analysis, the topological clusters

in a cone of $\Delta R = 0.3$ around the electron are considered and the resulting variable is labeled topoEtConeCor(0.3). Electrons with too much energy in the surrounding calorimeter clusters are vetoed.

Isolation criteria are also applied to electron tracks in the inner detector in a similar manner. For the inner detector a cone of $\Delta R = 0.3$ or $\Delta R = 0.4$ is used depending on the p_T of the electron and the resulting variables are labeled Ptcone(0.3) and Ptcone(0.4).

Finally, to eliminate electrons which do not originate from the primary vertex, cuts on the longitudinal impact parameter (z_0) and transverse impact parameter (d_0) are applied. A cut is made on $z_0 sin\theta$ to account for the fact that forward leptons tend to have a larger projection onto the z-axis and therefore a larger uncertainty. The full electron selection criteria is described in Table 4.1.

4.1.3 Muons

Muons are unique within the ATLAS experiment in that they are long-lived and do not decay within the ATLAS detector. In addition, muons lose only a small fraction of their energy within the ATLAS calorimeter systems. As a result, Muons are reconstructed using all of the ATLAS subdetectors. Primarily, muons are reconstructed by combining a track from the inner detector

$E_T GeV$	Electron ID	Calorimeter Isolation topoEtConeCorr	Track Isolation Ptcone	Impact Paramters
10-15		iso(0.3)$/E_T < 0.20$	iso(0.4)$/E_T < 0.06$	
15-20	Very Tight LH	iso(0.3)$/E_T < 0.24$	iso(0.3)$/E_T < 0.08$	$d_0/\sigma_{d_0} < 3.0,$
20-25		iso(0.3)$/E_T < 0.28$	iso(0.3)$/E_T < 0.10$	$z_0 sin\theta < 0.4mm$
> 25	Medium++			

Table 4.1: Electron Selection by E_T Bin

with a matching track in the muon spectrometer. Muons reconstructed in this manner are described as "combined" muons [30]. Muon candidates must pass a series of cuts to ensure the track is well-reconstructed. These cuts require that the muon track have hits in the pixel, SCT and TRT subdetectors as well as having a track in all three layers of the muon spectrometer.

Similar to the electron selection requirements, muons must pass isolation cuts in both the calorimeter and inner detector in addition to cuts on the impact parameter. The impact parameter and isolation cuts are described in table 4.2.

4.1.4 Missing Transverse Energy

The ATLAS detector is not designed to measure neutrinos due to the extremely small cross section of neutrino interactions. However, it is possible to

E_T (GeV)	Calo. Isolation topoEtConeCor	Track Isolation Ptcone	Impact Parameters
10-15	iso(0.3)/E_T < 0.06	iso(0.4)/E_T < 0.06	
15-20	iso(0.3)/E_T < 0.12	iso(0.3)/E_T < 0.08	$d_0/\sigma_{d_0} < 3.0$,
20-25	iso(0.3)/E_T < 0.18	iso(0.3)/E_T < 0.12	$z_0 sin\theta < 1.0mm$
> 25	iso(0.3)/E_T < 0.30		

Table 4.2: Table of muon isolation cuts by p_T.

estimate the energy of neutrinos using the concept of momentum conservation. For example, consider a W^- boson at rest decaying into an electron and an electron anti-neutrino. Using the ATLAS detector, it is possible to measure the transverse momentum of the electron. We can then use two pieces of information to determine the transverse momentum of the neutrino. First, we know that the initial transverse momentum of the W boson is zero, and second we know that momentum conservation dictates that the transverse momentum of the final electron-neutrino system must equal the transverse momentum of the initial W boson. We can therefore calculate the transverse momentum of the neutrino as having the same magnitude as the electron's momentum, but in the opposite direction such that the final electron-neutrino system has zero total transverse momentum.

In the simple W boson system, we know that the final energy of the system should be zero. We call the difference between our measured transverse energy

$(E_T^{measured})$ and our expected transverse energy $(E_T^{expected} = 0)$ the missing transverse energy (E_T^{miss}). We can calculate the missing transverse energy as

$$E_T^{measured} + E_T^{miss} = E_T^{expected} = 0. \tag{4.1}$$

We can easily extend this concept to more complicated events involving multiple particles. We know that the $p-p$ collisions within the ATLAS detector occur approximately head-on, resulting in zero transverse momentum in the initial system. Therefore, the total transverse momentum of the final state objects should also be zero and we can calculate the E_T^{miss} using equation (4.1).

For this analysis, the E_T^{miss} is reconstructed from calorimeter and inner detector track information with jet objects replaced with information from the calorimeter [27]. This missing energy definition is denoted by $E_T^{miss,J-TRK}$.

4.1.5 Overlap Removal

In the case that two objects overlap within the $\eta - \phi$ plane, the following overlap removal has been applied.

- If an electron and muon overlap within $\Delta R < 0.1$, the muon is kept and the electron is rejected.

- If an electron is within $\Delta R < 0.05$ of any muon, the event is rejected.

- If two electrons overlap within $\Delta R < 0.1$, the higher p_T lepton is kept.

- If a jet and electron overlap within $\Delta R < 0.3$, the electron is kept and the jet is rejected.

- If a jet and muon overlap within $\Delta R < 0.3$, the jet is kept and the muon is rejected.

ΔR is defined as $\Delta R = \sqrt{(\Delta\eta^2) + (\Delta\phi^2)}$.

4.2   Observables and Event Selection

4.2.0.1   Observables

The analysis described in this thesis relies on support vector machines to separate signal and background events. In order to properly distinguish events, the support vector machines rely on well-defined input parameters designed to exploit differences in signal and background events. In addition, before events are processed by the SVM analysis, it is important to remove as many background events as possible using a cut-based approach. This section will describe the observables used in both the cut-based and SVM portions of our analysis as well as detailing the cuts used to create the SVM input samples.

As described in chapter 2, the topology of VBF $H \rightarrow WW^* \rightarrow l\nu l\nu$ events is quite unique. VBF HWW events are characterized by two high-energy jets in forward regions of the detector along with two central leptons and missing energy. The parameters chosen for pre-selection cuts and SVM inputs are

designed to exploit this unique topology. The parameters used for preselection and SVM input are as follows:

- $m_{\tau\tau}$: In order to reduce background events due to $Z \to \tau\tau$ decays, a cut on the invariant mass of the two τ system is imposed. By assuming that each measured lepton is the product of a $\tau \to l\nu\nu$ decay and by using the missing energy to estimate the energy of the neutrinos, $m_{\tau\tau}$ can be calculated. Any event in which the calculated value of $m_{\tau\tau}$ is within 25 GeV of the Z mass is vetoed.

- $N_{b-jet} = 0$: VBF HWW events involve two initial state quarks interacting to form a Higgs Boson. These initial state quarks are provided by the interacting protons and are therefore unlikely to be b-quarks. Background top quark and $t\bar{t}$ events, on the other hand, are extremely likely to contain at least one b jet. Therefore, eliminating events containing a b-tagged jet is effective at removing background events while preserving most signal events. b-tagged jets are defined as any jet passing the MV1 b-tagging criteria at an 85% operating point and having $p_T > 25 GeV$.

- Central Jet Veto (CJV): VBF Higgs events ideally contain two forward jets and no extra hadronic activity. The presence of additional high-pt jets indicates background interactions within the event. Therefore, any event containing a third jet with $p_T > 20 GeV$ and with η between those of the two tagged-jets is vetoed.

- m_{ll}, $\Delta\phi_{ll}$: Due to the spin-zero nature of the Higgs boson, the final state leptons in $H \rightarrow WW^* \rightarrow l\nu l\nu$ decays are preferentially emitted in the same direction. Due to this effect, leptons resulting from $H \rightarrow WW^*$ decays tend to have low invariant mass, m_{ll}, and be close together in ϕ space (low $\Delta\phi_{ll}$).

- Outside Lepton Veto (OLV): HWW VBF events are expected to have two forward jets and two central leptons. As such, any event in which one of the good leptons has pseudo-rapidity not between the pseudo-rapidity of the two tagged-jets is vetoed.

- ΔY_{jj} and m_{jj}: The two tagged jets in VBF HWW events tend to be found in opposite forward regions of the detector. This results in signal events having higher difference in rapidity, ΔY_{jj} and invariant di-jet mass, m_{jj}, than background events.

- m_T: m_T is an expansion of the variable initially developed to measure the mass of the W boson, m_T^W. By assuming that the measured lepton and missing energy are due to the decay of a W boson, the transverse mass of the W can be determined as

$$(m_T^W)^2 = m_l^2 + m_\nu^2 + 2(E_l E_\nu - \mathbf{p}_l \cdot \mathbf{p}_\nu) \qquad (4.2)$$

where l and ν subscripts refer to the final state lepton and neutrino and E and $\mathbf{p}$ are energy and momentum in the transverse direction. This

expression can be expanded to the $H \to WW^* \to l\nu l\nu$ system as

$$m_T = \sqrt{(E_T^{ll} + E_T^{miss})^2 - |p_T^{ll} + E_T^{miss}|^2} \qquad (4.3)$$

Note that we have assumed that the neutrino mass is zero such that $p_T^{miss} = E_T^{miss}$. For signal events, m_T should peak below the mass of the Higgs boson with no events having $m_T > m_H$. Due to imperfections in the reconstruction of leptons and missing energy, this upper bound does not hold for all signal events.

- p_T^{total}: p_T^{total} is the magnitude of the vector sum of the p_T of all objects in the event and is defined as:

$$p_T^{total} = \sqrt{(p_{1_x} + p_{2_x} + E_x^{miss} + \sum_{jets} p_x)^2 + (p_{1_y} + p_{2_y} + E_y^{miss} + \sum_{jets} p_y)^2},$$
$$(4.4)$$

where p_1 and p_2 refer to the momentum of the two final state leptons. VBF signal events tend to have a lower p_T^{total} than background events. p_T^{total} is particularly useful in removing background events with soft gluon radiation but no additional high-p_T jets.

- $\sum_{l,j} m_{lj}$: $\sum_{l,j} m_{lj}$ is the sum of the invariant mass of all four possible pairs of leptons and tagged-jets. The VBF topology results in jets in the forward region and leptons in the central region of the detector. This causes large opening angles between leptons and jets in the VBF signal

79

events and therefore signal events tend to have higher $\sum\limits_{l,j} m_{lj}$ on average than background events.

- η_{lep} Centrality: η_{lep} Centrality is yet another way of using the fact VBF signal events tend to have leptons in the central region of the detector, with tagged jets in the forward region to separate signal from background events. η_{lep} centrality is defined as:

$$\eta_{l_1}^{cent} = 2 \cdot \left| \frac{\eta_{l_1} - \bar{\eta}}{\eta_{j_1} - \eta_{j_2}} \right| \tag{4.5}$$

$$\eta_{l_2}^{cent} = 2 \cdot \left| \frac{\eta_{l_2} - \bar{\eta}}{\eta_{j_1} - \eta_{j_2}} \right| \tag{4.6}$$

$$\eta_{lep}^{centrality} = \eta_{l_1}^{cent} + \eta_{l_2}^{cent} \tag{4.7}$$

Where $\bar{\eta}$ is the average pseudo-rapidity of the two tagged jets, $\bar{\eta} = \frac{\eta_{j_1} + \eta_{j_2}}{2}$. For each individual lepton, a value of η_{lep} centrality greater than 1 indicates that the lepton is not in the region between the two tagged jets. While a η_{lep} centrality value near zero indicates that the lepton is close to the center of the region between the two jets. VBF signal events tend to have a smaller η_{lep} centrality than background events.

4.2.1 Event Selection and Input Parameters

Before applying the SVM analysis, events are screened in order to eliminate as many background events as possible. In order to accomplish this, a set of preselection cuts are applied as outlined in Table 4.3.

Object	Cut Value
lepton	Exactly 1 electron and 1 muon
lepton charge	opposite charged leptons
lepton p_T	$p_{T_1} > 22 GeV$, $p_{T_2} > 15 GeV$
m_{ll}	$m_{ll} > 10 GeV$
Z/DY$->\tau\tau-$ Veto	$m_{\tau\tau} < 66 GeV$
n_{jets}	$n_{jets} \geq 2$
$b - veto$	$n_{jets}^{MV1,85\%} = 0$
CJV	$n_{central\ jets}^{p_T>20} = 0$
OLV	OLV$= 0$

Table 4.3: Event Preselection

OLV$= 0$ indicates that there are no leptons with pseudo-rapidity outside of the tagged jets. After preselection, the following SVM input parameters are stored for each event:

- p_T^{total}, η_{lep} centrality, ΔY_{jj}, m_{jj}, m_{ll}, $\Delta\phi_{ll}$, m_T, $\sum_{l,j} m_{lj}$

These parameters are specifically chosen to exploit the topological character-istics of $H->WW$ VBF events as described in the previous section.

4.3 Signal-Background Comparison Plots

The section presents a comparison between the signal and background distributions of SVM input variables. Background samples are separated into light and heavy classes as described in Chapter 8. The processes and associated MC samples contained in each class are listed in Tables 5.1 - 5.5.
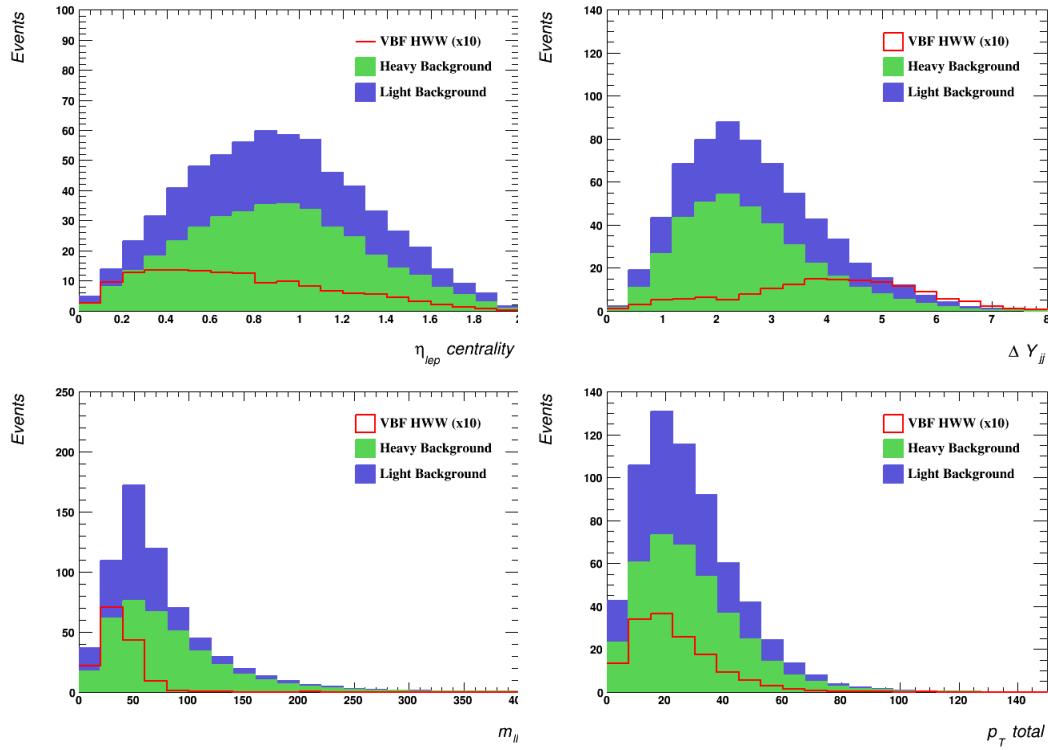
Figure 4.2: Signal-background comparison plots for SVM input parameters.
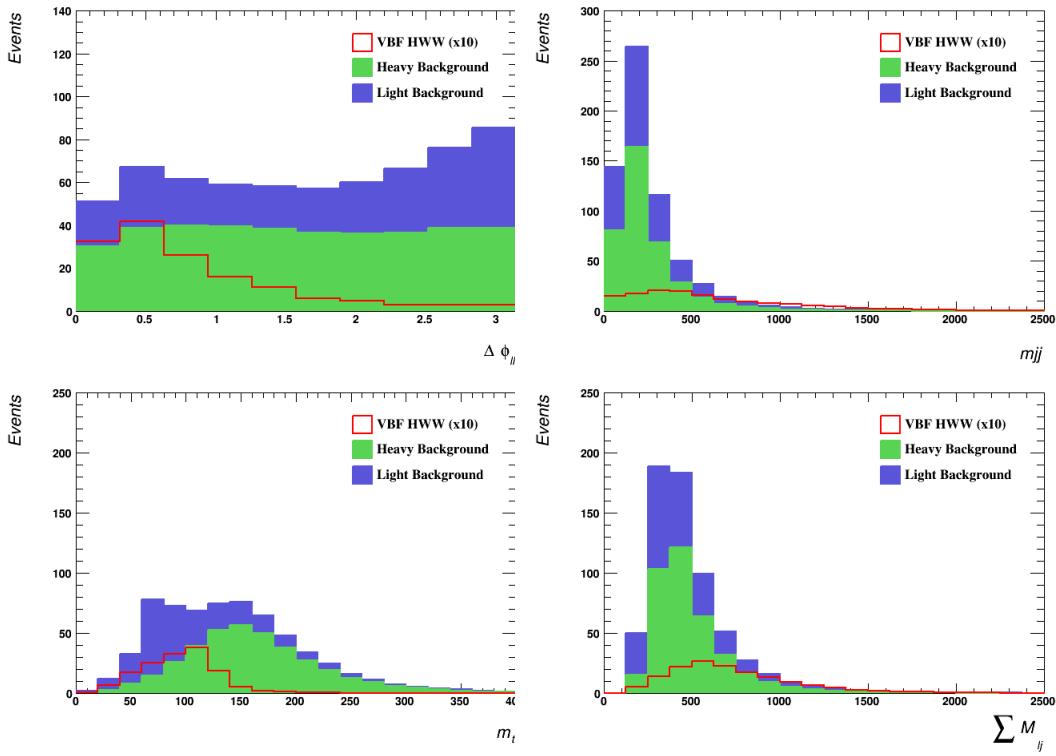
Figure 4.3: Signal-background comparison plots for SVM input parameters.

4.4 Expected Yields

Table 4.4 indicates the expected number of events for VBF signal and all background processes in the 2012 ATLAS data set after preselection. These numbers are taken from Monte Carlo studies using the MC samples listed in Tables 5.1 - 5.5.

Process	Yield
VBF HWW	12.7
Z+jets	103.7
Wgamma	11.5
ggf HWW	11.8
ttbar	331.2
WW	95.9
WZ	15.2
W+Jets	18.1
Single Top	41.9
QCD	12.8

Table 4.4: Expected yields after preselection for signal and background processes.

Chapter 5

Signal and Background

Estimation

The analysis presented in this thesis uses $20\,fb^{-1}$ of data collected in pp collisions at the ATLAS detector with a center of mass energy of $\sqrt{s} = 8\ TeV$. Monte Carlo (MC) simulated events representing signal and background were used to test analysis techniques, create SVM templates and analyze systematic uncertainties. In the case of $W + $ jets and QCD hadronic events, data-driven techniques were used to estimate backgrounds due to possible improper fake lepton predictions in Monte Carlo samples.

In order to better replicate real data, scale factors are applied to the Monte Carlo events. These scale factors include MC generator scaling, which compensates for discrepancies between MC generator and real data distributions in variables such as lepton multiplicity and p_T. In addition, scale factors cor-

recting for object energy, such as Jet Energy Scale (JES) and object detection efficiency, are applied on an event-by-event basis.

Finally, in order to generate conditions inside the actual detector in which dozens of interactions can occur per bunch crossing, MC events are generated with multiple pp events superimposed over the event of interest. These overlapping events are referred to as pileup and generally consist of soft (i.e. low energy) hadrons and leptons. An additional scale factor is applied globally to all Monte Carlo events in order to compensate for the difference in pileup conditions between the real detector and Monte Carlo simulations.

5.1 HWW VBF Signal

The VBF $H \to WW^* \to l\nu l\nu$ Monte Carlo used in this analysis was generated using Powheg and showered using Pythia with the next-to-leading order parton distribution function (PDF) set CT10 [16].

5.2 Backgrounds

5.2.1 W+jets

W+jets events constitute a minor background to $H \to WW^* \to l\nu l\nu$ signal events. W+jets events in which the W boson decays leptonically have one true high-p_T lepton in the final state, along with a number of jets. Though

normally W+jets events are rejected by the requirement that $N_{lep} \geq 2$, it is possible for W+jets events to pass our analysis cutflow when one of the jets is incorrectly reconstructed as a lepton.

Since the rate of jet misidentification may not be accurately described by the Monte Carlo, a data-driven technique is employed to estimate the W+jets background [17]. This technique involves the creation of two W+jets control samples. The first control sample requires two fully identified leptons, where one of the leptons is actually a jet misidentified as a lepton. The second control sample requires one fully identified lepton and one jet which passes a separate loosened lepton cutflow designed to enrich the sample in jets misidentified as leptons. Events in the second control sample are referred to as anti-id events. To estimate the number of W+jets events passing the full two-lepton cutflow, a fake factor defined as

$$ f_l = \frac{N_{id}}{N_{anti-id}}, \tag{5.1} $$

where N_{id} represents the rate at which jets pass the full lepton cutflow and $N_{anti-id}$ represents the rate at which jets pass the modified anti-id cutflow. The number of W+jets events in the signal sample is therefore given as

$$ N_{id+id}^{W+jets} = f_l \cdot N_{id+anti-id}^{W+jets}. \tag{5.2} $$

N_{id+id}^{W+jets} represents events with two fully identified leptons where one of the leptons is actually a misidentified jet. $N_{id+anti-id}^{W+jets}$ represents the number of

events with one fully identified lepton and one jet passing the anti-id lepton cutflow. We are able to calculate $N_{id+anti-id}^{W+jets}$ by creating an anti-id control sample using real data events and we are then able to estimate the W+jets contribution to the signal sample using equation (5.2).

5.2.2 QCD

QCD events involve hadronic decays with no real leptons. Similar to W+jets events, QCD events are normally rejected by the requirement that $N_{lep} \geq 2$. However, it is possible for two jets to be misidentified as leptons. Although the probability of two jets being misidentified is very small, the cross section of QCD events is very large, resulting in a non-negligible QCD background to $H \rightarrow WW^* \rightarrow l\nu l\nu$ events. Jets and misidentified leptons within QCD events have no preferred orientation, making the topological variables used in the SVM analysis ideal for distinguishing QCD and signal events. The QCD background was estimated using a data-driven technique similar to the one used for the W+jets events, except that 2 jets were assumed to be misidentified as leptons.

5.2.3 Z+jets

Z+jets events involve the decay of a Z boson with additional jets due to intial state radiation, final state radiation or pileup effects. For this analysis, Z bosons decaying into a pair of τ leptons are of primary interest as the τ

leptons can decay into two different flavor leptons through the process $Z \rightarrow \tau\tau \rightarrow e\nu_e\nu_\tau\mu\nu_\mu\nu_\tau$. In the case that there are at least two high energy jets, $Z \rightarrow \tau\tau$ events can mimic VBF HWW events.

Though $Z \rightarrow \tau\tau$ events may have the same final state particles as VBF HWW events, the topology of the events are quite different. Most importantly, leptons and neutrinos created by τ decays are preferentially emitted in the direction of p_τ. Since τ leptons resulting from Z boson decays are highly relativistic, the final state leptons and neutrinos are clumped together within the detector. This is in sharp contrast to HWW decays in which leptons and their associated neutrinos are emitted in opposite directions within the detector's rest frame. Knowledge of the τ lepton decays can be used to reconstruct the invariant mass of the two τ system. Events are rejected if $m_{\tau\tau}$ is within 25 GeV of the mass of the Z boson.

Z+jets backgrounds were modeled using MC samples generated with Alpgen [21] and showered with Herwig [19], with the exception of Z/γ samples which were created using Sherpa [22].

5.2.4 Top Quark

Top quark events constitute the largest background to the VBF $H \rightarrow WW^* \rightarrow l\nu l\nu$ signal. Specifically, $t\bar{t}$ events, which decay into two W bosons with two associated jets as shown in Figure 5.1, constitute approximately 50% of the total background after the full analysis cutflow. The major difference between

$t\bar{t}$ events and HWW VBF events is the presence of a b-quark in the $t\bar{t}$ final state. The proper tagging of b-jets is therefore imperative for rejection of top quark backgrounds.
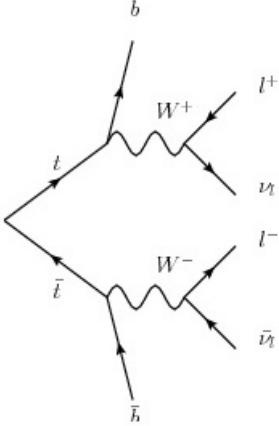
Figure 5.1: Feynman diagram of a $t\bar{t}$ pair decaying leptonically.

Wt events, in which a single top quark and W boson are produced, as shown in Figure 5.2, also constitute a major background for this analysis. In the case where the W boson created by the top decay, as well as the independent W boson, both decay leptonically, the final state will contain two leptons and one b-jet. If a second jet is found in the final state, either from initial state radition (ISR), final state radiation (FSR) or pileup effects, the event will have the same composition as a true VBF $HWW^* \to l\nu l\nu$.
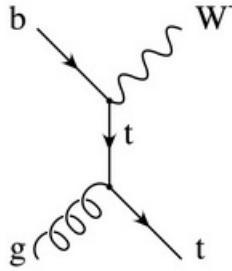
Figure 5.2: Feynman diagram of W boson and top quark associated production in the t-channel

The majority of top quark events are rejected by the requirement of $N^{jet}_{b-tag} = 0$. In addition, top quark events do not share much of the topology of VBF HWW events. Specifically, jets produced through top quark decay are not preferentially emitted in opposite forward regions of the detector and leptons produced by the decay of the two W bosons are not emitted in the same direction. These differences allow SVM input variables, such as m_{jj} and m_{ll}, to effectively separate HWW and $t\bar{t}$ events.

The $t\bar{t}$ and Wt samples were both generated using Powheg [20] and showered using Pythia [18].

5.2.5   ggf Higgs

Higgs bosons created through the gluon-gluon fusion process and decaying leptonicaly through two W bosons are considered as background to VBF events. Gluon-gluon fusion involves the creation of a Higgs boson through the fusion of two gluons via a top quark pair as shown in Figure 5.3. Assuming the Higgs
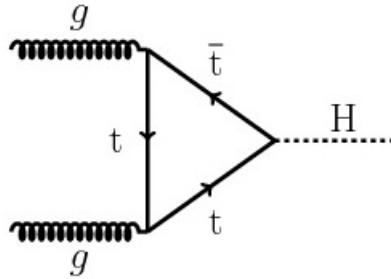
Figure 5.3: Feynman diagram of Higgs production via gluon-gluon fusion

then decays leptonically in the WW channel, the resulting final state will have two high p_T leptons, missing energy and no jets. Due to the lack of jets in the tree level final state, ggf events are generally removed by the requirement of at least two high-p_T jets. It is possible, however, for high-p_T jets to be found in the final state due to ISR, FSR or pileup. In these cases, ggf events will pass the analysis cutflow, resulting in a background on the order of the size of the expected signal.

ggf HWW events share the exact same lepton topology as VBF HWW events, making differentiation based on the properties of the final state jets of paramount importance. Jets from ISR, FSR and pileup are not oriented in any specific manner within the detector and will therefore have a lower m_{jj} and $\sum m_{lj}$ than VBF signal events. These properties are used by the SVM machinery to distinguish ggf and VBF events.

ggf background events were generated using Powheg and showered using Pythia.

5.2.6 $W\gamma$

$W\gamma$ events, in which the W boson decays leptonically, make up a very small background to VBF HWW signal events. In order to pass the analysis cutflow, at least two extra jets from ISR, FSR or pileup must be present in the final state and the photon or a jet must be misidentified as a lepton.

Two sets of $W\gamma$ Monte Carlo samples were used, one of which was generated and showered using Sherpa and the other was generated using Alpgen and showered using Herwig.

5.2.7   Diboson Backgrounds

The largest diboson background is due to WW events, with additional jets due to ISR, FSR or pileup, in which both Ws decay leptonically. These events share the same final state particles as VBF HWW signal events. However, they do not have the same topology, which makes the SVM input variables excellent at differentiating signal and WW background events. In addition, the small cross section of this background leads to only about 10 WW events passing preselection.

WZ events make up a very small background to HWW signal events. Since Z bosons can only decay leptonically into two leptons of the same flavor, WZ events are generally removed by the requirement of exactly two different flavor leptons.

Electroweak WW and $qq \rightarrow WW$ samples were generated and showered

using Sherpa. $gg \rightarrow WW$ samples were created using gg2WW [23]. WZ electoweak samples were generated and showered using Sherpa, while non-electroweak WZ samples were generated using Powheg and showered using Pythia.

5.3 MC Data Agreement

This section presents a comparison of data and Monte Carlo distributions for the observables used in the VBF $H \rightarrow WW^*$ analysis. The MC samples used in the creation of these plots are listed in Tables 5.1 - 5.5 and include both VBF HWW signal and data-driven W+jets and QCD backgrounds. Error bars include both statistical and systematic errors.
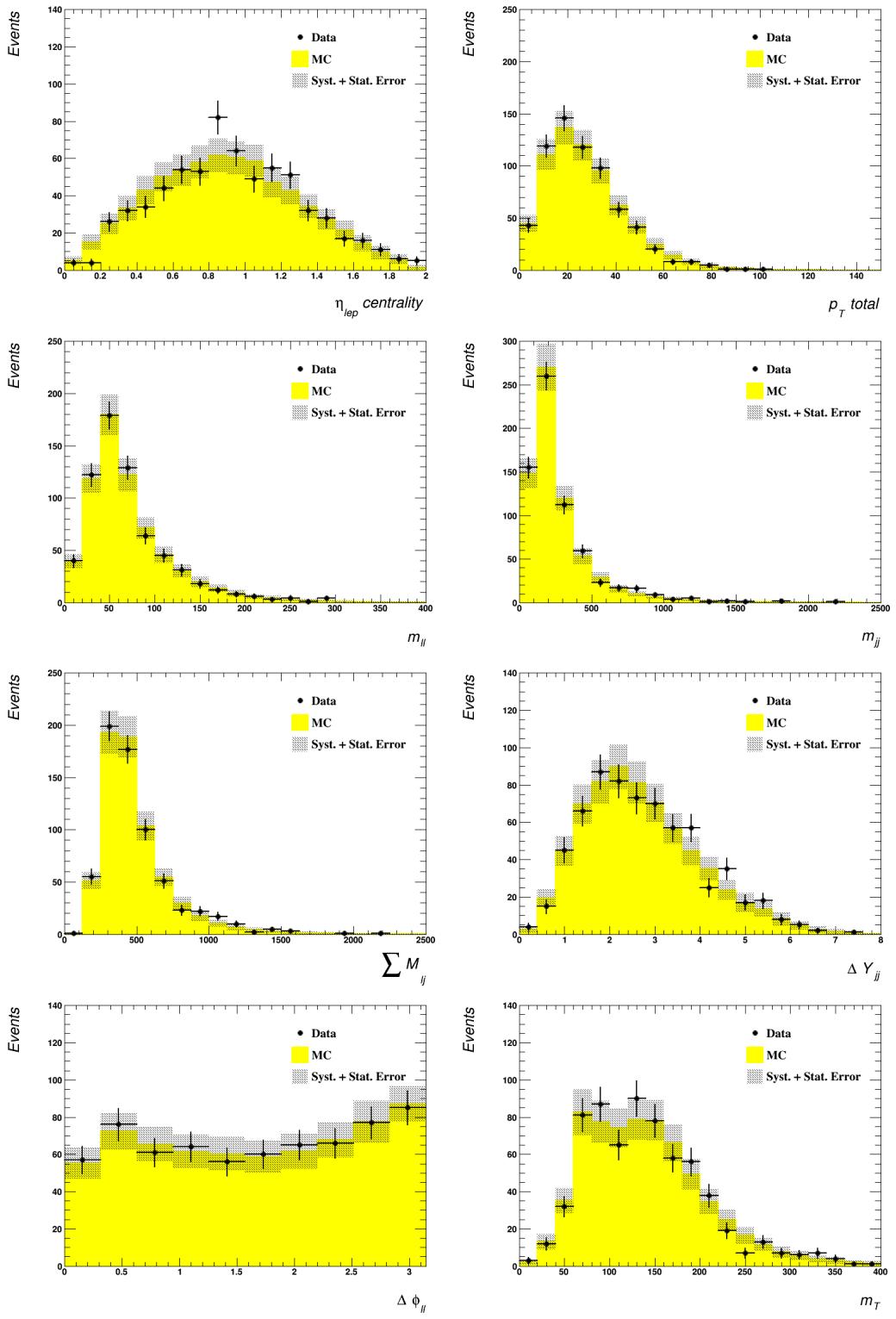
Figure 5.4: Comparison plots of Monte Carlo and data events for SVM input variables.
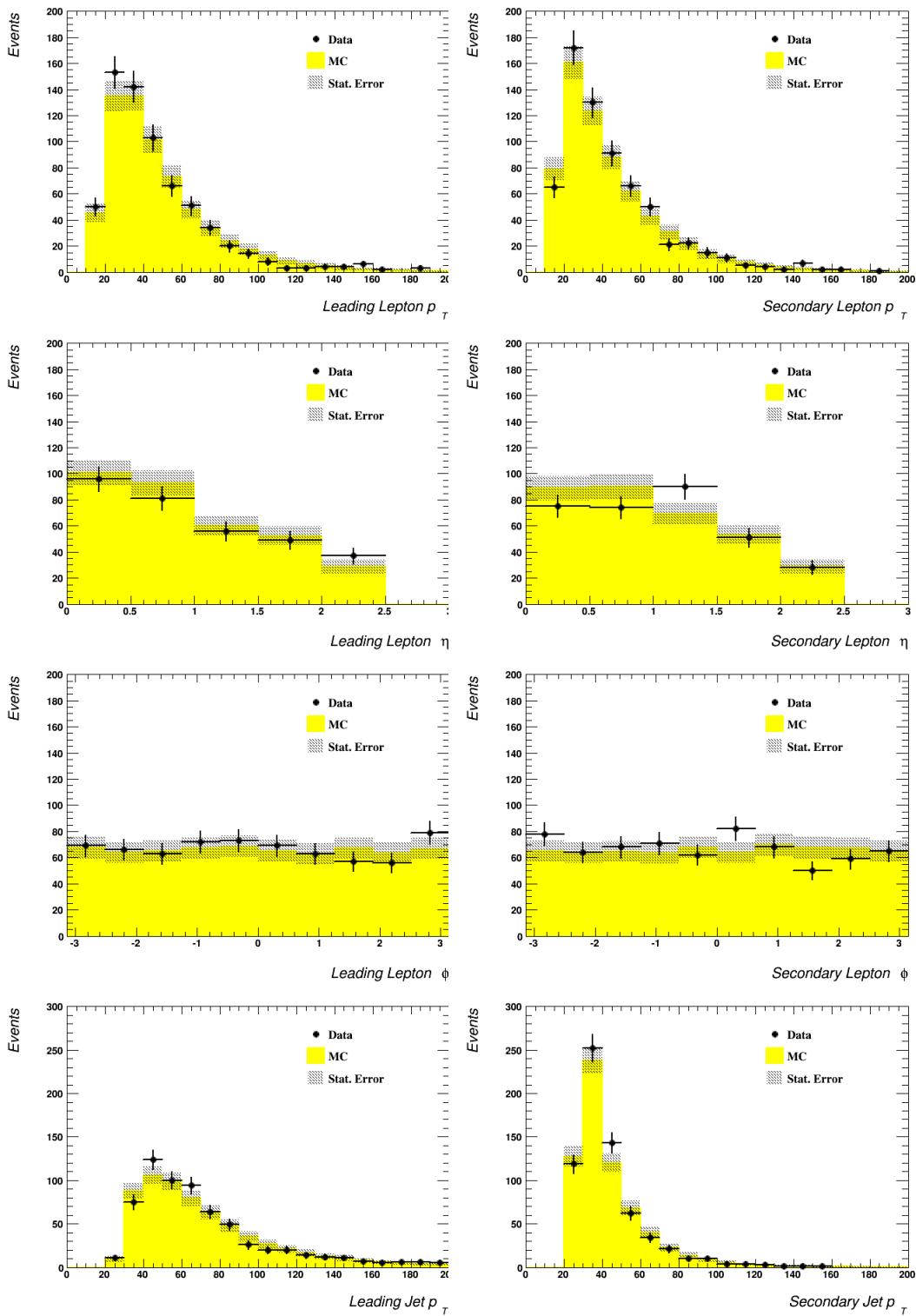
Figure 5.5: Comparison plots of Monte Carlo and data events for SVM input variables.
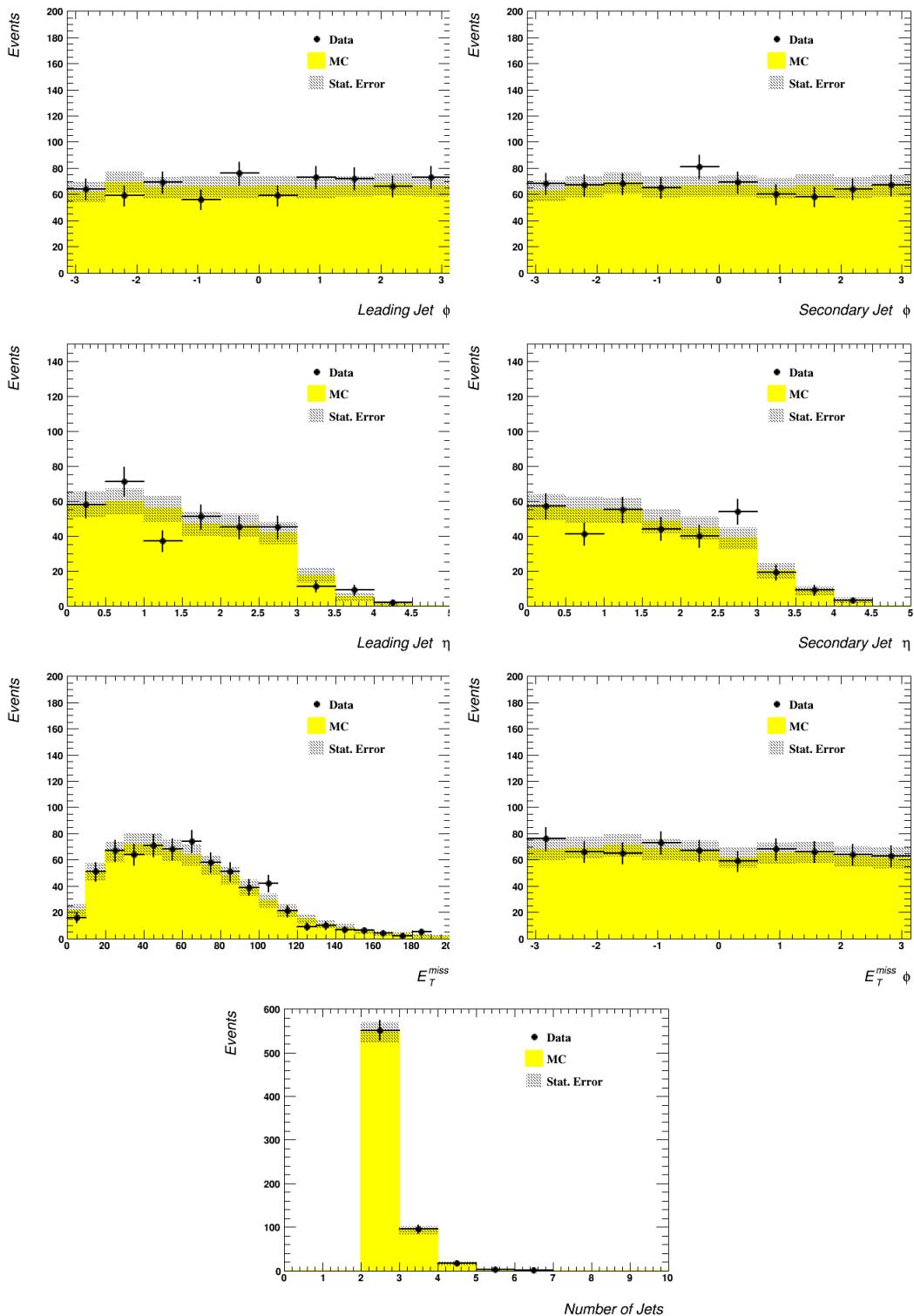
Figure 5.6: Comparison plots of Monte Carlo and data events for SVM input variables.

5.4　List of Monte Carlo Samples

This section presents a list of all Monte Carlo samples used in the VBF $H \to$ WW^* analysis. The expected yield associated with each MC sample is listed. This yield corresponds to the number of events of a given process expected to pass the analysis pre-selection cutflow in the ATLAS 2012 data set. MC samples are split into signal, light and heavy classes. These class assignments are used in the 3-class support vector machine analysis as discussed in Chapter 8.

Process	Sample #	x-sec(pb)	Generator	Simulation	Expected Yield After Preselection
VBF $H \to WW \to leplep$	161055	0.0356152	PowHeg+Pythia8	Full	12.7

Table 5.1: Table of MC Signal samples.

Process	Sample #	x-sec(pb)	Generator	Simulation	Expected Yield After Preselection
schan→lepton	110119	1.818	Powheg+Pythia6	AFII	0.0171
Wtchan→dilepton(DR)	110141	2.349	Powheg+Pythia6	AFII	40.8
tchan→lepton	110101	28.44	AcerMC+Pythia6	AFII	1.07
ttbar(dilepton)	181087	252.89	Powheg+Pythia	AFII	331

Table 5.2: Table of the MC samples used for the heavy class.

Process	Sample #	x-sec(pb)	Generator	Simulation	Expected Yield After Preselection
$gg \to WpWm \to e\mu\nu\nu$	169472	0.0208	gg2WW3.1.2	Full	2.56
$gg \to WpWm \to e\tau\nu\nu$	169473	0.0208	gg2WW3.1.2	Full	0.159
$gg \to WpWm \to \nu\nu$	169475	0.0208	gg2WW3.1.2	Full	2.44
$gg \to WpWm \to \mu\tau\nu\nu$	169476	0.0208	gg2WW3.1.2	Full	0.0503
$gg \to WpWm \to \tau\tau\nu\nu$	169477	0.0236	gg2WW3.1.2	Full	0.0247
$gg \to WpWm \to \nu\nu$	169478	0.0208	gg2WW3.1.2	Full	0.0612
$gg \to WpWm \to \tau\mu\nu\nu$	169479	0.0208	gg2WW3.1.2	Full	0.113
$WW \to l\nu(MassiveB/C)$	177997	5.679	Sherpa	Full	80.2
$Wgamma* \to l$	181452	10.175	Sherpa	Full	2.81
$Wgamma* \to l\nu\mu\mu$	181453	2.535	Sherpa	Full	1.84
$Wgamma + 4p$	146434	2.1224	Alpgen+Herwig	Full	0.469
$Wgamma + 5p$	146435	0.46612	Alpgen+Herwig	Full	0.443
$Wgamma + 0p$ LepPhoEF	146436	229.88	Alpgen+Herwig	Full	0.00256
$Wgamma + 1p$ LepPhoEF	146437	59.518	Alpgen+Herwig	Full	0.803
$Wgamma + 2p$ LepPhoEF	146438	21.39	Alpgen+Herwig	Full	3.7
$Wgamma + 3p$ LepPhoEF	146439	7.1203	Alpgen+Herwig	Full	1.41
$WW \to l\nu + jj$ (6EWcoupling)	161985	0.039	Sherpa	Full	10.3
$ZZ \to ll\nu\nu + jj$ (6EWcoupling)	161986	0.0012314	Sherpa	Full	0.00642
$WZ \to lll\nu + jj$(6EWcoupling)	161987	0.012559	Sherpa	Full	0.594
$ZZ \to llll + jj$(6EWcoupling)	161988	0.00073568	Sherpa	Full	0.0129
$W + Z \to e\nu\mu\mu$	129478	0.9382	Powheg+PYthia8	Full	1.66
$W + Z \to e\nu\tau\tau$	129479	0.1746	Powheg+PYthia8	Full	0.395
$W + Z \to \mu\nu ee$	129480	1.399	Powheg+PYthia8	Full	4.37
$W + Z \to \mu\nu\tau\tau$	129482	0.1746	Powheg+PYthia8	Full	0.382
$W + Z \to \tau\nu ee$	129483	1.399	Powheg+PYthia8	Full	0.265
$W + Z \to \tau\nu\mu\mu$	129484	0.9382	Powheg+PYthia8	Full	0.0633
$W + Z \to \tau\nu\tau\tau$	129485	0.1719	Powheg+PYthia8	Full	0.107
$W - Z \to e\nu\mu\mu$	129487	0.639	Powheg+PYthia8	Full	0.796

Table 5.3: Table of the MC samples used for the light class. Part 1.

Process	Sample #	x-sec(pb)	Generator	Simulation	Expected Yield After Preselection
$W - Z \to e\nu\mu\mu$	129487	0.639	Powheg+PYthia8	Full	0.796
$W - Z \to e\nu\tau\tau$	129488	0.1125	Powheg+PYthia8	Full	0.202
$W - Z \to \mu\nu ee$	129489	0.9359	Powheg+PYthia8	Full	4
$W - Z \to \mu\nu\tau\tau$	129491	0.1125	Powheg+PYthia8	Full	0.15
$W - Z \to \tau\nu ee$	129492	0.9359	Powheg+PYthia8	Full	0.219
$W - Z \to \tau\nu\tau\tau$	129494	0.1107	Powheg+PYthia8	Full	0.0399
$ZZ \to 2e2\mu(m_{ll} > 4GeV)$	126938	0.1708	Powheg+Pythia8	Full	1.51
$ZZ \to 2e2\tau(m_{ll} > 4GeV)$	126939	0.1708	Powheg+Pythia8	Full	0.157
$ZZ \to 2\mu2\tau(m_{ll} > 4GeV)$	126941	0.1708	Powheg+Pythia8	Full	0.0974
$ZZ \to 4\tau(m_{ll} > 4GeV)$	126942	0.0735	Powheg+Pythia8	Full	0.022
$ZZ \to ll\nu\nu(MassiveB/C)$	177999	0.494	Sherpa	Full	0.0804
$gg \to ZZ \to 2e2\mu$	116603	0.00135	gg2ZZ	Full	0.0712
$Z \to mm2jEW(40 < m_{ll}, \text{ inc.t-ch})$	129925	1.054	Sherpa	Full	0.036
$Z\mu\mu\gamma(M > 10GeV)$	161998	54.341	Sherpa	AFII	1.45
$Z\mu\mu + 1p(60 < M < 2000GeV)$	167331	155	Alpgen+Herwig	Full	0.0778
$Z\mu\mu + 2p(60 < M < 2000GeV)$	167332	48.945	Alpgen+Herwig	Full	0.238
$Z\mu\mu + 2p(10 < M < 60GeV)2l$	181322	52.879	Alpgen+Herwig	Full	0.0318
$Z \to tt2jEW(40 < m_{ll}, \text{inc.t-ch})$	167724	1.0637	Sherpa	Full	1.25
$Z \to tt2jEW(7 < m_{ll} < 40, \text{inc.t-ch})$	181349	0.72663	Sherpa	Full	0.111
$Z\tau\tau + 0p(60 < M < 2000GeV)$	146930	712.1	Alpgen+Herwig	Full	8.65
$Z\tau\tau + 1p(60 < M < 2000GeV)$	146931	154.95	Alpgen+Herwig	Full	13.2
$Z\tau\tau + 2p(60 < M < 2000GeV)$	146932	48.767	Alpgen+Herwig	Full	26.7
$Z\tau\tau + 3p(60 < M < 2000GeV)$	146933	14.184	Alpgen+Herwig	Full	8.15
$Z\tau\tau + 4p(60 < M < 2000GeV)$	146934	3.7959	Alpgen+Herwig	Full	0.899
$Z\tau\tau + 5p(60 < M < 2000GeV)$	147094	1.1366	Alpgen+Herwig	Full	0.122
$Z\tau\tau + 4p(10 < M < 60GeV)$	146854	2.592	Alpgen+Herwig	Full	0.309
$Z\tau\tau + 1p(10 < M < 60GeV)$	146881	108.71	Alpgen+Herwig	Full	0.411

Table 5.4: Table of the MC samples used for the light class. Part 2.

Process	Sample #	x-sec(pb)	Generator	Simulation	Expected Yield After Preselection
$Z\tau\tau + 2p(10 < M < 60GeV)$	146882	52.827	Alpgen+Herwig	Full	3.84
$Z\tau\tau + 3p(10 < M < 60GeV)$	146883	11.311	Alpgen+Herwig	Full	1.18
$Z\tau\tau + 0p(60 < M < 2000GeV)2l$	169450	712.14	Alpgen+Herwig	Full	2.18
$Z\tau\tau + 1p(60 < M < 2000GeV)2l$	169451	154.97	Alpgen+Herwig	Full	2.67
$Z\tau\tau + 2p(60 < M < 2000GeV)2l$	169452	48.805	Alpgen+Herwig	Full	16.5
$Z\tau\tau + 3p(60 < M < 2000GeV)2l$	169453	14.189	Alpgen+Herwig	Full	9.02
$Z\tau\tau + 4p(60 < M < 2000GeV)2l$	169454	3.8005	Alpgen+Herwig	Full	2.97
$Z\tau\tau + 5p(60 < M < 2000GeV)2l$	169455	1.1441	Alpgen+Herwig	Full	0.748
$Z\tau(M > 10GeV)$	126854	32.317	Sherpa	Full	2.95

Table 5.5: Table of the MC samples used for the light class. Part 2.

Chapter 6

Support Vector Machines

The goal of the analysis presented in the thesis is to search for evidence of vector boson fusion Higgs production in the $H \rightarrow WW* \rightarrow l\nu l\nu$ channel in the $20\,fb^{-1}$ 2012 ATLAS data set. This data set contains millions of events, most of which are not Higgs boson events. Our task is to remove as many non-Higgs boson events as possible from the data set and then estimate the percentage of VBF HWW events in the remaining data set. The removal of most background events is accomplished through a cut-based approach and the support vector machine analysis is then used to estimate the percentage of Higgs boson events in the remaining data set. The analysis presented in this thesis is therefore a combination of a cut-based pre-analysis and a final support vector machine analysis.

A cut-based analysis uses manually-created thresholds on relevant parameters in order to remove events from the data set. For example, in this analysis

we are specifically looking for events with two leptons. Therefore, any event which does not contain two leptons is removed from the data set. This process is continued using cuts on various event parameters until no more cuts can be made without removing significant amounts of signal events from the data set. The full set of cuts for a specific analysis is referred to as the cutflow and is described is Chapter 4. In the case of this analysis, we refer to the cutflow as the preselection cutflow, or simply preselection. In a fully cut-based analysis, once all relevant cuts have been made, the number of events in the real data set is compared to expected results based on Monte Carlo studies.

Cut-based analyses are effective at removing background events, particularly when there are very clear constraints on the events of interest, such as the number of leptons or the number of high-energy jets. However, cut-based analyses do not take into account correlations between parameters and need to be tuned manually. In addition, cut-based analyses do not allow for estimation of the probability that an individual event is a signal or background event. In order to properly account for correlations between parameters and to estimate the signal-event probability on an event-by-event basis, multivariate analyses are employed. A multivariate analysis is designed to take into account many input parameters at the same time in order to determine the relative probability of a given event being a signal or background event. In this thesis we use a specific type of multivariate analysis known as support vector machines (SVM).

6.1 Introduction to Support Vector Machines

A support vector machine is a linear event classifier designed to differentiate between two classes of events, in our case signal and background. In order to accomplish this differentiation, support vector machines create a separating hyperplane between signal and background events in feature space. Feature space is the n-dimensional space in which events are plotted using n distinguishing parameters, such as lepton energy and E_T^{miss}. As a simple example, let's assume that we can parameterize events in our analysis by two variables, H_1 and H_2 in feature space. We first need to train the SVM by supplying events of known type, either signal or background. The support vector machine plots these events in a 2-dimensional space and attempts to draw a hyperplane, in this case a line, between the signal and background events as shown in Figure 6.1. Once the separating line is determined, the SVM is said to be trained and can be used to distinguish events of unknown type. The position of the unknown event relative to the separating line determines the best guess as to whether the event is signal or background. For each unknown event the SVM returns a number, referred to as the SVM discriminant, which represents the distance of the event from the separating hyperplane in feature space. The sign of the SVM discriminant represents the best guess as to the class of the unknown event.

As can be seen from Figure 6.1, support vector machines are excellent at taking into account correlations between independent parameters. No indi-

vidual cut on H_1 or H_2 would effectively remove background while preserving signal events. Rather, it is only by determining the proper relationship between H_1 and H_2 that effective differentiation of signal and background events can be accomplished. In this simple example, it would be possible to create a cut on a linear combination of H_1 and H_2 which would perform the same function as the SVM. However, support vector machines are capable of solving much more complicated problems involving many parameters with non-linear separating hyperplanes. This ability to distinguish classes of events in complex feature spaces makes support vector machines and other multivariate analysis techniques crucial for high energy physics analyses.

6.2 Determination of the Separating Hyperplane

When given a set of data points of known type (signal or background), referred to as a training set, there are many possible hyperplanes separating the two classes, as seen in Figure 6.2. In order to determine the optimal hyperplane, we introduce the idea of a margin and choose our hyperplane such that the margin is maximized. The margin is the distance between the hyperplane and the points closest to the hyperplane, represented as half of the distance d in Figure 6.1. In order to determine the width of the margin for a given hyperplane, we first must be able to write down an equation representing the
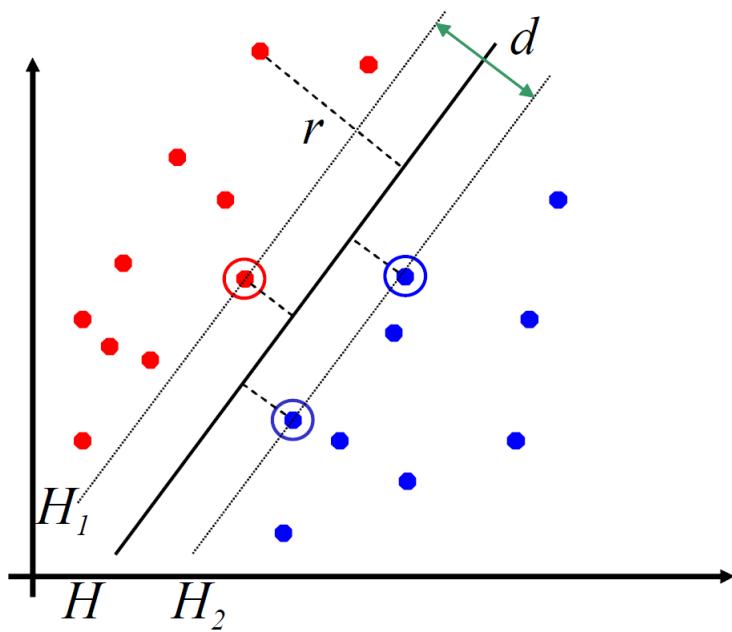
Figure 6.1: Red signal and blue background events, parameterized by example variables H_1 and H_2, separated by a hyperplane. d represents twice the width of the margin, while r shows the distance from the hyperplane to an arbitrary point.
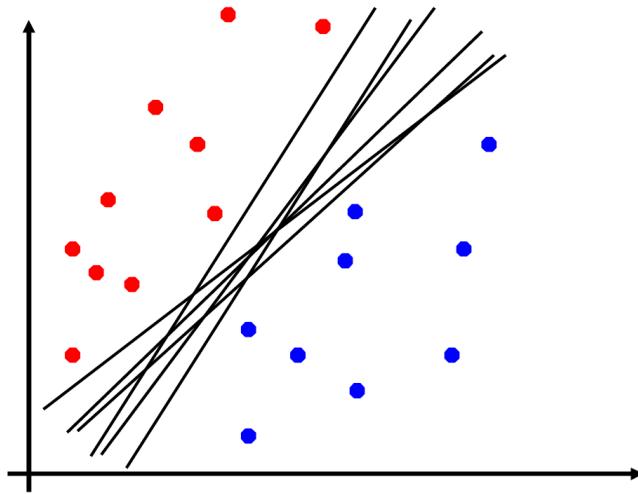
Figure 6.2: Red signal and blue background events separated by several possible hyperplanes. The goal of the SVM is to find the optimal hyperplane that maximizes the distance between the hyperplane and the points closest to the hyperplane.

hyperplane.

The general equation of a separating hyperplane in N-dimensional space is given by

$$0 = \vec{w} \cdot \vec{x} - b, \tag{6.1}$$

where $\vec{x}$ is the coordinate vector, $\vec{w}$ is a vector normal to the plane and b is the offset of the plane from the origin. It is easy to see that equation (6.1) represents a line in 2-dimensions with the form $y = mx + b$, while in three dimensions it represents a plane. Once the hyperplane has been determined by fixing $\vec{w}$ and b, we can write the learned function as

$$f(\vec{x}) = \vec{w} \cdot \vec{x} - b. \tag{6.2}$$

The sign of the learned function for an arbitrary input vector $\vec{x}$ represents our best guess as to which class an unknown event belongs. This can be seen by noting that all blue points in Figure 6.1 have $f(\vec{x}) < 0$, since they lie below the hyperplane, while all red points have $f(\vec{x}) > 0$. From equation (6.1), we find that the distance between the separating hyperplane and a point of interest, $\vec{x}$, is given by

$$r = \frac{\vec{x} \cdot \vec{w} - b}{|\vec{w}|}. \tag{6.3}$$

We have a problem however, in that there are an infinite number of normal vectors $\vec{w}$ which specify the exact same hyperplane. We therefore choose a normalization of $\vec{w}$ such that the learned function is equal to ± 1 at the margin. Inserting this requirement into equation (6.3), we find that the total width of the margin is given by

$$M = \frac{1}{|\vec{w}|}. \tag{6.4}$$

We have therefore specified an equation relating the margin, which we wish to maximize, to the normal vector defining the hyperplane, $\vec{w}$. From equation (6.4) we can see that minimizing $\vec{w}$ is equivalent to maximizing M. With this in mind, we characterize our goal as the minimization of $|\vec{w}|$ under the constraint that

$$y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1 \quad \text{for every } \vec{x}_i \in \text{training set.} \tag{6.5}$$

y_i is a manually set variable which provides information on the class associated with a given point: $y_i = +1$ for signal points and $y_i = -1$ for background points. This minimization can be accomplished using the method of Lagrange multipliers. The Lagrangian of interest can be written as

$$\mathcal{L} = \frac{1}{2}|w|^2 - \sum_{i=1}^{N} \alpha_i[y_i(\vec{w} \cdot \vec{x}_i - b) - 1], \tag{6.6}$$

where α_i are Lagrange multipliers satisfying the conditions

$$\alpha_i[y_i f(\vec{x}_i) - 1] = 0. \tag{6.7}$$

It is important to note that any point not lying along the margin such that $y_i f(\vec{x}_i) = 1$ will have $\alpha_i = 0$. This means that only points along the margin contribute to the definition of our hyperplane. These points are referred to as support vectors.

In order to minimize our Lagrangian, (6.6), we need to find a point that is a minimum of $\mathcal{L}$ with respect to $\vec{w}$ and a maximum with respect to α_i. This point satisfies the conditions

$$\frac{\partial \mathcal{L}}{\partial \vec{w}} = 0 \implies \vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x}_i, \tag{6.8}$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \implies \sum_{i=0}^{N} \alpha_i y_i = 0. \tag{6.9}$$

Inserting these conditions into our Lagrangian, (6.6), we find the dual Lagrangian

$$\mathcal{L}_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j. \tag{6.10}$$

Note that this dual Lagrangian depends only on the α_is and the dot product of position vectors $\vec{x}_i \cdot \vec{x}_j$. This formulation is important for complex systems in which the training points may not be linearly separable. Once the α_is are known, $\vec{w}$ can be recovered through equation (6.8) and b can be recovered through the requirement on the support vectors of

$$y_i(\vec{w} \cdot \vec{x}_i - b) = 1 \tag{6.11}$$

6.3   Non-Linear Problems

The support vector machine formalism that we have outlined so far is only appropriate for linearly separable training sets, such as those shown in Figure 6.1. In our discussions up to this point, we have directly translated parameters taken from the events of interest into points in feature space. However, in non-linearly separable systems, this method is ineffective as there are no possible separating hyperplanes. In these cases, we use Mercer's theorem in order
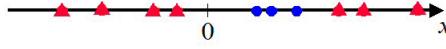
Figure 6.3: Two classes of training events, red and blue, described by a single parameter, x. These two classes are not linearly separable in feature space.

to map our original training points in feature space into higher-dimensional decisions space.

As an example of this idea, let's look at a one dimensional feature space as shown in Figure 6.3. This training set is described in feature space by a single parameter, x, but is unable to be separated by a single hyperplane (in this case a point). In order to find a separating hyperplane we need to map our original feature space into a higher dimensional decision space. We chose the mapping $y = x^2$ such that each point is now described by two parameters, (x, x^2), as shown in Figure 6.4. In this higher dimensional decision space, a separating hyperplane does exist which separates the two classes of events. Therefore, by mapping our initial problem into a higher dimensional space we have been able to isolate the two classes of events and determine the optimal separating hyperplane.

In this simple example, we were able to manually chose a mapping function such that our 1-dimensional problem was easily separable in 2-d space. In general, however, manually mapping from feature space into decision space is
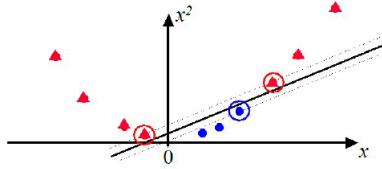
Figure 6.4: Two classes of training events, red and blue, described by a single parameter, x, mapped into a 2-dimensional space by the mapping $y = x^2$. This mapping allows a separating hyperplane to be created between the two sets of events. The circled points are the support vectors.

impossible. In order to avoid this problem we employ a 'kernel trick' based on Mercer's theorem. Mercer's theorem states that a symmetric kernel function $K(x_i, x_j)$ can be expressed as an inner product

$$K(\vec{x}_i \cdot \vec{x}_j) = \varphi(\vec{x}_i) \cdot \varphi(\vec{x}_j). \tag{6.12}$$

if and only if $K(x_i, x_j)$ is positive semidefinite, that is if

$$\int K(x_i, x_j) g(x_i) g(x_j) dx dy \geq 0 \tag{6.13}$$

for any g. Conversely, Mercer's theorem allows us to write a dot product of mapped vectors in terms of a kernel function. To illustrate this, let us define the particular mapping of a point $\vec{x}_i$ in feature space to a point in decisions space as $x_i \to \varphi(\vec{x}_i)$. We can write our dual Lagrangian after this mapping as

$$\mathcal{L}_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \varphi(\vec{x}_i) \cdot \varphi(\vec{x}_j). \tag{6.14}$$

112

As we can see from equation (6.14), $\varphi(\vec{x}_i)$ only enters into the dual Lagrangian when dotted with itself $(\varphi(\vec{x}_i) \cdot \varphi(\vec{x}_j))$. We can therefore replace $\varphi(\vec{x}_i) \cdot \varphi(\vec{x}_j)$ in our Lagrangian with a kernel function and write the dual Lagragian as

$$\mathcal{L}_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(\vec{x}_i \cdot \vec{x}_j). \qquad (6.15)$$

Equation (6.15) represents the exact same Lagrangian as equation (6.14), except that we have replaced the dot product of the unknown mapping function, $\varphi(\vec{x}_i)$ with a kernel function K. We are now in a position to minimize the Lagrangian (6.15), which is equivalent to minimizing (6.14). This means that we are able to minimize a Lagrangian in which our training points are mapped into a higher dimensional space without knowing the details of the mapping. To reiterate, the use of a kernel function allows us to determine the equation of an optimal separating hyperplane between two classes in a higher dimensional decision space without knowing the mapping function φ. The minimization of the Lagrangian, (6.15), determines the values of α_i, which can be used to find $\vec{w}$ and b and thus fix the learned function, (6.2). Figure 6.5 illustrates an example of a 2-dimensional problem which is not linearly separable in feature space, but which can be separated using a kernel function.

There are several possible kernels that can be used in SVM analyses. For the purposes of the analysis discussed in this thesis, a Gaussian kernel function of the form
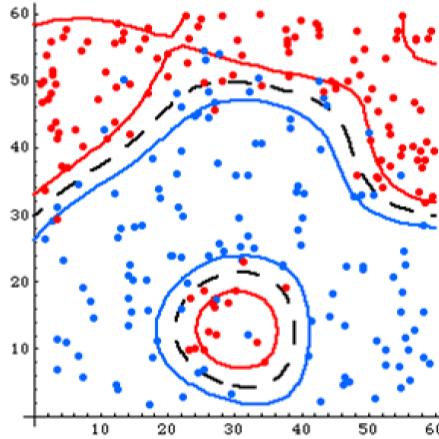
Figure 6.5: Two classes separated by the a non-linear learned function.

$$K(\vec{x}_i \cdot \vec{x}_j) = e^{-|\vec{x}_i - \vec{x}_j|^2 / 2\sigma^2} \qquad (6.16)$$

was used. The Gaussian kernel is somewhat of an all purpose kernel function, in that it provides good results for a wide variety of problems. The Gaussian kernel has one free parameter, σ, which determines the width of the Gaussian kernel function. σ affects the rigidity of the learned function in unmapped feature space. Large values of σ limit the ability of the learned function to curve, while small values of σ allow for more bends in the learned function. The value of σ is generally determined through empirical studies conducted on the problem of interest. However, σ should generally have approximately the same magnitude as the length of vectors used for training.

6.4 Soft Margin

In the previous sections we have set a goal of finding an optimal hyperplane separating two classes of events. We simplified this problem by introducing the idea of a margin, defined as the distance from the hyperplane to the closest points of each class. We then determined the optimal separating hyperplane by maximizing the margin. This system is appropriate for very well separated data sets; however, most real world problems are not so neat. Most data sets contain overlapping data points in which the separation between classes is somewhat murky. In order to find the optimal separating hyperplane in these situations, we need to slightly modify our formalism.

The use of a so-called hard margin, in which M is defined to be the distance from the hyperplane to the closest point, allows for a small number of points to dominate the definition of the hyperplane. In the case where these points are not characteristic of their respective class, or where the two classes overlap in a certain region, the hyperplane defined by these points may separate the two classes very poorly. In order to allow for a more robust hyperplane optimziation, we introduce a soft margin. The idea of a soft margin is that points are allowed to penetrate the margin, but they are penalized for doing so such that margins containing large numbers of points are disfavored.

In order to allow for points to penetrate the margin, we must relax the margin condition, (6.5), and instead require

$$y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1 - \xi_i, \quad \xi_i \geq 0. \tag{6.17}$$

ξ_i is a free parameter, which allows for points to enter the margin. Terms outside of the margin have $\xi_i = 0$, while for terms inside the margin the value of ξ_i is determined by the distance inside of the margin that the point lies. With our modified margin condition, we can write our Lagrangian in feature space as

$$\mathcal{L} = \frac{1}{2}|w|^2 + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i [y_i(\vec{w} \cdot \vec{x}_i - b) - 1 + \xi_i]. \tag{6.18}$$

Comparing equation (6.18) with (6.6) we note the addition of the term $C \sum_{i=1}^{N} \xi_i$. C is a positive constant, and since ξ_i is defined to be greater than zero, this term is positive for all points. Since our goal is to minimize $\mathcal{L}$, this term acts as a penalty for points that invade the margin, increasing the minimum value of $\mathcal{L}$ for each point inside of the margin. Larger values of C impose higher penalties on terms penetrating the margin while smaller values of C allow for many points within the margin. The value of C can be manually adjusted and is generally set by empirical studies.

In order to simplify our Lagrangian, it is useful to rewrite equation (6.18) as

$$\mathcal{L} = \frac{1}{2}|w|^2 - \sum_{i=1}^{N} \alpha_i y_i(\vec{w} \cdot \vec{x}_i - b) + \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N} \xi_i(C - \alpha_i). \tag{6.19}$$

116

Since the last term in the Lagrangian (6.19) is positive, we know that the minimum of $\mathcal{L}$ will be a minimum of ξ. We can therefore use the condition

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = 0 \implies C - \alpha_i = C, \qquad (6.20)$$

along with conditions (6.8) and (6.9) to write the dual soft margin Lagrangian as

$$\mathcal{L}_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j, \quad 0 \le \alpha_i \le C. \qquad (6.21)$$

Note that this is exactly the same as dual Lagrangian as (6.10) except that we have imposed a new condition on α_i.

We have now succeeded in finding a form of the dual Lagrangian that when minimized will yield the optimal separating hyperplane between two classes of events. This Lagrangian is capable of separating events in a higher dimensional decision space without requiring that we know the mapping function between feature space and decision space. In addition, we have been able to allow for terms to penetrate the margin without modifying the form of our Lagrangian. In the following chapters we will discuss the use of the SVM technique in the analysis of VBF HWW events along with expanding this formalism to allow for multiple support vector machines within a single analysis.

Chapter 7

2-Class Support Vector Machine Analysis

As a preliminary test of the support vector machine method for use in $H \rightarrow WW^* \rightarrow l\nu l\nu$ VBF analyses, a 2-class SVM analysis was performed. Though we hope to gain increased separating power using 3-class SVM systems, 2-class analyses offer several advantages. These advantages include having a readily definable signal region, quick training and classification and easy comparison with Boosted Decision Tree analyses. Due to these properties, much of the refinement of the SVM free parameters, input parameters and scale factors was performed using 2-class SVMs.

2-class SVM analyses are designed to separate between two differing classes of events, in this case HWW signal events and background events. The mathematical formalism of 2-class SVMs is discussed in Chapter 6 and we will

therefore restrict our discussions to the methodology of this particular analysis. Before beginning the SVM analysis, a set of Monte Carlo events is compiled that includes VBF *HWW* signal and all relevant backgrounds. These events are required to pass all preselection cuts as outlined in Chapter 4.

In order to function properly, support vector machines require two separate sets of data: a training set on which to define the optimal hyperplane and determine the learned function; and a test set. The test set must be completely independent of the training set and is used to test the discriminating power of the SVM. In order to train an SVM, training events are provided which include the parameter information and the known class of each event. Using this information, the SVM is able to create an optimal separating hyperplane between the two sets of events.

One danger of multivariate analysis techniques is that they will be trained to separate two specific sets of events, not two distinct classes of events. That is, the analysis may only be useful for separating the events used for training and not be able to determine the class of new events. This problem is referred to as over-tuning. In order to safeguard our analysis against over-tuning, no events used in SVM training are used for testing. In general, support vector machines are robust against over-tuning as only a small percentage of training events are used to define the separating hyperplane.

7.1 SVM Training Samples

In order to create independent training and testing data sets, Monte Carlo events passing all preselection cuts are separated into two sets based on event number. The events are then further subdivided based on their class, resulting in four independent data sets labeled as signal-even, signal-odd, background-even and background-odd. The first word in each category refers to the sample type and the second refers to the event number. Using these four data sets, two independent SVMs are trained to separate signal from background events. One SVM is trained using signal-even and background-even events and the another using signal-odd and background-odd events. The two trained SVMs are hereafter referred to as even-trained and odd-trained. This procedure allows for odd-numbered events to be used to test the even-trained SVM and vice versa, resulting in completely independent training and testing data sets. In addition, this method allows for all available Monte Carlo events to be utilized for both training and testing.

In order to properly train an SVM to distinguish between two classes of events in real data, training samples for the signal and background classes must accurately represent events found in real data. In order to accomplish this, SVM training data sets are designed to contain the correct fraction of Monte Carlo events for each background process. As an example, the creation of the training data sets for the even-trained SVM will be discussed. The signal-even training sample is composed of events in the signal-even data set,

corresponding to approximately 5000 events. The background-even training sample is designed to accurately represent the composition of background in the real data and to have approximately 5000 events. The number of events of each background type contained in the background training sample is proportional to the expected yield of the given background. For example, if $t\bar{t}$ events account for 30% of the expected background, then $5000 \cdot 0.3 = 1500$ $t\bar{t}$ events are taken at random from the $t\bar{t}$ background-even data set and placed into the background-even training sample. This procedure is repeated for all relevant background processes until the full background training sample is created.

After creation of the training data sets, the two SVM's are trained using a Gaussian kernel. This training is completed using a C++ implementation of the SVM formalism developed by Benjamin Whitehouse and Jacob Borgman and modified by Jeffrey Wetter with help from Noah Kurinsky.

7.2 Determination of the Signal Region

Once the support vector machines are trained, their performance is assessed using the testing data sets. All odd events make up the testing data set for the even-trained SVM and are therefore classified using the even-trained SVM. Likewise, all even events are classified using the odd-trained SVM. This event classification results in each event being described by a single variable known as the SVM discriminant, d. The sign of SVM discriminant indicates the SVM's best guess as to which class the event belongs. Positive discriminants

are associated with signal events and negative discriminants with background events. The separating power of a given trained SVM is determined by measuring the ratio of the number of signal events to the square root of the number of background events in the signal region. This ratio is referred to as the significance,

$$s = \frac{N_{sig}}{\sqrt{N_{bkg}}}.$$ (7.1)

We have assumed that the number of background events in the signal region is described by a normal distribution, such that the standard deviation is given by $\sigma = \sqrt{N_{bkg}}$.

The signal region is defined by a single cut, ζ, on the SVM discriminant, with all events having $d \geq \zeta$ falling into the signal region. In order to determine the signal region that optimizes the significance, we perform a search over all possible values of ζ between $\zeta = 0$ and $\zeta = 3.5$, in increments of 0.1. The significance of the corresponding signal region is calculated for each value of ζ. The value of ζ that yields the highest significance while maintaining at least 4 signal events in the signal region is chosen as the signal region cutoff. The cutoff of at least 4 events in the signal region was used in order to avoid impractically small numbers of VBF $H \rightarrow WW$ events in the signal region.

Figure 7.1 shows the SVM discriminant for signal and background Monte Carlo events along with the optimized signal region. This plot was created using both even and odd numbered Monte Carlo events with each event scaled
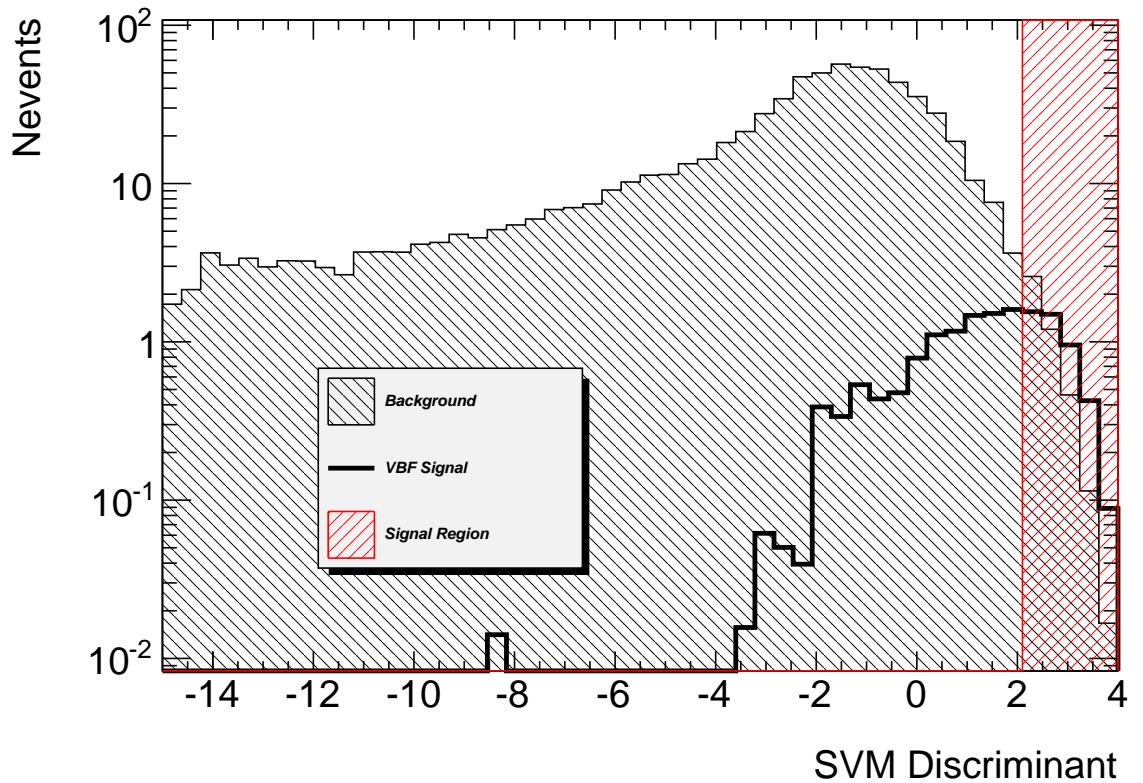
Figure 7.1: Signal-background discrimination or the 2-class SVM. The expected significance in the signal region is 2.22σ.

to represent the expected yield in real data. This scaling allows for less than one total event to be present in a given bin since each event has a very small weight.

7.3　Parameter Optimization

7.3.1　SVM Parameters

The SVMs described in this analysis allow for a soft margin and are trained using a Gaussian kernel. As described in section 6, the soft margin introduces the free parameter, C, which describes the rigidity of the margin. In addition, the Gaussian kernel, (6.16), relies on the free parameter σ, which describes the width of the Gaussian distribution. σ should in general be on the order of the size of the input parameters, in this case $\sigma \approx 1$. The value of C is varied logarithmically, as small variations in C do not affect SVM performance. Optimal values of C and σ were found by performing a grid scan using the significance of the signal region as a figure of merit.

The grid scan was performed by training a series of SVMs with different values of C and σ and evaluating the separating power of each SVM. The values of C and σ were allowed to vary between 0.5 and 100. In order to maintain stability between SVMs, the training samples were kept constant between trainings. The optimal values of the two free parameters were determined to be $\sigma = 1.2$, $C = 4$. Once set, these values were not changed throughout the remainder of the analysis.
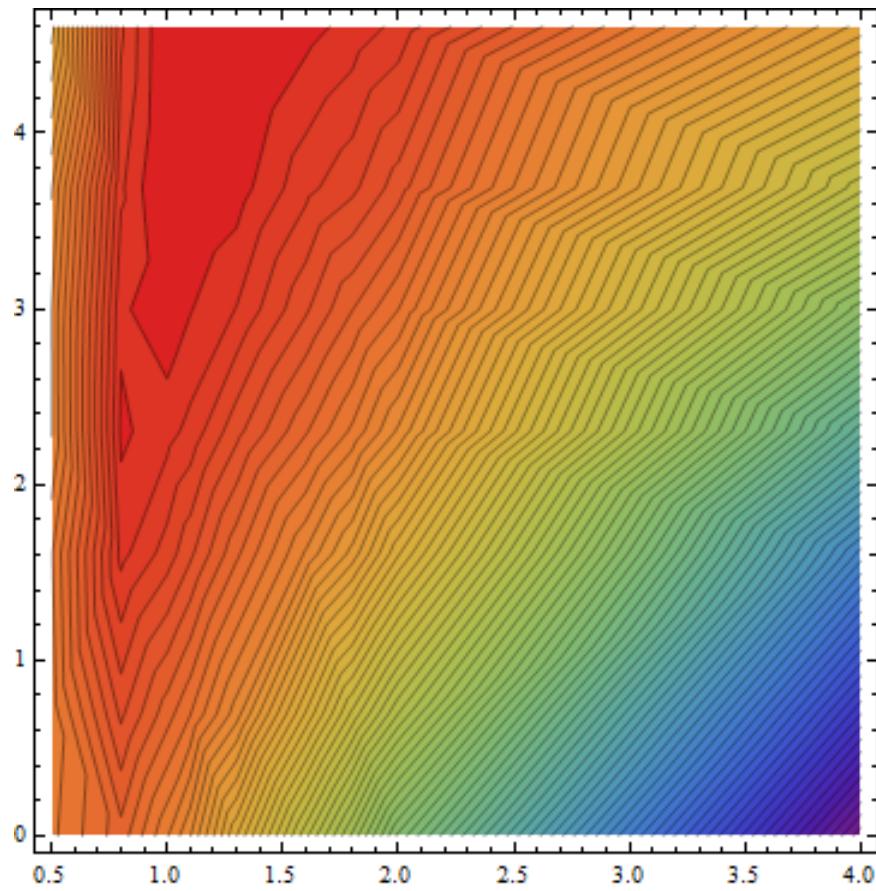
Figure 7.2: Grid scan over values of C and σ. Red indicates areas of better SVM performance.

7.3.2 Input Parameters

The design of a support vector machine analysis requires that the input parameters be mapped into an n-dimensional parameter space, where n is the number of input parameters. The input parameters are treated as components of vectors in this space. As such, it is advisable to scale all input parameters to be of the same order of magnitude to prevent parameters with naturally larger values, p_T for example, from dominating parameters with naturally smaller values, $\Delta\phi$ for instance. For this analysis, each input parameter was initially given a scale factor to bring its value to between approximately 0 and 1. Modifying the scale factors on a parameter-by-parameter basis influences the result by prioritizing certain parameters over others.

In order to explore the effect of the parameter scaling on the SVM training, a Markov Chain Monte Carlo (MCMC) study was conducted using significance in the signal region as a figure of merit. The Markov Chain Monte Carlo technique involves using a random walk to modify the scale factor associated with each input parameter.

Initially, the value of each scale factor was set manually such that each input parameter had a maximum value near 1. Once the scale factors are set, two SVMs (even-trained and odd-trained) are trained and the significance of the optimal signal region is determined. The MCMC then slightly modifies the values of individual scale factors and the SVMs are retrained. The significance of the new SVMs, trained using the modified parameter scaling, are then

Parameter	p_T^{total}	η_{lep} cent.	ΔY_{jj}	m_{jj}	m_{ll}	$\Delta\phi_{ll}$	m_T	$\Sigma_{l,j}m_{lj}$
Scaling	1/100	1/2	1/12	1/3000	1/200	$1/\pi$	1/200	1/3000

Table 7.1: Scale factors applied to SVM input parameters

compared with the significance of the previous SVMs. If the new SVMs result in a higher significance, then the new scale factors are kept. If the new scale factors do not result in an increase in significance, then the previous scale factors are retained. Once the comparison between the new and old SVMs is completed, the scale factors are again modified slightly and the process is repeated. To avoid being stuck in a false maximum in the scale factor phase space, the MCMC has a small probability to accept a new set of scale factors even if the significance of the new SVM training is lower.

An MCMC study was conducted using 10,000 trials. Figure 7.3 shows the expected significance values obtained over the 10,000 trials. The final scale factor values used for the 2-class analysis were chosen to give a significance of approximately 2.25σ. A set of scale factors were chosen which are stable under small variations in value and which provided a 10% increase in expected significance over the baseline scaling. Table 7.1 lists an example of the scale factors used for each input parameter.
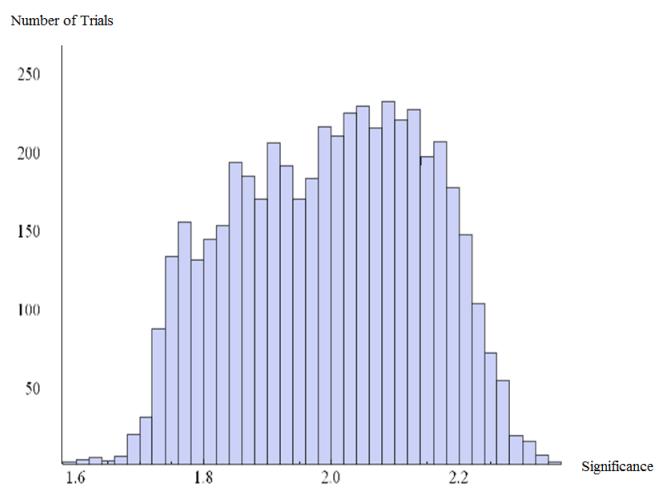
Figure 7.3: Plot of expected significance in the 2-class SVM analysis over 10,000 MCMC trials.

Chapter 8

3-Class Support Vector Machine Analysis

In addition to the 2-class analysis described in chapter 7, a 3-class SVM analysis was performed. 3-class SVM analyses involve splitting background samples into two separate classes referred to as light and heavy. A separate SVM is trained to distinguish between each of the three classes, resulting in a total of 3 SVMs: Signal vs. Light (SvL), Signal vs. Heavy (SvH) and Light vs. Heavy (LvH). Once these SVMs are trained, each event is classified using all 3 SVMs. The result of this procedure is that all events are described by three SVM discriminants and can be plotted in 3-space.

The goal the 3-class SVM procedure is to allow better discrimination between different types of background. We hypothesize that grouping all background events into the same class effectively removes information from the

analysis and may lead to non-optimal class separation. Figure 8.1 illustrates a simple example of a 3 class SVM system in which each event in characterized by 2 parameters, (H_1, H_2). The red points indicate signal events while the green and blue points represent two separate background classes. We can see that the blue background points are distinct from both red and green points and that a simple hyperplane can be formed between the blue points and all other points. By allowing blue events to constitute a separate class, a simple separation can be made between the red signal events and all blue events, leaving only the problem of separating the red and green events. This simple separation might not be possible in the case of a 2-class analysis, as the blue background events would be grouped with the green background events. One goal of our analysis is to test for improvements in signal-to-background separation by using a 3-class SVM system in comparison with a 2-class analysis.

For this analysis, each background type would ideally have its own class, allowing the SVM to detect differences between individual background processes. However, the machinery and computing power required to create this many classes is impractical for our analysis. As such, the choice was made to separate the background processes into two classes, with all top quark backgrounds being placed into the heavy class and all other background being placed into the light class. Top quark events make up over 50% of all background events after preselection. As such, placing all top quark events into the heavy class results in approximately the same number of expected events in the light and
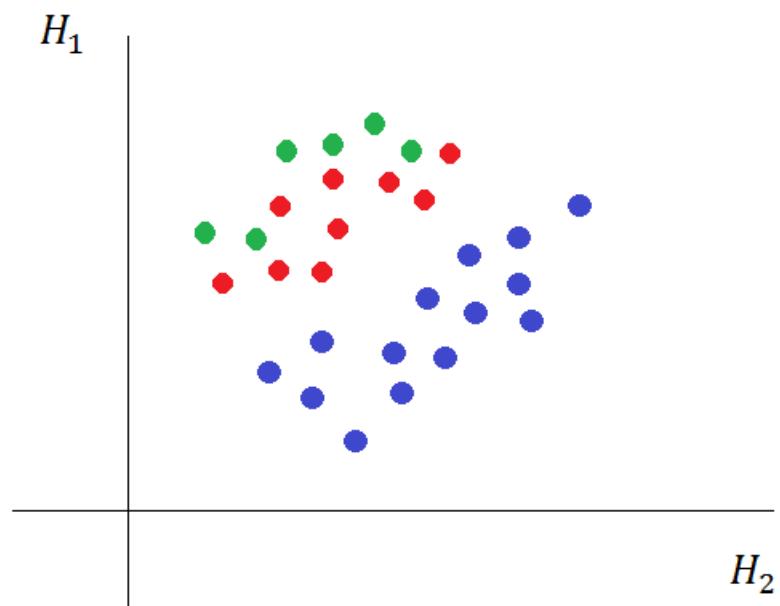
Figure 8.1: Example of a 3-class analysis with red points representing signal events and green and blue points representing two separate classes of background events.

heavy classes. In addition, top quark events ($t\bar{t}$ and tW) are unique from other backgrounds in that they contain two W bosons and high-p_T jets at tree level.

This section will discuss the training and classification procedure of events in the 3-class SVM analysis. In addition, the analysis of test data sets and real data sets will be discussed.

8.1   SVM Training and Classification

All events used in this analysis are required to pass the preselection cutflow as described in Chapter 4. Once selected, Monte Carlo events are split into 6 independent data sets. First, the events are split into two independent data sets based on event number, with all even-numbered events placed in one set and all odd-numbered events placed in another. The two data sets are then further subdivided into 3 independent classes labeled as signal (S), heavy (H) and light (L). The signal class contains only $H \rightarrow WW \rightarrow l\nu l\nu$ VBF events, while top quark events, including $t\bar{t}$ and single top backgrounds, are placed into the heavy class. All other backgrounds, including vector-boson, ggf Higgs and qcd backgrounds are placed into the light class. Vector-Higgs samples are not included in this analysis as their topological signature differs significantly from that of the VBF signal events and they constitute less than 1% of the predicted Standard Model yield after preselection.

The even-numbered and odd-numbered data sets provide completely independent samples with which to train two sets of independent SVMs in parallel.

For a single analysis, only one set of trained SVMs is required, as such, the even and odd-numbered data sets are used as a cross check. As our support vector machine formalism does not support negative event weights, any event with negative weight is ignored in training and classification. Events with negative weight are, however, included in Standard Model yield predictions used in the construction of training and pseudo-data sets. The same procedure is followed to train and classify the even and odd-numbered data sets. However, for the sake of clarity, we will discuss only the training of the even-numbered events.

In order to train the three separate SVMs, 3 sets of training samples must be created. Each training sample is designed to accurately represent the composition of real data for a specific class. For the purposes of SVM training, each event is considered to have a weight of unity. It is therefore essential to create training data sets which contain the correct number of events of each process. To accomplish this goal, training sets are created such that the percentage of events for each background process is equal Standard Model predictions. To illustrate this process, the creation of the heavy training sample, containing only top quark events, will be discussed. The total expected yield after preselection of all samples included in the heavy class is approximately 373 events. Of those expected events, approximately 331 are $t\bar{t}$ events. Therefore, $331/373 = 89\%$ of the events in the even-numbered heavy training sample are taken from the even-numbered $t\bar{t}$ data set.

133

The same procedure is followed for all other samples in the heavy and light classes. The total number of events in the heavy and light training classes is set to match the number of events in the signal training sample. The signal training sample is made up of only $H \rightarrow WW^* \rightarrow l\nu l\nu$ VBF events. In order to utilize the largest possible training sample, all even-numbered VBF events are included in the signal training sample, corresponding to approximately 5000 events.

Once training samples have been created representing the light, heavy and signal classes, two sets of three SVMs are created. A set of SVMs is trained using a Gaussian kernel to separate between each of the three classes using only even numbered training samples, resulting in three separate SVMs: SvL, SvH, LvH. The same procedure is followed for the odd number training samples, resulting in a total of 6 trained SVMs. After each SVM has been trained, all events are classified using the opposite numbered SVMs (i.e. all even numbered events are classified using all three odd-trained SVMs and all odd numbered events are classified using all three even-trained SVMs). This training and classification method allows all Monte Carlo events passing preselection to be parameterized by three numbers corresponding to the discriminants of each of the three SVMs.

8.2 Gram-Schmidt Orthogonalization

After event classification, each event can be described in vector notation using the SVM discriminants from the 3 SVMs, (SvL, SvH, LvH). These discriminants tell us about the distance from each SVM's separating hyperplane to the point of interest in feature space. However, we can not simply plot these three SVM discriminants as points in 3-space, as the separating hyperplanes are not necessarily orthogonal in feature space. As an example, Figure 8.2 shows three classes of points plotted in feature space with hyperplanes separating each of the three classes. For the purposes of this illustration, we have chosen a simple example in which the separating hyperplanes are straight lines. We can see in Figure 8.2 that moving perpendicularly away from one hyperplane necessitates moving toward or away from another hyperplane. The distance of a point of interest from any of the three separating hyperplanes is dependent on the distance of that point from the other two hyperplanes. This is clearly a problem if we would like to plot our events in standard 3-space, since moving in the y direction should not affect our x position. Therefore, plotting our 3 SVM discriminants of classified events, (SvL, SvH, LvH), in 3-space would misrepresent the actual relationship between classified events and the three separating hyperplanes. To avoid this problem we utilize the Gram-Schmidt orthogonalization procedure to determine an orthogonal basis in which to plot our events.

The Gram-Schmidt procedure is based on the projection of one vector onto
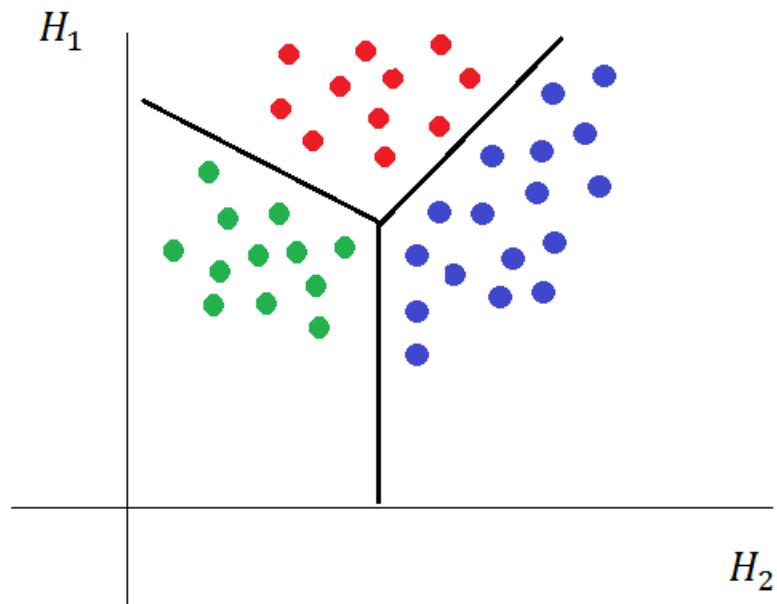
Figure 8.2: 3 classes of events represented by red, green and blue points. Each event is parameterized by two values (H_1, H_2). The three hyperplanes, each separating two classes of events, are non-orthogonal. This prevents the svm disrciminants from being plotted independently in 3-space.

another. For example, given two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, we say that the projection of $\mathbf{u}_2$ onto $\mathbf{u}_1$ is given by

$$\text{proj}_{\mathbf{u}_1} \mathbf{u}_2 = \frac{<\mathbf{u}_1, \mathbf{u}_2>}{<\mathbf{u}_1, \mathbf{u}_1>} \mathbf{u}_1, \tag{8.1}$$

where $<\mathbf{u}_1, \mathbf{u}_2>$ indicates the inner product of vectors $\mathbf{u}_1$ and $\mathbf{u}_2$. Given three vectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$, the Gram-Schmidt procedure allows us to write three orthogonal bases as

$$\mathbf{u}_1' = \mathbf{u}_1, \tag{8.2}$$

$$\mathbf{u}_2' = \mathbf{u}_2 - \text{proj}_{\mathbf{u}_1} \mathbf{u}_2, \tag{8.3}$$

$$\mathbf{u}_3' = \mathbf{u}_3 - \text{proj}_{\mathbf{u}_1} \mathbf{u}_2 - \text{proj}_{\mathbf{u}_2'} \mathbf{u}_3. \tag{8.4}$$

In the case of this analysis, we are interested in finding an orthogonal basis for the separating hyperplanes of three distinct SVMs (SvL, SvH, LvH). From Chapter 6 we recall that each separating hyperplane is parameterized in terms of a normal vector $\mathbf{w}$. We can parameterize $\mathbf{w}$ for each trained SVM as

$$\mathbf{w}_{SvL} = \sum_{i}^{SV_{SvL}} y_i \alpha_i \varphi(\mathbf{x}_i). \tag{8.5}$$

$$\mathbf{w}_{SvH} = \sum_{i}^{SV_{SvH}} y_i \alpha_i \varphi(\mathbf{x}_i). \tag{8.6}$$

$$\mathbf{w}_{LvH} = \sum_{i}^{SV_{LvH}} y_i \alpha_i \varphi(\mathbf{x}_i). \tag{8.7}$$

The sum is taken over all support vectors for the given trained SVM. y_i represents the known class of the event x_i, and α_i is the corresponding Lagrange multiplier. φ is the mapping function from feature space to decision space. Applying the Gram-Schmidt procedure to the normal vectors, $\mathbf{w}$, we can define the (x, y, z) coordinates of classified events in an orthogonal basis as

$$X(\mathbf{x}) = \frac{<\mathbf{x}, \mathbf{w}'_{SvL}>}{<\mathbf{w}'_{SvL}, \mathbf{w}'_{SvL}>}, \tag{8.8}$$

$$Y(\mathbf{x}) = \frac{<\mathbf{x}, \mathbf{w}'_{SvH}>}{<\mathbf{w}'_{SvH}, \mathbf{w}'_{SvH}>}, \tag{8.9}$$

$$Z(\mathbf{x}) = \frac{<\mathbf{x}, \mathbf{w}'_{LvH}>}{<\mathbf{w}'_{LvH}, \mathbf{w}'_{LvH}>}. \tag{8.10}$$

We have therefore found a means of translating events and their SVM discriminants into points in an orthogonal 3-space. Using this procedure, it is possible to directly compare the SVM discriminant vectors of events, allowing for the creation of class templates and probability distribution functions as discussed in the next section.

8.3   Template and PDF Creation

After event classification, each event can be described in vector notation using the SVM discriminants from the 3 SVMs (SvL, SvH, LvH). A scatter plot of the Gram-Schmidt orthogonalized SVM discriminants for all even-numbered events is shown in Figure 8.3. As can be seen from this scatter plot, each class inhabits its own phase space, with distinct regions of overlap.

Our goal is to establish the probability that an event of unknown type belongs to each of the three classes. Logically, we would expect that events falling into the top-right corner of Figure 8.3 would most likely be signal, while events in the bottom left corner would be light or heavy. In order to quantify this probability, we convert the scatter plot of events associated with each class into a binned 3-dimensional histogram known as a template. Each bin in the template is characterized by the total weight of events falling into that bin. The entire histogram is normalized such that the total weight of each template is unity. For the creation of templates, all available MC events are used. Each event is weighted by Monte Carlo scale factors as described in Chapter 5 and normalized such that the total weight of each background process is proportional to Standard Model predictions.

In order to more accurately represent the true weight associated with individual points, the binned templates are interpolated into smooth functions. In order to find the interpolated weight associated with a specific point, a weighted average is taken of the bins closest to the point of interest. Figure 8.4 shows a contour plot of the signal, light and heavy class interpolated templates. The three contour plots are superimposed in order to show class separations. We interpret an individual interpolated template as a probability distribution function (pdf) relating points described by three SVM discriminants (SvL, SvH, LvH) to the probability that an event described by this point belongs to a specific class.

At this point in our analysis we have trained two separate sets of SVMs, one on all even events and one on all odd events. We then classified all events using SVMs trained on opposite-numbered events. Finally we have created interpolated templates which we can use to estimate the probability that a given event is a member of any of the three classes. We are now prepared to analyze a set of data in which each event comes from an unknown class. Our goal is to determine the number of events from each of the three classes contained in the sample data set. In order to accomplish this goal, we use a likelihood technique.

8.4   Likelihoods

We have now developed a method for determining the relative probability that an unknown event belongs to each of the three classes. Using this information we can write down the likelihood function,

$$L = \prod_i^N (\theta S_i + \phi L_i + \eta H_i). \tag{8.11}$$

The index i runs over all events in the sample and S_i, L_i and H_i are the PDF values of the signal, light and heavy classes for a given event. θ, ϕ and η are the percentage of signal, light and heavy events in the unknown sample. η is a dependent parameter, defined to be $\eta = (1 - \theta - \phi)$ such that the total percentage of signal, light and heavy events sums to unity. This likelihood
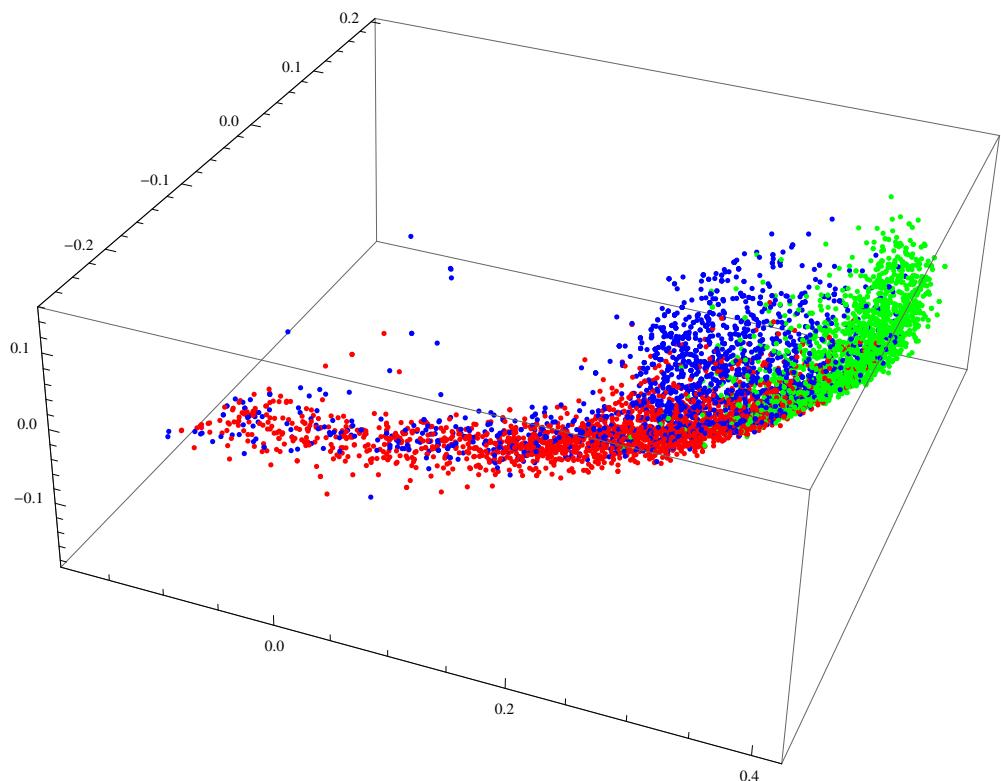
Figure 8.3: 3-dimensional plot showing the (SvL, SvH, LvH) SVM discrimi-
nants for each event. Events are color-coded by class as Signal, Heavy and
Light.
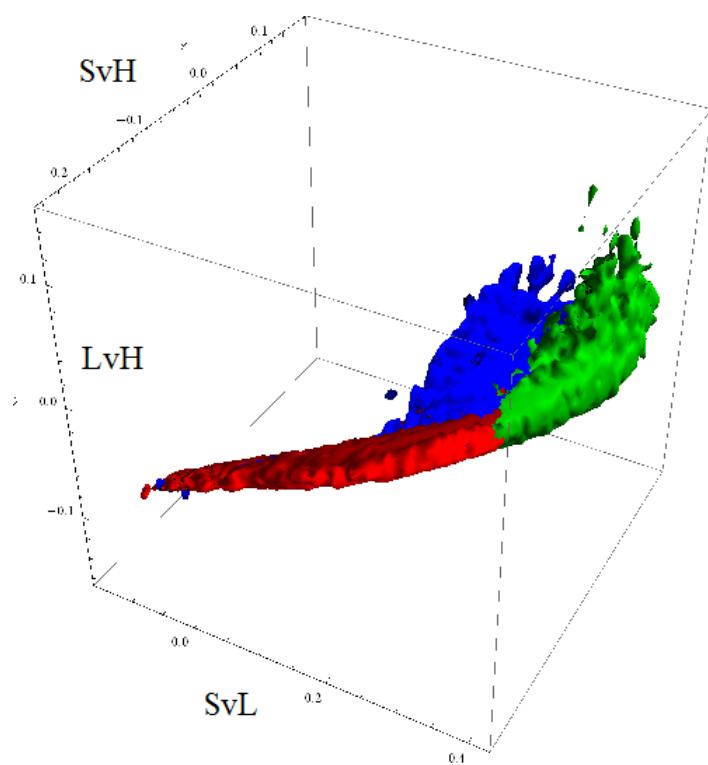
Figure 8.4: A contour plot of PDF templates created from Figure 8.3. Events
are color-coded by class as Signal, Heavy and Light.

function will have a maximum at values of θ, ϕ and η that correspond to our best guess as to the actual signal, light and heavy event fractions in the unknown sample.

Though this likelihood formulation has simplified our problem, maximizing a product over hundreds or thousands of events is computationally expensive. In order to further simplify our problem, we replace the likelihood function, (8.11), with a log-likelihood,

$$\Lambda = ln(L) = \sum_i ln(\theta S_i + \phi L_i + \eta H_i). \tag{8.12}$$

The log-likelihood function will have a maximum at the same values of θ, ϕ and η as the likelihood function, but has the advantage of using a sum over i instead of a product.

In order to maximize the log-likelihood function, a C++ version of the MINUIT package was used. However, MINUIT is not designed for problems involving mutual restrictions as is the case in (8.12). Specifically, the value of θ and ϕ must satisfy the requirement,

$$\theta < 1, \quad \phi < 1, \quad \theta + \phi \leq 1. \tag{8.13}$$

In order to avoid the problem of mutual restrictions within MINUIT, the variable transform

$$\theta = x_1, \quad \phi = (1 - x_1)x_2, \quad \eta = (1 - x_1)(1 - x_2) \tag{8.14}$$

143

was used. The variables x_1 and x_2 were restricted to lie in the range $(0 < x_1 < 1, 0 < x_2 < 1)$. In this manner, MINUIT was only responsible for maximizing the log-likelihood, (8.12), with respect to two free parameters, while preserving the mutual restrictions imposed on the likelihood.

8.5 Pseudo-Experiments

We have now created a methodology for determining the fraction of events of each class in a test sample. To accomplish this, we first classified all Monte Carlo events passing preselection cuts using three trained SVMs. We then used these classified events to create templates, allowing us to estimate the probability that a given event belongs to each of the three classes. Finally, we developed a likelihood function designed to determine the fraction of events of each class in a sample containing events of unknown type. The ultimate goal is to use this likelihood procedure to estimate the number of HWW signal events in a real data sample. However, first we need to test the accuracy of our method by conducting pseudo-experiments using Monte Carlo events.

Pseudo-experiments involve the use of pseudo-data sets containing Monte Carlo events designed to recreate real data samples. In order to create a pseudo-data set, Monte Carlo events are taken from all relevant background samples in proportion to their expected yields. The number of events from each background type is allowed to fluctuate according to a Poisson distribution in order to simulate uncertainty in the number of background events in the real

data. The mean number of events in each pseudo-data set must match the number of real data events from the $20\,fb^{-1}$ ATLAS data set which pass all preselection cuts. The number of VBF HWW signal events in each pseudo-data set was varied in order to test the robustness of the SVM methodology against various signal strengths.

Each pseudo-data set is analyzed using the likelihood method described in section 8.4. The maximized likelihood provides our best estimate of the fraction of signal, light and heavy events in a given pseudo-data set. Since the exact fraction of each class in each pseudo-data set is known, we can compare the likelihood estimate of the fraction of signal events in the pseudo-data set with the true number of signal events. The true number of signal events in a given data set is labeled ρ and the likelihood estimated fraction of signal events is labeled as R. We can compare these two numbers to find the bias, defined as

$$\beta = R - \rho. \tag{8.15}$$

Assuming a perfect analysis, the bias should be zero for every pseudo-data set. However, for a real analysis this is generally unattainable. If the bias is stable across pseudo-data sets with various signal fractions, we can compensate for this bias when analyzing the real data. We correct our best guess as to the fraction of signal events in the real data set using the equation

$$\theta_{true} = \theta_{fit} - \beta. \tag{8.16}$$

θ_{true} represents the bias-corrected best guess as to the fraction of signal events in the real data sample and θ_{fit} is the fraction of signal events returned by the SVM analysis procedure.

Chapter 9

Systematics

Unfortunately, it is not possible for us to open up the detector and peer in upon the Higgs boson as it is created. We must instead rely on the ATLAS detector to observe the decay products of the Higgs boson and on reconstruction algorithms to turn information from the detector into physics objects. Throughout this process, we rely of many assumptions about the detector, including the rate at which various objects are correctly identified, the rate at which objects are not detected and the way in which an object's measured energy relates to the object's true energy. In addition, we rely on assumptions about our knowledge of object interactions, decay rates and momenta in order to create accurate Monte Carlo simulations of the Higgs boson decay as well as background processes. Each of these assumptions brings with it an associated uncertainty referred to as a systematic error. In order to properly assess the error on our measurement of the Higgs boson, we must identify and

asses as many sources of systematic uncertainty as possible. This section will describe the most important sources of systematic uncertainty for the VBF *HWW* analysis as well as the methods used to analyze systematic errors.

9.1   Systematics Errors within the 3-class Analysis

We are interested in determining the effect of each source of systematic uncertainty on our estimate of the number of signal events in the 2012 ATLAS data set. To this end, we use dedicated Monte Carlo data sets specifically designed to vary parameters associated with specific systematic uncertainties. For example, it is possible that the MC samples used in this analysis overestimate the electron identification efficiency. In order to estimate the effect that this would have on the real data analysis, a specific systematic Monte Carlo sample is created in which the electron identification rate is decreased. By analyzing the difference in results between the nominal and systematic SVM analyses, it is possible to estimate the uncertainty due to each source of systematic error.

In order to accurately estimate the effect of each type systematic uncertainty on the real data analysis, it is important that systematic samples are treated in the same manner as real data samples. This means that systematic samples are classified using SVMs trained with nominal MC samples and systematic pseudo-experiments must be performed using nominal pdf templates.

A set of pseudo-experiments is performed for each systematic MC sample. The difference in the estimated number of signal events in the nominal and systematic pseudo-data sets is taken to be the uncertainty for each systematic source.

9.2   Detector Modeling Systematics

After Monte Carlo samples are generated, showered and reconstructed, scale factors are applied in order to match the distributions of important variables between Monte Carlo and real data samples. These scale factors can be applied in two different ways, the first being to apply an individual weight to each event. Event weights are taken to be the value of a given event when plotting histograms or determining yields. By applying event weights, it is possible to adjust the shape of various distributions to better match data. For example, if the Monte Carlo simulated data is not producing enough events with 4 jets, a scale factor could be applied to events with $N_{jets} = 4$ in order to boost the number of expected 4-jet events in Monte Carlo samples.

The second method of modifying the distribution of Monte Carlo samples is to apply scale factors directly to the parameter of interest. This technique is primarily used for modifying the expected energy of objects such as leptons and jets. By applying scale factors to specific parameters, it is possible to influence a single distribution without affecting the distributions of other event parameters.

Each method of scaling Monte Carlo events comes with an associated uncertainty. In order to asses this uncertainty, specific Monte Carlo data sets are created in which a single scale factor is modified up or down by one standard deviation. All other scale factors are left at their nominal values, except in the case of correlations. In this way, each source of systematic uncertainty can be evaluated independently. In the following sections, the event and parameter scalings associated with various objects and their systematic assessments will be discussed.

9.2.0.1 Lepton Uncertainties

In order to match the trigger, reconstruction and identification [31][32][33][34] rates of leptons between Monte Carlo and data samples, scale factors are applied to the weight of each Monte Carlo event. These scale factors are determined by comparison of lepton trigger, identification and reconstruction rates between data and Monte Carlo samples in Z/DY events. Z/DY events are unique in that their final state contains two high-energy leptons with invariant mass close to the Z mass and no tree-level jets or missing energy. This simple final state allows for the creation of a very pure set of real data Z/DY events, allowing for the direct comparison of MC and real data events.

Once a set of data and Monte Carlo Z/DY events have been created, the tag and probe technique is used to identify the rate at which leptons are triggered, reconstructed and identified. The tag and probe technique involves

the use of a well-identified tag lepton and a less-well identified probe lepton. The rate at which the probe lepton is able to be identified and reconstructed is compared between the MC and data samples. The associated lepton scale factor is taken to be the ratio of the data and MC rates. For example, the electron identification scale factor is taken to be the rate at which the probe electron is identified in real data divided by the rate at which the probe electron is identified in Monte Carlo samples.

In addition to scale factors applied to the overall event weight, lepton energies and resolution are also corrected in order to better match Monte Carlo samples to real data [35][36]. These scale factors are determined by comparing the width and mean of di-lepton invariant mass distributions in Z/DY events. In order to evaluate systematic uncertainties based on all lepton scale factors, each scale factor is independently varied up and down by one standard deviation.

9.2.1 Jet Energy Uncertainties

In order to account for differences in jet energy and resolution between MC and data events, scale factors are applied directly to jet objects. The systematic uncertainties due to the jet energy scale (JES) corrections are divided into 12 separate systematics as per the recommendation of the Higgs working group. These 12 systematic uncertainties are:

- in-situ: detector

- in-situ: modeling

- η inter-calibration: modeling

- η inter-calibration: stat+method

- high p_T jets

- pileup: μ

- pileup: number of primary vertices (NPV)

- pileup: p_T

- pileup: ρ topology

- flavor composition

- flavor response

- b-JES

These 12 uncertainties include four systematics dedicated to the modeling of pileup. The pileup conditions within the detector are important for the correct modeling of jet energies. This is due to the fact that jets are reconstructed by adding up all of the energy deposited within a cone surrounding the jet within the calorimeter. Pileup objects such as soft gluon radiation may be present within the jet's cone and therefore contribute to the total reconstructed jet energy. The reconstructed jet energy is therefore scaled to account

for the pileup conditions. To asses the uncertainty associated with this scaling, four separate JES systematics are analyzed. Pileup: μ accounts for possible differences in the pileup conditions between MC and data events. The pileup p_T and pileup ρ systematics estimate the uncertainty in the p_T and total event energy due to pileup events. Finally, pileup NPV accounts for the number of additional primary vertices due to additional collisions within the detector.

There are 2 systematics dedicated to the in-situ jet calibration: in-situ: detector, which estimates uncertainties in how jets are reconstructed within the detector, and in-situ: modeling, which accounts for uncertainties in jet modeling. In addition, there are 2 η inter-calibration systematics which estimate the uncertainty in parton showering models. In particular, these systematics account for potential mismodeling of additional radiation within the hadronic calorimeter.

The flavor composition systematic accounts for uncertainties in the fraction of light quarks and gluons within jets, while the flavor response systematic estimates uncertainties specifically due to gluon jets. The jet energy scale of b jets is accounted for independently using the b-JES systematic.

In addition to the JES systematics, a jet energy resolution (JER) systematic is also evaluated by smearing the energy of reconstructed jets.

9.2.2 Missing Transverse Momentum Systematics

Missing transverse momentum (E_T^{miss}) is calculated by summing the p_T of all hard objects emitted from the primary vertex as well as soft (i.e. low energy) objects. As such, the calculation of the E_T^{miss} relies on proper scaling of hard jets and leptons as well as the scale and resolution of soft objects. In order to estimate the systematic uncertainty of the missing transverse momentum due to hard objects, the jet and lepton uncertainties discussed above are propagated into the E_T^{miss}. In addition, separate systematic uncertainties are estimated by modifying the scale and resolution of soft objects used in the calculation of E_T^{miss}.

9.2.3 b-tagging Systematics

Within this analysis, b-jets are identified using the MV1 tagging algorithm with an 85% working efficiency. In order to estimate the systematic uncertainty associated with b-tagging, a set of 6 uncorrelated systematic variations are used, corresponding to the number of p_T bins used for b-tag calibration. These systematics vary the efficiency of b-tagging in Monte Carlo samples by modifying the event weight on an event-by-event basis. In addition, separate systematics are used in order to estimate the uncertainty in the mistagging of light and c jets as b jets.

9.3 Theory Uncertainties

Monte Carlo simulations represent our best estimate of the real particle interactions within the ATLAS detector. However, these simulations are in no way perfect representations of actual physical interactions and different Monte Carlo simulations may vary in important ways. In order to estimate the uncertainty due to possible differences between Monte Carlo and real data events, multiple Monte Carlo simulations are compared. As with detector systematics, dedicated pseudo-experiments are preformed using alternative Monte Carlo simulations and the difference in estimated signal fraction between nominal and systematic pseudo-experiments is taken to be the systematic error.

9.3.1 $t\bar{t}$ Systematics

$t\bar{t}$ events make up the largest single background to VBF HWW events. As such, proper generator modeling of $t\bar{t}$ events is of particular importance. In order to estimate the uncertainty due to the choice of $t\bar{t}$ generator, analysis results were compared using the nominal Powheg+Phythia generator and parton shower and the MC@NLO+Herwig generator and parton shower. For this systematic analysis, all Monte Carlo samples except $t\bar{t}$ were unchanged and the nominal Powheg $t\bar{t}$ sample was replaced with an MC@NLO sample. The difference in the final result between pseudo-experiments using Powheg and MC@NLO $t\bar{t}$ samples was then taken to be the $t\bar{t}$ generator uncertainty.

In addition to the generators, the systematic uncertainty due to the choice

of parton distribution function (PDF) within the generators was also analyzed. Using the MC@NLO generator, pseudo-experiments were conducted using various PDF sets. It was determined that the systematic variation due to the choice of PDF set was negligible in comparison with the generator systematic. Therefore, systematic variations due to the choice of $t\bar{t}$ parton distribution functions were neglected.

Chapter 10

Results

The goals of the analyses outlined in this thesis are to validate the support vector machine analysis method in high energy physics searches as well as to search for evidence of vector boson Higgs production in the $H \rightarrow WW^* \rightarrow l\nu l\nu$ channel. In order to accomplish these goals, several analyses were performed. First, in order to offer direct comparisons with a parallel VBF $H \rightarrow WW^*$ analysis using boosted decision trees, a 2-class support vector machine analysis was performed. The 2-class analysis was designed to validate the support vector machine methodology as well as to allow for tuning of analysis parameters. After tuning, SVM analysis results using Monte Carlo data sets were compared with results from the boosted decision tree analysis. In addition, a VBF $H \rightarrow WW^*$ search was conducted on the 20 fb^{-1} 2012 ATLAS data set.

After completion of the 2-class SVM analysis, a 3-class SVM analysis was conducted. For this analysis, background events were split into two separate

classes and a set of 3 support vector machines was trained to distinguish between each of the three classes. The goal of this analysis was to investigate improvements in expected significance in multi-SVM systems as well as to conduct a VBF Higgs search in the 2012 ATLAS data.

All Monte Carlo samples used for the analyses described in this thesis are listed in Chapter 5. All Monte Carlo events were normalized to match expected yields in the 2012 ATLAS data set. Events with negative event weights were ignored, as negative event weights are not supported by our SVM machinery. For the nominal analysis, only the W+jets and qcd samples are allowed to have negative event weights and only a small fraction of events are negatively weighted. In addition, the $t\bar{t}$ generator systematic samples include events with negative weights due to the weighting scheme employed by MC@NLO. Studies were conducted both ignoring events with negative weight and using the absolute values of events weights. In all cases, the effect on final fit results was negligible.

After pre-selection, the 2012 ATLAS data set used in this analysis contained 667 events with an expected 12.7 VBF $H \rightarrow WW^*$ signal events based on Monte Carlo studies. In order to facilitate comparison with other analyses, all ntuples used by the 2 and 3-class SVM analysis were created using the ATLAS HWW group's centrally-produced ntuple creation code.

10.1 2-Class SVM Analysis Results

10.1.1 Expected Results and Comparison with BDT

The 2-class support vector machine analysis was performed using two separate SVMs, each trained to distinguish between VBF $H \rightarrow WW^*$ and background events. This procedure is designed to replicate the procedure for the ATLAS HWW subgroup's boosted decision tree analysis. One support vector machine was trained using only events with even event numbers and the other using only odd-numbered events. In order to train the machines, four separate training data sets were created as described in Chapter 7. The four data sets are labeled as signal-even, signal-odd, background-even and background-odd. Each training set consisted of approximately 5,000 events. After training, all events were classified such that all even-numbered events were classified by the SVM trained with odd-numbered events and vice versa. After classification, all Monte Carlo events were joined into a single data set for analysis.

In order to measure the ability of support vector machines to separate signal and background events, the significance, defined as $S/\sqrt{B}$ in the signal region, was used as a figure of merit. The full signal region selection procedure is described in Chapter 7. Using Monte Carlo samples described in Chapter 5, the expected significance of the 2-class SVM analysis was found to be 2.23σ with 4.9 signal and 4.9 background events expected in the signal region. This result was compared to the expected significance using the HWW group's

BDT analysis. Running on the same data set, the expected significance for the BDT analysis was 2.37σ with 5.1 signal and 4.7 background events expected in the signal region. These findings indicate that the SVM analysis performed slightly worse than BDT analysis. With further tuning, it is likely that the SVM analysis technique would yield equivalent expected significance to the BDT analysis. Figures 10.1 and 10.2 show the signal-background separation for SVM and BDT analysis.
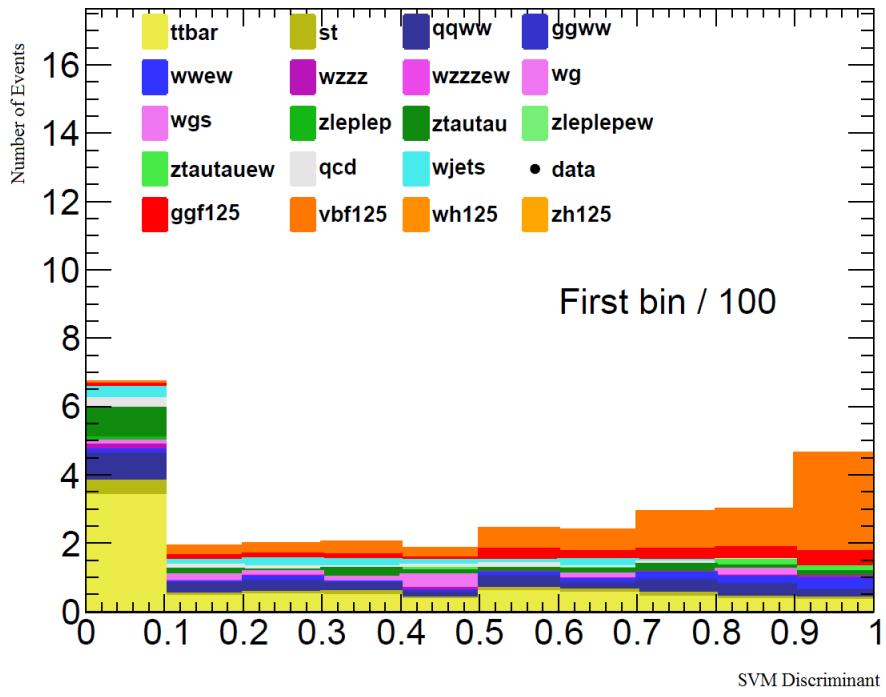
Figure 10.1: Plot of the SVM discriminant for all signal and background Monte Carlo events. The SVM discriminant has been scaled such that 98% of all background are in the first bin. The first bin has been scaled by 1/100 in order to maintain continuity.
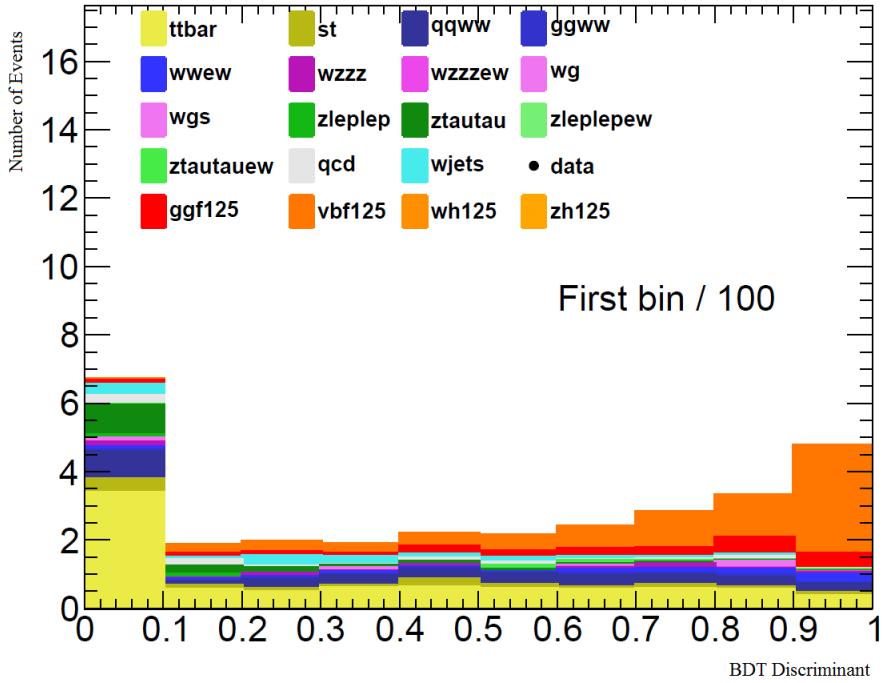
Figure 10.2: Plot of the BDT discriminant for all signal and background Monte Carlo events. The BDT discriminant has been scaled such that 98% of all background are in the first bin. The first bin has been scaled by 1/100 in order to maintain continuity.

10.1.2   Data Results

After completing Monte Carlo studies, a 2-class SVM analysis was performed on the $20\,fb^{-1}$ 2012 ATLAS data set. The same trained SVMs used in the 2-class Monte Carlo analysis were used to classify all data events. From the 667 events in the final data set, the 2-class SVM analysis yielded 9 data events in the signal region, leading to an observed significance of 1.85σ. As described in the preceding section, SVM Monte Carlo studies yielded an expected 9.8

total events in the signal region, with 4.9 signal events and 4.9 background events. The real data result is therefore consistent with Standard Model Higgs boson predictions. However, the small statistical significance of the result is insufficient to claim evidence for VBF $H \to WW^* \to l\nu l\nu$ decays. Figure 10.3 shows the distribution of real data events plotted by SVM discriminant such that the distribution of background MC events can be seen. Figure 10.4 shows the same data plotted such that the signal region is clearly distinguishable.
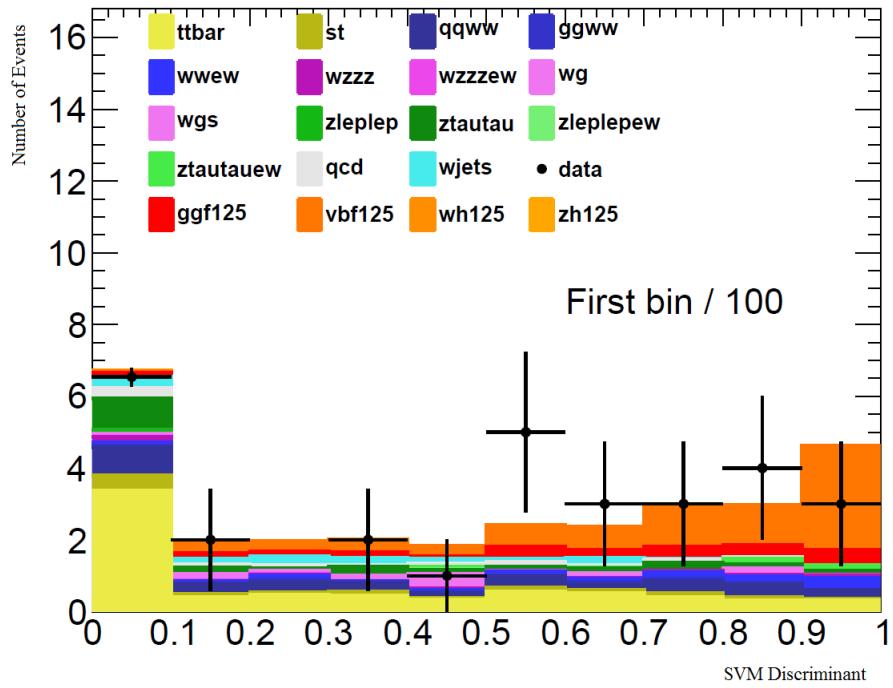
Figure 10.3: Plot of MC and data events for the $20\,fb^{-1}$ 2012 ATLAS data set. The SVM discriminant has been scaled such that 98% of all background are in the first bin. The first bin has been scaled by 1/100 in order to maintain continuity.
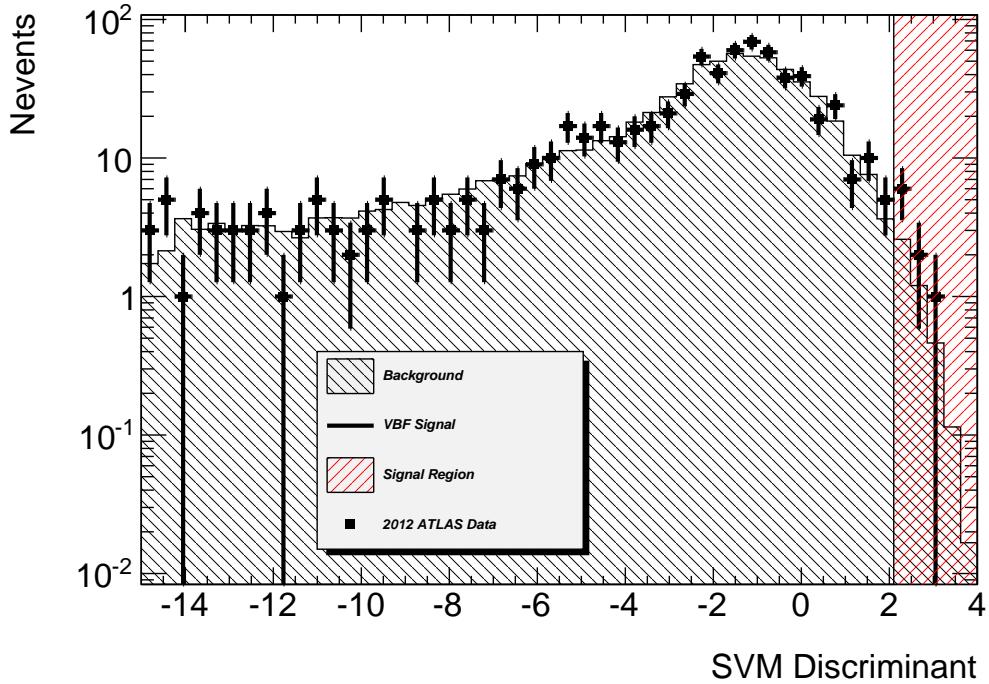
Figure 10.4: SVM discriminant for MC and 2012 ATLAS data events. The red shaded area indicates the signal region.

10.2   3-Class SVM Analysis Results

In addition to the 2-class analysis, a 3-class analysis was performed in which the background samples were split into two classes. The two background classes consisted of a 'heavy' class containing all top quark backgrounds and a 'light' class containing all other backgrounds. Using training samples created from each of the three classes, two sets of three support vector machines were trained. One set of three SVMs was trained using only even-numbered Monte Carlo events and another set of three was trained using only odd-numbered

Monte Carlo events. The training, classification and analysis procedure for the 3-class SVM analysis is detailed in Chapter 8. All MC results shown in this section are from SVMs trained on even-numbered events, however, the results are equivalent for SVMs trained using odd-numbered events. Data results are taken to be the average of the even-trained and odd-trained SVM analysis results for the 2012 ATLAS data set.

10.2.1 Signal-Background Separation

Figures 10.5-10.7 show the separation of signal, light and heavy events using the signal vs. light (SvL), signal vs. heavy (SvH), and light vs. heavy (LvH) support vector machines. Figures 10.5 and 10.6 show good discrimination of signal events from both heavy and light background events. Figure 10.7 displays some discriminating power, with red signal events having more positive values on average than the light and heavy backgrounds. The lack of separation between event classes in the LvH support vector machine discriminant is most likely due to topological similarities between light and heavy events. The SVM input variables used in this analysis were specifically chosen to exploit differences between VBF $H \to WW^*$ events and background events. Topological signatures such as two forward jets and two leptons close together in ϕ are unique to VBF $H \to WW^*$ events and are generally not shared by light or heavy background processes. Thus the distributions of many of the SVM input parameters are very similar for heavy and light background events as
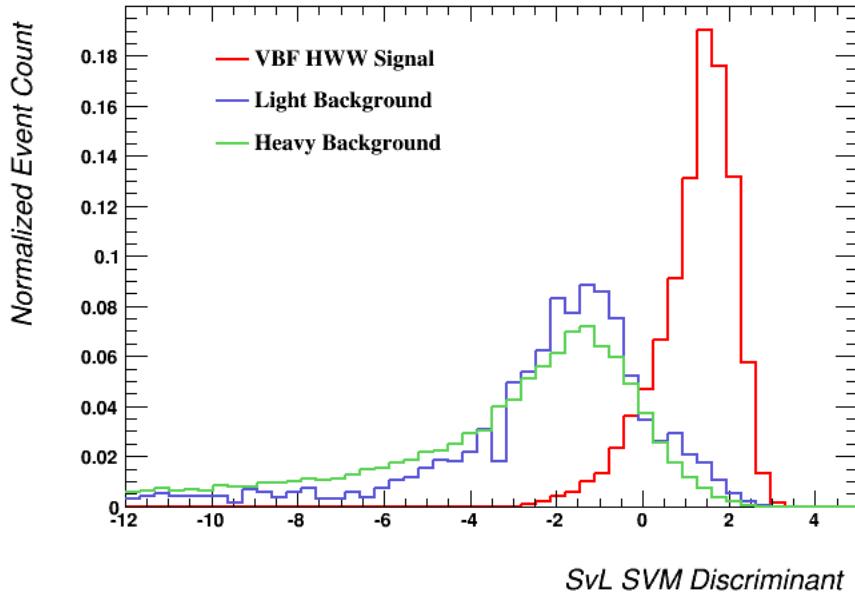
Figure 10.5: Distribution of SvL support vector machine discriminants for signal, light and heavy events. The total integral of each class has been normalized to unity.

can be seen by the plots in Chapter 4. This similarity makes it extremely difficult for a support vector machine to distinguish between light and heavy events based on the input parameters provided.

10.2.2 Statistical Uncertainty and Bias

In order to determine the expected statistical significance of the 3-class support vector machine search for VBF $H \rightarrow WW^*$ events, a set of pseudo-experiments was performed using Monte Carlo samples. Each pseudo-experiment consisted of the creation and analysis of a pseudo-data set, designed to replicate the ATLAS 2012 data set. Each pseudo-data set was created with a known number
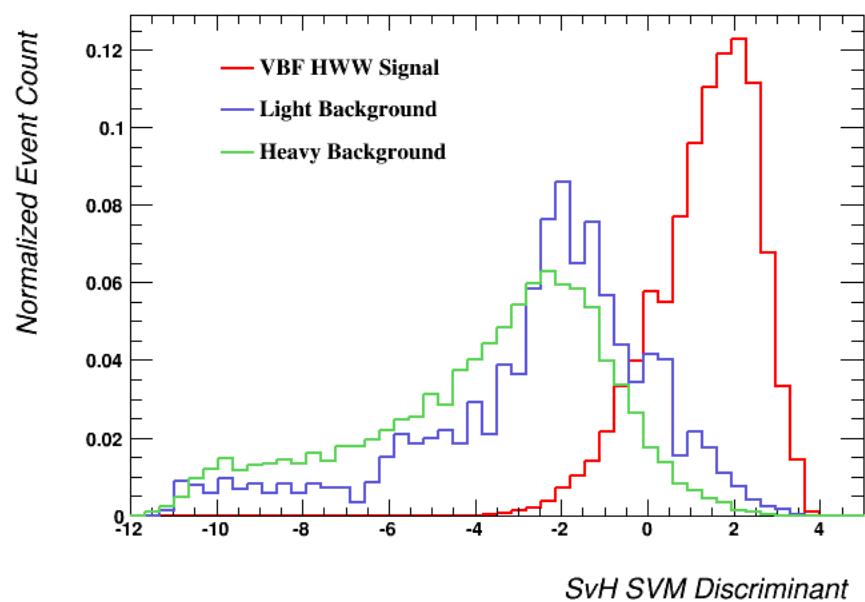
Figure 10.6: Distribution of SvH support vector machine discriminants for signal, light and heavy events. The total integral of each class has been normalized to unity.
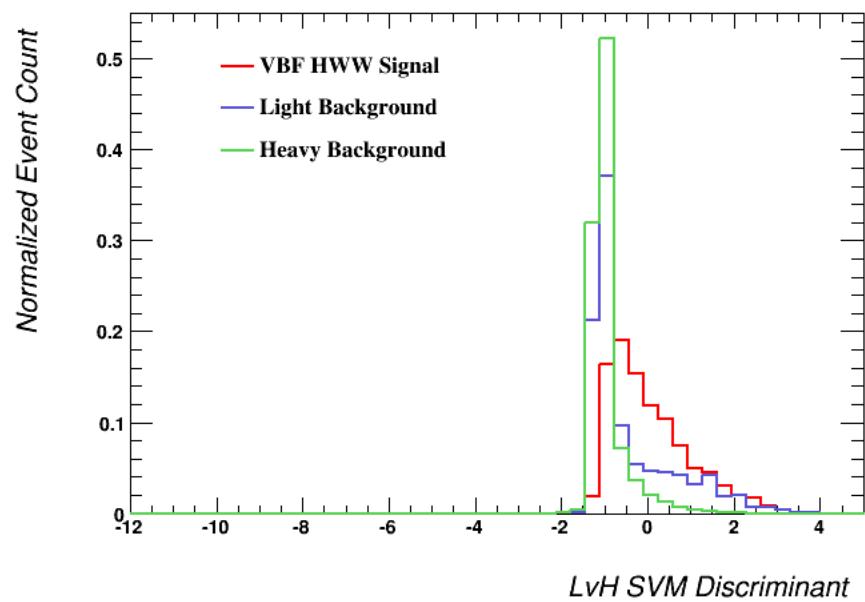
Figure 10.7: Distribution of LvH support vector machine discriminants for signal, light and heavy events. The total integral of each class has been normalized to unity.

of signal events. The SVM analysis was then performed on the pseudo-data set and the analysis best-fit result for the number of signal events was compared to the known number of signal events. The difference between the best-fit result and the known number of signal events is referred to as the bias (see equation 8.15).

In order to analyze the bias, a set of 2000 pseudo-experiments was performed for each of 16 input signal fractions ranging from 1.5% signal to 3% signal. The average bias and standard deviation of the bias was recorded for each of the 16 signal fractions. The results are shown in Figures 10.8 and 10.9. It is important to note that the bias is small in comparison to the true signal fraction and that the bias is stable across a wide range of input signal fractions. This result indicates that the SVM analysis technique is robust with respect to varying signal strengths. Due to these factors, it is possible to correct the real data best-fit result with respect to the average bias using equation 8.16.

The statistical error of the best-fit 3-class analysis result is taken to be the standard deviation of the pseudo-experiment results over all pseudo-experiments. Since the pseudo-experiments are designed to replicate the real data analysis in event composition and total number of events, the statistical error can be carried over from Monte Carlo studies to the real data result.
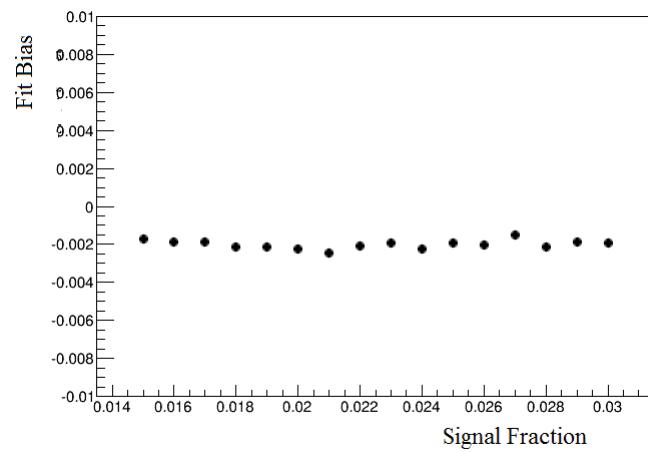
Figure 10.8: Average bias of the likelihood fit over 2000 pseudo-experiments with each of 16 different input signal percentages
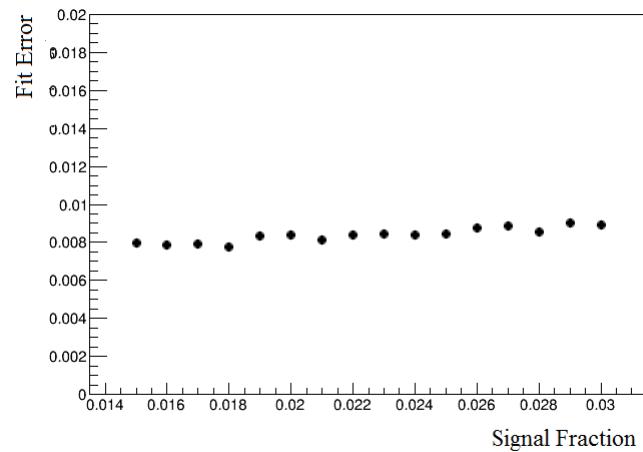
Figure 10.9: Error of the likelihood fit for 2000 pseudo-experiments with each of 16 different input signal percentages

10.2.3 Systematic Uncertainties

Systematic uncertainties, as described in Chapter 9, were analyzed for the 3-class analysis. All systematic uncertainties are added in quadrature to determine the total systematic error. Table 10.1 lists the largest systematic uncertainties. The most important systematic uncertainties are the $t\bar{t}$ generator modeling along with the jet and MET energy scales. A full list of systematic uncertainties analyzed is given in appendix A.

10.2.4 Results

Monte Carlo studies indicate an expected significance using Standard Model predictions for the 3-class SVM analysis of 2.26σ including only statistical uncertainties and 2.02σ when statistical and systematic uncertainties are included. This result is consistent with the expected significance in the 2-class SVM analysis and with the BDT analysis [38].

Analysis of the $20\,fb^{-1}$ 2012 ATLAS data set found $15.6 \pm 5.6(\text{stat}) \pm 2.9(\text{syst})$ VBF HWW events in the 2012 data set. This result is consistent with the Standard Model prediction of 12.7 VBF HWW events and represents a statistical significance of 2.47σ over the background-only hypothesis. The signal strength (N_{obs}/N_{SM}) associated with this result is $\mu = 1.22 \pm .44\,(\text{stat}) \pm .23\,(\text{syst})$.

Systematic Type	Percent Var.
TTBar MCNLO	13
ATLAS JES Eta Modelling	9
ATLAS JER	8
ATLAS TRACKMET SCALESOFT	5
ATLAS TRACKMET RESOPARASOFT	5
ATLAS JES 2012 Modelling1	4
ATLAS JES NPV	3
ATLAS EL ESCALE	3
ATLAS JES 2012 Detector1	3
ATLAS BTag B6EFF	3
ATLAS JES NonClosure AFII	3
ATLAS JES 2012 Eta StatMethod	3
ATLAS JES BJET	2
FakeRateOther QCD HWW	2

Table 10.1: Table of the largest systematic variations for the 3-class SVM analysis. The percent variation in the expected number of signal events is listed for each systematic.

10.3 2-Class vs. 3-Class SVM Analysis Comparison

The 2 and 3-class SVM analyses resulted in similar findings, both of which agreed with the Standard Model predictions and with the findings of the AT-LAS *HWW* group's boosted decision tree analysis. As hoped, the 3-class analysis performed slightly better in several ways. Of primary importance, the expected significance using only statistical error was marginally larger for the 3-class analysis. Since no dedicated optimization studies were performed on SVM input parameters, it is likely that the 3-class SVM result could be improved considerably. In addition, it may be possible to increase the statistical significance of the 3-class analysis by providing additional input variables in order to better distinguish the signal, light and heavy classes. At the moment, studies show that the addition of more SVM input parameters decreases the statistical significance of both the 2-class and 3-class analysis. However, we believe that this is a product of insufficient MC events and not an inherent feature of the analysis. Hopefully we will be able to test this hypothesis in the upcoming LHC run.

In addition to the comparison of expected significance, a signal region analysis similar to that described in Chapter 7 was performed on the 3-class SVM discriminants using Monte Carlo samples. For this analysis, a signal region was created in the SVM discriminant space parameterized by the three

SVM discriminants (SvL, SvH, LvH). Just as the 2-class SVM signal region is characterized by a single cut on the 2-class SVM discriminant, the 3-class signal region is characterized by a cut on each of the three SVM discriminants. The 3-class SVM signal region was optimized in order to maximize the significance ($S/\sqrt{B}$) of events within the signal region.

The 3-class signal region analysis was performed prior to much of the SVM input optimization, resulting in an expected significance of only 2.03σ. Using the same unoptimized MC input data set, the 2-class analysis also found a statistical significance of 2.03σ in the signal region. Though the 3-class and 2-class signal regions yielded the same significance, the 3-class analysis had an expected 5.54 signal events in the signal region compared to only 4.66 events in the 2-class signal region. The 3-class analysis therefore offered a 20% increase in the expected number of signal events.

10.4   Analysis Limitations Due to Data and MC Statistics

Both the 2 and 3-class analyses were heavily limited by statistics. The most obvious statistical issue was a lack of expected signal events in the real data set. MC studies predict that our final data set, containing only events passing pre-selection, should contain only 12.7 real signal events with approximately 650 background events. This limitation can be seen clearly in comparing the

statistical and systematic errors in the 3-class analysis, where the statistical error dominates the systematic error 2-to-1. This statistical error is due to two primary factors in the 3-class analysis: first, small numbers of signal events and inherently limited event separation necessarily lead to small signal-to-background ratios; and second, with such a limited number of data events, the VBF *HWW* analysis is susceptible to significant statistical fluctuations within both the pseudo-data sets and the real data set. The issue of real data statistics will be mitigated during the LHC run 2, which is expected to yield much higher luminosity.

In addition to the limited number of real data events, the size of Monte Carlo simulated event samples were also insufficient. The training of SVMs in our analysis requires the correct proportion of each background type as described in Chapter 8. For several MC samples, there were an insufficient number of background events to properly train the support vector machines. The limited MC statistics also became a problem when creating pseudo-data sets, causing certain backgrounds to be under represented in pseudo-experiments. In order to mitigate the problem of limited MC statistics, we have requested much larger MC data sets for the next LHC run.

10.5   Template Binning

It was discovered in the course of the 3-class analysis that the binning used in the creation of the 3-d templates had a large affect on the statistical errors

as well as on the real data result. The bin size used for the 3-class SVM templates resulted in approximately 50 bins in the x, y and z dimensions of each template. As expected, increasing the bin size reduced the resolution of the templates and lead to higher statistical errors. The estimated number of signal events in the real $20\,fb^{-1}$ data set was not significantly changed by increasing the template bin size.

In contrast, decreasing the bin size of the 3-d templates led to unexpected behavior. Specifically, decreasing the bin size beyond approximately 0.02 (in units of SVM discriminant) led to a significant increase in the best fit estimate of the fraction of signal events in the real 2012 ATLAS data set. This increase can be seen in the first three data points in Figure 10.10. As can be seen in Figure 10.11, the increase in estimated signal fraction corresponds to an increase in the statistical error of the fit.

It is believed that this increase in the real data signal fit result is due to insufficient data statistics in comparison with bin size. For small bin sizes, most bins contain very few data events, allowing for small fluctuations in the distribution of data events to drastically alter the fit result. In order to prevent these statistical fluctuations from influencing the final result, a bin size greater than 0.02 was chosen. As can be seen in Figure 10.10, both the fit result and fit error are stable over a wide range of binning between 0.02 and 0.05. This stability provides confidence that the fit results are robust with respect to changes in binning and statistical fluctuations in the distribution of
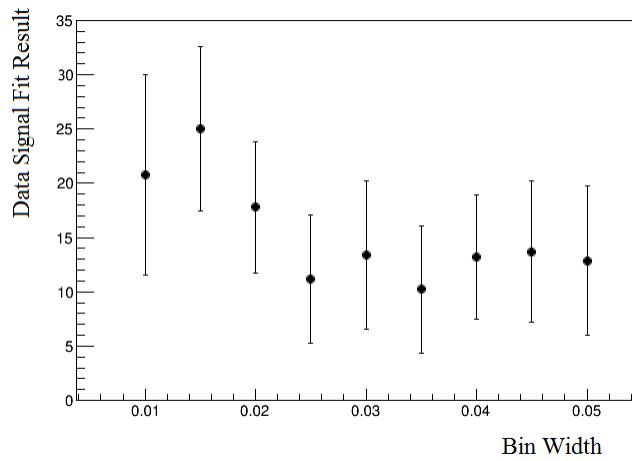
Figure 10.10: Best fit estimate of the fraction of VBF *HWW* events in the 2012 ATLAS data set after pre-selection for multiple template binnings. Error bars include only statistical error.
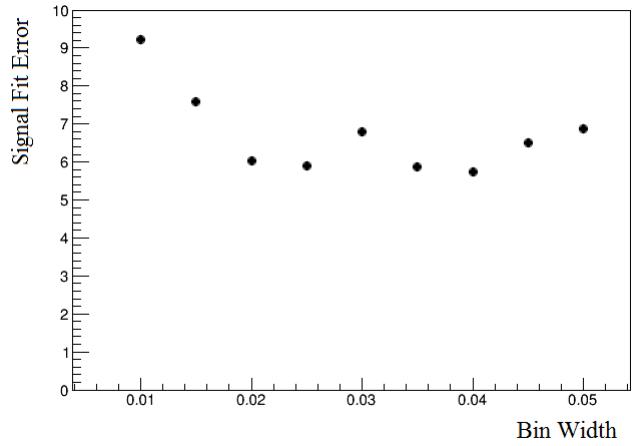

Figure 10.11: Statistical error associated with the best fit estimate of the fraction of VBF *HWW* events in the 2012 ATLAS data set after pre-selection for multiple template binnings.

data events.

10.6 Conclusion

The analysis presented in this thesis was successful in accomplishing two main objectives. First, to search for evidence of vector boson production of the Higgs, and second to study the effectiveness of support vector machines in high energy physics analyses.

The search for definitive evidence for VBF $H \rightarrow WW^*$ decays is severely hindered by the small cross section of VBF Higgs production along with the small branching ratio of $W \rightarrow l\nu$ decays. Despite the unique topology of VBF $H \rightarrow WW^*$ decays, the limited number of signal events coupled with the large cross sections of many background processes necessarily lead limited signal-background ratios. Despite these challenges, the analysis presented in this thesis found evidence at the 2σ level of VBF Higgs production. This result is consistent within 1 standard deviation of the Standard Model prediction and matches the results of other ATLAS analyses. The forthcoming LHC run at $13\,TeV$ center of mass energy should provide much larger statistics and therefore allow for definitive evidence of VBF $H \rightarrow WW^*$ decays.

In addition to lending support to the observational evidence of the $125\,GeV$ Higgs boson, our analysis verified the effectiveness of the support vector machine analysis technique. Using a 2-class SVM analysis we were able to compare directly with the ATLAS HWW group's boosted decision tree analysis.

We found that with limited tuning our the 2-class SVM analysis was competitive with BDT results. It is likely that with additional tuning, the SVM analysis could give equivalent discrimination power to the BDT analysis. It is possible, that given the large statistics expected in the next LHC run that the SVM technique could yield advantages with respect to BDT analyses due to the resistance of SVMs to over-tuning.

Finally, the analysis presented in this thesis found marginal improvements in expected significance using a multi-SVM analysis. The effectiveness of the 3-class SVM analysis was most likely hindered by the similar topology of events in the light and heavy classes. This similarity prevented good discrimination of events in the LvH support vector machine. It is possible that in the next LHC run, additional MC and data statistics will allow for larger gains using the 3-class SVM analysis through larger training sets. In addition, it may be possible to add new input variables to help distinguish light and heavy events. Finally, this thesis did not contain an optimization study on the backgrounds contained in each class. It may be possible to more optimally chose the backgrounds included in each class in order to aid in SVM training. In the opinion of the author, the 3-class SVM method has shown promise as an improvement over 2-class SVM analyses and should be investigated further in order to determine its potential.

Appendix A

Systematic Uncertainties for the

3-Class SVM Analysis

Systematic Type	Percent Var.
TTBar MCNLO	13
ATLAS JES Eta Modelling	9
ATLAS JER	8
ATLAS TRACKMET SCALESOFT	5
ATLAS TRACKMET RESOPARASOFT	5
ATLAS JES 2012 Modelling1	4
ATLAS JES NPV	3
ATLAS EL ESCALE	3
ATLAS JES 2012 Detector1	3
ATLAS BTag B6EFF	3
ATLAS JES NonClosure AFII	3
ATLAS JES 2012 Eta StatMethod	3
ATLAS JES BJET	2
FakeRateOther QCD HWW	2
ATLAS JES MU	2
ATLAS JES 2012 PileRho HWW	2
ATLAS JES FlavComp HWW other	2
ATLAS BTag LEFF	2
ATLAS BTag Pythia6 LEFF	2
ATLAS BTag B5EFF	2
ATLAS TRACKMET RESOPERPSOFT	1

Table A.1: Table of systematic uncertainties for the 3-class SVM analysis. Percent var. indicates the percent variation in the expected signal result due to the listed systematic. Part 1.

Systematic Type	Percent Var.
FakeRateCorr QCD HWW	1
FakeRateStat QCD HWW	1
FakeRate MU Stat GT25 HWW	1
FakeRate MU Stat 10 15 HWW	1
ATLAS EL RES	1
ATLAS JES 2012 PilePt	1
ATLAS EL EFF ID HIGHPT	1
ATLAS JES HighPt	1
ATLAS TRIGGER HWW	1
ATLAS MU EFF	1
ATLAS JES FlavResp	1
FakeRate EL Stat GT25 HWW	≤ 1
ATLAS BTag CEFF	≤ 1
ATLAS BTag B2EFF	≤ 1

Table A.2: Table of systematic uncertainties for the 3-class SVM analysis. Percent var. indicates the percent variation in the expected signal result due to the listed systematic. Part 2.

Systematic Type	Percent Var.
FakeRate MU Other HWW	≤ 1
ATLAS EL EFF	≤ 1
ATLAS JES FlavComp HWW WW	≤ 1
ATLAS MU ISO	≤ 1
ATLAS BTag B3EFF	≤ 1
FakeRate EL Uncorrl OS HWW	≤ 1
ATLAS MET RESOSOFT	≤ 1
ATLAS EL EFF ID CORRLOW	≤ 1
ATLAS MU RESCALE lvlv 2012	≤ 1
ATLAS BTag B4EFF	≤ 1
ATLAS EL TRIGGER HWW	≤ 1
FakeRate EL Stat 20 25 HWW	≤ 1
ATLAS BTag B1EFF	≤ 1
FakeRate MU Stat 15 20 HWW	≤ 1
FakeRate MU Uncorrl OS HWW	≤ 1
FakeRate EL Other HWW	≤ 1
ATLAS EL EFF RECOID80010	≤ 1
FakeRate MU Stat 20 25 HWW	≤ 1
ATLAS EL EFF RECO CORRLOW	≤ 1

Table A.3: Table of systematic uncertainties for the 3-class SVM analysis. Percent var. indicates the percent variation in the expected signal result due to the listed systematic. Part 3.

Systematic Type	Percent Var.
FakeRate EL Stat 15 20 HWW	≤ 1
ATLAS EL EFF RECOID80015	≤ 1
FakeRate EL Corrl OS HWW	≤ 1
ATLAS DIL TRIGGER HWW	≤ 1
FakeRate EL Stat 10 15 HWW	≤ 1
ATLAS MU ESCALE	≤ 1
ATLAS MET SCALESOFT	≤ 1
ATLAS MU ID RES	≤ 1
ATLAS MU TRIGGER HWW	≤ 1
ATLAS ISO	≤ 1
ATLAS EL EFF RECO CORR	≤ 1
ATLAS EL ISO	≤ 1
FakeRate MU Corrl OS HWW	≤ 1
ATLAS MU MS RES	≤ 1

Table A.4: Table of systematic uncertainties for the 3-class SVM analysis. Percent var. indicates the percent variation in the expected signal result due to the listed systematic. Part 4.

Bibliography

[1] Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, ATLAS Collaboration, Phys. Lett. B **716** (2012) 1.

[2] Observation of a new boson at a mass of 125 GeV with the SVM experiment at the LHC, CMS Collaboration, Phys. Lett. B **716** (2012) 30.

[3] Study of the spin of the new boson with up to 25 fb-1 of ATLAS data, ATLAS-CONF-2013-040, CERN, Geneva, April 2013.

[4] Measurements of the Higgs boson production and couplings in the diboson final states with the ATLAS detector at the LHC, ATLAS Collaboration, CERN-PH-EP-2013-103, Submited to Phys. Lett. B.

[5] Observation and measurement of the Higgs boson decays in WW* in the ATLAS detector, ATLAS Collaboration, CERN-PH-EP-2014-270, Submitted to Phys. Rev. D.

[6] "Standard Model of Elementary Particles" by MissMJ - Own work by uploader, PBS NOVA, Fermilab, Office of Science, United States Department of Energy, Particle Data Group. Licensed under CC BY 3.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg #mediaviewer/File:Standard_Model_of_Elementary_Particles.svg

[7] "Right left helicity" by en:User;HEL, User:Stannered - en:Image:Right left helicity.jpg. Licensed under Public Domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Right_left_helicity.svg#mediaviewer/ File:Right_left_helicity.svg

[8] "Mecanismo de Higgs PH" by Gonis from es. Licensed under CC BY-SA 3.0 via Wikimedia Commons - http://commons.wikimedia.org /wiki/File:Mecanismo_de_Higgs_PH.png#mediaviewer/File:Mecanismo_de_Higgs_PH.png

[9] Measurement of the Higgs boson mass from the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow llll$ channels in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector, ATLAS Collaboration, Phys. Rev. D **90** 052004 (2014).

[10] The ATLAS Collaboration, Science 21, Vol. 338 (2012) 1232005.

[11] Courtesy of the CMS experiment.

[12] Courtesy of the ALICE Experiment.

[13] Courtesy of the ATLAS Experiment.

[14] The ATLAS Experiment at the CERN Large Hadron Collider. ATLAS Collaboration, JINST, 3:S08003, 2008.

[15] ATLAS Muon Spectrometer Technical Design Report. ATLAS Collaboration, CERN, Geneva, 1997.

[16] H.L. Lai et al., Phys. Rev. **D82** (2010) 074024, arXiv:1007.2241.

[17] The ATLAS Collaboration, Phys. Rev. Lett. **107** (2011) 041802, arXiv:1104.5225.

[18] T. Sjostrand, S.Mrenna, and P.Z. Skands, JHEP **0605** (2006) 026.

[19] G. Corcella et al., JHEP **0101** 010 (2001).

[20] S. Frixione, P. Nason, and C. Oleari, JHEP **0711** (2007) 070, arXiv:0709.2092.

[21] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A.D. Polosa, JHEP **0307** (2003) 001, arXiv:hep-ph/0206293.

[22] T. Gleisberg et al., JHEP **0902** (2009) 007.

[23] N. Kauer and G. Passarino, Inadequacy of zero-width approximation for a light Higgs boson signal, JHEP **1208**, 116 (2012) arXiv:1206.4803.

[24] Jet energy scale and its systematic uncertainty in proton-proton collisions at srqt(s)=7 TeV with the ATLAS 2011 data, ATLAS-CONF-2013-004, CERN, Geneva, Jan 2013

[25] M. Cacciari, G.P. Salam, and G. Soyez, JHEP 0804 (2008) 063, arXiv:0802.1189

[26] Selection of jets produced in proton-proton collisions at the ATLAS detector using 2011 data, ATLAS-CONF-2012-020, CERN, Geneva, Mar 2012

[27] Object selection for the HWW-¿WW search with the ATLAS detector at sqrt(s) = 8 TeV, ATLAS-COM-2012-861, CERN, Geneva, Feb 2013

[28] Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fbˆ1 of Data from the ATLAS Detector, ATLAS-CONF-2012-043, CERN, Geneva Mar 2012

[29] Description and Performance of the Electron Likelihood Tool at ATLAS using 2012 LHC Data, ATLAS-COM-PHYS-2013-378, CERN, Geneva, Aug 2013

[30] Performance of the muon identification and reconstruction with the ATLAS detector, ATLAS Collaboration, To be submitted to the Eur. Phys. J.

[31] Performance of the ATLAS muon trigger in pp collisions at sqrt(s) = 8 TeV, ATLAS Collaboration, Submitted to the European Physics Journal C.

[32] Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector, ATLAS-CONF-2011-063, CERN, Geneva, 2011.

[33] Performance of the ATLAS Electron and Photon Triggers in p-p Collisions at swrt(s) = 8 TeV in 2012, ATLAS-COM-DAQ-2013-121, CERN, Geneva, December 2014.

[34] Supporting document on electron efficiency measurements using 2012 LHC proton-proton collision data, ATLAS-COM-PHYS-2013-1295, CERN, Geneva, May 2014.

[35] ATLAS Muon Momentum Resolution in the First Pass Reconstruction of the 2010 p-p Collision Data at sqrt(s) = 7 TeV, ATLAS-CONF-2011-046, CERN, Geneva, March 2011.

[36] Electron performance measurements with the ATLAS detector using 2010 LHC proton-proton collision data. European Physics Journal C, 72:1909, 2012.

[37] ontinuous b-tagging for the ATLAS experiment, ATL-COM-PHYS-2014-035, CERN, Geneva, Jan. 2014.

[38] Observation and measurement of the Higgs boson decays to WW* with the ATLAS detector, ATLAS Collaboration, CERN-PHY-EP-2014-270, Submitted to Phys. Rev. D.