Search for VH and Technicolor Production in the $qqbb$ Final State
Using the RunII DØ Detector

by

Justace R. Clutter, PhD(c)

Submitted to the graduate degree program in Physics and the Graduate
Faculty of the University of Kansas in partial fulfillment of the
requirements for the degree of Doctor of Philosophy

Committee:

_____

Dr. Alice Bean, Chairperson

_____

Dr. Philip Baringer

_____

Dr. Graham Wilson

_____

Dr. John Ralston

_____

Dr. David Lerner

Date defended: _____

The Dissertation Committee for Justace R. Clutter, PhD(c) certifies that
this is the approved version of the following dissertation:

Search for VH and Technicolor Production in the $qqbb$ Final State
Using the RunII DØ Detector

Committee:

—————————————————

Dr. Alice Bean, Chairperson

—————————————————

Dr. Philip Baringer

—————————————————

Dr. Graham Wilson

—————————————————

Dr. John Ralston

—————————————————

Dr. David Lerner

Date approved: ——————————

i

Contents

List of Tables

List of Figures

xiii

Acknowledgements

The work that is presented in this dissertation is the culmination of many years of effort. While my name fills the byline, it should be understood that I could not have accomplished this alone. I have behind me some of the greatest people I know, this section along with the remainder of this dissertation is dedicated to those people.

While I can not specifically thank each and every person that has helped me along this long road, I would like to bring to the front a number of people who, without their help, I would not have achieved this goal.

First and foremost, I would like to thank my parents and my family. My family has stood by me in times of happiness and ease, but more importantly, they have stood by me in those times of aggravation and difficulty. My parents, Pat and Randy Clutter, have taken the time to build a foundation within me that has ultimately allowed me to complete this task. I can only hope to be half the parent that my mother and father are when the time comes. I would like to thank Dr. Alice Bean for giving me the opportunity to join the high energy experimental group. She has helped to guide me though my studies and shown great restraint when I was being stubborn. For all of those times, I apologize Alice. Like my parents, you have stood by me, and for that I thank you. I could not have even got close to what I have achieved without you. I hope that there will be some way that I can repay you in the future for all of the time and effort that you have placed in me.

I would like to thank my wonderful girlfriend, Huan Zhu, who stood by me while I wrote this dissertation. She was there to act as a sounding board and overall general support. She was instrumental in keeping me sane during this process. I can only hope to repay her for these actions when the comes for her to finish and submit her dissertation.

There have been many times that I have called upon the help of a friend to explain something to me, even in the odd hours of the night (or to some the odd hours of the

Abstract

A search for dijet resonance production in a four-jet all-hadronic final state from the DØ detector at Fermilab's Tevatron is presented. The data set, acquired at a $p\bar{p}$ center-of-mass energy of $\sqrt{s} = 1.96$ TeV, contains primarily multijet events and represents approximately 1 fb^{-1} of data. The cross section limits for associated Higgs production and Technicolor processes are determined through a background subtraction method using data to estimate the background. This four-jet channel is potentially very powerful, but is extremely challenging due to the large multijet background from QCD processes. Background rejection is performed by utilizing b-tagging, pre-selection cuts, a multi-variate boosted decision tree discriminant, and the correlated information contained in the $M(bb)$ and $M(jj)$ dijet invariant masses. The search for VH ($WH+ZH$) processes yields a 95% confidence level observed upper limit of 20.4 pb on the VH cross section for a Higgs mass of 115 GeV/c^2. Additionally, a 95% confidence level observed upper limit of 16.7 pb was set for a Higgs boson mass of 125 GeV/c^2 and 24.6 pb was set for a Higgs boson mass of 135 GeV/c^2. The same data set was used to place limits on the Technicolor process $\rho_{\mathrm{TC}} \rightarrow W\pi_{\mathrm{TC}}$ where the technirho mass was fixed to 240 GeV/c^2. For a technipion mass of 115 GeV/c^2 we find a 95% confidence level observed upper limit on the cross section of 49 pb. The technipion masses of 125 GeV/c^2 and 140 GeV/c^2, the 95% confidence level observed upper limits are 57 pb and 71 pb, respectively.

1 Introduction

Since the late 1970's, the Standard Model has been the primary workhorse in describing high energy particle physics to an exceptional degree of precision. An example of these successes is shown in the agreement between the predicted and measured values of the W and Z boson masses [1]. Theories prior to the Standard Model were unable to describe the masses of the fundamental particles. Weak force bosons have mass, which is evidenced by the long time scale of the corresponding interactions. In addition to the weak force bosons, other particles have measured masses which were not originally described in the theory.

Many attempts were made to extract particle masses from these earlier mathematical formalisms. The brute force addition of massive terms was prevented due to a loss in gauge invariance [2]. In 1964, three papers from three different research groups were published in Physical Review Letters which highlighted a new technique to introduce mass in the Standard Model. François Englert and Robert Brout [3], Peter Higgs [4], and the research group composed of Gerald Guralnik, C. R. Hagen, and Tom Kibble [5] each independently produced a solution that year. This new technique now bears the name of Peter Higgs and is called the Higgs mechanism. While other techniques have been proposed, such as Technicolor, the Higgs mechanism has become an integral component of what defines the Standard Model.

The Higgs mechanism was put into to the Standard Model through the work of Glashow [6], Salam [7], and Weinberg [8] between the years of 1961 and 1972. With the Higgs mechanism, a complex scalar doublet is added with specific properties imposed on the field potential. With the field initially having four degrees of freedom, three are used to impart mass to the gauge bosons of the weak sector. The remaining degree of freedom presents itself as a detectable particle called the Higgs boson which has not yet been observed. Finding the Higgs boson is difficult due to the small predicted cross-section and the unknown particle mass. In the Standard Model formalism, the Higgs

boson mass is parameterized by two quantities, the vacuum expectation value, or the *vev*, and the quartic self coupling, λ. While the value of the *vev* has been constrained to be 246 GeV through the Fermi coupling, the quartic self coupling term (λ) is still unknown [9]. Through direct experimentation and indirect evidence, the mass of the Higgs boson is currently thought to be between 114.4 and 185 GeV/c^2 at 95% confidence level. Studies have been designed to search for this particle in different regions of the allowed mass range. Below 135 GeV/c^2, the Higgs boson prefers a decay to b quarks while for masses above this point it decays preferentially to W boson pairs [9].

This difference in the decay products of the Higgs boson creates a natural division in analysis strategies. Searches sensitive to a high mass Higgs boson will primarily observe two or more W bosons and their associated decay products. Those searches focusing on the light Higgs boson will normally observe two or more jets from b quarks in the final state. This study represents one such search using a four jet final state with the DØ detector located at Fermilab. The full Feynman diagram depicting the associated Higgs production and specific decay processes used in this study is presented in Figure 1. While there is a Higgs production process with a higher predicted production rate, $gg \rightarrow H$, the large backgrounds present when looking in the low mass Higgs boson range would dramatically reduce the sensitivity of the study.
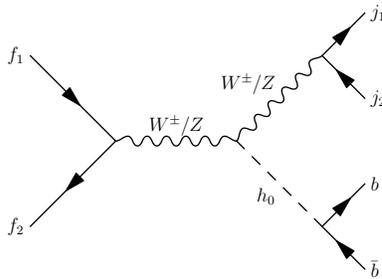
Figure 1: Associated Higgs boson production showing the hadronic decay of the W and H bosons

The majority of the current searches for a light Higgs boson produced in association with a W boson take advantage of the final state leptons from the W decay. These final state leptons can be utilized to enhance the signal significance in the search. This

is accomplished at the cost of the reduced decay fraction for the W boson since the individual leptonic decays only account for $\simeq 11\%$ of the total possible decays. While multiple final state lepton searches can be combined, the maximum available branching fraction is only $\simeq 33\%$ [9]. The W boson decays into hadrons $\simeq 67\%$ of the time. A study based on hadronic final states has the advantage of the large decay fraction but the added challenge of large backgrounds from multijet processes which are difficult to model correctly through Monte Carlo simulations. The analysis presented here addresses the issue by utilizing data to estimate the background contribution.

The $qqbb$ (the q is sometimes referred to as a j in this paper) final state is shared between many different physics processes. Two of these types of processes, WZ and Technicolor production, are investigated in this study. While the WZ has been observed in the tri-lepton decay mode [10], the semi-leptonic analysis has not been able to produce the required sensitivity to cite an observation. The analysis presented in this paper is the first all hadronic search for WZ production where both the W and Z decay to two quarks producing a four jet final state.

We investigate Technicolor production using three points outside the current DØ Technicolor exclusion zone in the M(π_{TC}) vs M(ρ_{TC}) parameter space. Technicolor is an alternative theoretical framework designed to introduce mass into the Standard Model though a dynamical symmetry breaking scheme. In this theory, a new set of fermions, called technifermions, are added to the Standard Model and a bound pair of these fermions, the technipion, takes the place of the Higgs boson. The charged Technicolor signal process is produced in an s-channel configuration with a ρ_{TC} as the intermediate particle. The ρ_{TC} then decays to a $W^{\pm}$ and a π_{TC} with the π_{TC} decaying into two b quarks. The Feynman diagram representing this process is shown in Figure 2.

Using approximately 1 fb^{-1} of data taken from the DØ detector between 2002 and 2006, this analysis utilizes a multivariate boosted decision tree after preliminary cuts and then exploits the correlation between the dijet invariant masses of the qq and the

3
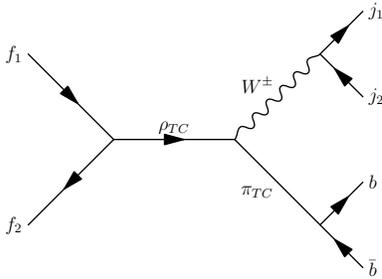
Figure 2: Technicolor production showing the hadronic decay of the W boson and the π_{TC}

bb in order to derive a result. The background model for this analysis is derived directly from data with no Monte Carlo inputs.

The preliminary cuts of this analysis are strongly designed around the contents of the final state and require two tightly-tagged b jets in events containing exactly four jets. Each of these jets have a $p_T > 15$ GeV/c, $|\eta| \leq 2.5$. A b-tagging method was used and required exactly two b-tagged jets. I benefit from previous studies by my collaborators on b-tagging, triggers, jet energy scale, to name a few, and proceed using their proven methods. For these searches, in the four-jet channel, there is no ambiguity or combinatoric problem determining which jets are associated with the W, H, Z, or π_{TC} since the b-tagging information is present.

After the preliminary cuts, the large multijet background is further reduced by employing Boosted Decision Trees (DTs). The decision trees were trained with 28 variables which included, for example, jet-based quantities such as the jet p_T's and event shape variables such as sphericity. The decision trees were trained using an independent data sample which only required one b-tagged jet instead of two, to represent the background.

In order to improve the sensitivity after the decision tree, a combination of invariant mass cuts are applied. The first cut is applied on the $M(jj)$ variable and is designed to favor a W boson. The second invariant mass cut is done on the $M(bb) - M(jj)$ mass. The specific cuts depend on the analysis channel under study in addition to the mass of the Higgs boson or the mass of the technipion.

A matrix technique is used to determine the number of signal events required to

derive a cross section. A decision tree (DT) sideband (or "control" region) from data is used to determine the background contribution. The amount of background in the DT signal region is estimated from the control region by scaling the content with a "Scale Factor" (SF). This SF is determined using an ensemble technique applied to a data sample composed of events that did not pass the second invariant mass cut. Signal present in the control region is taken into account when determining the background normalization.

Chapter 2 discusses the theory behind the Standard Model, the Higgs mechanism, and the specific physics processes that are the focus of this study. Following the theory, Chapter 3 will introduce you to the detector and the accelerator complex. Chapter 4 will establish the basics in performing a high energy physics analysis in addition to the physics object reconstruction and analysis tools used in this study. An overview of the history and current status for the VH and the Technicolor searches is presented in Chapter 5. Chapters 6, 7, and 8, discuss the inputs and preliminary cuts, the background rejection phase, and the final results, respectively. Specifically, information on the decision tree can be found in Section 7.1 while more information on the invariant mass cuts can be found in Section 7.2. A search for the WZ diboson signal is described in Chapter 9. The concluding remarks along with suggestions for future improvements to this analysis are presented in Chapter 10.

2 Theory

Quantum field theory is the mathematical model used to describe the fundamental particles of our world. The Standard Model and Super symmetry are two examples of QFT. QFT does not allow for the simple addition of mass terms into the Lagrangian. The focus of this chapter will be to review the Higgs mechanism which allows for the introduction of massive particles via spontaneous symmetry breaking while not requiring the inclusion of a brute force mass term in the Lagrangian. It will become evident through this chapter that the mass of the Higgs Boson, the particle remaining from the spontaneous symmetry breaking, is unknown due to an ignorance of the Higgs self coupling term. A brief overview of upper and lower theoretical bounds on the Higgs boson mass will also be presented in this chapter. Analyses designed to search for VH, $WH + ZH$, production have the capacity to search for additional signals from theories beyond the Standard Model. In addition to a VH search, this analysis also performs a search for WZ production along with a search for Technicolor signatures. Both of these additional analyses utilize a final state compatible with the VH analysis consisting of four jets two of which come from a b quark. A brief introduction to Technicolor is presented along with a discussion of the $WZ/ZZ/WW$ diboson processes.

2.1 Standard Model

In an attempt to better understand the world, it is common to dissect matter and objects into their smaller and more basic building blocks. The Standard Model of particle physics accomplishes this in the most aggressive manner by introducing a collection of fundamental particles in addition to the interactions between them. At its core, it is a quantum field theory that is based on an $SU(3) \times SU(2) \times U(1)$ gauge symmetry [11]. This structure creates two distinct classes of particles: Fermions and Bosons.

The fermions are subdivided into two further groups called quarks and leptons. The quarks consist of six particles and are grouped into three families. Figure 3 depicts these

six quarks, along with the other particles of the Standard Model. Each quark carries four main differentiable properties: an electric charge, a color charge, weak isospin, and a mass. The up type quarks all carry an electric charge of +2/3 while the down type quarks carry an electric charge of -1/3. Each quark can have one of three possible color values: red, green, or blue.

Figure 3: The fundamental particles of the Standard Model

With all the possible permutations of the stated properties, there are 18 different quarks. Each quark has an anti-particle form, called and anti-quark. While the anti-quark has the same mass as the quark, all of its quantum numbers are inverted. For instance, an anti-down quark will have a positive charge of $2/3e$ along with having the color altered to be the anti-color. Taking into account the anti-quarks, the total number of quarks is doubled giving a total of 36 possible quarks. The different particles

constructed by combining quarks, called hadrons, can be grouped into two categories: mesons and baryons. The mesons are combinations of a quark and an anti-quark while the baryons are combinations of three quarks. A strict combination rule, for both mesons and baryons, requires that the resulting quark combination be colorless [12]. A colorless baryon state is one in which each quark carries a different color. An example of this would be the proton which is composed of two up quarks and one down quark. In this proton, the colorless requirement is met by having one of the quarks red, another blue, and finally the last quark green. It does not matter which is which, but rather that each of the colors are present. It is also possible to have an anti-baryon where each of the quarks carry a different anti-color. The colorless state is achieved in a meson by ensuring that the two quarks are anti-colors of each other so that they cancel each other out yielding no color in the combination. For instance, a pion composed of an up quark and an anti-down quark is colorless if the up quark is red and the anti-down quark has the anti-red color. An example of a configuration that is not allowed would be to have the up quark a green color but the anti-down quark to have an anti-red color. The collection of different composite particles that can be constructed in this manner are vast yielding a complex hierarchy of bound states.

Each quark family has an associated set of leptons. For the up and the down quark, the two leptons are the familiar electron and electron neutrino. Moving up in mass through the families, one will find the muon and the tau leptons in conjunction with their associated neutrinos. The electron, muon, and tau have a charge of -1e while their associated neutrino's have a neutral charge. The leptons do not interact via the strong force and are therefore color blind. Each has an associated anti-particle providing a total of 12 leptons. The leptons interact through electro-weak processes only.

The remaining components of the Standard Model are the individual force carriers. The carrier of the electromagnetic force is the photon. Required to be precisely massless, the electromagnetic interaction is infinite in range. The weak force utilizes three gauge bosons: W^+, W^-, and the Z. The weak force bosons have a non-trivial mass, shown

in Figure 3, and, therefore, the range and timescale for the electroweak interactions are not infinite but on the order of 10^{-18} meters and 10^{-10} seconds, respectively [13]. With the force carriers for the electroweak interactions introduced, the only remaining force is the strong force. The mediators of the strong force are gluons. Each gluon carries both a color and an anti-color. An example of this would be a red / anti-green gluon. The strong force behaves quite differently than the other forces in that the strength of the interaction does not diminish with larger distances. At close range the quarks feel free to move about, but as they move further apart from each other, they feel a restoring force. This unique behavior is called asymptotic freedom and confinement, respectively

While not present in Figure 3, the Higgs boson, a by-product of the Higgs mechanism as applied to the electroweak sector, is predicted to exist and is considered a component of the Standard Model even though it has yet to be observed. The Standard Model has done a phenomenal job describing the world around us on a sub-atomic scale, though, the theory is known to be incomplete. An example of this lack of completeness is the non-inclusion of the gravitational interaction.

2.2   Higgs Mechanism

In order to generate particle masses, a complex scalar field is added to the Standard Model Lagrangian, see Equation 1 [2]. The interplay of this field with the covariant derivative generates mass terms for what will become the electroweak gauge bosons. In addition to the introduction of masses, this interplay also describes a series of predicted interactions between the Higgs boson and the electroweak bosons. The same could be said for the fermionic portion of the full Standard Model Lagrangian where the masses and the couplings for the quarks and leptons would be derived.

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi) \tag{1}$$

Here, ϕ represents the Higgs field while the V represents the potential corresponding

to the field. These two components, with specific conditions on the potential, are the vital elements to describe mass in the Standard Model. The Higgs field is shown in Equation 2 while the potential has the form shown in Equation 3.

$$\phi = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2}$$

$$V(\phi^\dagger \phi) = -\mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2 \tag{3}$$

Consider the meaning of different allowed values of both μ^2 and λ. In the event that μ^2 is positive, there are three possibilities. The first is that λ is less than zero. In this event the potential is unbounded from below and would not represent a physical potential. If λ was set identically equal to zero, the potential would represent a parabola centered on zero. This would not normally be a problem if it were not for the fact that the field minimum would be maximally degenerate and would therefore not allow any masses to appear in the Standard Model. Permitting μ^2 to be negative leads to a similar situation. This leaves only the case that both μ^2 and λ are positive which is the configuration presented in Figure 4. In the work that has been accomplished so far a one-dimensional potential has been used. In the actual Standard Model treatment the potential is actually a two-dimensional function which can be constructed by rotating the function in Figure 4 about the V axis. This will produce the famed sombrero potential seen in many articles and books on the subject. With the two-dimensional case, there are a continuum of solutions for the minima instead of just two as is represented in Figure 4.

It can be easily noted that the minimum of the potential in this simple one-dimensional example is not at the zero point in the field. While the desire of the field is to exist in its ground state, it is of interest to study the properties of the field by investigating its behavior in the neighborhood of its ground state. The ground state, the point at which the potential is a minimum, can be found by studying the derivative of the potential
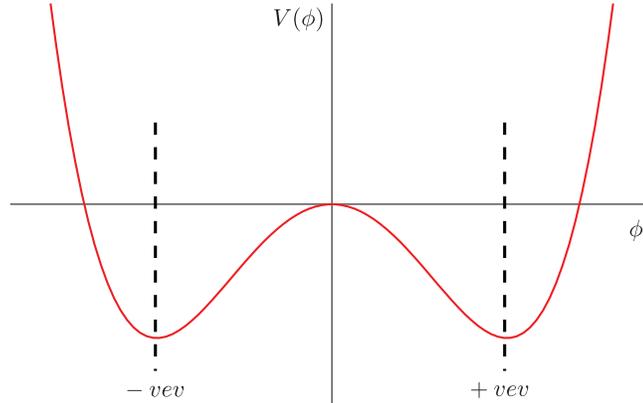
Figure 4: The Higgs field potential with μ^2 and λ both forced to positive values

which is presented in Equation 4.

$$\frac{\mathrm{d}V}{\mathrm{d}|\phi|^2} = -\mu^2 + 2\lambda|\phi|^2 = 0$$

$$\therefore \quad \langle\phi\rangle = \pm\frac{\mu}{\sqrt{2\lambda}} = v \tag{4}$$

A simple observation of Figure 4 will show that there are multiple possible ground states for the Higgs field. The minimum of the potential, located at the vacuum expectation value (*vev*), has two possible choices, positive or negative. Conventionally, the positive value is chosen. As a reminder, it should be noted that the *vev* is equal to the magnitude of the field (Equation 5). Since the field has four components, the value of the *vev* can be associated with any one of these. It is through the simple act of choosing a particular *vev* (one of the two in this example) that spontaneous symmetry breaking has been accomplished. One should be careful that the *vev* is not placed in the charged Higgs location, determined by hypercharge requirements, since this causes other inconsistencies such as the non-conservation of charge [2].

$$\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 = v^2 \tag{5}$$

For this purpose I will choose to place the *vev* in the ϕ_3 position, the real component of the neutral Higgs boson. With this in place, the value of the field in the ground state is shown in Equation 6.

$$\phi = \begin{pmatrix} 0 + i0 \\ v + i0 \end{pmatrix} \tag{6}$$

With the field in the ground state, performing perturbations of the field will allow its properties to be determined. This is accomplished by adding a small parameter to each of the field components. It is found that the small field components in place of ϕ_1, ϕ_2, and ϕ_4 represent goldstone bosons and have the ability to be removed through a gauge transformation. In this new gauge, called the unitary gauge, the Higgs field is represented in Equation 7 where the imaginary components have been removed since they are identically equal to zero.

$$\phi = \begin{pmatrix} 0 \\ v + h \end{pmatrix} \tag{7}$$

The mass of the Higgs boson, contained in the expanded potential, is determined by collecting terms that are quadratic in the field h. It should be noted that only terms quadratic in the field are retained in the expansion of the Lagrangian potential resulting in Equation 8.

$$\mathcal{L}_{\text{pot}} = -\mu^2(v^2 + vh + h^2) + \lambda(v^2 + vh + h^2)(v^2 + vh + h^2)$$

$$= -\mu^2 h^2 + \lambda(3v^2 h^2)$$

$$= (-\mu^2 + 3\lambda v^2)h^2$$

$$= (-v^2 2\lambda + 3\lambda v^2)h^2$$

$$= \lambda v^2 h^2 \tag{8}$$

Since the mass term is identified as any term with the form of $\frac{1}{2}m^2 h^2$, the mass of the Higgs particle can be extracted from Equation 8 as $m_H = v\sqrt{2\lambda}$. It is clear from Equation 8 that the Higgs boson mass depends on both the value of the *vev* and the quartic coupling from the field potential. As stated previously, the *vev* has been constrainted through the Fermi coupling to be 246 GeV [9]. While we are aware of the experimental value of the *vev*, there is currently no known experimental handle on the value of the quartic coupling. Our ignorance of the Standard Model Higgs boson mass is a direct result of our lack of knowledge of the quartic coupling.

2.3   Higgs Boson Production Modes and Decay Rates

The discussion of Higgs boson production and decay rates requires one to extract the interaction vertices present in the Standard Model Lagrangian that contain a Higgs boson. These are found by looking for Lagrangian terms that contain three field terms with at least one being a Higgs boson. There are four different vertices that correspond to this requirement that are shown in Figure 5 [2].

While the Higgs boson is produced through several different modes the two dominant modes used in the search analyses are gluon fusion and associated production. The production rates of these processes at current hadron colliders along with the additional production modes as a function of the Higgs boson mass can be seen in Figure 6.

The gluon fusion process dominates the production rate at both the Tevatron and

(a) $H \to W^+ W^-$ (b) $H \to ZZ$ (c) $H \to HH$ (d) $H \to ff$
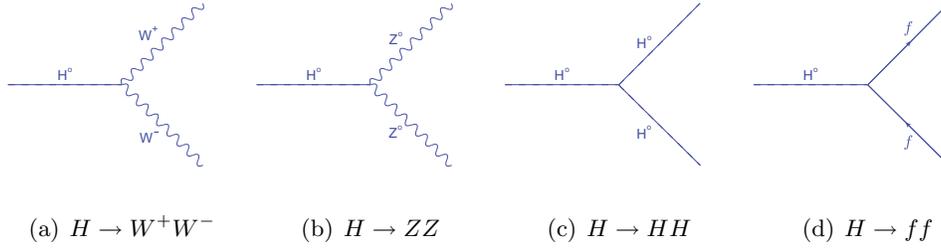
Figure 5: Tree level three particle interaction vertices involving the Higgs boson in the Standard Model implementation

(a) Proton on antiproton collisions at the Tevatron ($\sqrt{s} = 1.96$ TeV)

(b) Proton on proton collisions at the LHC ($\sqrt{s} = 14$ TeV)
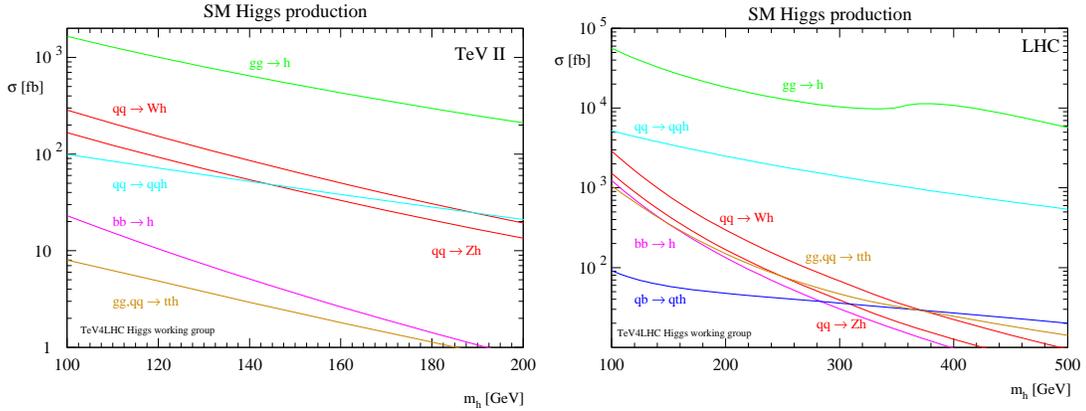
Figure 6: The production cross-sections for various Higgs production schemes [9].

the LHC. For this process, two gluons couple to quarks which then combine to produce a Higgs boson, as shown in Figure 7.
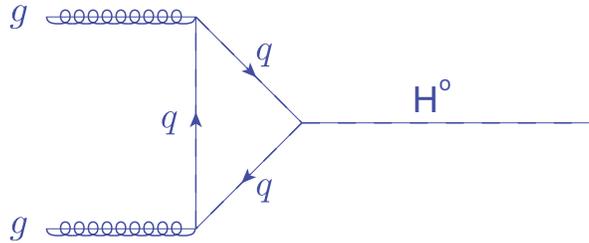
Figure 7: Gluon fusion production mode for the Higgs boson

While the gluon fusion process dominates the production rates, it is an undesired

14

mode for a light Higgs boson search due to the large backgrounds in the analysis. The production process used in this analysis, called associated production, is described by having two fermions combine to form a vector boson. This vector boson then radiates a Higgs boson leaving a H and a $W^{\pm}$ or a Z in the final state. This process has been presented in Figure 1. The associated production rates for the three mass points used in this analysis can be seen in Table 1 [14].

Higgs Mass (GeV/c^2)	$p\bar{p} \rightarrow WH$ (pb)		$p\bar{p} \rightarrow ZH$ (pb)		$p\bar{p} \rightarrow VH$ (pb)
115	0.186	+	0.108	=	0.293
125	0.138	+	0.081	=	0.219
135	0.104	+	0.062	=	0.166

Table 1: The production rates in pb for a Higgs boson though the associated channel at the Tevatron using a center of mass energy of 1.96 TeV [14]. The VH cross section is just the sum of the WH and ZH modes.

The decay of the Higgs boson follows from the same vertices shown in Figure 5. While there are several different decay modes, the most important modes are the Higgs bosons to two quarks, the Higgs boson to two W bosons, and the Higgs boson to two Z bosons. The partial widths of these three decays are shown in Equations 9, 10, and 11 [2].

$$\Gamma(H \rightarrow ff) = 2\frac{G_F}{4\pi\sqrt{2}}m_f^2 m_H (1 - 4\frac{m_f^2}{m_H^2})^{\frac{3}{2}} \tag{9}$$

$$\Gamma(H \rightarrow WW) = \frac{G_F}{8\pi\sqrt{2}}m_H^3 (1 - 4\frac{m_W^2}{m_H^2})^{\frac{1}{2}} (12\frac{m_W^4}{m_H^4} - 4\frac{m_W^2}{m_H^2} + 1) \tag{10}$$

$$\Gamma(H \rightarrow ZZ) = \frac{G_F}{16\pi\sqrt{2}}m_H^3 (1 - 4\frac{m_Z^2}{m_H^2})^{\frac{1}{2}} (12\frac{m_Z^4}{m_H^4} - 4\frac{m_Z^2}{m_H^2} + 1) \tag{11}$$

The form of the partial decay widths for the Higgs boson shown in Equations 9 - 11 show a dependance on the mass of the Higgs boson. The dependence is such that the primary decay for a low mass Higgs boson will be to two quarks while a high mass Higgs boson will decay to two W bosons. The transition between these two decays is at a Higgs boson mass of approximately 135 GeV/c^2 [9]. Figure 8 displays the branching

ratios for the various decay modes that the Standard Model Higgs boson can decay to. It is seen in Figure 8 that the $b\bar{b}$ decay dominates below 135 GeV/c^2 while the WW decay becomes dominant in the high mass region. In the low mass region the decay to the $b\bar{b}$ dominates due to the square of the fermion mass in the partial width, see Equation 9. Since the t quark is more massive than the b quark, it might be thought that the $t\bar{t}$ decay would dominate. The $t\bar{t}$ decay does not become kinematically available until the mass of the Higgs boson reaches twice the mass of the t quark, approximately 340 GeV/c^2. There are several decay processes shown in Figure 8 that have no direct corresponding vertex, an example being $H \rightarrow \gamma\gamma$. The $H \rightarrow \gamma\gamma$ decay process takes place though a virtual triangle of particles that have a photon vertex. These are suppressed from the direct tree level decays due to the complexity of the Feynman diagram.
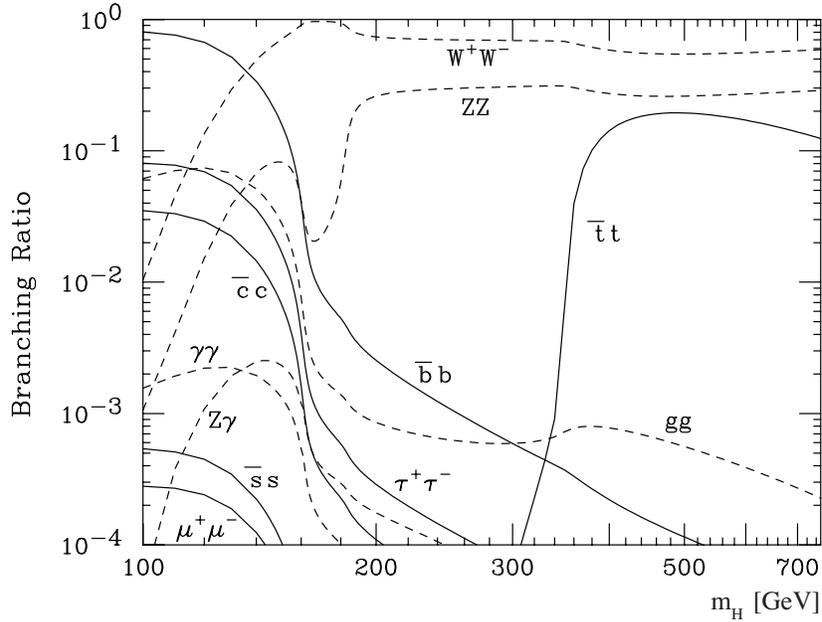
Figure 8: Decay rates of the Higgs particle as a function of the Higgs mass [9].

The values of the $b\bar{b}$ branching ratios used in this analysis are 0.732 for a Higgs boson mass of 115 GeV/c^2, 0.610 for a mass of 125 GeV/c^2, and 0.463 for a Higgs boson mass of 135 GeV/c^2 [15].

2.4 Theoretical Higgs Mass Bounds

2.4.1 Effective Potential

In the previous section, the mass of the Higgs particle was derived using a specific form of the field potential. In order to study the different theoretical mass bounds imposed on the Higgs particle, the effective one loop potential will be used. The loops that are inserted are due to the Higgs propagator adding creation / annihilation events that couple to Higgs particles, other bosons, and the fermions. It is through this step that the mass of the Higgs boson can be bound and controlled by the masses of other particles. Of all the fermion interactions, the top quark is the strongest and therefore, the influence of the other fermions are normally ignored.

2.4.2 High Mass Bounds

In Section 2.2, a discussion was presented regarding the different possibilities for the allowed values of μ^2 and λ [16]. In this and the following section, the conditions on λ are used to find both an upper and lower bound on the Higgs boson mass. For the case on the upper bound, the quartic coupling is allowed to run to a specific cutoff point representing the need for new physics. The bound, set in this way, is often called the triviality bound in the literature and is executed by considering the running equation for λ. In the event that only self interactions are observed, (i.e. suppress the fermion loops in addition to the $W^\pm$, Z and γ loops in the effective potential) the running equation is shown in Equation 12 [16].

$$\frac{d\lambda}{dt} = \frac{3\lambda^2}{4\pi^2} \qquad t \equiv \log\left(\frac{Q^2}{Q_0^2}\right) \tag{12}$$

This equation is solvable through simple integration due to its separable quality.

$$\lambda(Q) = \frac{\lambda(Q_0)}{1 - \frac{3\lambda(Q_0)}{4\pi^2}\log\left(\frac{Q^2}{Q_0^2}\right)} \tag{13}$$

The requirement used to establish the theoretical upper limit translates to an obligation that the $\lambda(Q)$ coupling, represented by Equation 13, always be finite. By using this requirement and choosing that the reference scale Q_0 be equal to the *vev*, the bound is found in Equation 14.

$$0 < \frac{1 - \frac{3\lambda(Q_0)}{4\pi^2} \log\left(\frac{Q^2}{Q_0^2}\right)}{\lambda(Q_0)}$$
$$0 < 1 - \frac{3M_h^2}{2v^2 4\pi^2} \log\left(\frac{Q^2}{v^2}\right)$$
$$\therefore \quad M_h^2 < \frac{8v^2\pi^2}{3} \left[\log\left(\frac{Q^2}{v^2}\right)\right]^{-1} \tag{14}$$

It should be noted that $\lambda(v) = \frac{M_h^2}{2v^2}$ was used in the previous equations to simplify notation. The upper bound from this equation can be seen in Figure 9.
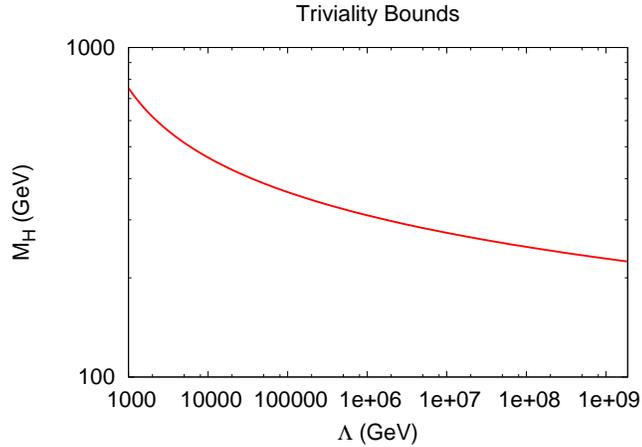
Figure 9: The upper bound on the Higgs mass using the triviality argument taking into account only one loop self interactions [16]

If it is assumed that there is no new physics up to the GUT scale of $Q = 10^{16}$ GeV, the relevant upper bound on the Higgs mass is 160 GeV/c^2. It should be stressed again that this is valid only for a self-interacting one loop running coupling. When altering the coupling evolution equations to account for the added boson and fermion loops, the

18

equation for λ becomes slightly more complicated [16].

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \frac{1}{16\pi^2}\left[12\lambda^2 + 12\lambda g_t^2 - 12g_t^4 - \frac{3}{4}\lambda\left(3g^2 + g'^2\right)\right.$$
$$\left. + \frac{3}{16}\left(2g^4 + (g^2 + g'^2)^2\right)\right] \tag{15}$$

The g and g' couplings in Equation 15 are the standard couplings for the electroweak sector, while g_t represents the coupling to the top quark. Due to the specific type of bound, it is relevant to only retain the terms that contain λ. In this case Equation 15 then becomes Equation 16.

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \frac{\lambda}{16\pi^2}\left[12\lambda + 12g_t^2 - \frac{3}{2}(3g^2 + g'^2)\right] \tag{16}$$

Running to the GUT scale again and setting the top quark mass to 175 GeV/c^2, the relevant upper bound on the Higgs mass is adjusted to 170 GeV/c^2.

2.4.3 Low Mass Bounds

In order to derive the lower mass bounds, the requirement changes to ensure that λ remains positive [16]. This bound, sometimes called the vacuum stability bound, refers to the possibility that a true ground state exists at some other point in the field instead of the previously determined *vev*. This bound is determined by looking at Equation 15 and only considering terms that do not contain λ. The terms that contain λ explicitly would only inflate the lower value giving an incorrect lower bound. The running equation with these terms explicitly removed is presented in Equation 17.

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \frac{1}{16\pi^2}\left[-12g_t^4 + \frac{3}{16}(2g^4 + (g^2 + g'^2)^2)\right] \tag{17}$$

Once this equation is solved, through simple integration, the limit on the square of the Higgs boson mass is shown in Equation 18 [16].

19

$$M_h^2 > -\frac{v^2}{8\pi^2}\left[-12g_t^4 + \frac{3}{16}(2g^4 + (g^2 + g'^2)^2)\right]\log\left(\frac{Q^2}{v^2}\right) \qquad (18)$$

Running again to the GUT scale and using a top quark mass of 175 GeV/c^2, the lower limit on the Higgs mass is 130 GeV/c^2. By combining these two results one can obtain a plot that provides a theoretical constraint on the Higgs mass with respect to the scale at which new physics is required. This plot, known by many as the chimney plot, is shown in Figure 10.
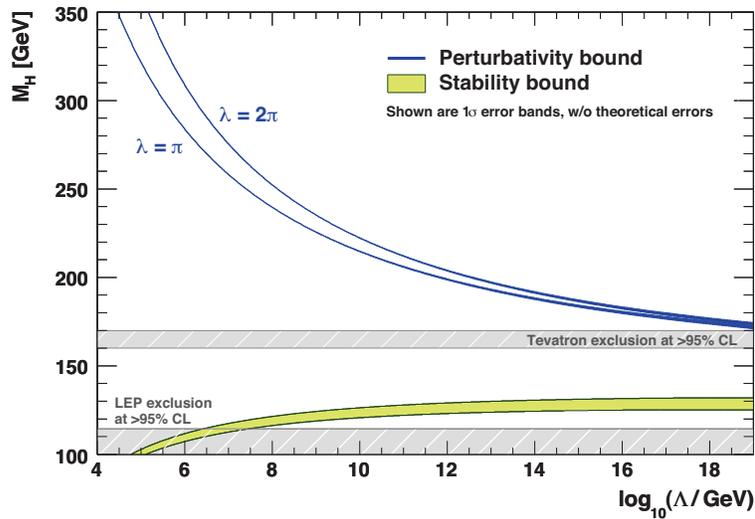
Figure 10: Combination of the triviality and the vacuum stability theoretical constraints on the Higgs mass as a function of the scale of new physics [17].

2.5 Technicolor

The Higgs mechanism, as implemented in the Standard Model, contains a large number of free parameters in order to provide different masses to the fermions. In an effort to construct a theory that provides mass to the weak sector bosons in addition to the fermions without adding a large number of free parameters, the concept of a dynamical electroweak symmetry breaking was introduced. In 1979 two independent papers were submitted to Physical Review by Weinberg and Susskind [18, 19]. These papers outlined

the beginning stages of what is known today as Technicolor.

Technicolor is seen, at the phenomenological level, as a low energy QCD implementation [20]. In this theory a new set of fermions, called Technifermions, and an asymptotically free gauge interaction are added into the Standard Model. These new particles, which are massless, exhibit an exact chiral flavor symmetry. When the energy of the system falls to a low enough level, such that the gauge interaction coupling becomes an appreciable value, the Technifermions condense into Technimesons. The production of these condensates spontaneously breaks the chiral symmetry and in accordance with the Goldstone Theorem a set of goldstone bosons is produced. Just as is done in the Standard Model Higgs implementation, three of these goldstone bosons are used to provide mass for the $W^{\pm}$ and the Z^0 of the electroweak sector.

The simple explanation given above does not explain the masses of the individual fermions. This is accomplished by introducing Extended Technicolor. This is a modification of the above theory such that a new set of interactions are introduced which explicitly break the chiral flavor symmetry of the Standard Model fermions in addition to the newly introduced Technifermions. This new interaction is embedded in a gauge group (G_{ETC}) that contains the chiral gauge group of the Technifermions in addition to the regular Standard Model gauge structure. This new gauge group is broken at an energy level Λ_{ETC} and the gauge description of the new model is returned to the initial Technicolor gauge plus the regular Standard Model gauge description. This breaking then provides masses to the Standard Model fermions though single loop diagrams with the newly introduced gauge bosons of the broken G_{ETC} group.

Just as the WZ and the Higgs boson production have followed from Figure 2, Technicolor production does the same. For Technicolor, the $W^{\pm}$ intermediate boson propagator is replaced with a technirho particle. This particle then propagates and subsequently decays to a technipion and a $W^{\pm}$ in the final state. Just as the Higgs boson production rates depended on the Higgs boson mass, the technicolor production rates depend on two masses, $M(\rho_{\text{TC}})$ and $M(\pi_{\text{TC}})$. The mass of the technirho was set to 240 GeV/c^2,

due to Monte Carlo availability, for this study and the mass of the technipion was allowed to change between the values of 115 GeV/c^2, 125 GeV/c^2, and 140 GeV/c^2. The production cross sections for these mass points using Pythia is found to be 0.753 pb, 1.081, pb, and 0.949 pb, respectively for the energies of the Tevatron [21]. The production rates do not follow a continuous pattern with the increasing of the technipion mass. This is because the coupling of the technirho to two technipions is larger than the coupling for it to decay to a technipion and a $W^\pm$ boson [22]. When the technirho mass is larger than twice the technipion mass, that decay is available and dominates resulting in a lower cross section for the process depicted in Figure 2. The production rates for the case where the technirho mass is less than twice the technipion mass follow a decaying pattern as the technipion mass increases, as one would expect.

The technipion can decay to four primary modes: $b\bar{b}$, $c\bar{c}$, $s\bar{s}$, and $\tau^+\tau^-$. The decay rates are proportional to the square of the decay product's mass [22]. The branching ratio for the $b\bar{b}$ decay is 85.7% while the $c\bar{c}$ and $s\bar{s}$ decay rates are 5.3% and 1.8%, respectively. The remaining 7.2% is reserved for the $\tau^+\tau^-$ decay [21].

3 Accelerator Complex and Detector

3.1 Accelerators

The protons and anti-protons used in the collisions at the DØ detector are generated by a collection of accelerators, each one stepping the energy of the particles closer to their final energy of 980 GeV. The complex, shown in Figure 11, is split into five individual systems: the pre-accelerator, linear accelerator, booster, main injector, and finally the Tevatron.
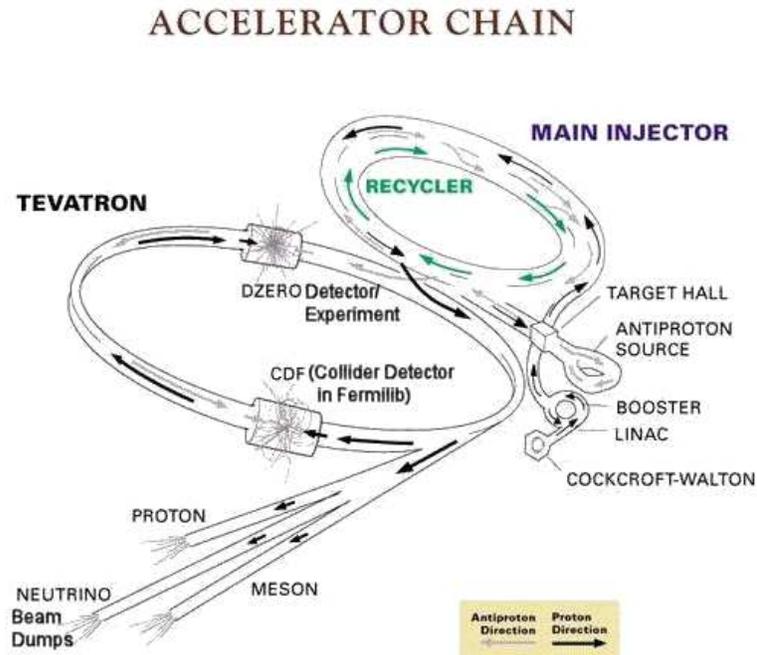
Figure 11: Overview of the Fermilab accelerator complex

The pre-accelerator (preacc), or the source, is a Cockcroft-Walton accelerator. Here, a hydrogen gas passes over a heated Cesium brick where it picks up an additional electron to form a negative ion [23]. This negatively charged ion is then accelerated through a 750 kV potential difference and is then passed into the first of two linear accelerators.

The linear acceleration (linac) portion of the complex is actually split into two different accelerators. The first is composed of several Alvarez cavities [24] and brings the energy of the hydrogen ions to 116 MeV [23]. After the first linear accelerator, the

ions are then passed to the second side coupled RF cavity accelerator. In this final linear stage, the energy of the ions is brought to 400 MeV [23].

After the preacc and the linac, the ions pass into the booster ring. The goal of the booster is to bring the energy of the protons up to 8 GeV [25]. Injection of the ions in this manner is preferred due to the reduced complexity of the injection site. Since the ions are oppositely charged, the same steering magnet can be used to bend the already circulating proton beam and the negative ion beam together. As the ions are injected into the circular booster ring, the two electrons are stripped off by a thin Carbon film leaving only the proton.

The next stage in the acceleration process is the passage of the protons from the booster ring into the Main Injector. The Main Injector performs two primary functions. The first of these (in the order of the Tevatron startup) is to raise the energy of the protons to 150 GeV [26]. These 150 GeV protons are then redirected to a nickel target in order to produce anti-protons. The rate of anti-proton generation is a limiting factor at the complex. The anti-protons are sent to an anti-proton storage ring until enough have been produced to inject into the Tevatron. The second task of the main injector is to raise the energy of the protons and anti-protons to 180 GeV, at which point they are transferred back to the main injector. Once the particles are at this energy they are injected into the Tevatron ring.

The final stage in the particle acceleration occurs in the Tevatron ring. The ring has a circumference of 6.28 km and brings the particle energies from 180 GeV to 980 GeV. This acceleration is provided from electric fields produced in a series of radio frequency cavities. Upon injection from the main injector, the protons and anti-protons are split into 36 individual bunches each containing approximately 10^{10} particles [27]. These bunches are spread out in the ring such that there are two large gaps which serve as abort regions. With the spacing taken into account, these bunches interact in the center of the DØ detector every 396 ns. On either end of the detector are a series of focusing magnets called "Low Beta Quads" which reduce the cross-sectional area of the beam in

order to increase the probability of a collision. The beams are maintained in the ring by a series of powerful dipole magnets that act to steer the particle bunches and a series of quadrupole magnets designed to aid in beam focusing.

3.2   Luminosity

Of critical importance is the rate at which particles are produced in the interactions. This is a function of the cross-section of the production process, σ, and the instantaneous luminosity, $\mathcal{L}$, of the beam. The instantaneous luminosity is a measure of the number of particles per unit time per unit area at a particular point and can be expressed in the form of Equation 19 [9].

$$\mathcal{L} = f\frac{n_1 n_2}{\sigma_x \sigma_y} \tag{19}$$

In Equation 19, f represents the rate of collisions, n_i the number of particles in each bunch, and σ_x and σ_y the horizontal and vertical Gaussian beam widths. While the instantaneous luminosity is a good measure of what is happening from moment to moment in the detector, it is the integrated luminosity, L, that represents the total amount of data that the experiment has collected over a period of time. Using the integrated luminosity, the number of events generated during a period of time is shown in Equation 20, where σ is the process cross section and N is the total number of produced events corresponding to the given cross section.

$$N = \sigma \int \mathcal{L} dt = \sigma L \tag{20}$$

The instantaneous luminosity is the highest at the beginning of a store and falls exponentially until the next store is initiated. The record initial instantaneous luminosity is 3.81×10^{32} cm^{-2}s^{-2} taken in March of 2010 [28]. A common unit of area measurement used in high energy physics is the barn. One barn is equal to 1×10^{-28} cm^2 and is denoted by the symbol b. The total integrated luminosity used for this study

is approximately 1007 pb^{-1} which corresponds to just over 115M individual recorded data events.

3.3 Detector

The data used in this study has been acquired from the DØ RunIIa detector [29] between 2002 and 2006. A schematic cross-section of this detector is shown in Figure 12. Ten meters square on the end and twenty meters long, the detector houses a collection of different tracking and calorimetry systems. The design of the detector is hermetic at its core in an effort to capture all of the particles in the aftermath of the collisions. This design feature also enables the experimenters to detect neutrinos indirectly as unbalanced transverse energy in the collisions. The individual detectors and support electronics have been designed with the ability to handle bunch crossings every 396 ns with an average of two collisions per crossing.
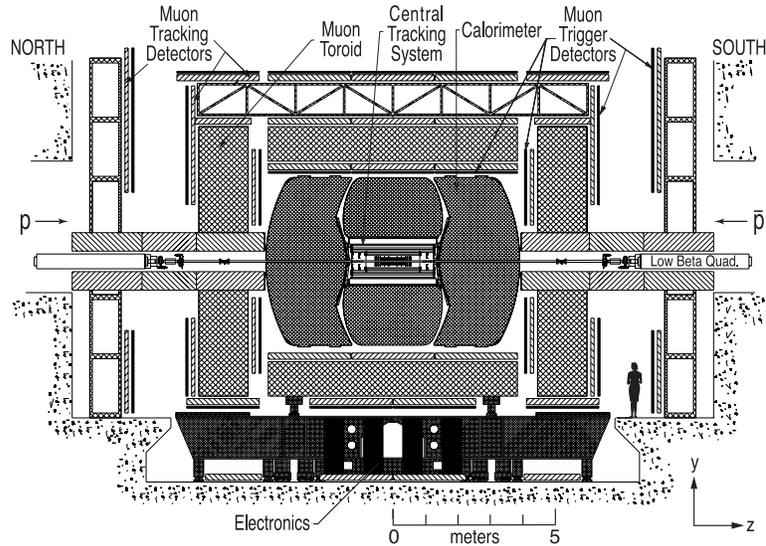
Figure 12: A visual representation of the DØ RunII detector used to acquire the data for this study [29].

The detector can be split into three major sections: tracking, calorimetry, and muon detection. This chapter will focus on introducing each of these major sections. Before moving forward, it is prudent to discuss the coordinate system in use in addition to

26

several quantities that will be used in the analysis.

3.3.1 Coordinates and Useful Quantities

A common set of coordinates and variable definitions are used throughout the DØ environment and other analyses to help describe the motion of the particles both before and after the collision. To begin, a simple Cartesian coordinate system is defined such that the origin is placed at the geometric center of the detector. The y axis points upward toward the zenith and the x axis points toward the outside of the ring. The z axis travels along the beam-line such that the positive direction follows the protons which is clockwise when the ring is viewed from above [29].
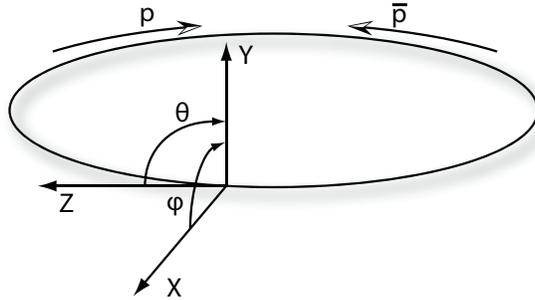
Figure 13: Definition of the Cartesian coordinate system for the DØ detector

Due to the nature of the collisions and the geometry of the detector, the spherical coordinate system is preferred over the Cartesian system. Here the radial quantity is, of course, measured from the Cartesian origin while the ϕ coordinate is measured from the positive x-axis toward the positive y-axis and ranges from 0 to 2π. The θ angle is measured from the positive z axis to the displacement vector and ranges from 0 to π.

In addition to the two coordinate systems defined above there are additional quantities that are used from analysis to analysis to describe the motion and location of the particles. The first, called the rapidity, is often used in systems where there is an excessive amount of energy over the rest energy of the system. Denoted by y and shown in Equation 21, it is a relinearization of the velocity and allows for simple treatments of relativistic particle velocities.

$$y = \frac{1}{2} \ln \frac{E + p_L}{E - p_L} = \tanh^{-1}(\beta_L) \qquad (21)$$

In this equation E represents the total energy of the particle while p_L is the longitudinal component of the total momentum. In the second form, β_L is the longitudinal component of the velocity relative to that of light. A more common experimental quantity, strongly related to the rapidity, is the pseudorapidity. This is an approximation of the rapidity which is allowed when the total momentum of the particle is much larger than the particle's rest mass. Represented by η and shown in Equation 22, this quantity is useful due to the invariance of the difference between two particle's pseudorapidity in boosted frames of reference [30].

$$\eta = \frac{1}{2} \ln \frac{|\vec{p}| + p_L}{|\vec{p}| - p_L} = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \qquad (22)$$

The pseudorapidity takes on the value of 0 when $\theta = \frac{\pi}{2}$ and moves to ∞ when $\theta = 0$ and $-\infty$ when $\theta = \pi$. A final set of useful variables are presented below:

$$E_T = E \sin \theta \qquad\qquad p_x = p_t \cos \phi \qquad (23)$$

$$p_t = \sqrt{p_x^2 + p_y^2} \qquad\qquad p_y = p_t \sin \phi \qquad (24)$$

$$\Delta R = \sqrt{\Delta \phi^2 + \Delta \eta^2} \qquad\qquad p_z = p_t / \tan \theta \qquad (25)$$

3.3.2 Tracking

The tracking detector is the first detector the particles from the collision interact with. Positioned just outside the beryllium beam-pipe covering the radial region from 2.6 cm to 52 cm and the longitudinal region ± 126 cm from the geometric center of the detector, the tracker is composed of two sub-detectors: the Silicon Microstrip Tracker (SMT) and the Central Fiber Tracker (CFT) [29]. A cross section of the tracking volume is shown in Figure 14. These trackers, which could operate independent of each

other, work in concert to yield a high precision momentum measurement of the charged particles emanating from the collision. The measurement of momentum is accomplished by determining the curvature of the particle track due to the Lorentz force as it moves through a uniform magnetic field. The magnetic field is provided by a superconducting solenoid that rests between the tracking detectors and the preshower elements. This solenoid is driven by a current of 4,749 A producing a uniform magnetic field of 2 T in the tracking volume. As the particle moves through the tracking volume, it interacts with the detector leaving behind small deposits of energy, called clusters, in specific locations. By looking at the clusters correlated to a particle, a track can be found.
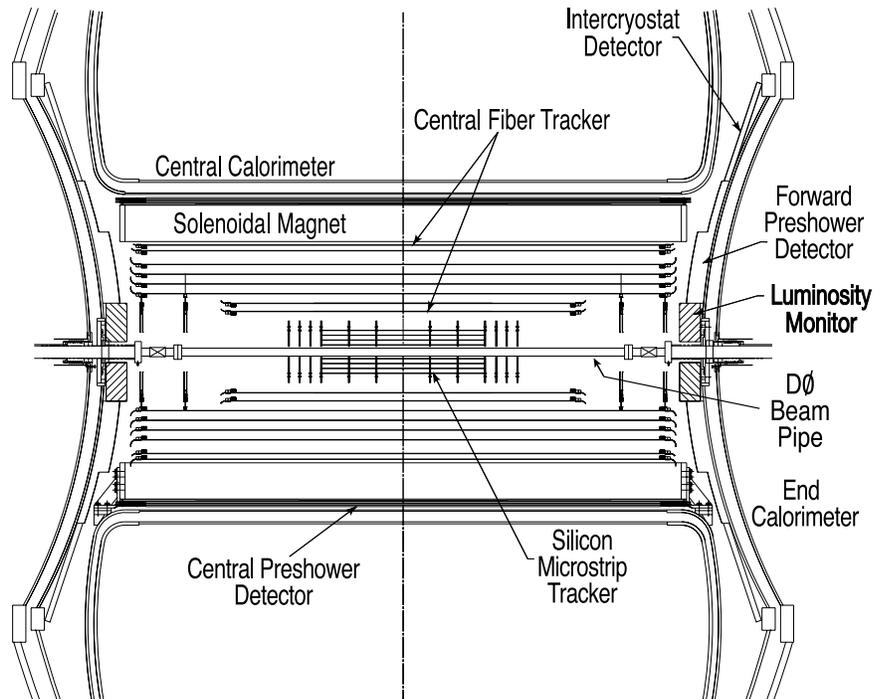
Figure 14: A cross section of the DØ tracking elements [29].

3.3.3 Silicon Microstrip Tracker

The Silicon Microstrip Tracker, or SMT, is a vast array of 792,576 silicon PIN strip diode sensors [29]. Each sensor functions by collecting the electron-hole pairs produced when a charged particle traverses the bulk material. This collection is facilitated by an

imposed electric field produced via a potential known as the bias voltage. The charge carriers are collected by a capacitor and then converted into a digital signal by the custom designed SVXIIe readout chip [31].
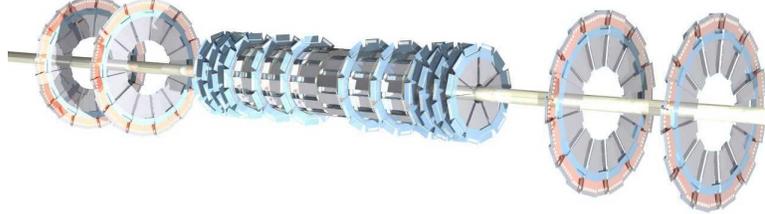
Figure 15: A rendering of the SMT showing the three different geometries. The H-Disks are on the outside with orange edges while the F-Disks are closer to the core with blue edges. The disk structures position the PIN diodes perpendicular to the beam line. In between the first several F-Disks is the barrel structure that holds the silicon parallel to the beam direction [29].

The SMT is split into three main geometries in an effort to achieve good detection coverage with detector surfaces that are perpendicular to particle trajectories, as shown in Figure 15. Starting from the center and moving in the positive z direction, the first detector geometry is called the barrel. Each barrel is composed of four silicon readout layers mounted in concentric circles around the beam-pipe with each layer having an array of sensors called ladders. These ladders are positioned such that the sensor is parallel to the beam line. The first two layers contain 12 ladders while the third and fourth layers have 24 ladders each. Immediately past the first barrel is the second geometry called the F-Disk. The F-Disk is positioned perpendicular to the beam-line and is composed of 12 overlapping sensors called F-Wedges. This arrangement of barrel and F-Disk repeats itself three times in rapid succession. Just past the last F-Disk in the barrel/F-Disk groups, there are three more closely spaced F-Disks. Placed in the far forward region there are two H-Disks which represent the third and final geometry. Each H-Disk is composed of 24 individual overlapping sensors called H-Wedges, much like the F-Disks. This collection of detector elements is mirrored on the other side of the origin to give a total of 6 barrels, 12 F-Disks, and 4 H-Disks.

30

The individual sensors come in a variety of configurations. Layers 2 and 4 of the barrels are double-sided sensors with one side of the sensors set at a $2°$ offset with respect to the other side. This angle offset provides the ability to determine the z position of the hit on the sensor. Layers 1 and 3 of the inner four barrels have a double-sided double-metal configuration. One side of the sensor has an additional layer of metal to allow the sensor strips to be oriented $90°$ to the opposite side. The outer two barrels have two single-sided sensors mounted back-to-back in layers 1 and 3. The F-Wedge sensors are double-sided with a full stereo angle of $30°$. Similar to the case for the double-sided ladders, this stereo angle aids in the ability to determine a full 3D position of the cluster. The H-Wedges are composed of two single-sided sensors attached to each other with an effective stereo angle of 15 degrees.

The longitudinal centers of the barrels are positioned at $|z|=$(6.2, 19.0, and 31.8) cm while the centers of the F-Disks are positioned at $|z|=$(12.5, 25.3, 38.2, 43.1, 48.1, and 53.1) cm. The H-Disks are centered at (100.4 and 121.0) cm away from the physical center of the tracker. The positioning of the individual elements of the SMT allow it to track particles that have pseudorapidities of less than 3. The impact parameter resolution of the tracker is a function of the p_T of the track in question. For low p_T tracks, lower than about 5 GeV, the resolution is approximately 30 μm. As the p_T increases, the resolution improves and plateaus near 20 μm [32].

3.3.4 Central Fiber Tracker

The Central Fiber Tracker, or CFT, surrounds and works in concert with the SMT. Together these detectors provide a position resolution of 35 μm in the longitudinal direction and 15 μm in the transverse plane [29]. The CFT is composed of eight concentric rings of scintillating fibers. Of these eight rings, the inner two rings are 1.66 m long while the outer six rings are 2.52 m long. This difference in length is to allow the SMT H-Disks to rest inside of the CFT. The overall dimensions of the CFT allow it to handle particles with an $|\eta|$ of up to 1.7. Each of the rings is composed of two layers of fibers,

the axial layer and the stereo layer. An example of a fiber ribbon is shown in Figure 16.
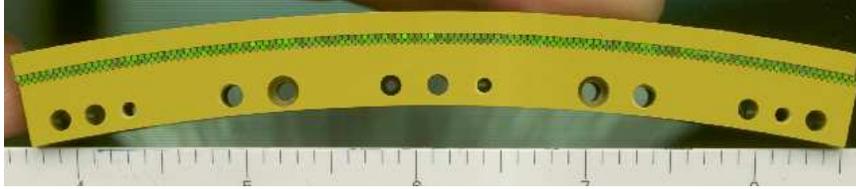
Figure 16: Example of a CFT fiber ribbon [29].

The axial layer is oriented along the z direction while the stereo layer is mounted at an offset angle of $\pm3°$. Each of the scintillating fibers is 835 μm in diameter and is composed of a combination of polystyrene, paraterphenyl, and 3-hydroxyflavone. This combination of materials and their respective proportions produce photons at a wavelength of 530 nm which efficiently propagate through polystyrene light guides to the photon counters. The sensors, called visible light photon counters, or VLPC's, are positioned at the base of the detector in a cryostat designed to maintain an operating temperature of $\sim$9K. This low temperature both increases the efficiency and reduces the noise enabling the detection of individual photons. With the positions of each fiber in the CFT known to better than 50 μm, the resolution of the subdetector is on the order of 100 μm.

3.3.5 Calorimeter

While the tracking detector provides a measurement of the momentum for charged particles, the calorimeter, shown in Figure 17, completes the picture by determining the energy of both charged and uncharged particles. The construction of the calorimeter is based on a sampling design and uses several layers of both stopping and active material. While the stopping material changes throughout the detector, the active material is liquid argon maintained at a temperature of $\simeq$ 90K. The energy measurement of a particle is determined by how the initial particle showers through the different layers of stopping material.

The calorimeter of the DØ detector physically encapsulates the entire tracking and
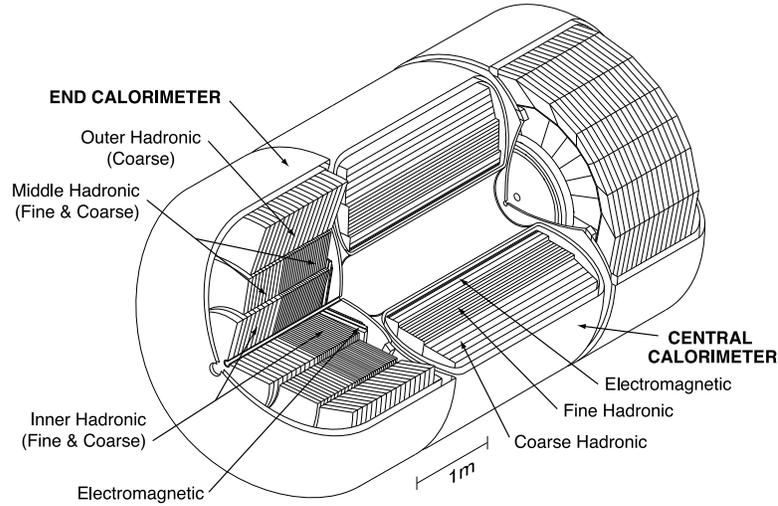
Figure 17: The DØ calorimeter [29].

preshower core leaving only small gaps and openings for the beam pipe and other re-
quired connections. It can be separated into three major regions: two in the forward η
directions, $1 \leq |\eta| < 4$, called the end caps, and a central region which covers $|\eta| < 1$. In
each of these three regions, the calorimeter can be further separated into three specific
types: Electromagnetic, Fine Hadronic, and Coarse Hadronic. While there are three
different types of calorimeters in each η region, a common calorimeter cell design is
utilized, see Figure 18.

One distinction between the cells in the different calorimeter types is based on the
stopping material that is utilized. The electromagnetic cells use nearly pure depleted
uranium. The fine hadronic cells use an alloy of depleted uranium and niobium at a con-
centration of approximately 2%. Finally, the coarse hadronic cells utilize copper plates.
The motivation for the different absorber materials is to equalize the measurement
efficiency between the electromagnetically interacting particles in the electromagnetic
region and the strongly interacting particles in the hadronic portions of the calorimeter.
A calorimeter with an equalized measurement efficiency, called a normalized response,
between these two regions is called a compensating colorimeter.

Each cell is composed of an absorber plate, active material, and finally a signal board.

33
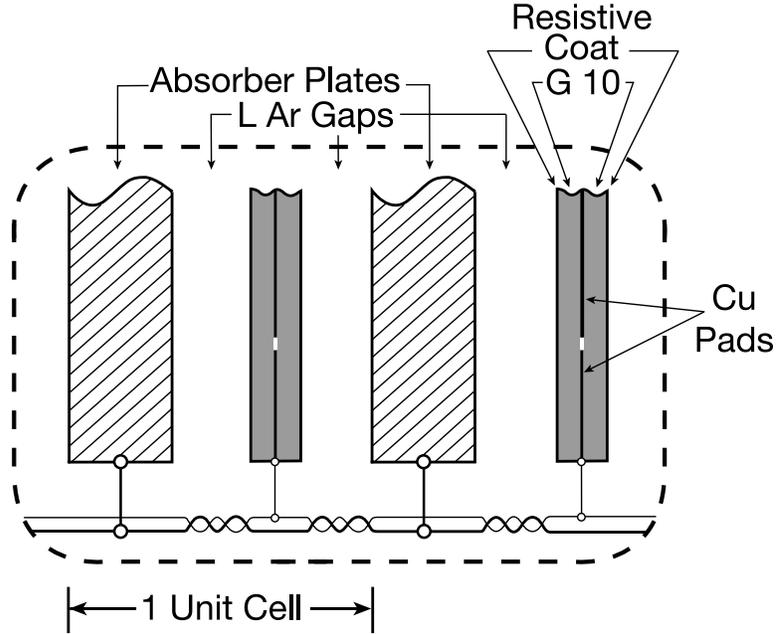
Figure 18: A schematic representation of a calorimeter cell [29].

As particles, both charged and uncharged, traverse the active medium, liquid argon in this case, electron production occurs through ionizing collisions. Similar to what is done with the SMT, a bias potential is established in order to move the electrons toward a collector plate on the sensor. The bias voltages in the DØ calorimeter are approximately 2 kilovolts. With a gap of 2.3 mm, the average drift time for the liberated electrons is 450 ns. The cells are sized and organized to form pseudo-projective towers in constant η. This can be visualized in Figure 19.

The resolution of the central calorimeter has been determined and is shown in Figure 20 and ranges from 10% to 20% depending on the measured transverse energy.

3.3.6 Luminosity Monitor

In an effort to determine the luminosity acquired by the detector, subdetectors called the luminosity monitors are placed in the forward regions on the calorimeter endcaps. Each luminosity monitor is composed of four optically isolated scintillators and covers the range of $2.7 < |\eta| < 4.4$. The instantaneous luminosity is approximated using the
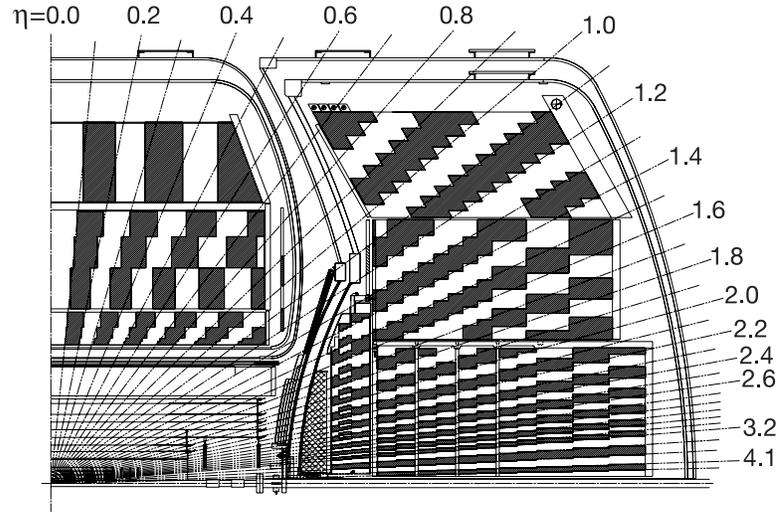
Figure 19: A quadrant of the DØ calorimeter highlighting the pseudo-projective towers in η [29].
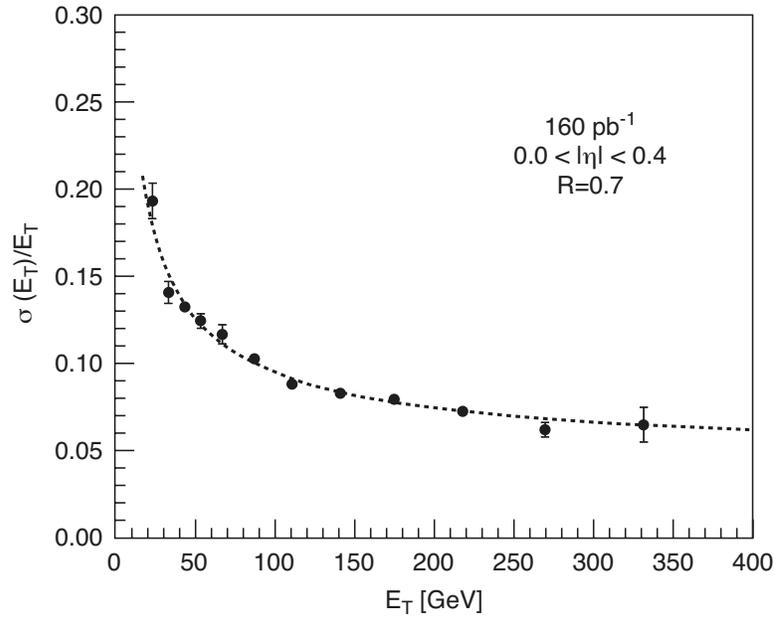
Figure 20: Measured relative transverse energy resolution for jets in the central region of the central calorimeter as a function of transverse energy. The dashed line is a fit to the data [29].

formula shown in Equation 26.

$$\mathcal{L} = f\frac{N_{LM}}{\sigma_{LM}} \qquad (26)$$

The f in Equation 26 is the same as the f in Equation 19 and represents the collision rate. N_{LM} is the number of interactions per crossing and σ_{LM} is the effective cross-section for the luminosity monitor. It is understood that σ_{LM} takes into account detection inefficiencies and acceptance. The luminosity monitor acts as a trigger input by measuring the coincidence between the north and the south detectors. This is done by using the timing information and can be used to estimate the z position of the interaction point. By requiring this interaction point to be less than 100 cm from the center of the detector, backgrounds from beam halo effects can be reduced to negligible amounts.

3.3.7 Muons

While a very large portion of the particles that are produced in the initial collision and secondary interactions are stopped by either the tracking detector or the calorimeter, muons and neutrinos survive. In the physical region beyond the calorimeter, the detector places a three layer deep (A, B, and C) assortment of scintillation detectors called pixels and drift tubes [29]. Figures 21(a) and 21(b) show the orientation and location of the different sensors. The muon detector, like the other components of the overall detector, can be split into a central and forward region. The central region covers $|\eta| < 1$ and is composed of proportional drift tubes (PDT). The two innermost layers (A and B) in the central region also have trigger scintillation counters. The forward regions of the muon detectors cover $1.0 < |\eta| < 2.0$ and are composed of scintillation counters (FSC) and mini-drift tubes (MDT). In order to gain an independent momentum measurement, a 1.8 T magnetic field is produced by the toroid which is placed between the A and B layers of the muon detector. The position resolution of the MDT is ~ 0.8mm [33].

(a) Drift tube's   (b) Scintillation pixel's
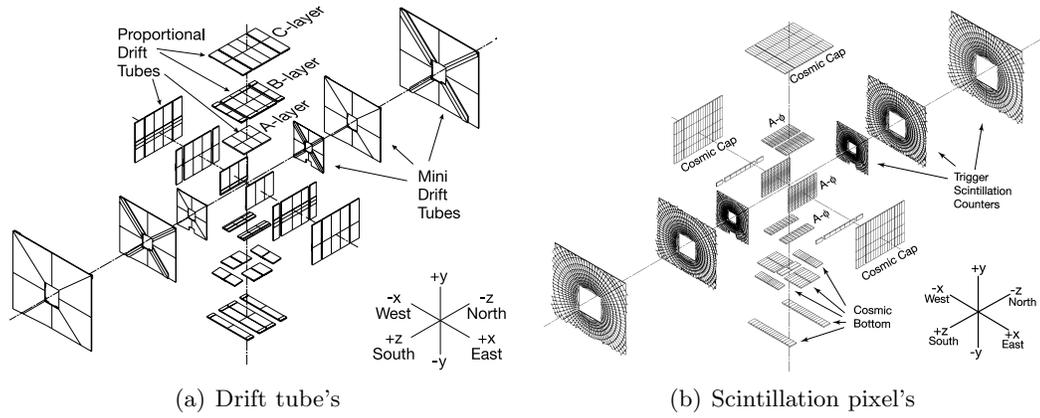
Figure 21: Muon detector sub-component layouts

3.3.8 Trigger

As stated previously, bunch crossings occur every 396 ns. Taking the abort gaps into account, there is an effective collision rate of 1.7 MHz. It is unrealistic to record every collision due to the size of the events and the deadtime that occurs during the readout process. Apart from timing and size arguments against recording each event, an effort should be made to store only those events that will benefit an analysis. DØ has implemented a three level trigger system to aid in selecting events that contain interesting physics and has implemented it in such a way that the final recording rate is within the capacity of the recording hardware.
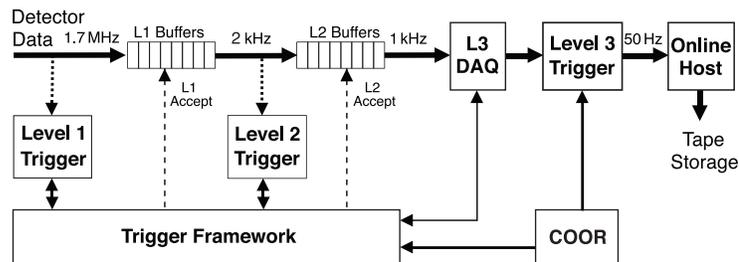
Figure 22: A schematic overview of the DØ trigger system [29].

Figure 22 shows an overview of the full trigger system [29]. A trigger is defined as a unique combination of level 1, 2, and 3 requirements. For an event to be selected for recording on a specific trigger, it must pass each of these requirements. Each level of

the trigger system is implemented in a different manner to optimize the performance of that level.

The level 1 trigger system is designed to reduce the initial collection of collisions to approximately 2000 events per second. This is accomplished by passing data through a series of hardware-only triggers designed to select events consistent with the various DØ physics programs. The inputs to the level 1 trigger system are the calorimeter, the central track trigger, the muon system, and the forward proton detectors. At this stage in the triggering system, actual physics objects have not been reconstructed. The decisions to either keep or discard an event are based primarily on the transverse energy for the individual calorimeter towers and transverse momentum for the tracks that are reconstructed in the tracking system and delivered via the central track trigger system. If an event passes any of the level 1 conditions, a level 1 accept signal is issued to the trigger framework and the data continues into the level 2 trigger system.

The level 2 trigger takes the data that was used in the level 1 trigger and adds the SMT data along with more comprehensive muon system information. Unlike the level 1 trigger, which was hardware only, the level 2 system combines a hardware system with a custom reprogrammable series of processors to allow for programmable triggers. Due to this unique combination of hardware and software, the level 2 system can actually reconstruct pseudo-physics objects in order to make trigger decisions. The level 2 system was designed to accept inputs at a maximum rate of 10 kHz and perform a rejection to reduce this to ~1 kHz. Since the level 1 rate is 2 kHz, the level 2 system is not near its design capacity.

If an event passes both level 1 and level 2 requirements, a level 2 accept signal is propagated to the trigger framework and the data for the event is moved into the level 3 trigger. Following the general trend from level 1 to level 2, the level 3 trigger is a fully programmable trigger. The level 3 trigger system is a collection of commodity computer systems that performs a coarse reconstruction of the event based on the actual offline reconstruction code. With full physics objects available, a large assortment of different

and complex trigger requirements can be placed on the data. The level 3 trigger system is designed to reduce the input rate of 1 kHz to $\simeq$100 Hz. This final rate is dictated by the capacity to write data to tape.

4 Physics Object Reconstruction and Analysis Tools

4.1 Analysis Workflow

The ultimate goal of a high energy physics analysis of this type is to determine the production cross section for a particular subset of physics processes. This is accomplished by determining the number of events in a given data set that correspond to the physics processes under investigation. Figure 23 outlines the generalized steps of a typical analysis flow. This section will motivate and describe each step in an effort to allow the reader to better understand the historical survey that follows in addition to the analysis presented in this paper.
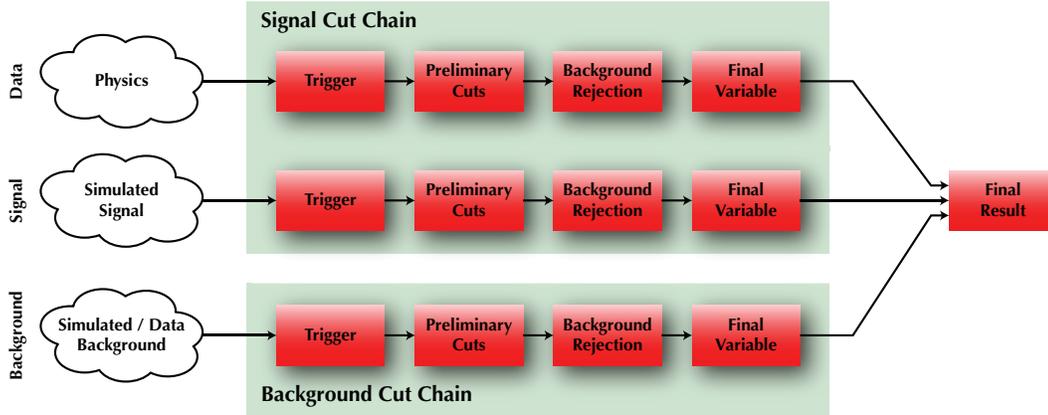
Figure 23: A generalized flow of a high energy physics analysis

To begin it is necessary to define three different sets of events: the data, the signal model, and the background model. Each set of events has a specific purpose in the analysis. The data set is a collection of unclassified physics events taken from the detector with an unknown fraction of signal and background content. It is from this data set that the experimenter will ultimately derive a result. The signal model, normally a set of Monte Carlo generated events, is a description of the signal process that is under investigation. The signal content of the analysis data can be estimated, $N_{S,\mathrm{est}}$, by scaling the signal model efficiency (ε_S) by the integrated luminosity ($\mathcal{L}$), signal process cross section (σ_S), and the signal process decay branching ratio (BR_S), shown in Equation

27.

$$N_{S,\text{est}} = \mathcal{L}\sigma_S \text{BR}_S \varepsilon_S \tag{27}$$

Where the signal model is built almost entirely from Monte Carlo simulated events, the background estimate can be derived using Monte Carlo, constructed using real data, or a combination of the two. Analyses using a background estimate constructed from data are said to be data-driven analyses. These three different data sets are represented in Figure 23 as three rows with the data at the top, signal in the middle, and the background on the bottom.

With these data sets defined, there are two main paths that each data set can traverse. The first of these is the signal cut path. This is a path representing a collection of data set modifications designed to favor the signal model. Both the data from the detector and the signal model are propagated though this chain. Opposing the signal cut chain, the background cut chain is designed to both favor the background processes and create an independent data set from either the signal or the detector data.

The first step is to understand the source data set that is acquired by the detector. There are a large number of collisions produced in the core of the detector each second. While each of these events represents a physics process, only a subset of these desired events are consistent with the physics goals of the experiment. In order to select these desired events, an automated selection tool called a trigger is designed and implemented. The triggering system has two main functions: selecting interesting events and to reduce the final recording rate to a manageable level consistent with the capabilities of the recording hardware. The trigger is designed to identify specific patterns of energy deposition in the different detector elements. Once this pattern is found, the event is flagged as an interesting event and then said to be triggered. In an experiment such as DØ with a rich physics program, there are many different triggers designed to target different physics processes. Once the data has been recorded it is beneficial to categorize

the data into subsets called data skims. The criteria to belong to a particular skim can be based on kinematic details of the physics objects or just the fact that it passed a particular set of triggers. As an example, the data skim used in this analysis has a loose requirement of at least three jets with a high transverse momentum.

At this point there has been no mention of the signal content of this data set. This is estimated by applying these steps to a set of Monte Carlo simulated events that represent your signal(s). By determining the efficiency of these simulated events along with the production cross section, the number of events in the data set due to the signal can be estimated. If you assume that the data represents the sum of the background and the signal events, the sensitivity can be determined by simply dividing the number of predicted signal events by an estimation of the signal error which is represented by the square root of the number of data events. If the number of data events is large enough, a gaussian distribution can be assumed and the error in the data can be represented by the square root of the number of data events.

Since the events are not conveniently pre-tagged as background and signal, the analyzer is required to pare the data set down to a manageable level of events that are biased to be signal like. These events are then called candidate events, since they could be either signal or background, and it is this data set that is used to determine the final result.

In order to increase the sensitivity of the analysis, a set of cuts, called preselection cuts, are applied in order to remove background events while retaining a larger fraction of the signal events. An example of a preselection cut is a minimum transverse momentum requirement for the jets in the analysis. By setting the threshold value appropriately it is possible to remove an entire class of physics events from the analysis resulting in a dramatic increase to the sensitivity. These preselection cuts are, to first order, designed to characterize the final state of the physics process. In this analysis, the preselection cuts focus on selecting a four-jet final state where two of the jets are identified as b quarks. Additional cuts can be added at this stage with the sole motivation of increasing

42

the quality of the reconstructed objects. An example of this style of cut requires the z position of the primary vertex of the event, the physical location along the beamline at which the collision occurred, to be within 35 cm of the fiducial center of the detector. This ensures that the objects created in the events have an ample amount of tracking volume available in order to gain a good momentum measurement.

With the preselection applied and the sensitivity improved, a series of additional cuts can be applied on the event variables or quantities derived from these variables such as a decision tree discriminant. As with the preselection cuts, the focus of this set of cuts is to further increase the sensitivity. At the point of the analysis where the sensitivity from the cuts has reached a plateau, the final result can be calculated by measuring the overall signal efficiency, background estimation, and numbers of events after the cuts.

4.2   Analysis Figure of Merit

A quantity to compare one analysis with a different analysis, or even to determine the performance of an analysis at different stages, is called the sensitivity. This is simply the ratio of the predicted signal content with respect to the error on the number of signal events. An estimate of this quantity is the predicted number of signal events divided by the square root of the number of data events at that point. This is a valid estimate for signals with a low cross section and a significantly higher number of data events than predicted signal events. While this estimate is used throughout this analysis, the final sensitivity is the number of predicted signal events divided by the error on the derived signal content and is not simply divided by the square root of the number of data events after the final cuts.

4.3   Physics Object Reconstruction

In order to perform an analysis at DØ, the raw data from the detector is passed through an application called `D0Reco` that reconstructs the full event. This full event recon-

struction finds and determines the properties of physics objects and places them in a convenient format that the analyzer can utilize. For this analysis, the physics objects that are important are: tracks, calorimeter jets, and muons. While it is the calorimeter jets that are used to generate the final quantities, the tracks are needed for b-jet identification and the muons are used for corrections to the jet energy scale.

4.3.1   Tracks

The tracks, which show charged particle trajectories in 3D space, are determined from hits in the tracking detectors. DØ uses a combination of two different track reconstruction algorithms. These two algorithms are applied to the data and the duplicates removed. The remaining collection of tracks are then used for the remainder of the analysis. The first algorithm, called the Alternative Algorithm [34], works by finding an initial track from hits in the first layers of the SMT. Other clusters are added to the track if they satisfy a series of conditions. It is possible for the algorithm to have several different candidate tracks sharing an initial set of tracking clusters. In this case the track with the least number of misses is favored over the others. A miss is defined as a vacant cluster in the track path where a detector element exists.

The second algorithm, called the Histograming Track Finder [35], determines a set of track parameters by looking at all the possible tracks that would intersect a series of clusters. By using a mathematical tool called the Hough Transformation, all circles of curvature ρ and angle ϕ can be mapped into a single straight line in $\rho-\phi$ space. For each cluster this is accomplished and the lines are added to a 2D histogram. Since clusters coming from the same track will have the same ρ and ϕ parameters, the histogram bin corresponding to these track parameters will exhibit a peak. This peak can be extracted from the histogram and the clusters that are contained in that peak will belong to the same track. Once the tracks have been found from this algorithm, a kalman filter is applied to remove noisy or otherwise bad tracks from the collection. With this technique the tracking efficiency is found to be $\simeq 93\%$ while $\simeq 3\%$ of the found tracks are fake [36].

4.3.2 Jets

The majority of this analysis is based on jet based quantities derived during the jet reconstruction. The jet reconstruction algorithm used in this analysis is called the Run II Cone Algorithm [37]. The process for finding the jets can be divided into three major procedures: generation of proto-jets, application of the midpoint algorithm, and finally, a splitting and merging procedure. To begin, a simple cone algorithm is applied to the raw calorimeter data in order to derive a collection of preclusters composed of reconstructed calorimeter towers. These clusters, ranked in transverse momentum, are used to form proto-jets. A proto-jet is simply a collection of preclusters that fall into a simple cone. The proto-jets are started from a seed precluster that is greater than half of the simple cone radius. An iterative process scans the region for other preclusters that should be included in the proto-jet. Once all the proto-jets have been found, the midpoint algorithm is applied. The midpoint algorithm is used to suppress jets that come from soft radiation. The procedure simply finds the midpoint between two near proto-jets and then uses that as a seed location. The above proto-jet determination algorithm is then applied to this midpoint. The list of proto-jets from the seeds and the midpoints is then passed into the merging/splitting routine. The goal of this routine is to ensure that each tower belongs to one jet and one jet only. Towers that are shared between two proto-jets are located and the total transverse momentum is measured for the entire overlapping region of the two proto-jets. If the total transverse momentum sum in this region is greater than 50% of the transverse momentum of the second proto-jet, the two proto-jets are combined into a single jet. Conversely, if the sum in the shared region is less than 50%, then the towers are split and assigned to the nearest proto-jet such that there are no more shared items.

4.3.3 Jet Energy Scale Correction

The energy that is determined for the jet in the jet reconstruction does not correspond to the actual physics energy that was present during the final state of the collision. This

raw calorimeter energy (E_Raw) is corrected back to the original true physics energy (E^Corr) though the jet energy scale correction, or JES [38]. (The form of the correction is shown in Equation 28.)

$$E^\text{Corr} = \frac{E^\text{Raw} - O}{F_\eta \cdot R \cdot S} \tag{28}$$

There are several different components to this correction and I will briefly discuss each one. The first and largest single correction is the energy offset correction represented by the O in Equation 28. The O embodies the background energy in the detector not related to the actual hard scatter. This is due to other scattering events in that particular bunch crossing. This correction is determined by looking at two data sets. The first of these uses a set of data that was triggered by the luminosity monitor but did not trigger a physics event and is called the minimum bias, or min-bias, data set. In the second data set, called zero-bias, the events are generated at each bunch crossing. The min-bias data set is correlated with background physics events while the zero-bias is related to the residual energy in the calorimeter and other detector-level effects. Since the min-bias data set contains the background multiple collision data, the offset correction is luminosity dependent. This dependence is carried in through the higher number of primary vertices when the luminosity is increased. The first step in determining the offset correction is to observe the level of energy from the min-bias data set taken at a specific luminosity and subtract from it the energy from a min-bias data set with exactly one primary vertex. The energy of the jet with the min-bias energy subtracted now represents the energy of a single hard scatter event. This does not contain the detector-level effects such as noise due to the uranium. The zero-bias energy, representing these detector effects, is then added in and the new total is now called the offset correction.

The denominator of Equation 28 contains three terms, F_η, S, and R. These terms represent the response of the detector. While the calorimeter was designed to be com-

pensating, there is still a measurable amount of difference between the EM and the hadronic response. The R quantity is the absolute response correction and is computed using the Missing E_T Projection method. For the absolute response, a collection of $\gamma + \text{Jet}$ events in the central region of the detector is studied. Using the kinematics of the event, the difference in the hadronic response (seen with the jet) can be compared with the response of the photon which is known. From this a correction F_η can be determined. Due to instrumentation in the detector as a function of η, a relative response, also called the η-intercalibration, is required to be performed. This is done in the same way as the absolute response with the exception that the photon is required to be observed in the central calorimeter and the jet is then observed as a function of η. In addition to $\gamma + \text{Jet}$ events, the relative response uses dijet events for the high energy regime.

The last component in the denominator in Equation 28, S, is the showering correction. It is possible for energy related to the jet to be absent in the reconstruction due to it being geometrically outside of the cone definition. In addition to this aspect, the opposite can be true and energy from other jets can be associated with the jet since it is in the reconstruction cone. This is corrected by studying the energy density profile of the jets in Monte Carlo and comparing the energy due to the particles that generated the jet and the energy due to all of the particles in the jet. This is done for Monte Carlo that is and is not passed through the detector simulation. Observing the difference in these two cases allows for a correction due to only the detector-level effects and is called the showering correction.

4.3.4 Muons

The muons used in this analysis, for the jet energy correction, are found by selecting tracks from the muon system that meet certain quality criterion. In addition to requiring a match to a track found in the main tracking volume, there are three basic requirements in order to be selected as a muon. The first is that there are at least two hits in both the

drift tube and the pixel detectors in the A layer. The second condition requires three drift tube hits in either the B or the C layer, while the final requirement is at least one pixel hit in either the B or the C layer. An object that meets these criterion is called a Medium NSeg3 muon [39]. The muon transverse momentum resolution was found to be $\simeq 10\%$.

4.4 Analysis Tools

As seen in Figure 6 on page 14, the production cross-section for the higher mass Higgs falls to a point that the search sensitivity becomes prohibitively low due to overwhelming backgrounds. The application of mathematical multivariate discriminating tools and b-tagging has given experimenters the ability to reduce these backgrounds improving the sensitivity of the analysis in question. Multivariate discriminating tools use a collection of variables in concert to establish the classification of an event as either signal or background. While this can be done by performing a cut on a single variable, it is the correlation between several different variables in the event that produces the effectiveness of this tool. In the search for heavy particles such as the Higgs boson and the top quark, a common theme in the particle decay is the large branching fraction to include a b quark. It is possible to distinguish a jet, a collimated deposition of energy in the detector, produced by a b quark, and to a lesser extent a c quark, due to the extended lifetime of the B or D mesons produced in the hadronization. The extended lifetime of a B meson, on the order of a picosecond, is directly related to the long time scale of the weak interaction upon which the B meson is forced to decay [9]. The ability to identify which jets in an event are produced via heavy flavor quarks, called tagging, allows the experimenter to remove the jets that are produced via QCD production which are dominated by light quarks.

4.4.1 Boosted Decision Trees

Currently there are several different multivariate discrimination tools available to the experimenter. The Neural Net (NN) and the Decision Tree (DT) based tools fall under this category and have been heavily utilized. In addition to each of these tools being used individually, some analyses have begun to combine the NN and the DT together to further enhance the sensitivity through background reduction. The study presented in this paper is based on a variant of the Decision Tree called a Boosted Decision Tree and is implemented from the Toolkit for Multivariate Data Analysis, or TMVA, package of ROOT [40]. An introduction to the decision tree is presented followed by a discussion on the concept of boosting.

A decision tree assigns a value, called the discriminant, between -1 and 1 to an event based on whether it is more signal-like or background-like. If the event represents a signal, the value that it is assigned tends to be closer to 1 whereas an event representing the background will be closer to -1. This pattern in the decision tree output can be seen in Figure 24 where the background sample (shown in red) is shifted lower than the signal sample (shown in blue). Before the decision tree can be utilized, it must be trained on a set of known signal and background samples that are well separated through a collection of discriminating variables. If only one variable is provided, the decision tree would act as a single well tuned cut. With the addition of more variables, the decision tree becomes a very powerful tool taking into account the individual shape differences spanning several different quantities. While both the decision tree and the neural net work in this fashion, a distinction needs to be made between the two with respect to the input variables. When choosing the input variables for a neural net, the analyzer must be careful to not choose variables that are correlated with each other. Doing so will damage the effectiveness of the neural net. This requirement is relaxed in the decision tree such that there is no concern over variable correlations [41]. A second distinction is related to the overall number of variables. If one chooses too many input variables for the neural net, the effectiveness will again be reduced. Like the correlation,

this is not a concern for the decision tree. By and large, the decision tree gives one the power of a multivariate discriminant without the concern about overhead related to the input variables normally required by the neural net [42].

Boosted Decision Tree Discriminant Distributons
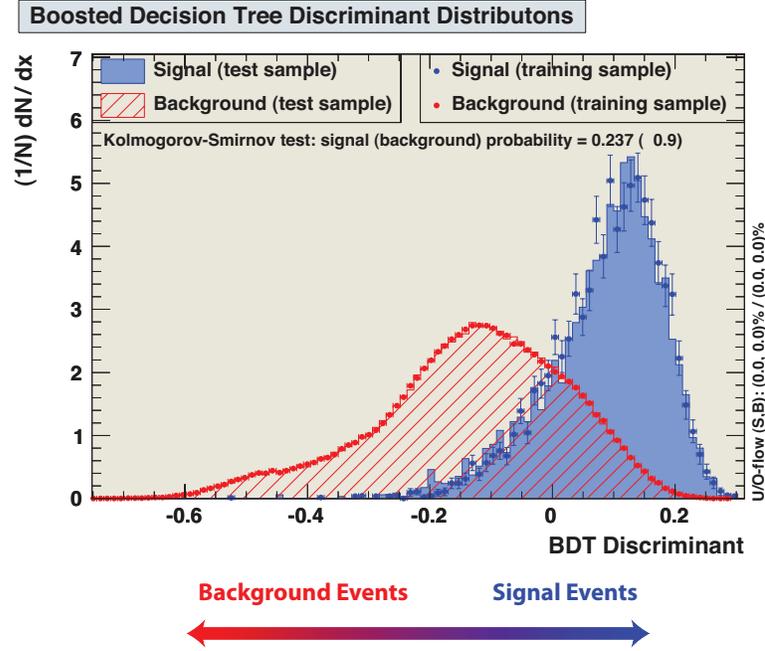
Figure 24: The decision tree discriminant distribution for a set of background (red points) and signal training events (blue points). The testing variants of the data sets are shown as the shaded regions of the same color.

The decision tree operates by determining the optimum set of successive cuts based on the input variables during the training phase [42]. Before the training of the decision tree is executed, the input samples are each split into three different equally sized samples. The splitting is performed by looking at three contiguous events at a time; the first is placed in a data set called training, the second in a data set called testing or validation, and the last in a set called distribution. The training and testing samples are used in the training and performance testing stages of the decision tree analysis while the distribution sample is used to evaluate the discriminant distributions and efficiencies after the training has been completed. Each of these data sets is independent and therefore, one avoids the possibility of an introduced bias. In the training phase,

the background and signal training samples are merged together into a single data set and associated with the initial or top most node, which is defined as the root node. At this point the purity of this set is defined as the number of signal events divided by the total number of events. When a cut is applied, this data set is split into two new data sets: one which passed the cut and one which did not. Each of these new data sets has associated with it a new purity. The goal is to increase the separation of signal and background events by studying how the purity changes with a given variable and cut. After each cut, the data is split into two different collections which are called nodes. These new nodes are sometimes called the children nodes while the original node is called the parent node. At each node a quantity called G_{ini} is defined which encodes information about the purity of the events associated with the node. Written explicitly in terms of this purity, p, and in the case where only two categories of events, signal and background, are defined, the G_{ini} is shown in Equation 29.

$$G_{ini} = 2p(1 - p) \tag{29}$$

From Equation 29 it is easy to see that as the purity approaches either zero or one, the G_{ini} approaches zero while it is at its maximum when the purity is 50%. In order to select the optimum variable and cut point to use, the difference between the parent G_{ini} and the weighted sum of the children's G_{ini}, Equation 30, is maximized. The weighting, p_i in Equation 30, is the fraction of events in that node from the total number of events in the parent node. In this way, if the majority of the events go to only one child node, that node will principally determine if the split should be used or not.

$$\Delta i = G_{ini}(\text{parent}) - [p_1 G_{ini}(\text{child 1}) + p_2 G_{ini}(\text{child 2})] \tag{30}$$

The process of choosing a variable and cut point is then iteratively applied for each of the new nodes from the previous splitting. The splitting procedure is suspended when one of two conditions are met. The first of these conditions, based on the improvement,

will halt the splitting if the improvement does not change by a significant amount. Splitting is also halted when the node to be split has less than a minimum amount of events available. For this study the minimum number of events in a node is 100. Nodes that are not split are called leaf, or terminal, nodes. The discriminant value assigned to an event is a function of the signal purity of the leaf node in which the event landed.
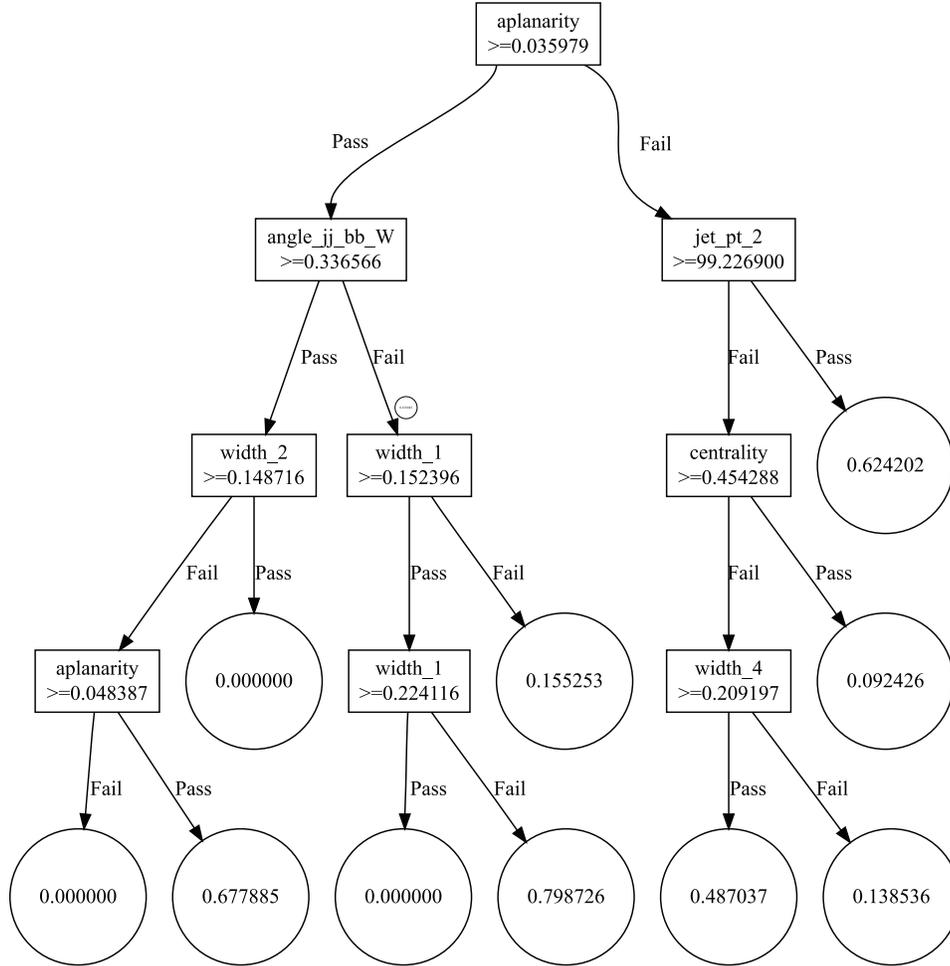
Figure 25: A subset of a trained decision tree. The rectangular boxes represent a rectangular cut on a single variable. The circular nodes represent the leaf or terminal nodes of the decision tree. The number in the leaf nodes is the unscaled signal purity.

A portion of a trained decision tree is presented in Figure 25. The full form of the decision trees can become quite complex spanning several hundred nodes over several levels. One must be careful to not over-train the tree. An over-trained tree will have

a quantization effect in the output discriminant. This quantization effect comes from having low statistics when calculating the purity of the leaf node. There are two methods of combating an over-trained tree used in this analysis. The first is to not allow it to become over-trained in the first place. This is accomplished by requiring a minimum number of unweighted events in the leaf node. This restricts the splitting from occurring and will provide a statistically correct leaf node purity. As stated previously, the minimum number of events in the leaf node was set to 100 events. The second method is the tree pruning technique. In this method, the tree is allowed to split until there is no significant improvement in the G_{ini}. Once the training is accomplished and the tree is over-trained, the pruning algorithm removes leaf nodes that are considered over-trained [42].

While there are several different pruning techniques defined in the literature, the procedure implemented for this analysis uses the weakest link cost complexity variant. This algorithm looks at each node in the tree and determines how effectively the tree below that node level is helping the signal / background separation. The unit of measure is principally based on what is called the resubstitution rate, denoted here as $R_{t,n}$ where t denotes the tree and n denotes the node under question and is shown in Equation 31.

$$R_{t,n} = 1 - \max(p_{t,n}, 1 - p_{t,n}) \tag{31}$$

If a node in a decision tree had a purity of 1, i.e. it contained only signal events, then the resubstitution rate would be equal to zero. This would be a perfect node, as long as there were enough events in the node for it to be statistically significant. If the situation were reversed and the purity were zero, the situation would be the same. The resubstitution rate attempts to quantify the amount that the node would have to have added to it in order for it to be a perfect node irrespective of the node class, signal or background. Up to this point, the resubstitution rate has been defined as a function of a single node in a decision tree. It is possible, and required, to define a new resubstitution

form that represents a full or partial decision tree. This is simply defined as the sum of the individual resubstitutions from each terminal node in the tree or subtree, as in Equation 32.

$$R'_{t,n} = \sum_{n=0}^{\text{NTerm}} 1 - \max(p_{t,n}, 1 - p_{t,n}) \qquad (32)$$

It is now possible to have two resubstitution rates defined for one node, the first for the node itself and the second defined for the subtree below it. If the two resubstitution rates are similar, the difference will be small and it will be understood that the portion of the decision tree below the node in question is not producing a large gain in performance. The difference in these resubstitution rates is not the final quantity that is used to determine which branches of the tree will be removed. This difference is rescaled by the number of terminal nodes in the subtree minus one. The functional form of the final metric used to determine which branches to remove is shown in Equation 33. Equation 33 can be seen as the terminal node density of a subtree that fails to improve the separation between signal and background. Due to the training algorithm of the decision tree, it is always true that this is a positive definite quantity and is a result of the split requiring an improvement in the G_{ini}.

$$\rho_{t,n} = \frac{R_{t,n} - R'_{t,n}}{\text{Num Term Nodes} - 1} \qquad (33)$$

The cost complexity pruning algorithm begins by determining the individual node resubstitution rates for all the nodes in the tree. As a second step, it calculates the subtree resubstitution rate for the tree below each node. With these rates determined and $\rho_{t,n}$ derived for each node, the algorithm removes the tree below the node with the lowest $\rho_{t,n}$. It is this characteristic that gives this variant of the cost complexity algorithm the name of "Weakest Link". This procedure is reapplied iteratively until the result would be to remove the tree below the root node or $\rho_{t,n}$ is larger than a threshold value set in the configuration.

54

TMVA has an additional algorithm to determine the optimum value of the threshold and is implemented by initially setting the threshold value to a negative number. In this optimization procedure, the training sample is split into two equal components, one for training and the other for validation. The system prunes, one at a time, the tree below the node with the weakest $\rho_{t,n}$ and then redetermines the performance of the tree with the validation sample at each step. This is repeated until the root node would be removed. The value of $\rho_{t,n}$ that corresponds to the highest performing tree is then chosen as the threshold value.

Each decision tree that was trained in this analysis was checked for overtraining by comparing the discriminator output of the training and the testing samples of the background and signal data sets. If the decision tree is overtrained, then the training and the testing sample will have different distributions and therefore different efficiencies. Specifically, in the case of overtraining, the training sample will have a higher efficiency [42]. The comparison of the training and testing samples for a trained decision tree can be seen in Figure 24. In this example, a look at both the background, shown in red, and the signal distributions, shown in blue, show good agreement between the training and the testing shapes and therefore this tree is not considered as overtrained.

It is possible for the decision tree to mis-categorize events. The information from the mis-categorization can be used to improve the performance of the tree through a technique called boosting. A leaf is flagged as either background or signal based on its signal purity. It is possible that background events ended up in a signal leaf node, or vice-versa. These mis-categorized events can be altered, by increasing the weight associated with the event, so that the training mode of the decision tree works harder to correctly categorize them. With the initial decision tree (t) trained, a new decision tree ($t + 1$) is trained with the entire data set along with the modified event weights for the mis-categorized events. This new tree is then added to the collection of decision trees, called the forest, with an overall weight (W_t) determined by the mis-classification rate. The final decision tree discriminant is then the weighted average of the discriminant

55

taken from each individual tree. This procedure, called adaBoost, can be recursively applied an arbitrary number of times and for this analysis it is applied 400 times [43].

The modifications to the event weight ($\omega_{i,t}$) and the tree weight (W_t) are based primarily on the mis-classification rate of the tree. The mis-classification rate is determined by calculating the weighted efficiency for the mis-categorized events. This is shown in Equation 34.

$$\text{err}_t = \left[\sum_{i=0}^{N_{\text{Events}}} \omega_{t,i} \times \text{isMisclassified}(t,i) \right] / \sum_{i=0}^{N_{\text{events}}} \omega_{t,i} \tag{34}$$

With the error on the tree determined, a new quantity, called α_t, is derived. Shown in Equation 35, this quantity combines the error rate of the tree with an additional parameter called β. The β parameter, set to the default value of 1.0 for this analysis, allows the experimenter to alter the boosting strength.

$$\alpha_t = \left(\frac{1 - \text{err}_t}{\text{err}_t} \right)^{\beta} \tag{35}$$

When the boosting algorithm is applied, the decision tree is given a weight, W_t, set to α_t and each event is checked to see if it was mis-classified. If the event was mis-classified, the event weight is modified using Equation 36 and the entire process is repeated to derive an additional decision tree to add to the forest.

$$\omega_{i,t+1} = \omega_{i,t} \alpha_t \tag{36}$$

The final decision tree discriminate for an event (D_i) is based on the response of the event through all of the different trees that were trained in the boosting algorithm ($D_{t,i}$) and the weights of the individual trees. The method of combination used in the TMVA implementation for this analysis is the weighted sum of the individual tree responses where the weighting is done on the overall tree weight (W_t) and is shown in Equation 37.

$$D_i = \frac{\sum\limits_t^{N\text{trees}} \ln(W_t) D_{t,i}}{\sum\limits_t^{N\text{trees}} \ln(W_t)} \tag{37}$$

One of the important questions that can be asked after a decision tree is trained is how the input variables ranked in performance. Which variable produced the strongest separation power? TMVA implements a variable ranking metric as suggested by Breiman [42]. In order to see the importance of a particular variable, the method uses the G_{ini} for each node that used that variable to discriminate between signal and background. In addition to the separating power at each node, the technique takes into account the total number of events in that node in an effort to give variables that separated a large amount of data more importance. The functional form of the importance for variable i is shown in Equation 38.

$$\text{Imp}_i = \sum_N G_{ini}^2 \left(\sum_{Events} \omega_{t,i} \right)^2 \tag{38}$$

The final importance assigned to a variable used in a decision tree is the fractional of the total importance which is found by summing all of the individual variable importances.

4.4.2 b-Tagging

Of critical importance for the ongoing Higgs boson searches is the ability to determine the flavor of the quark that produced a reconstructed jet in the detector. Analyses designed to look for the light Higgs boson will need to specifically look for two jets that both originate from a b-quark. The ability to add a b-tagging cut in an analysis with heavy flavor quarks in the final state will reduce the background considerably and will increase the analysis sensitivity. One of the important characteristics that separate the b-quark from the lighter quarks is the amount of distance it will travel before it decays. Upon generation, the b-quark quickly hadronizes into a B meson, the B meson is then

forced to decay via the weak force. It is this decay through the weak force that drives its long lifetime. In light quark jets most of the initial hadrons decay via the strong force to long lived particles such as π mesons. Before the B meson can decay, it propagates approximately 3 mm while D mesons only propagate approximately 2 mm, see Figure 26. The decay distance of the B meson is long enough for it to be well separated from the primary vertex. When the B meson decays, the daughter particles produce tracks which are found to not be consistent with the primary vertex and can be more strongly associated with a new vertex called a secondary vertex which corresponds to the B meson decay point [44].
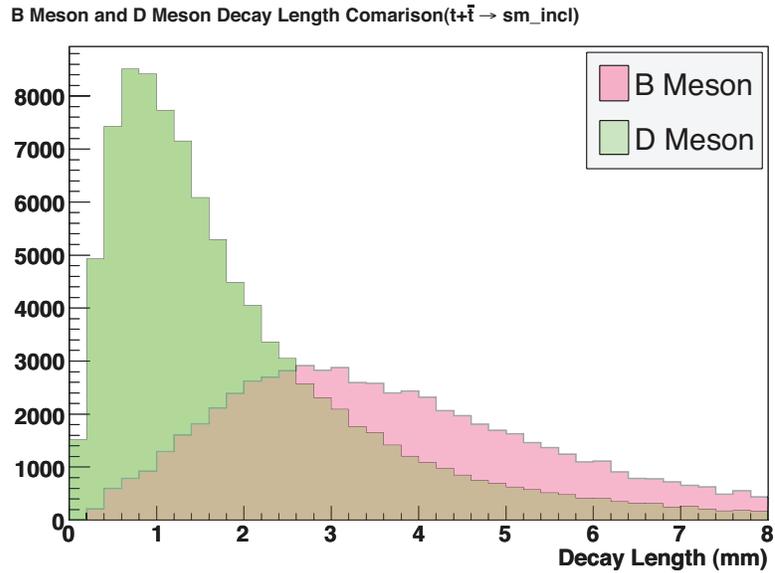
Figure 26: The distribution of decay lengths for both B and D mesons from an inclusive $t\bar{t}$ Monte Carlo sample.

Several algorithms have been devised to separate the b-quark jets from the jets resulting from lighter quarks. One of the earliest methods, called the soft lepton tagger, looks for a correlated muon in the jet cone. The B meson will decay to an electron or muon 20% of the time, 11% directly and 9% after a cascade decay through an intermediate particle [9]. A later method, called the Secondary Vertex Tagger or SVT, takes advantage of the tracking resolution and searches for the actual displaced vertex.

The algorithm currently employed in this analysis utilizes a neural net to combine the inputs from several of the previous methods. By using seven variables, the majority of which come from the SVT algorithm, a cut on the neural net discriminant was able to produce a b-tagging efficiency of $\simeq 50\%$ while maintaining a fake tagging rate of only 0.5%. A fake b-tagged jet is a light jet that has been inadvertently positively tagged as a jet from a B meson. Jets that are b-tagged with this cut are called tight neural net b-tags. A plot representing the fake rate vs the efficiency for the NN b-tagger in two different kinematic regions is shown in Figure 27 [45].
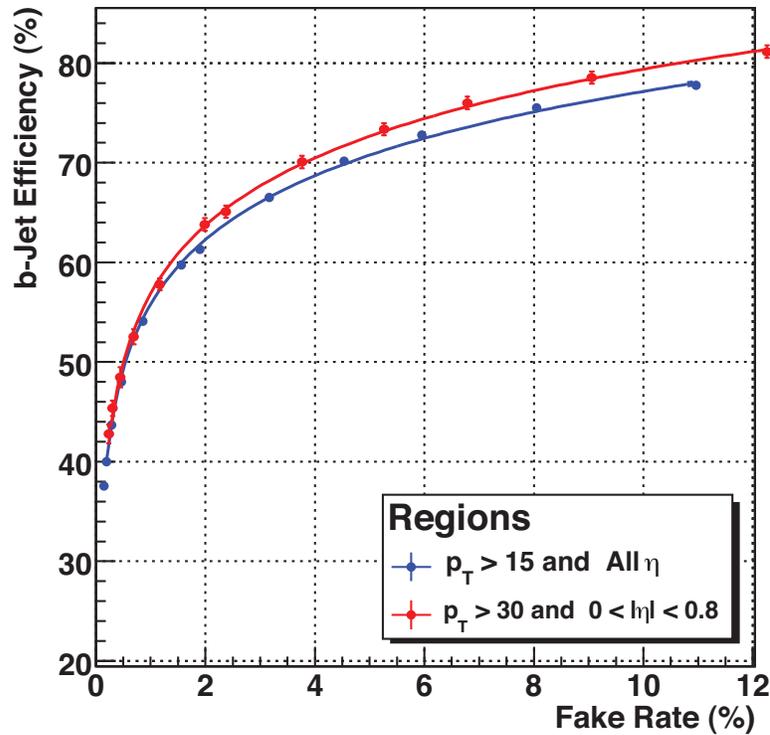
Figure 27: The performance of the NN b-tagging algorithm in two different kinematic regions [45].

5 Experimental Contexts

5.1 Historical Survey

In this section a review of both current and past analyses for related physics processes is presented. The first review will be on Higgs boson searches while the second review will be related to Technicolor.

5.1.1 Higgs Boson

The search for the Higgs boson has been an ongoing pursuit for the last several decades. The data for these searches were acquired primarily at two different laboratories: the Large Electron/Positron Collider (LEP) at the European Organization for Nuclear Research (CERN) in Europe, and the Tevatron, discussed in Section 3.1, at Fermilab near Chicago. The early analyses performed at the LEP accelerator were sensitive to low mass Higgs bosons. These searches culminated in an exclusion zone extending up to approximately 115 GeV/c^2. The Tevatron extended the available mass search range to include the region up to the theoretical upper limit. Searches at the Tevatron have recently been able to exclude a small mass range near 160 GeV/c^2. The remainder of this section will present a brief synopsis of the analyses performed using the data from these laboratories.

One of the most often quoted lower limits for the Higgs boson mass comes from the LEP collaborations: ALEPH, DELPHI, L3, and OPAL [46]. Published in the year 2002, the final limit takes into account approximately 2600 pb^{-1} of electron-positron collision data at a variety of center of mass energies extending to 209 GeV when the accelerator ceased operation in 2000 [47]. Production of the Higgs boson at the LEP accelerator was accomplished primarily through association with the Z boson. The e^+ and e^- collide producing a virtual Z boson that then radiates a Higgs boson leaving a Z and an H in the final state. With this production mechanism, the decays of the HZ final state can be divided into four different categories: hadronic, leptonic, missing

energy, and the τ channels. The decay modes are:

- $e^+e^- \rightarrow (H^0 \rightarrow b\bar{b})(Z^0 \rightarrow q\bar{q})$

- $e^+e^- \rightarrow (H^0 \rightarrow b\bar{b})(Z^0 \rightarrow \nu\bar{\nu})$

- $e^+e^- \rightarrow (H^0 \rightarrow b\bar{b})(Z^0 \rightarrow e^+e^-, \mu^+\mu^-)$

- $e^+e^- \rightarrow (H^0 \rightarrow \tau^+\tau^-)(Z^0 \rightarrow q\bar{q})$, and
  $e^+e^- \rightarrow (H^0 \rightarrow q\bar{q})(Z^0 \rightarrow \tau^+\tau^-)$

The τ channels are split into two different searches, the first of which has the Higgs decaying to two τ leptons and the second has the Z decaying to two τ's. The LEP accelerator was run in three different phases, each with different center of mass energies. The first run, called LEP I, ran from 1989 to 1995 with a center-of-mass energy of 91 GeV, the Z pole. After LEP I concluded in 1995, a small run called LEP 1.5 was taken between 1995 and 1996 at center-of-mass energies greater than or equal to twice the $W^\pm$ mass. The final run, called LEP II, was run at the highest energies achievable, up to 209 GeV, and contributed $\simeq 2500$ pb^{-1} of data. The lepton-based channels were heavily utilized during LEP I while the hadronic channels were used in the LEP II analyses. The hadronic analyses were not effective during the LEP I era due to prohibitive backgrounds [9]. Each of the four LEP experiments performed Higgs boson searches in each of the four major decay groups. Over the course of the data acquisition, the effective search range of LEP extended up to Higgs masses of 115 GeV/c^2. This ceiling on the mass search was driven by the kinematically allowed range based on the highest center of mass energy.

Because of the imposed kinematic limit and the preferred decay of the light H to $b\bar{b}$ pairs, the individual analyses expected only $b\bar{b}$ pairs from the Higgs boson decays. There is, however, an exception to this for the $H \rightarrow \tau^+\tau^-$ based study. The differences in the LEP analyses mainly stem from the decay of the Z boson. In the hadronic decay, the Z decays to two quarks. This decay channel for the Z boson captures $\simeq 70\%$ of

the decay space while the decay to two neutrinos occurs $\simeq 20\%$ of the time [9]. The $Z \to \nu\nu$ decay was used for the missing energy analysis. In the leptonic analysis the Z boson decays into an e^+e^- or an $\mu^+\mu^-$. For each of these leptonic Z decays, the branching ratio is approximately 3.4% [9]. The τ decays were specifically removed from the leptonic decay analyses due to the different procedure needed to study a final state with τ leptons. The backgrounds for these analyses consisted of three main processes: WW, ZZ, and two-photon processes. The WW and ZZ backgrounds were reduced through a combination of kinematic cuts and b-tagging while the two-photon processes were reduced with pre-selection cuts.

Each of the collaborations using the LEP accelerator (ALEPH, DELPHI, L3, and OPAL) maintained their own versions of the general analyses listed above. The 95% C.L. lower limit on the Higgs boson mass resulting from the LEP I analyses alone was found to be 65 GeV/c^2 [9]. With the increased energies of LEP II, these analyses produced a combined final lower limit on the Higgs boson mass of 114.4 GeV/c^2 at a confidence level of 95%. In addition to placing a new direct lower limit in the Higgs mass, the combined analyses observed an excess in the data at a Higgs mass of 115.3 GeV/c^2 [47]. This excess, which was originally seen only in the hadronic analysis from the ALEPH experiment, prompted an additional month of data acquisition [47]. After re-evaluating the analyses, it was found that the L3 experiment also observed an excess. Several systematic studies were accomplished in order to better understand the significance of the excess and the location of the 95% lower limit on the Higgs mass. The general level of systematic uncertainty was found to be $\simeq 5\%$ for the signal and $\simeq 10\%$ for the background samples. A fluctuation of 13% on the background would be required to reduce the excess while a fluctuation of 26% would be required to remove the excess completely and force the background to be consistent with data [47]. Hence, the excess seen in data is not consistent with a statistical or systematic fluctuation with that probability.

The limit setting technique, called the CL$_\mathrm{s}$ method, was devised by the LEP working

group [48]. Questions were raised regarding the validity of this technique when setting a limit so close to the kinematic limit of the accelerator. This was addressed in a specific systematic study of the limit setting method where a subset of the entire data set was taken excluding center of mass energies greater than 189 GeV. The limit calculation was performed at this point and it was found that there was no evidence for a systematic effect due to the limit setting technique near the kinematic limit [47]. An additional systematic related to the technical tools used to derive the distributions and final result was performed. This was accomplished by using a variety of different tools to determine the result. The limits and data excess were correctly reproduced with each set of tools with a relative variation of only 5% [47]. The collaboration used a conservative approach and used the tool set that gave the lowest limit on the Higgs boson mass of 114.4 GeV/c^2 at a confidence level of 95%.

The DØ and CDF experiments at the Tevatron extended the effective mass search range to approximately 185 GeV/c^2 due to the higher center-of-mass energy operating point. While the production methods at the LEP accelerator were primarily limited to associated production via intermediate Z bosons, the Tevatron allows for several production processes such as gluon fusion ($gg \rightarrow H$) and associated production via intermediate vector bosons ($q\bar{q} \rightarrow W^{\pm}/Z\,H$, shown in Figure 1). The production cross-sections for these processes are shown in Figure 6. While the gluon fusion production process has the highest production rate, it is difficult to use due to the large backgrounds. Using the gluon fusion process in the light Higgs boson mass regime (less than 135 GeV/c^2) results in a two jet final state composed of two b jets. The cross section for two jet production is extremely high and the fake rate for the b-tagging algorithm allows too many events to pass though. Due to this low multiplicity multijet background the light Higgs boson is not currently identifiable in the gluon fusion production channel. When the Higgs boson mass extends into the high mass region, the W^+W^- decay becomes preferred over the $b\bar{b}$ final state. Due to the reconstruction and identification efficiency of the detector for leptons, the $W^{\pm} \rightarrow \ell\nu_{\ell}$ decay provides a clean handle to

reject events that do not conform to the fully leptonic or semi-leptonic decay modes of the $W^\pm$ boson. The ability to cleanly isolate the leptons in the $W^\pm$ decays allow the gluon fusion process to contribute to the limits in the high mass region.

The primary production workhorse of the Tevatron studies is associated production via an intermediate vector boson. In this mode, two fermions from the p and $\bar{p}$ interact to form either a Z or a $W^\pm$. This then propagates until it radiates a Higgs boson leaving the same $Z/W^\pm$ boson along with a Higgs boson in the final state. The different analyses begin with this production mode and then separate into different final state configurations. With the higher range of the mass search window, the analyses can now be split into two major regimes, low mass searches or high mass searches. The low mass analyses mirror those from the LEP experiments due to the similarities in the final state while the high mass analyses search for a $H \rightarrow W^+W^-$ final state. As mentioned previously, the W^+W^- final state provides leptons in the decay process that can be used to reject background contributions which would otherwise make the analysis inconceivable. The low mass searches utilize, like the LEP analyses, the power of b-tagging to reject the backgrounds. Just as the different LEP experiments performed similar analyses, the Tevatron follows the same pattern. The two Tevatron experiments, CDF and DØ, for the most part, maintain similar analyses with similar preselection and background rejection techniques. Table 2 lists the different analysis channels along with the integrated luminosities used in the final Higgs boson production combination from the CDF and the DØ experiments.

The $WH \rightarrow \ell\nu b\bar{b}$ analyses look for events with two jets, an isolated lepton, and missing energy. The missing energy is the neutrino's signature in the detector. In addition to requiring the above conditions, an explicit veto is placed on events with more than one isolated lepton. With this preselection, the analyses split into several different types. The DØ experiment separates these into exclusive two or three jet analyses with a further splitting level occurring for one or two b-tagged jets. The CDF experiment, like DØ, splits the preselection into specific groups. Instead of grouping

64

| | Int. Luminosity (fb^{-1}) | |
Analysis Channel	CDF	DØ
$WH \to \ell\nu b\bar{b}$	4.3	5.0
$ZH \to \nu\nu b\bar{b}$	3.6	5.2
$ZH \to \ell\ell b\bar{b}$	4.1	4.2
$H \to WW$	4.8	5.4
$WH \to WWW$	4.8	3.6
$H + X \to \tau^+\tau^- + jets$	2.0	N/A
$WH + ZH \to jjb\bar{b}$	2.0	N/A
$VH \to \tau\tau b\bar{b}/qq\tau\tau$	N/A	4.9
$H \to \gamma\gamma$	N/A	4.2
$t\bar{t}H \to t\bar{t}b\bar{b}$	N/A	2.1

Table 2: The different analysis channels along with the integrated luminosities used in the Tevatron combination [49]. Entries with a N/A for the integrated luminosity indicate that an analysis was not present from that experiment in that channel.

on the number of jets, the CDF experiment groups its analyses on the lepton type and then a second level of grouping is on the type of b-tagging that was used. Both groups utilize a multivariate discriminant tool to reduce backgrounds and subsequently determine signal cross-sections. A similar analysis, the $ZH \to \nu\bar{\nu}b\bar{b}$, uses comparable preselection cuts to the previous analyses. The exception is on the lepton requirements and the veto is altered to remove events with any isolated leptons. Since the final state contains only two jets with no lepton requirement, there is a larger multijet background. This background is reduced by applying either a neural net or a decision tree tool. The final associated vector boson analysis in the light Higgs regime is the $ZH \to \ell^+\ell^- b\bar{b}$ channel.

The CDF experiment is also providing a $VH \to jjb\bar{b}$ analysis [50]. This analysis requires four jets, two of which are considered b-tagged jets. With the background estimated directly from data, the final variable is a matrix element signal probability discriminant.

There are several different high Higgs mass analyses being pursued by each experiment. The first to mention is the $WH \to WW^+W^-$ analysis. Both DØ and CDF use a selection based on lepton charges. Candidate events must have at least two isolated

leptons with the same charge. The experiment has split this preselection into three different categories; like sign electrons, like sign muons, and one muon and one electron with the same sign. The third W boson in the event is allowed to decay inclusively and no subsequent requirements are added. These analyses from each experiment are sensitive to masses up to 200 GeV/c^2. Both experiments also provide an analysis utilizing the gluon production mode and a final state consisting of two W bosons decaying to leptons. As was stated before, using the $b\bar{b}$ decay with this production mechanism is too difficult to perform due to the excessive multijet background.

In order to gain the sensitivity to exclude the production of the Higgs boson, the Tevatron experiments CDF and DØ must combine their efforts [51]. In November of 2009 combinations were performed on 90 individual search channels using 5.4 pb^{-1} of data producing a new exclusion region between 163 and 166 GeV/c^2 which can be seen in Figure 28. The plot shows the 95% C.L. upper limit cross section divided by the Standard Model prediction as a function of the Higgs boson mass. Since the y axis represents a ratio normalized to the Standard Model predicted cross section, when the observed limit (solid black line) extends below the unity mark the corresponding mass is then excluded at 95% C.L. The plot in Figure 28 also presents the expected 95% C.L. upper limits and is represented by a dashed black line. Figure 29 shows the log likelihood, LLR, values for the background (LLR$_B$) and the signal plus background (LLR$_{S+B}$) scenarios. The larger the difference between LLR$_B$ and the LLR$_{S+B}$ the higher the sensitivity.

As the energy increases with the transition from the Tevatron to the Large Hadron Collider (LHC), the production cross-sections are enhanced. This is evident in Figure 6 which illustrates the different production mechanisms for the Higgs specifically for the LHC.

So far we have considered direct searches for the Higgs particle, however, there is at least one other important technique for experimentally constraining the Higgs mass. The mass of both the W and the t quark depend on the mass of the Higgs through
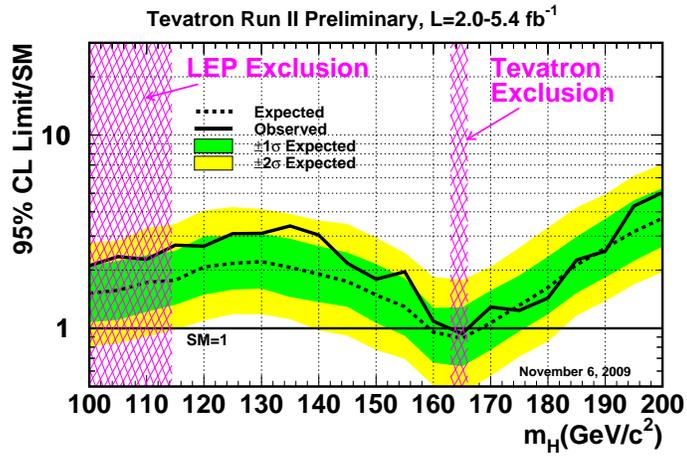
Figure 28: Current mass limits for the Higgs particle at the Tevatron. Please see text for full description [51].
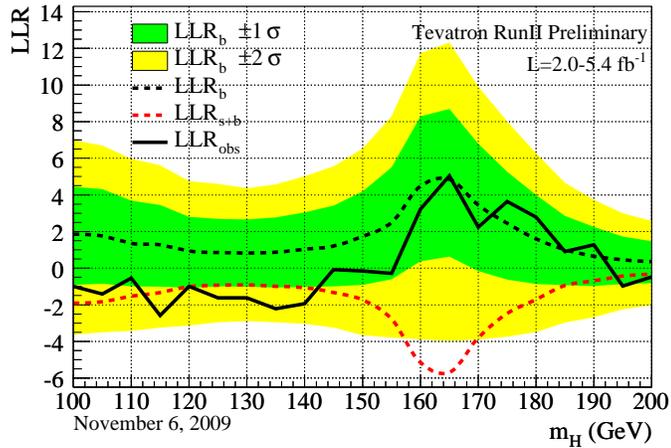
Figure 29: LLR distributions for the Higgs particle at a particular mass point at the Tevatron. The large distance between the background and the signal+background LLR distributions show a large sensitivity and therefore the ability to exclude this region as was seen in Figure 28. Please see text for full description [51].

radiative corrections. Due to this dependence, a plot, Figure 30, showing the mass of the top quark on the x-axis and the mass of the W boson on the y-axis will provide information about the allowed masses of the Higgs boson. It is seen that the red circle,

67

the combination of the direct and indirect evidence, barely overlaps the 117 GeV/c^2 mass Higgs boson line. By looking at the trending in the Higgs boson masses, getting higher as one moves to the lower right, this plot clearly prefers a lower mass Higgs boson. Although the dependence is weak, it is still an important contribution to the Higgs mass bounds.
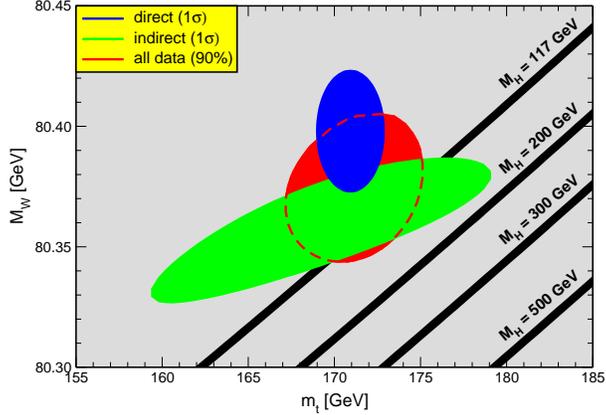
Figure 30: Current results of the W and t quark masses and their relationship to the logarithmic dependence of the Higgs boson mass [9].

In addition to the top quark vs W boson mass plane, it is possible to perform a fit to all of the electroweak parameters to determine the preferred mass of the Higgs boson. The LEP electroweak group performs this fit and provides the $\Delta\chi^2$ distribution which is shown in Figure 31. This plot presents the 95% C.L. upper limit on the Higgs boson mass at 185 GeV/c^2 by finding the $\Delta\chi^2 = 2.7$ point. If the exclusion regions are taken into account and only the non-excluded regions are used, the 95% C.L. upper limit on the Higgs boson mass becomes 185 GeV/c^2.

5.1.2 Technicolor

The WZ and VH systems all-hadronic decay yields a final state of four jets with two b-tagged jets. The production of a Technipion, π_{TC}, in association with a hadronically
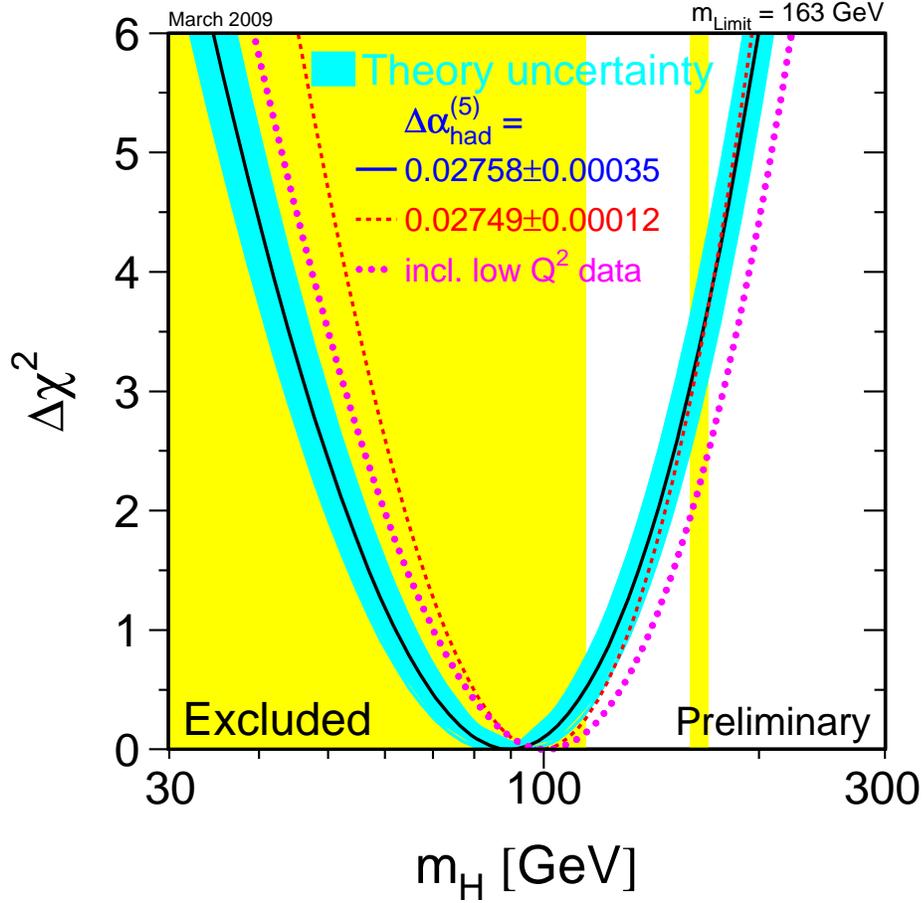
Figure 31: The $\Delta\chi^2$ from the least squares fit of the Higgs boson mass to all of the experimentally determined electroweak parameters [1].

decaying W boson also gives this same final state. The $W^{\pm}$ boson decays hadronically into two light jets while the Technipion decays to the heaviest kinematically-available fermions. In the case of the neutral Technipion, this is a $b\bar{b}$ state. Several searches have also looked for a charged Technipion that decays to a combination of a b and a c quark. No matter the charge of the Technipion in the final state, it is seen that b-tagging will be a strong tool to increase the sensitivity.

The analyses that have been performed at the LEP accelerator in the DELPHI ex-

periment and the two competing analyses at the DØ and CDF experiments principally utilize a leptonically decaying $W^{\pm}$ boson. The LEP result from the DELPHI experiment, published in 2001 using 452 pb^{-1} of data, excluded Technirho masses between 90 and 206.7 GeV/c^2 at a 95% C.L. irrespective of the Technipion mass that was studied. This was accomplished by using a combination of two separate analyses, one using a leptonic W decay and the second analysis using a hadronically decayed W producing a four-jet final state. This combination also allowed for the 95% C.L. exclusion of a Technipion mass lower than 79.8 GeV/c^2. [52] Of the two analyses accomplished at DELPHI, the semi-leptonic variant produced the lowest systematic uncertainties at 10% for the background and 2% for the signal with the largest single systematic error coming from the lepton ID. The four-jet mode increases these systematic levels and has an 11% background systematic with a 5% signal systematic uncertainty. The largest source of systematic error for the all-hadronic analysis is the b-tagging.

More recently, DØ has published an analysis, in 2007, with 388 pb^{-1} of data, that utilizes a leptonic decay of the $W^{\pm}$ [53]. In addition to using the b-tagging as a technique to increase sensitivity, a neural net multivariate discriminator was used in one version of the analysis. Of the two versions of the DØ analysis, the NN version provided the greatest exclusion zone in the $M(\pi_{\mathrm{TC}})$ vs $M(\rho_{\mathrm{TC}})$ plane. The backgrounds for this analysis range from W/Z+jets, both single top and top pair production, and multijet processes. The multijet processes present a background in the semi-leptonic analysis variant by having one of the jets fake a lepton. The NN version of this analysis presented a 10-20% background systematic with a 20% signal level systematic. The largest contribution to the signal systematic error is the jet energy scale fluctuation which has a magnitude of 11%. The results of this analysis can be seen in Figure 32.

The most recent publication from the Tevatron for a Technicolor search is provided by the CDF experiment utilizing the largest data set to date of 1.9 fb^{-1} [54]. This analysis, which looks for the $\rho_{\mathrm{TC}} \to W^{\pm}\pi_{\mathrm{TC}} \to \ell\nu b\bar{b}$, utilizes a combination of different b-tagging schemes to improve the sensitivity. In addition to the different b-tagging

schemes, the CDF experiment introduced a new variable, $Q = M(\ell\nu b\bar{b}) - M(b\bar{b}) - M(\ell\nu)$. Since the variable Q is a difference in invariant masses, a large portion of the systematics are reduced.

In addition to the reduction in systematics, the resolution on this quantity is improved over using just the individual invariant masses $M(b\bar{b})$ or $M(\ell\nu)$. The analysis backgrounds are shared with the previous analyses having W/Z+Jets and non-W multijet production as the largest contributors. Still present, but at a smaller level, are the single top and top pair production, diboson production, and $Z \to \tau\tau$. The systematics on this analysis are the largest of the three present analyses with 30% on the W+jets backgrounds from the heavy flavor determination alone. The signal level systematics change due to the b-tagging scheme in use and range from 4.7% to 17%. Using a 2D maximum binned likelihood technique in $M(\pi_{\text{TC}})$ vs $M(\rho_{\text{TC}})$, this analysis presents the largest exclusion region to date from both the Tevatron and the LEP complexes and can be seen in Figure 32.
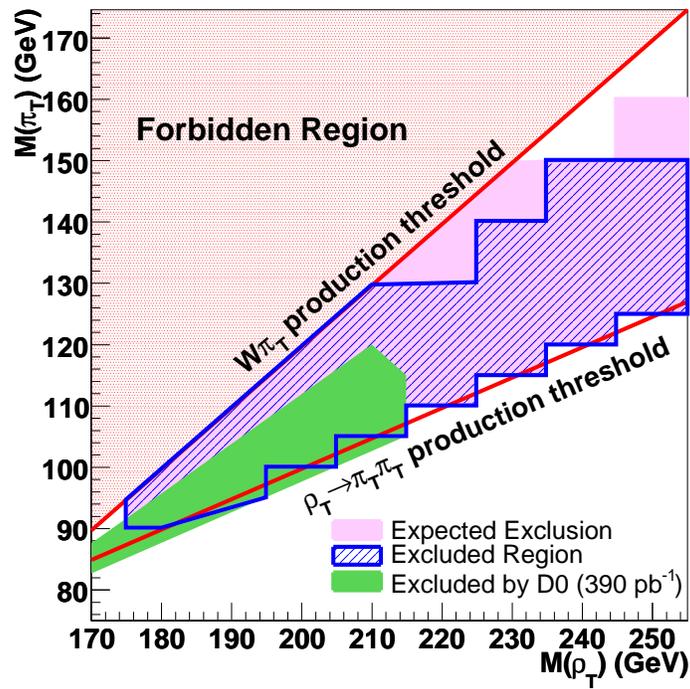
Figure 32: The CDF and DØ Technicolor exclusion regions in the $M(\pi_{\mathrm{TC}})$ vs $M(\rho_{\mathrm{TC}})$ plane [54].

6   Input Data Set and Preliminary Cuts Processing

This chapter details the requirements for the input data set along with the processing procedure for the Monte Carlo samples used for both signal evaluation and the background cross checks. The preliminary cuts are introduced and the performance after these initial cuts are also presented.

The input data set for this analysis is derived from the global DØ data set acquired during the RunIIa phase of the Tevatron between the years 2002 and 2006. This data set is called the p17 data set after the processing version of the reconstruction code. A loose cut on the event properties was performed requiring at least three jets with uncorrected energies above a threshold in addition to the requirement that they are in the central region of the detector to ensure high quality reconstruction. This collection of data, called the 3JET skim, is used in several analyses currently being performed at DØ. The specific set of cuts that define the 3JET skim are the following (it should be noted that the transverse momenta are observed before any corrections to the jet energy):

- 1 jet $p_T \geq 20$ GeV/c

- at least 3 jets $p_T \geq 15$ GeV/c

- each of 3 jets has to have $|\eta| < 2.6$

The cuts defined for this skim yield a total of 115.8M data events. Both the data and the simulated Monte Carlo samples are processed in order to remove events that correspond to improperly measured instantaneous luminosity or bad data quality flags that are assigned during the data acquisition phase. These instantaneous luminosity and quality cuts are called the data quality cuts. For more information on the input data set and the software processing chain, see Section A in the Appendix.

The total integrated luminosity was calculated from the 3JET skim data set with the additional requirement that the events pass an *or* of the four triggers listed in Table

Table 3: Trigger suite used for the analysis and the number of events that passed each criteria after the CAFe skim cuts.

Trigger	Trigger List	Luminosity [pb^{-1}] (Total Contribution)	Number of events (Total Contribution)
3JT15	v8-v10	44.4 (4.4%)	4.0M (12.7%)
3JT15_PV	v11	60.1 (6.0%)	3.5M (10.9%)
3J15_2J25_PVZ	v12	196.0 (19.5%)	15.5M (48.5%)
JT2_3JT15L_IP_VX	v13-v14	707.1 (70.2%)	8.9M (28.0%)

3. These triggers are shared between this analysis and an additional analysis using the Minimally Super-symmetric Standard Model (MSSM) [55]. The primary component shared between the triggers listed in Table 3 is a requirement for three jets with a significant amount of p_T in the event. The 3JT15 trigger used in version 8 though version 10 of the trigger system required 3 jets inside of an η of 3 with an uncorrected p_T of 15 GeV/c. The 3JT15_PV and 3J15_2j25_PVZ triggers used in versions 11 and 12 of the trigger system required that two of the jets have an uncorrected p_T of at least 25 GeV/c and the z coordinate of the primary vertex position be within 35 cm of the center of the detector in addition to the original three jets with an uncorrected p_T greater than or equal to 15 GeV/c. The final trigger used in this study is the JT2_3JT15L_IP_VX trigger which was used in versions 13 and 14 of the trigger system. This final trigger maintained the same requirements as the 3JT15_PV, but, added the requirement that the probability of the event to not have a b jet to be less than 5%. With these trigger requirements, the 3JET skim contains $\simeq$1007 pb^{-1} of data.

With the trigger and the 3JET skim cuts applied, the data is processed to apply data quality cuts, the jet energy scale correction, and the neural net b-tagging algorithm using the tight operating point [45]. The b-tagging at this operating point has a typical b-jet tagging efficiency of approximately 45% , a c-jet tagging efficiency of 10%, and a fake-rate (light jet tag efficiency) of around 0.4%. It should be stressed at this point that the b-tagging cut is not applied here, however, a boolean variable is stored in the custom skim for later use. In summary, the cuts, corrections, and algorithms applied

in the CAFe stage of the analysis are listed below. (CUT represents a cut while COR translates to an object level correction and ALG is the application of an algorithm on the objects in the event.)

- [CUT] Data quality cuts

- [COR] Jet Energy Scale

- [ALG] p17 Neural Net Tight b-tagging

The data set after the above cuts are applied will be referred to as the CAFe Skim.

The event properties and object kinematics are of interest after the preliminary cuts have been applied. Figure 33 shows the p_T distributions of the jets while Figure 34 shows the p_T distributions for the b jets and the non b jets separately. The η and ϕ distributions of the jets can be seen in Figure 35. The Jet Energy Scale (JES) corrections applied to each jet, for both the JES and JESMU algorithms, are shown in Figure 36 [56]. The JESMU correction is the same as the JES correction with the exception that special corrections are needed if the jet has a muon associated with it.
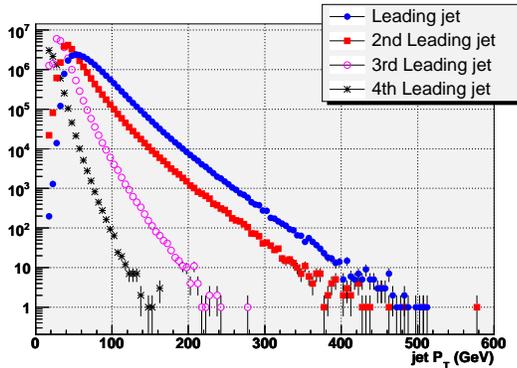
Figure 33: The p_T distributions of the leading four jets after the trigger and PV_z requirements. In order to be included in this plot, the jets need have a transverse momentum greater than or equal to 15 GeV/c, $|\eta| \leq 2.5$, and have a track-jet associated with it. The blue disks are the leading jets in order of p_T rank. The red squares and purple circles correspond to the second and the third jet in event, while it black asterisks represents the 4th leading jet.

75
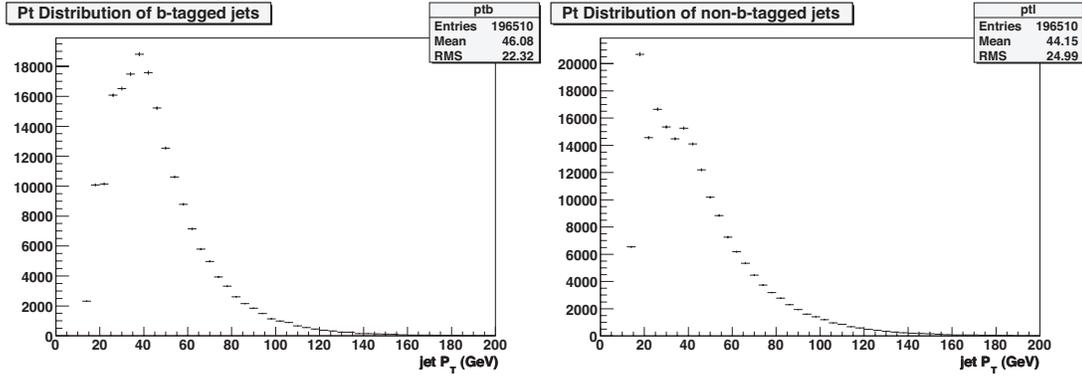
Figure 34: The p_T distributions of the b-tagged jets (left) and non-tagged jets (right) after preliminary cuts in the data set. (This includes all cuts including jet cuts and b-tagging cuts)
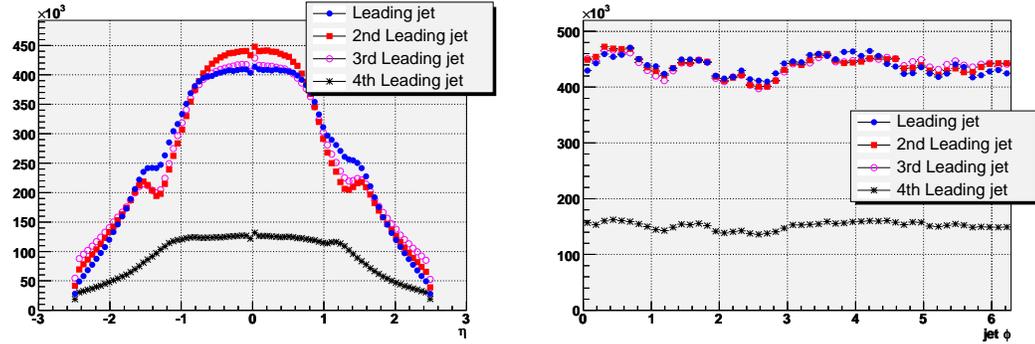
Figure 35: The η distributions of the leading four jets (left) and the ϕ distributions of the leading four jets (right) both after the trigger and PV_z requirements. Jets needed to pass the p_t requirement, η requirement, and taggability to be included in these plots.

6.1 Monte Carlo samples

The Monte Carlo signal samples are used to calculate the efficiency of the simulated events to pass certain cut criterion, which is needed to determine the final cross section. Other Monte Carlo sample sets were used to provide a handle on the background composition. However, it should be noted that the background used for the final result is derived solely from the data with no Monte Carlo input. Most of the event samples were from the DØ p17.09 full simulation chain which included a zero bias event overlay, p17.09.05 refixing, and the production of CAFe Trees [57]. The following generators were used to generate the Monte Carlo samples: PYTHIA 6.323 [21], ALPGEN 2.05 [58],
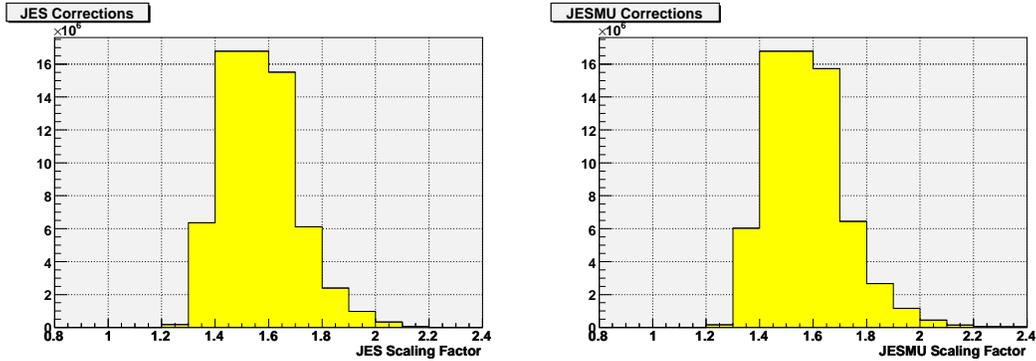
Figure 36: The factors for the JES (left plot) and the JESMU (right plot) corrections in the data jets found for jets after the trigger and PV_z requirements. Jets needed to pass the p_t requirement, η requirement, and taggability to be included in this plot.

CompHep 41.10 [59], and EVTGEN alpha-00-11-07 [60]. The standard GEANT simulation was used for the detector [61]. The list of Monte Carlo samples, including the DØ specific Monte Carlo request ID's, can be found in Table 21 in the Appendix - Section A.1.

A set of HW and HZ signal samples with different Higgs masses were generated using PYTHIA. Since this analysis focuses on the low mass Higgs boson, the generated masses were 115 GeV/c^2, 125 GeV/c^2, and 135 GeV/c^2. The H was forced to decay to $b\bar{b}$ while the W or Z was forced to decay into jj from the W. The W to jets branching ratio is $(67.60\pm0.27)\%$ while the branching ratio for the H decays depends on the generated mass [9]. For the 115 GeV/c^2, 125 GeV/c^2, and 135 GeV/c^2 samples, the branching ratios are 0.73, 0.61, and 0.44, respectively [15]. Combining the branching ratios for the Higgs boson decay and the $W^{\pm}$ decay, the total branching ratio for the $m_H = 115$ GeV/c^2 WH processes is approximately 0.49. Using the $Z \to q\bar{q}$ branching ratio of $(69.91 \pm 0.06)\%$, the combined $ZH(115)$ branching ratio becomes approximately 0.51.

A second set of Monte Carlo samples, from PYTHIA, was generated in order to accomplish the Technicolor search. In the Higgs samples, the experimenter needed only to provide the Higgs mass in order to generate the Monte Carlo sample. In contrast, each technicolor sample is characterized by two parameters: the mass of the technirho

and the mass of the technipion. The technicolor samples follow from a simple s-channel process where two initial state fermions combine to form a charged technirho particle. This technirho then couples to a $W^{\pm}$ and a technipion. For this analysis, the technipion was allowed to decay inclusively while the W boson was forced to decay into jets. In each of the technicolor Monte Carlo samples, the mass of the technirho was set to 240 GeV/c^2 while the technipion mass (similar to the Higgs boson) was allowed to take the values of 115 GeV/c^2, 125 GeV/c^2, and 140 GeV/c^2.

Each of these samples was generated with PYTHIA and decayed through EVTGEN [60].. The EVTGEN software is a package that takes the final states of the original process as generated by PYTHIA and correctly propagates these particles taking into account the decay properties. This is especially important for B mesons due to the complex nature of the cascade decays [9].

Other Monte Carlo samples with potential signals (resonances) include WZ, WW, and ZZ. The suite of top samples used in this analysis are s and t channel single top Monte Carlos along with a $t\bar{t} \to$ SM inclusive sample. The single top samples were generated using a combination of PYTHIA and CompHEP. The $t\bar{t}$ sample was generated from PYTHIA alone and EVTGEN was enabled to correctly decay the heavy quarks in these samples. The single top samples were forced to decay to an all-jet final state while the $t\bar{t}$ sample decayed inclusively within the Standard Model.

Other processes with a resonance that could contribute to the data sample include Wjj, Wcc, Wbb, and $\gamma Z \to \gamma bb$. These MC samples were generated with a combination of ALPGEN and PYTHIA. The samples with a W were forced to decay to an all-jet final state. For the γZ Monte Carlo, the $Z \to b\bar{b}$ decay is forced and $bbjj$ events can then be produced.

For this analysis, the dominant background is from QCD multijet production. While the valence quarks of the p and $\bar{p}$ are the up and down quarks, there are, of course, sea quarks with heavy flavor. Our background events represent the interactions of the sea quarks in a hard scatter and are composed mainly of simple s-channel processes. A

collection of Monte Carlo samples modeling these various QCD processes containing different heavy flavor combinations were produced for a Higgs boson analysis designed for the Minimally Super Symmetri Model (MSSM) [62]. Using a non-trivial technique, the MSSM analyzers determined the correct weighting for each of these samples to correctly represent this background. The weighting prescription from the MSSM study was used to compose a background which will be identified as the "QCD background" sample. This cocktail includes the heavy quark components bb, bbb, bbj, $bbjj$, bbc, cc, ccj, and $ccjj$. This sample is used for cross-checks and is not utilized as a primary background sample in any stage of this analysis. The background samples used in the various training and fitting stages of this study are derived solely from the actual data.

6.1.1 Jet Energy Scale and Jet Shifting, Smearing, and Removal

Just as the data had a jet energy scale (JES) correction (Section 4.3.3) applied to correct for the difference in the detector energy and the real physics energy, a similar correction is applied to the Monte Carlo. To parallel the data correction chain this jet energy scale correction was applied on jets with a reconstruction cone size set to $R = 0.5$, known as JCCB jets The JES correction was run with the inclusive muon mode enabled which utilized information from the muon, if present, to enhance the correction. It was found that the kinematic distributions, after the JES correction was applied in both data and Monte Carlo, did not match each other. In November of 2005 the Jet Energy Scale working group at DØ introduced a new correction to compensate for this deviation in distributions. This correction, identified as the Jet Smearing, Shifting, and Removal algorithm (JSSR), is originally described in reference [63]. This analysis uses the JSSR parametrization optimized for the final p17 JES correction and is discussed in reference [64].

In this JSSR algorithm, the Monte Carlo jets were initially smeared in p_T to compensate for the difference in the Monte Carlo and data resolutions. An energy shift was then applied which corrected for the relative differences in the jet energy scales between

Monte Carlo and data. The final step in the algorithm was to apply a cut of 15 GeV on the transverse momentum. This cut point was chosen by locating the beginning of the plateau in the reconstruction and identification efficiency plots for the jets. In this analysis all three corrections of the JSSR are enabled.

6.1.2   Trigger Simulation

The data set used in this analysis is required to pass one of the four triggers listed in Table 3. To have an effective description of the data from the Monte Carlo, this trigger list must be correctly simulated. The trigger simulation was originally used in a separate MSSM neutral Higgs boson search and is described in detail in Section 4 of reference [65]. The simulation took into account all three levels of the four triggers in the trigger list. At each trigger level for each trigger list, the trigger efficiency was parametrized as a function of the scalar sum of the corrected p_T for the jets in the event. A probability for the event to pass the trigger suite for each trigger list is calculated and stored in the event. The final trigger weight issued to the event is the luminosity weighted sum of these probabilities. An additional weighting factor was applied on the v13 and v14 trigger lists to compensate for an artificial bias due to the b-tagging implemented in the level 3 trigger term. The code from the MSSM analysis was used for this simulation.

6.1.3   3JET Skim Simulator

Th3 3JET skim cuts on the data need to be effectively simulated on the various Monte Carlo samples. In data, the cuts are applied to the uncorrected energies of the jets in the event. Due to the differences in the data and Monte Carlo jet energies at the uncorrected stage, the uncorrected form of the Monte Carlo jet energy can not be used for the cut. The data-like uncorrected energy for the Monte Carlo jets was determined by iteratively applying the data parametrization of the JES correction to the jet until the smeared energy returned by the JSSR algorithm was re-attained. The initial uncorrected energy for the jet is the MC-like uncorrected energy. At each stage of the iteration, the data-

like uncorrected energy was found by dividing the smeared MC energy by the data correction found from the JES calculation. This was then recursively fed back into the next iteration of the JES evaluation as the data-like uncorrected energy. After this algorithm was completed, currently after three iterations, the 3JET skim criterian were applied to the new collection of data-like uncorrected jet energies. Approximately 70% of the signal events that were processed though this algorithm passed the 3JET skim requirements.

6.1.4 Taggability Requirement

The data for this analysis required four "good" taggable jets. The good qualifier on this requirement is concerned with the kinematic requirements of the jets while the "taggable" portion of this requirement is related to the correlation of a track jet, which are jets constructed from the tracking system independent of the calorimeter, and the calorimeter jet. For a jet to be taggable the calorimeter jet must have a track jet matched to within 0.5 units of ΔR. The taggability rate can be measured in a data sample as the efficiency for taggable jets against all the jets in the sample.

It has been found that the taggability in Monte Carlo is not equivalent to the taggability of the data sample. Due to this, a correction to jet taggability in the Monte Carlo samples was required. The MSSM analysis faced a similar challenge and was able to implement this correction by directly altering the taggable state of the jets in a consistent way [62]. This is different from other analyses that apply an appropriate scale factor to the event to compensate. It was decided to utilize the taggability correction from the MSSM analysis.

For this technique, two sets of taggability scale factors were derived, one for light jets and the other for b jets. The light jet taggability scale factor was derived by dividing the taggability of the jets in the entire data sample by the taggability of light flavor jets extracted from the MSSM background Monte Carlo samples which strongly represents our largest background since it is composed of multijet MC samples. The b jet scale

factor was obtained in a similar way but using b-tagged jets from the same Monte Carlo samples and using data jets from the v12 trigger list. The b-tagged jets in the Monte Carlo sample were determined by applying the recommendation of the b-ID working group at DØ and used the neural net tagging algorithm at the tight operating point. Once the scale factors were derived, they can be used to determine whether the jet in question should change its taggability status. If the scale factor is less than one and the jet is originally taggable, then the jet is randomly toggled to the untaggable state based on the scale factor. In the event that the jet is originally untaggable and the scale factor is greater than one, then the inverse is done and the jet is randomly toggled to the taggable state. For more information on this algorithm, consult Section 3.2.3 in reference [62].

6.1.5 b-Tagging

The preliminary cuts of this analysis, optimized for the $H \rightarrow b\bar{b}$ decay, require two jets that are positively tagged as b-jets from the Neural Net Tight b-tagging algorithm [45]. While a large fraction of the different DØ analyses have chosen to use tag rate functions parametrized in p_T and η to determine which jets in the event are the b-jets along with the overall event weight, it was decided that this analysis would use the direct tagging technique. The direct tagging algorithm, in a broad sense, performs exactly what is done in the data on the Monte Carlo with the exception that the event is then weighted by a factor that corrects for the difference in the b-tagging rates between data and Monte Carlo. This factor, called the scale factor, is parametrized by p_T and η and determined by dividing the b-tagging rate in data by the b-tagging rate found in the Monte Carlo. The tagging rates are taken from the b-ID group though the standard b-ID tools available in the CAFe framework [57] and were extracted from these tools without the internal taggability since the taggability is processed separately.

The b-tagging weight applied to the event is presented in Equation 39 where the index i is for the N_b jets in the event that passed the Neural Net Tight b-tagging algorithm

and the index j is for the N_l jets that were not tagged. The weighting function depends on two quantities for the jet x, the scale factor (SF_x) and the tag rate function (TRF_x). The tag rate function is the rate at which real b jets are positively tagged as b jets and is normally parameterized as a function the jet p_T and η. There are two different types of tag rate functions that can be defined for a tagging algorithm, that determined from data (TRF) and that determined from Monte Carlo (TRF^{MC}). The tag rate function in Monte Carlo is normally higher than that found in data and can be scaled to that value found in data though the scale factor. The scale factor is defined as the tag rate function in data divided by the tag rate function found from the Monte Carlo.

$$W_b = \left(\prod_{i=1}^{N_b} \text{SF}_i \right) \left(\prod_{j=1}^{N_l} \frac{1 - \text{TRF}_j^{\text{MC}} \times \text{SF}_j}{1 - \text{TRF}_j^{\text{MC}}} \right) \tag{39}$$

This scale factor (W_b), shown in Equation 39, is split into two components. The first is simply a factor demoting the event since b-tagging in data is less efficient than in Monte Carlo. The product in the second component is the probability to not b-tag a jet in the data divided by the probability to not tag a jet in the Monte Carlo. The second product is over the jets that are not tagged in the event. This second factor tends to scale up instead of down and these two factors together create a balance. More information regarding the formulation of the b-tag weight can be found in reference [66].

6.2   Preliminary Cut Performance

The preliminary cuts are designed to remove the bulk of the background events in this analysis. These cuts, presented in Table 4, are designed to favor a four-jet final state with two b-tagged jets to further enhance the desired signal contribution (WH,ZH, WZ, and Technicolor(TC)).

Cut Name	Value	Description
$|PV_z|$	≤ 35 cm	z coordinate of the primary vertex position
3JETSkim	True	This cut ensures that this event would have passed the 3JET skim requirements that are imposed on the input 3JET data skim. This cut is only applied to the Monte Carlo samples since the data already has this cut imposed on it
Trigger	True	This cut ensures that the event would have passed the data trigger requirement
Jet p_T	≥ 15 GeV/c	All jets are required to have an energy-corrected transverse momentum greater than or equal to 15 GeV
Jet $|\eta|$	≤ 2.4	Require that the jets from the event be captured by the tracking volume for good reconstruction
# Jets	=4	Require exactly four good jets in the event
# b-tags	=2	Demand that two of the four jets in the event be considered b jets from the NN Tight b-tagging algorithm

Table 4: Preliminary cut descriptions

Figure 37 shows the distribution of the number of jets per event found in data before the CAFe skim as well as the distribution for various signal MC samples (described in the following section). An examination of Figure 37 shows that requiring exactly four jets in the event captures the highest efficiency ($\sim 40\%$) from the Monte Carlo.

The input data set, composed primarily of multijet background events, initially contains 80.3 M events after the data quality cuts. Applying the preliminary cuts to
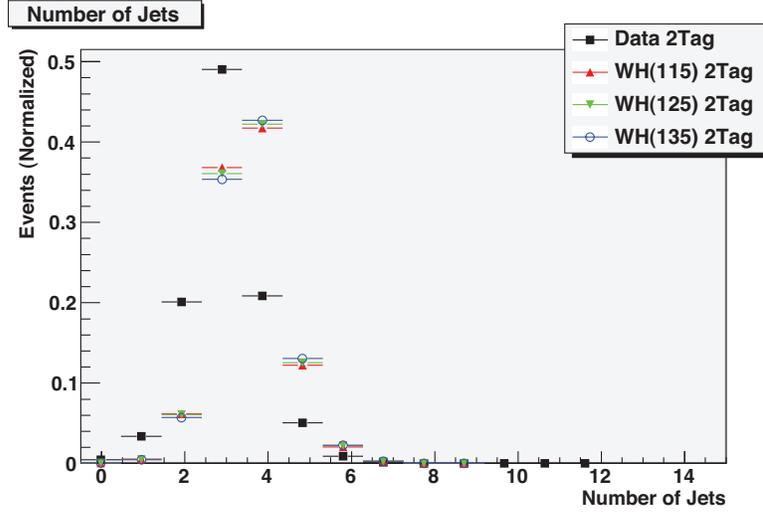
Figure 37: Number of jets per event for the 3Jet skim of the data compared to signal Monte Carlo samples. The red triangles represent the $WH(115)$ sample while the green triangles and the blue circles represent the samples $WH(125)$ and $WH(135)$, respectively.

this data set reduced this to 98,194 events. While the data have seen a substantial reduction in the number of events, the signal Monte Carlo samples see a preliminary cuts efficiency of approximately 3%. The Standard Model (SM) cross section for the WH signal with a Higgs boson mass of $m_H = 115$ GeV/c^2 is 0.19 pb[14] which yields 187 produced events. Taking into account the branching ratio for the W to decay hadronically, BR($W \rightarrow jj$)=0.68, and the rate for the Higgs boson to decay to $b\bar{b}$ with a mass of $m_H = 115$ GeV/c^2, BR($H \rightarrow bb$)=0.73[15], this 187 events is reduced to approximately 92 events. Propagating this Monte Carlo signal sample through the preliminary cuts yields an efficiency of (3.39±0.06)% giving a predicted signal content of 3.2 events. Performing the same procedure with the ZH sample of the same Higgs boson mass, a predicted ZH event count of 2.3 events is determined. By adding these two predicted event counts and dividing by the square root of the data content, which represents the error, a preliminary cuts sensitivity for the $m_H = 115$ GeV/c^2 Higgs boson is found to be 0.017. The preliminary cut flow for these two samples along with the other signal Monte Carlo samples and the data is shown in Table 5. The branching

ratios and cross sections used for this analysis are given in Table 6. This table also shows the expected numbers of events and sensitivities ($S/\sqrt{B}$) after the preliminary cuts.

Process	Initial Events	3JET Skim	Trigger	PV_z	=4 Jets	=2 b-tags	Total
Data	80M	N/A	30.9M	29.2M	6.1M	98K	98,194
$WH(115)$	94229.6	66.4%	45.7%	38.4%	15.7%	3.4%	$(3.39\pm0.06)\%$
$ZH(115)$	92197.4	69.2%	48.5%	40.6%	17.3%	4.1%	$(4.11\pm0.07)\%$
$WH(125)$	91827.0	70.2%	49.5%	41.6%	17.3%	3.6%	$(3.37\pm0.06)\%$
$ZH(125)$	92367.9	73.1%	52.4%	43.9%	18.7%	4.6%	$(4.58\pm0.07)\%$
$WH(135)$	89517.6	73.4%	52.8%	44.4%	18.7%	4.1%	$(4.11\pm0.07)\%$
$ZH(135)$	98364.1	75.9%	55.2%	46.3%	20.1%	4.9%	$(4.94\pm0.07\%$
WZ	465174.9	29.6%	17.8%	15.0%	5.3%	0.2%	$(0.230\pm0.007)\%$
TC(115)	186260.9	63.8%	43.8%	36.7%	13.8%	2.73%	$(2.73\pm0.04)\%$
TC(125)	184976.2	64.2%	43.8%	36.8%	13.6%	2.74%	$(2.74\pm0.04)\%$
TC(140)	173433.0	63.4%	42.9%	36.0%	13.3%	2.73%	$(2.73\pm0.04)\%$

Table 5: Preliminary cut flow for the data and different signal Monte Carlo samples used in the analysis. The = 4 jets cut includes the jet p_T, η, good jet quality, and taggability requirements. The efficiencies listed are total efficiencies to that point in the cut flow.

6.2.1 Post Preliminary Cuts Background Composition

This section describes the background composition based on a Monte Carlo study. The actual background used to derive the final limits (reported later) is based directly from the data. The background composition and magnitudes from Monte Carlo are presented here for informational purposes only. The preliminary cuts described previously substantially reduce the data from the initial 80.3M events to $\simeq 98K$ events. The signal

Analysis	Channel	σ (pb)	BR	Prod.	Prelim. Cuts Efficiency (%)	Pred. Prelim. Count	Sensitivity
$VH(115)$	WH	0.186	0.49	93.0	3.39±0.06	3.15±0.05	0.017
	ZH	0.108	0.51	55.6	4.11±0.07	2.28±0.04	
VH (125)	WH	0.138	0.42	57.5	3.73±0.06	2.14±0.04	0.012
	ZH	0.081	0.43	34.9	4.58±0.07	1.60±0.02	
VH (135)	WH	0.104	0.32	32.8	4.11±0.07	1.35±0.02	0.008
	ZH	0.062	0.32	20.3	4.94±0.07	1.00±0.01	
WZ	WZ	3.7	0.10	380.0	2.17±0.07	8.1±0.3	0.026
TC (115)	TC	0.753	0.68	515.8	2.73±0.04	14.1±0.2	0.045
TC (125)	TC	1.081	0.68	740.1	2.74±0.04	20.3±0.3	0.065
TC (140)	TC	0.949	0.68	649.5	2.73±0.04	17.7±0.3	0.056

Table 6: Preliminary cuts sensitivities for the various sub-analyses that are being performed in this paper. TC is short for technicolor and the "Prod." column is the number of produced events

WH MC also sees a reduction from the 93 produced events to $\simeq 3$ events. It is important to recall that these cuts require two tight b-tagged jets and equal to four hard jets.

Table 7 gives the input cross sections and expected numbers of events after the preliminary cuts assuming 1 fb^{-1} of data (the p17 data set). The background is dominated by the QCD multijet processes from the MSSM background sample. With the cuts of this analysis, the $bbjj$ multijet process dominates with ~80% of the QCD background, see Table 8. The individual components of this composite sample have been scaled to represent the correct relative portion of each sub-process present in the data set. This scaling was determined from data using a system of equations based on various b-tagging rates and various b-tagged jet multiplicities in data. In addition to this scal-

ing, the samples have a k-factor correction to compensate for the difference in Monte Carlo production and real QCD processes. Table 8 gives the relative amounts of each of the heavy flavor components in this cocktail after the preliminary cuts. There were $\simeq$84,700 QCD events that passed the preliminary cuts.

Non-Higgs processes with a W decay include: the di-boson decays, the top processes, and the Wjj processes. We expect a total of 304 ± 4 events with a W which will most likely appear in the invariant mass plot of the jj. Background processes with a Z include the WZ, ZZ, and $\gamma Z \rightarrow b\overline{b}$ where we expect a total of 66 ± 8 events. These are probably $Z \rightarrow b\overline{b}$ decays.

Table 7: Cross sections used and predicted numbers of events after the preliminary cuts for the p17 data set and background Monte Carlo samples. Branching ratio corrections are discussed in the text. Only statistical errors are given and no error is given to the cross sections. The preliminary cut efficiency is defined as the number of events that passed the preliminary cuts divided by the total number of Monte Carlo events generated. The SM Inclusive samples, such as $WZ \rightarrow$ SM Inclusive, have a degraded preliminary cut efficiency due to this definition.

Process	σ (pb)	Prelim Cut eff.(%)	Predicted Number
DATA			**98194**
$WW \rightarrow$ SM Inclusive	13.8	0.02	2.7 ± 0.3
$ZZ \rightarrow$ SM Inclusive	1.4	0.53	7.5 ± 0.2
$tb \rightarrow Wbb$ (s-channel)	0.88	4.02	25.1 ± 0.3
$tqb \rightarrow Wbq$ (t-channel)	1.98	0.89	12.0 ± 0.3
$t\bar{t} \rightarrow$ SM Inclusive($M_t = 175$ GeV/c^2)	6.7	3.53	238 ± 3
$Wbb \rightarrow jjbb$	19	0.13	16 ± 1
$Wcc \rightarrow jjcc$	48	0.004	1.2 ± 0.5
$Wjj \rightarrow jjjj$	621	0.0003	1 ± 3
$\gamma Z \rightarrow bb$	784	0.04	50 ± 8
$Zb \rightarrow bbb$	1.693	0.33	0.87 ± 0.05
QCD (MSSM) background		0.02	84700 ± 1500

Some of the important variables in this analysis are the invariant masses of the two b-jets and the remaining two jets. The invariant mass distributions for the two non-b-tagged jets, the b-tagged jets, and the 2D plot of the jj vs. $b\bar{b}$ mass plot are shown in Figure 38 for the data. Figure 39 shows the expected distributions from the $WH(115)$

Table 8: QCD multijet background sample relative contributions in % after the preliminary cuts and after the $WH(115)$ trained decision tree cut.

Process	Preliminary Cuts	DT cut
bbjj	80.9	95.0
bbc	5.9	3.9
ccjj	5.6	0
bbb	4.3	1.2
bbj	2.8	0
bb	0.3	0
ccj	0.2	0
cc	0	0

MC sample after preliminary cuts.
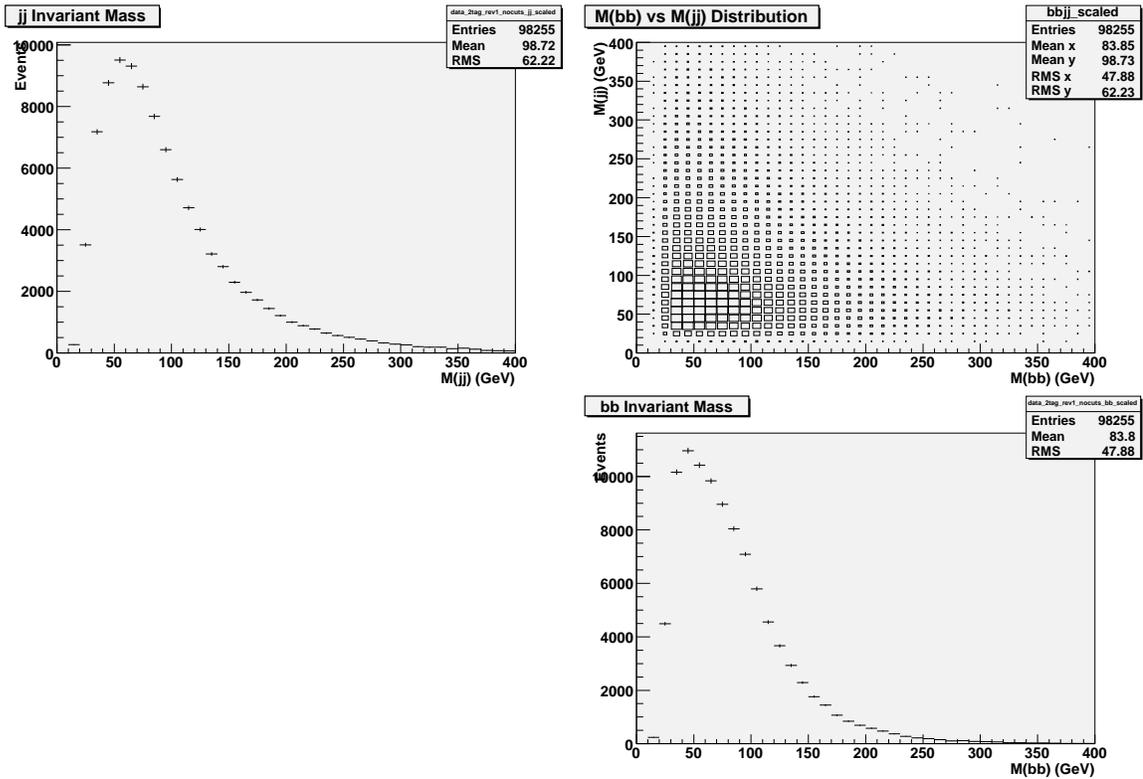
Figure 38: Invariant mass plot for events from the data after the preliminary cuts are applied for: the two non-tagged jets - jj Mass(top left), a 2D plot of the jj vs. $b\bar{b}$ jets (top right), and the two b-tagged jets (bottom right).
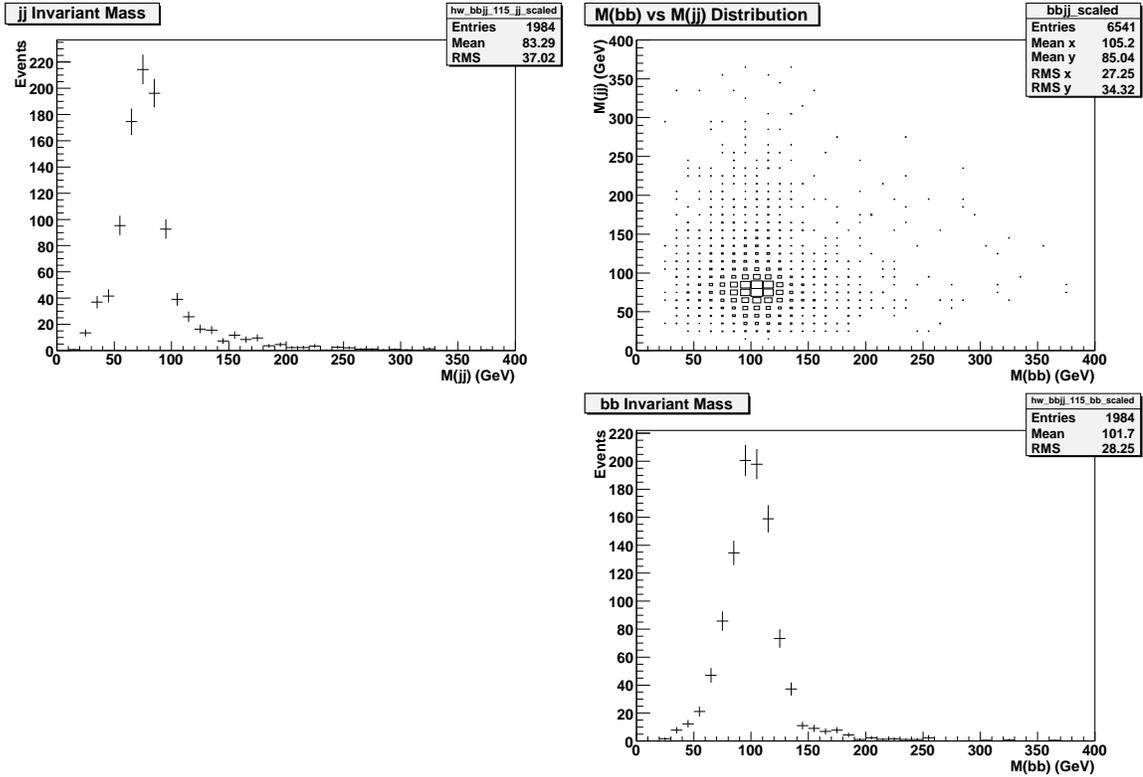
Figure 39: Invariant mass plot for events from the $WH(115)$ MC sample after the preliminary cuts are applied for: the two non-tagged jets - jj Mass (top left), a 2D plot of the jj vs. $b\bar{b}$ jets (top right), and the two b-tagged jets (bottom right).

6.2.2 Additional Post Preliminary Cuts Sample Distributions

In addition to the $=2$ b-tag sample, we used an $=1$ b-tag sample to represent the background in the decision tree training phase. Due to the combinatoric issues, a choice was made to identify which jets would be b-tagged. For the $=1$ b-tagged jet data set, the non b-tagged jet highest in p_T is used as the second b-tagged jet. Figures 40 and 41 show the bb and jj invariant mass distributions after the preliminary cuts for four-jet events where there were no b-tags and exactly one b-tag respectively. These invariant mass distributions are also shown in Figures 42, 43, 44, 45, 46, 47, 48, 49, and 50 for the Monte Carlo samples representing the processes $WH(125)$, $WH(135)$, $ZH(115)$, $ZH(125)$, $ZH(135)$, WZ, ZZ, $t\bar{t}$, and s-channel single top, respectively.
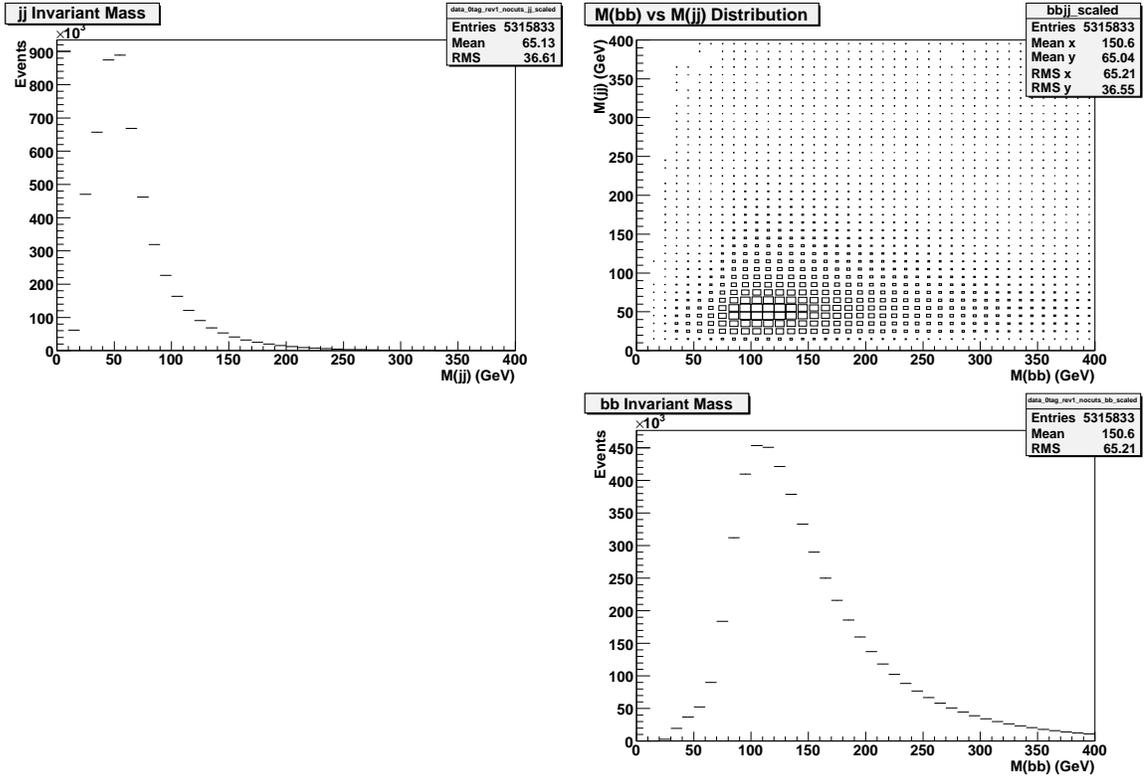
Figure 40: Invariant mass plot for events from the $= 0b$-tagged data after the preliminary cuts are applied for: the two lowest p_T jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two highest p_T jets-bb Mass (bottom right).

7   Background Reduction

The background reduction phase of this analysis is completed in two steps. The first of these utilizes a decision tree to increase the sensitivity and is discussed in Section 7.1. A second tool used in this analysis, presented in Section 7.2, is the application of invariant mass cuts designed to favor the final state of the process under investigation.

7.1   Decision Trees

A boosted decision tree is implemented in an effort to increase the sensitivity of the analysis. The decision tree is implemented in the TMVA Root package and specific details about its implementation can be found in Section 4.4.1 [40]. The decision tree for this analysis utilized 28 variables split into three broad categories: topological variables,
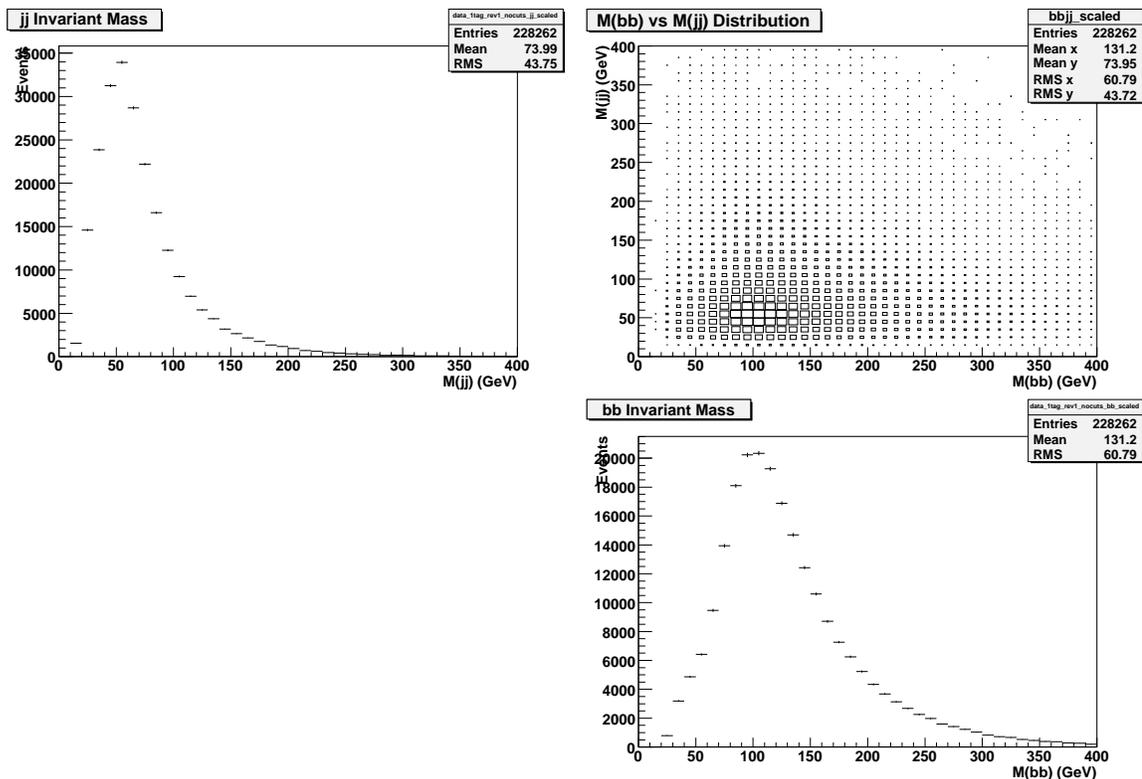
Figure 41: Invariant mass plot for events from the $= 1b$-tagged data after the preliminary cuts are applied for: the two non-tagged jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two 'b-tagged' jets (bottom right).

jet based quantities, and event related variables. Because we would like to examine the bb and jj dijet invariant masses after a decision tree cut, we exclude variables that could be highly correlated with the dijet masses such as the angles between the bb or jj jets. However, variables that discriminate signal from background are inherently related to the invariant mass in some form and will bias the distributions mostly towards having harder jets. We then have to make sure that we understand the background and signal bb and jj invariant mass distributions after the application of the DT.

We choose to use all possible variables to discriminate between signal and background. This is possible since there is no penalty in having more variables in a decision tree unlike a neural net discriminator. Many of the variables we use were inspired by the all-jets analysis from the top group at DØ [67]. The variables and their descriptions
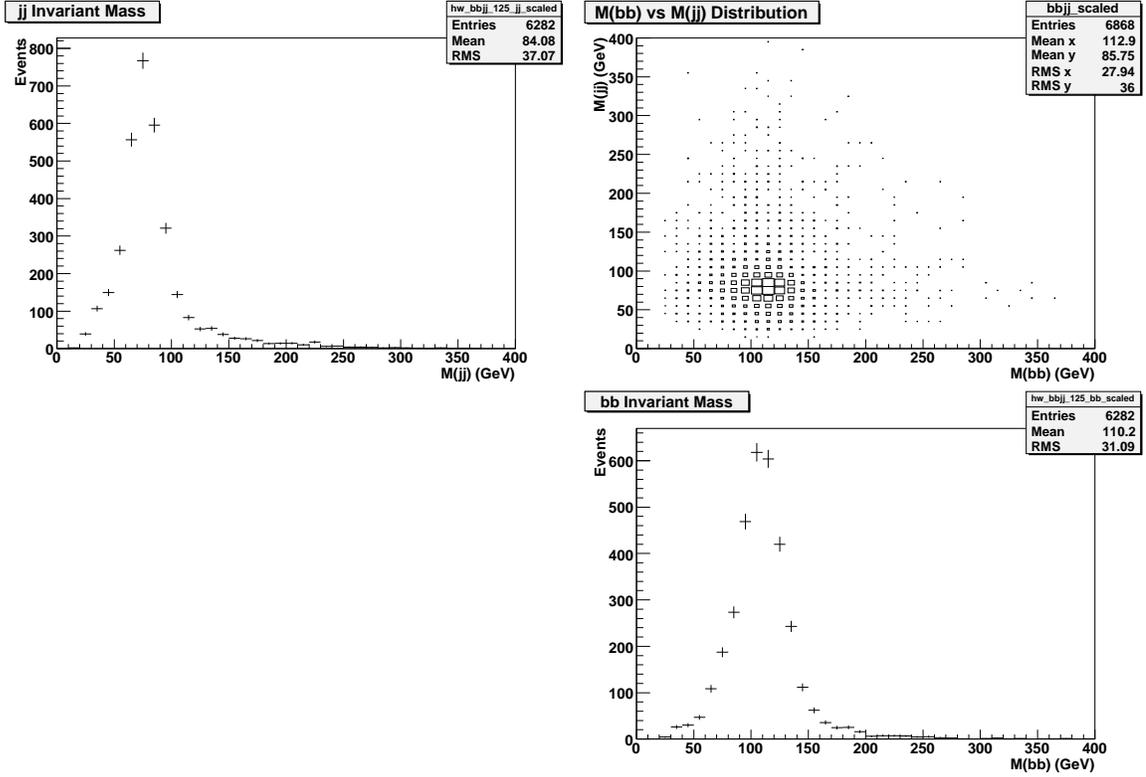
Figure 42: Invariant mass plot for events from the $WH(125)$ MC after the preliminary cuts are applied for: the two non-tagged jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two b-tagged jets (bottom right).

are listed in Table 9.

The topological variables contain quantities that aid in describing the shape and momentum distribution of the overall event and include: Sphericity, Centrality, and Aplanarity. Sphericity and Aplanarity are primarily concerned with the flow of momentum of the jets while the Centrality describes the overall shape of the jets in the η distribution. The Sphericity and the Aplanarity are found by deriving the eigenvalues, Q_i, of the normalized momentum tensor which is composed from the momenta of the jets in the event. Once the eigenvalues are found they are ordered from least to largest and numbered one to three, respectively. The Aplanarity is then defined as $A = \frac{3}{2}Q_1$ while the Sphericity is defined as $S = \frac{3}{2}(Q_1 + Q_2)$. The Sphericity is a measure of the summed p_T^2 with respect to the event axis and runs from zero to one, with S around

Variable	Description
$p_T \times 4$	The transverse momentum of the four jets that pass preliminary cuts
$\eta \times 4$	The pseudorapiditiy of the four jets that pass preliminary cuts
$\phi \times 4$	The ϕ of the four jets that pass preliminary cuts
Width $\times 4$	The width, $\sqrt{\sigma_\phi^2 + \sigma_\eta^2}$, of the four jets that pass the preliminary cuts
$CPF_0 \times 4$	The ratio of the total momentum of the tracks associated with both the jet and the primary vertex divided by the total momentum of the jet for each of the four jets that pass the preliminary cuts
Sphericity	The sum over the jets in the event of the transverse momentum squared with respect to the event axis
Centrality	The ratio of H_T^3 and H_T . The H_T is the sum of the transverse momentum of the jets in the event while H_T^3 is the sum of the transverse momentum of the jets in the event excluding the leading two jets.
Aplanarity	A measure of the transverse momentum component out of the event plane
H_T	The scalar sum of the transverse momentum of the four jets in the event
jj Charge	The sum of the charges from the two non b-tagged jets
bb Charge	The sum of the charges from the two b-tagged jets
Angle(H, W/Z)	The angle between the $b\bar{b}$ combination and the jj combination. In the case of the WZ channel this is the angle between the W and the Z.
Unweighted η	The unscaled RMS of the jet η for the four jets in the event, $\sum_{j=1}^{4} (\eta_j - \bar{\eta})^2$

Table 9: The input variables used in the boosted decision tree implementation
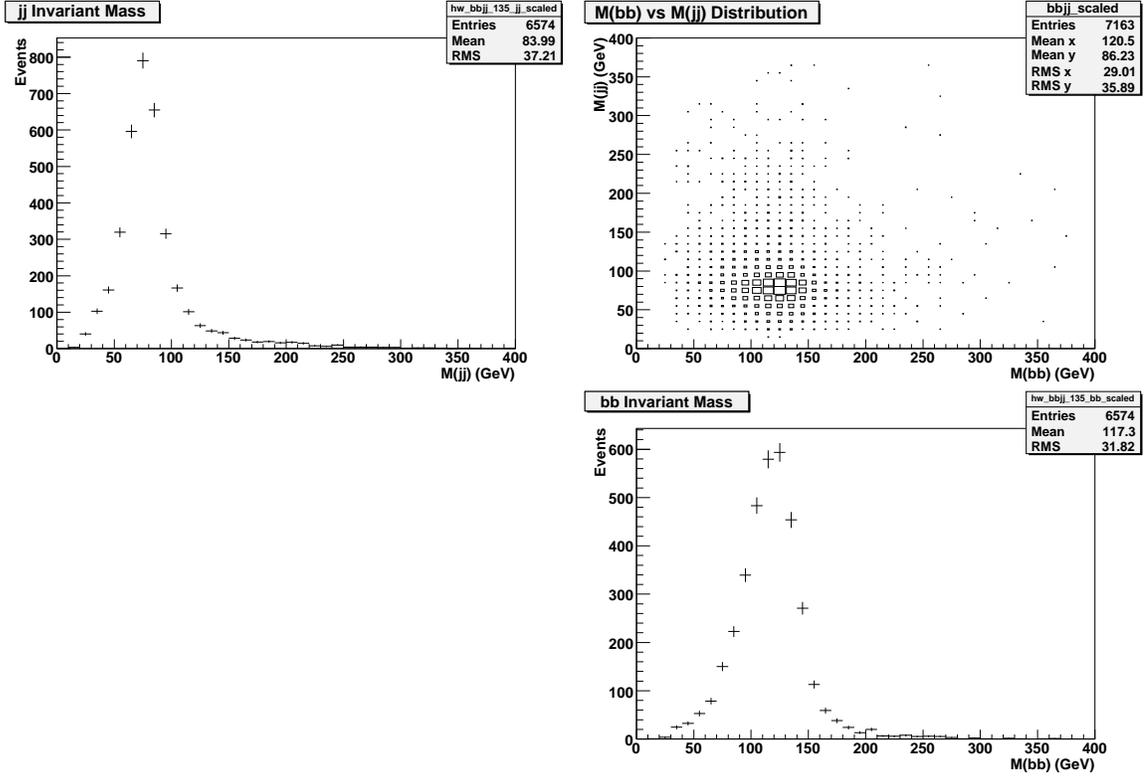
Figure 43: Invariant mass plot for events from the $WH(135)$ MC after the preliminary cuts are applied for: the two non-tagged jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two b-tagged jets (bottom right).

one corresponding to an isotropic event. Aplanarity is constrained between zero and $1/2$ with a planar event having $A \simeq 0$ and an isotropic event has $A \simeq 1/2$. Centrality is simply defined as the ratio of H_T^3 and H_T. The H_T is the sum of the transverse momentum of the jets in the event while H_T^3 is the sum of the transverse momentum of the jets in the event excluding the leading two jets.

The second set of variables focus on the kinematics, which consist of the properties of the individual jets. The jet width is found from width= $\sqrt{\sigma_\eta^2 + \sigma_\phi^2}$, where σ_η and σ_ϕ are the widths along those individual axes. The jet width tends to be smaller for quark jets than for gluon jets. The jet charge is taken as the sum of the individual charges of the two individual jets, where the individual jet charge is determined by looking at all of the tracks inside the jet cone.
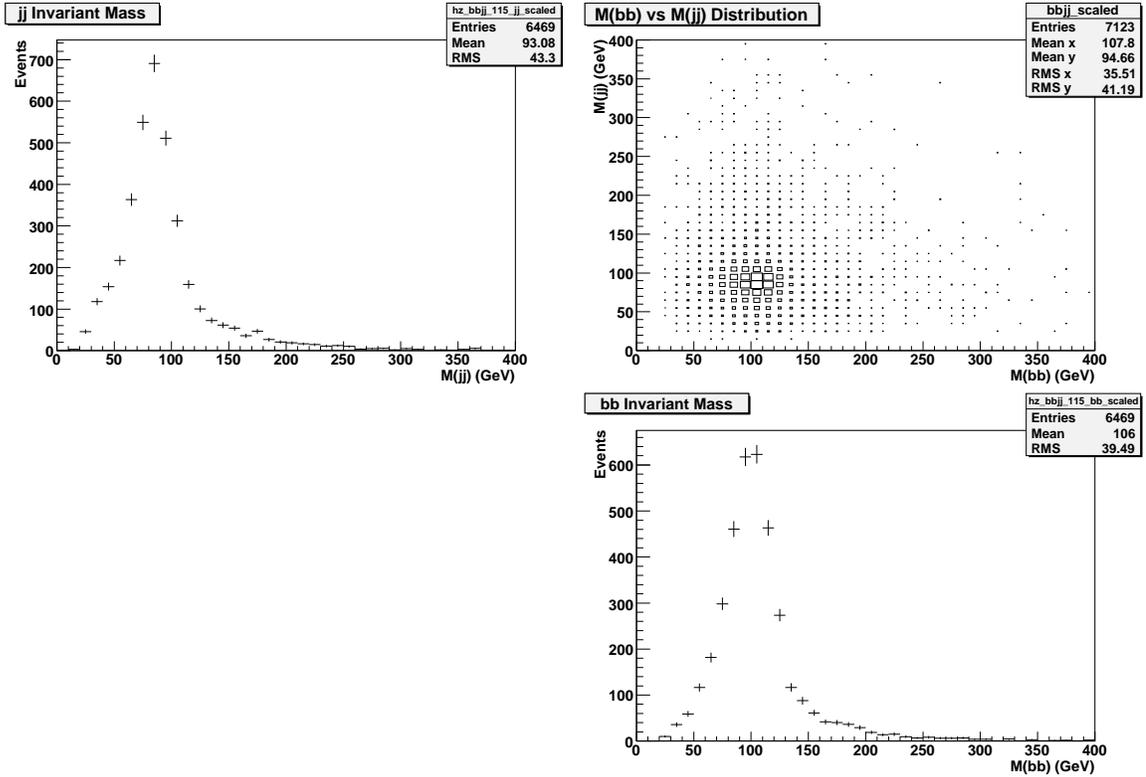
Figure 44: Invariant mass plot for events from the $ZH(115)$ MC after the preliminary cuts are applied for: the two non-tagged jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two b-tagged jets (bottom right).

Along with these quantities, the charged particle fraction with respect to the primary vertex is used. Here, we want each of the four jets to be consistent with coming from the primary vertex. The charged particle fraction is found by first summing the transverse momentum of the charged tracks found in the tracking system that are associated with the primary vertex and the jet. This sum is then divided by the total transverse momentum of the jet itself. Jets which are consistent with the primary vertex will tend to have their primary vertex charged particle fraction closer to one. The unweighted η variable is defined to be $\Sigma(\eta_i - \overline{\eta})^2$ where the i index is summed over the four jets and the $\bar{\eta}$ is the average of the η's of the four jets in the event. Higgs events will have a smaller unweighted η value than QCD events where the jets are less correlated with each other.

96
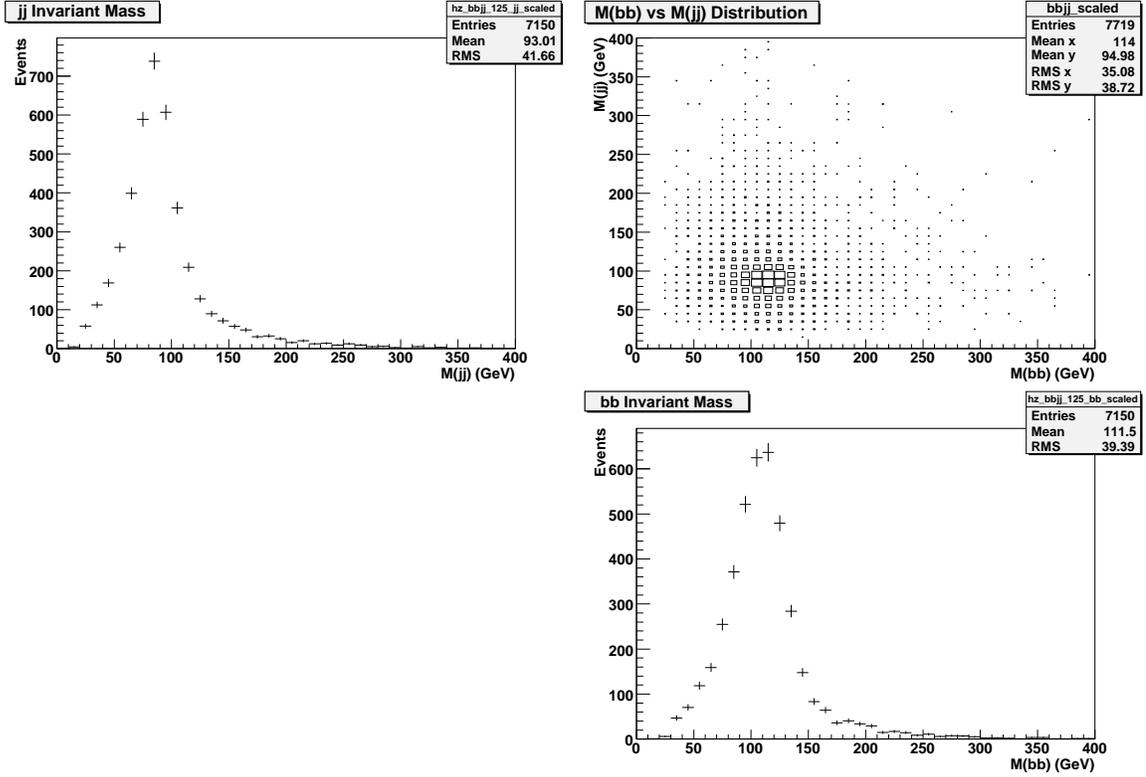
Figure 45: Invariant mass plot for events from the $ZH(125)$ MC after the preliminary cuts are applied for: the two non-tagged jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two b-tagged jets (bottom right).

We examine the distributions in each of these variables by comparing the data sideband distributions to those from the $WH(115)$ MC after the preliminary cuts. Here, the sidebands are defined to be $M_{jj} < 50$ GeV/c^2 or $M_{jj} > 100$ GeV/c^2 and $M_{bb} < 70$ GeV/c^2 or $M_{bb} > 130$ GeV/c^2. The two invariant mass distributions are shown in Figure 51. The comparisons of the input training distributions are found in the Appendix - Section B.5. Distributions of the MSSM background (QCD) Monte Carlo and $t\bar{t}$ Monte Carlo are also shown in the Appendix - Section B.6. The variables can be ranked in the order of their importance in separating the signal and the background training distributions. This metric used in the ranking is called the variable importance and is discussed in more detail in Section B.3.

Since we intend to use data sample sidebands in our final limit determination, we
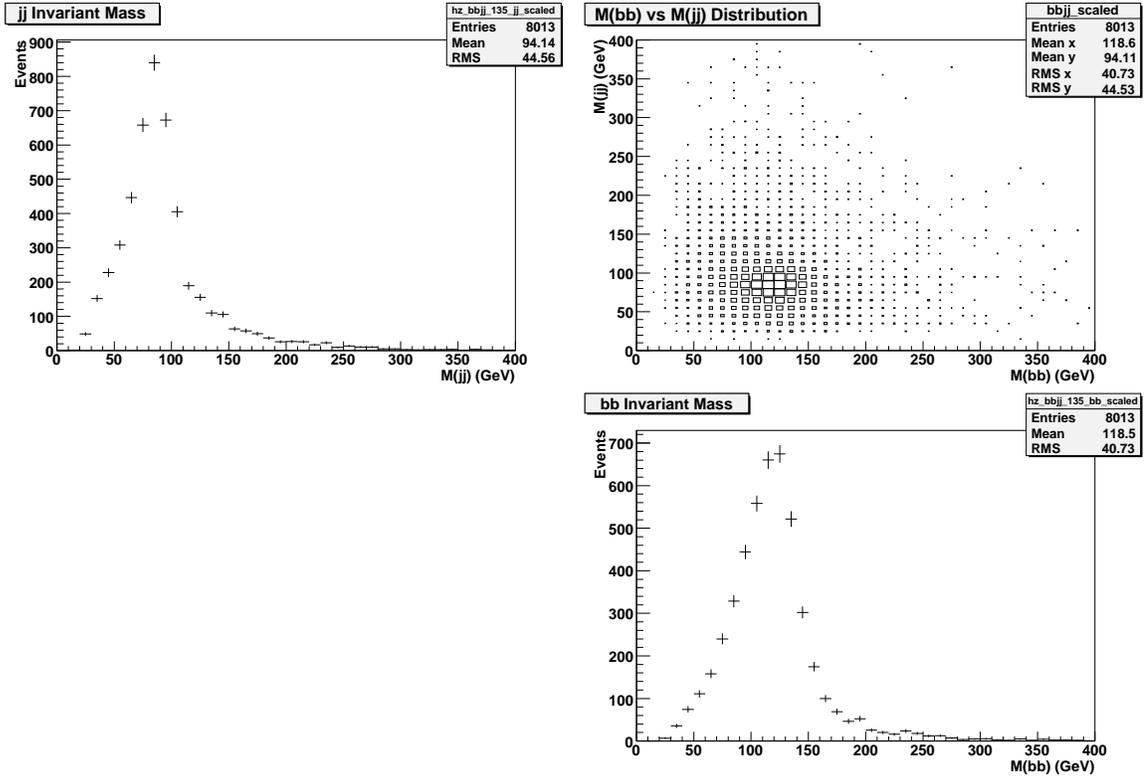
Figure 46: Invariant mass plot for events from the $ZH(135)$ MC after the preliminary cuts are applied for: the two non-tagged jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two b-tagged jets (bottom right).

have to find a sample to simulate the background for training the decision trees from another source. We use the $= 1b$-tagged data sample as our background training sample. For a couple of the variables (bb charge and the angle between the jj and bb), we are required to assign which of the three non b-tagged jets will be used as the second b-tagged jet for the event. We explored many different ideas for this, but found that using the hardest non-b tagged jet as the second b in the event is appropriate. In some cases the agreement is less than perfect. While this degrades the performance of the decision tree, it does not invalidate the tree as an effective tool to separate signal and background.

The decision trees were trained with the 28 variables listed above using the background sample set to the $= 1b$ tagged data sample and the $WH(115)$ MC samples as
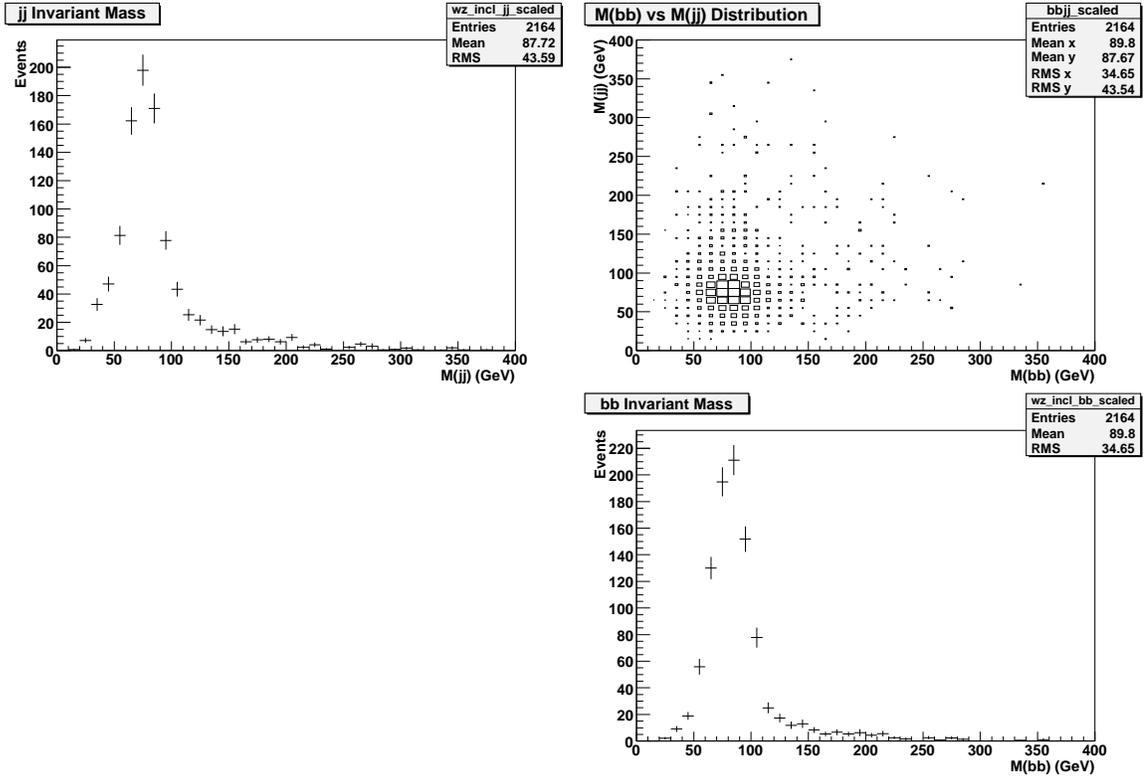
Figure 47: Invariant mass plot for events from the WZ MC after the preliminary cuts are applied for: the two non-tagged jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two b-tagged jets (bottom right).

the signal sample. The training used 1/3 of the background (over 200,000 events) and MC samples for training and 1/3 for testing. Therefore, the remaining 1/3 of the signal MC samples ($\simeq$ 2000 events) was used to calculate the efficiency for the analysis. The DT discriminator output is shown for the $WH(115)$ trained decision tree in Figure 52.

For our further analysis, we will make a cut on the decision tree variable and examine the jj and bb invariant mass distributions. The optimal DT cut point was chosen as that point that maximized the sensitivity. We scanned the decision tree output variable for the data set (=2 b-tags) and compared to the scaled signal MC expectations to determine the sensitivity.

We have measured the correlation coefficients between the DT variables, the DT output, and the jj and bb invariant masses in both the data and signal Monte Carlo
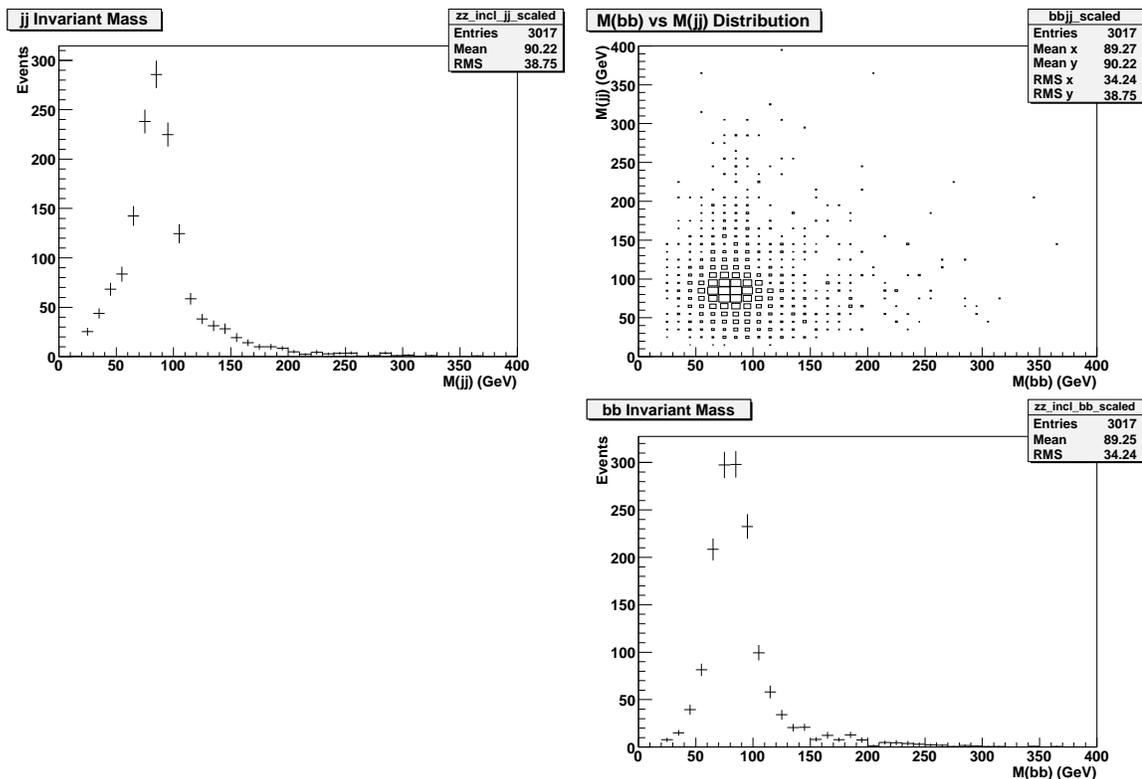
Figure 48: Invariant mass plot for events from the ZZ MC after the preliminary cuts are applied for: the two non-tagged jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two b-tagged jets (bottom right).

sample. The correlation coefficients were determined by dividing the covariance matrix by the product of the standard deviations in the two variables under consideration. Since the correlation values are normalized, a value of 1.0 would mean that the two variables are exactly correlated, which means that as one of the variables increases the other increases in the same way. Examples of exact correlations can be seen in the diagonal elements of the matrix shown in Figure 53(a). A correlation of -1.0 would be exactly anti-correlated and describes the case where one of the variables goes up and the other variable goes down by the same amount. A value of 0.0 for the correlation means that the two variables under consideration are not correlated and the first variable does not influence the second variable. We show these correlations in Figure 53(a) for the $WH(115)$ signal Monte Carlo and in Figure 53(b) for the data. As one might expect,

100
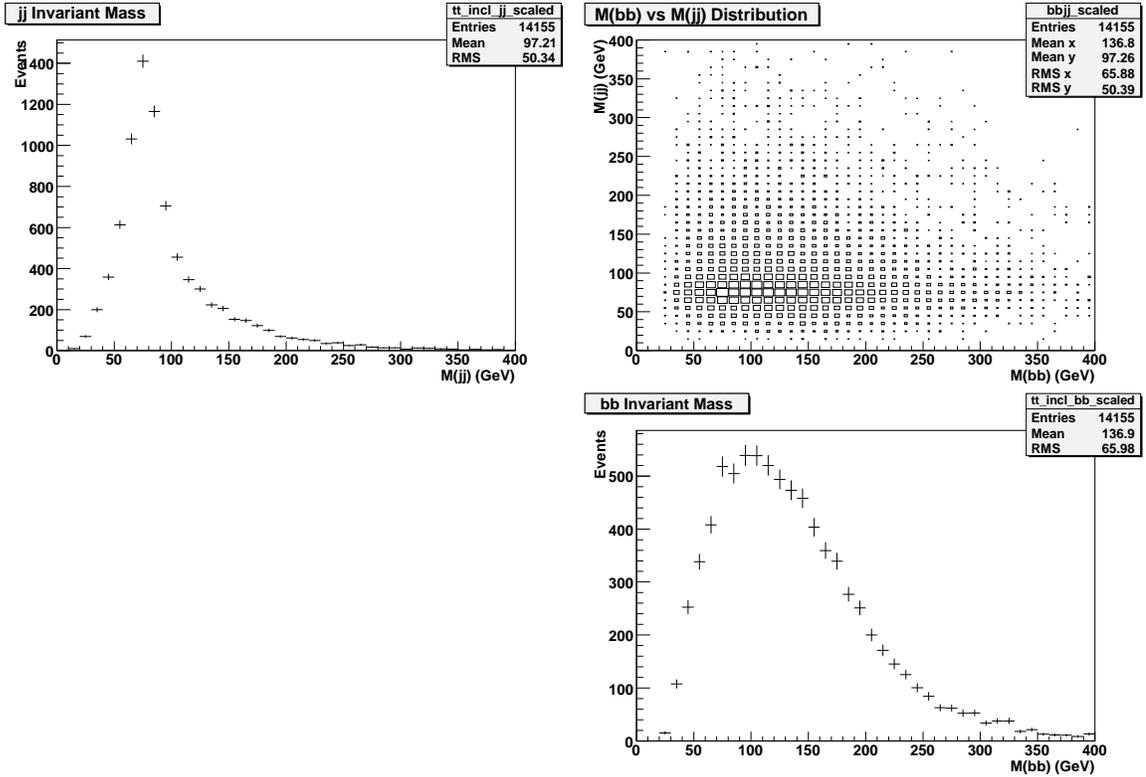
Figure 49: Invariant mass plot for events from the $t\bar{t}$ MC after the preliminary cuts are applied for: the two non-tagged jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two b-tagged jets (bottom right).

the largest correlations are found between the DT output and the unweighted eta which provides the most improvement. We see little correlation between the bb and jj invariant masses and the DT output. Due to the nature of the analysis a more in-depth look at the correlations between the DT discriminator and the invariant masses is warranted. These correlations, presented in Figure 54, reflect a weak relationship.

7.1.1 DT Results

The decision tree discriminant distributions for the signal Monte Carlo distributions and the data are shown in Figure 55. The cut point on the decision tree discriminant is chosen such that the WH sensitivity, $S/\sqrt{B}$, is optimized. This is accomplished by evaluating the improvement to the preliminary cuts sensitivity and choosing the point
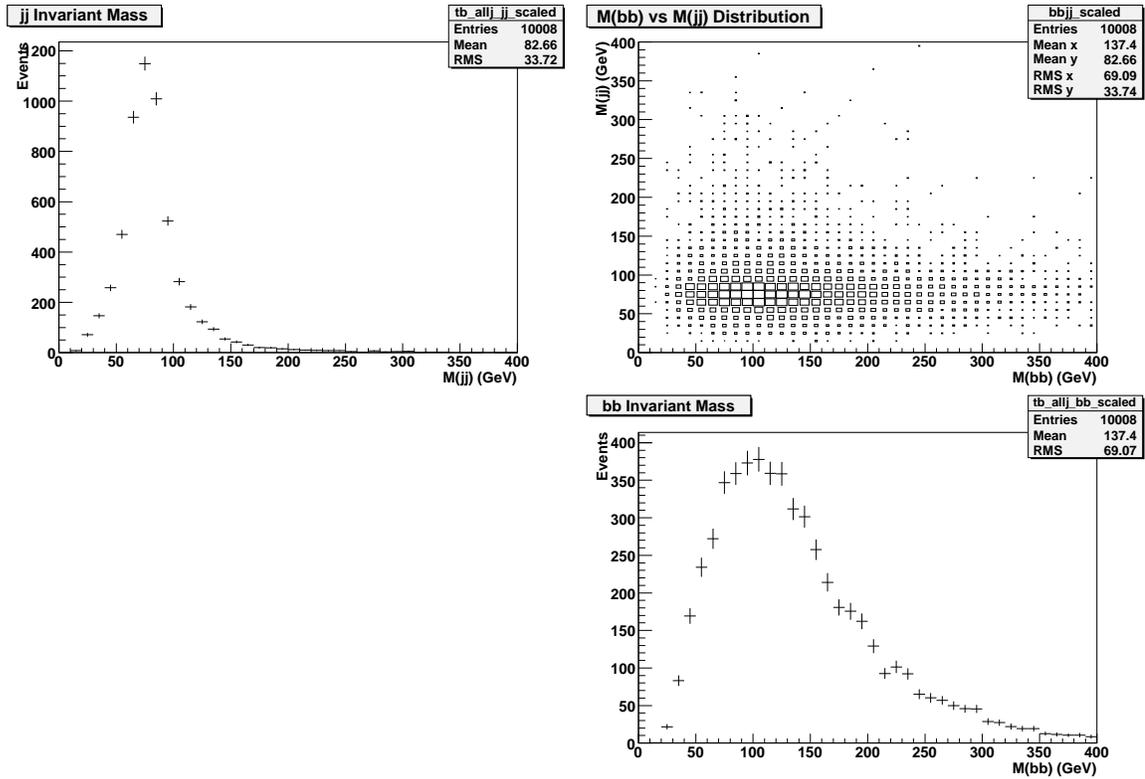
Figure 50: Invariant mass plot for events from the s-channel single top MC after the preliminary cuts are applied for: the two non-tagged jets-jj Mass(top left), a 2D plot of the jj vs. bb jets (top right), and the two b-tagged jets (bottom right).

(a) Invariant mass of the two non-tagged jets

(b) Invariant mass of the two b-tagged jets
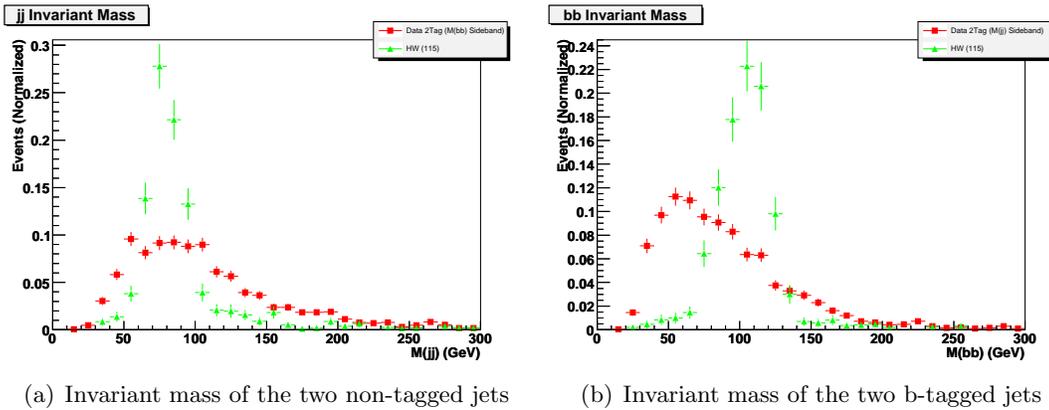
Figure 51: Invariant mass distributions for events from the data sidebands regions (red squares) and $WH(m_{\mathrm{H}} = 115$ GeV/c^2) signal Monte Carlo (green triangles) after the preliminary cuts are applied

102
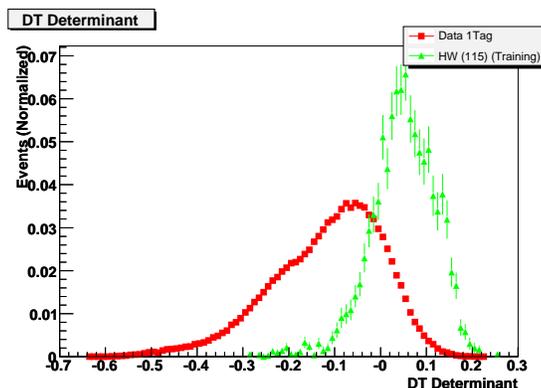
(a) $WH(115)$ training

Figure 52: The decision tree output for the background (=1 b-tag data) and WH(115) signal sample.

corresponding to the maximum value. Figure 56 presents the improvement scan for the $WH(115)$ decision tree along with the number of predicted events after a cut and the number of data events remaining after the same cut. The peak in the sensitivity improvement was fount to be at a DT discriminant cut point of 0.085.

The decision tree signal region for the $m_H = 115$ GeV/c^2 mass point is 0.085→1.0. While this was found using the WH distribution, the similarities in the WH and the ZH decision tree discriminant distributions allow us to use the same cut point for the ZH sample. It is found that applying this cut to the WH Monte Carlo for the $m_H = 115$ GeV/c^2 mass point yields a relative efficiency of 29% giving a total efficiency of (0.97±0.05)%. Applying this new total efficiency to the number of produced events there are an expected 0.90±0.05 WH events after the decision tree cut. Performing the same operation on the ZH sample for the same mass point a relative efficiency of 27% is found giving a total efficiency of (1.12±0.03)%. With this total efficiency, the number of ZH signal events expected is 0.62±0.02. Combining the expected numbers of events from the WH and the ZH channels, the total number of expected signal events is 1.5. Propagating the data though the decision tree and applying the same cut it is found that 3064 events remain. Combining the expected number of signal events and the number of data events that pass the cut, the sensitivity is increased to 0.027 which

Table 10: Predicted number of events in the samples listed in Table 7 and the data both before and after the DT cut. All samples are processed though the WH(115) trained tree using a cut of 0.085. An exception to this is the (W/Z)H(125) and (W/Z)H(135) samples which have been processed and evaluated though the WH(125) and WH(135) DT respectively. The DT cut on the WH(125) tree is 0.09 and the DT cut on the WH(135) tree is 0.07. The preliminary cut efficiency is defined as the number of events that passed the preliminary cuts divided by the total number of Monte Carlo events generated. The Post DT eff(%) column is generated by dividing the number of events that passed the DT cut by the total number of events that entered DT, scaled by 100.0 to represent a percent. The SM Inclusive samples, such as $WZ \rightarrow$ SM Inclusive, have a degraded preliminary cut efficiency due to this definition. The Monte Carlo samples above the horizontal line represent di-boson signal samples.

Process	Prelim cuts Number	Relative DT cut eff(%)	Total eff(%)	Post DT Number
DATA	**98194**			**3064**
$WH(115) \rightarrow jjbb$	3.1	29	0.97 ± 0.05	0.90 ± 0.05
$WH(125) \rightarrow jjbb$	2.1	37	1.38 ± 0.06	0.80 ± 0.03
$WH(135) \rightarrow jjbb$	1.6	42	1.74 ± 0.06	0.57 ± 0.02
$ZH(115) \rightarrow jjbb$	2.3	27	1.12 ± 0.03	0.62 ± 0.02
$ZH(125) \rightarrow jjbb$	1.6	35	1.62 ± 0.04	0.57 ± 0.01
$ZH(135) \rightarrow jjbb$	1.0	41	2.02 ± 0.04	0.410 ± 0.009
$TC(115) \rightarrow jjbb$	14.1	23	0.62 ± 0.02	3.17 ± 0.09
$TC(125) \rightarrow jjbb$	20.3	26	0.71 ± 0.02	5.3 ± 0.1
$TC(140) \rightarrow jjbb$	17.7	27	0.74 ± 0.02	4.8 ± 0.1
$ZW \rightarrow$ Inclusive	8.1	13.6	0.030 ± 0.003	1.10 ± 0.09
$WW \rightarrow$ Inclusive	2.7	20	0.0040 ± 0.0009	0.5 ± 0.1
$ZZ \rightarrow$ Inclusive	7.5	17	0.092 ± 0.006	1.29 ± 0.08
$tb \rightarrow Wbb$ (s-channel)	25.1	26.8	1.12 ± 0.03	6.7 ± 0.2
$tqb \rightarrow Wbq$ (t-channel)	12.1	15.4	0.14 ± 0.01	1.9 ± 0.1
$t\bar{t}(M_t = 175 \text{ GeV/c}^2)$	238	30.9	1.09 ± 0.02	74 ± 1
$Wbb \rightarrow jjbb$	16.3	16.8	0.021 ± 0.003	2.7 ± 0.4
$Wcc \rightarrow jjcc$	1.2	0	0 ± 0	0 ± 0
$Wjj \rightarrow jjjj$	1.2	0	0 ± 0	0 ± 0
$\gamma Z \rightarrow bb$	50	6	0.002 ± 0.002	3 ± 2
$Zb \rightarrow bbb$	0.9	6	0.021 ± 0.005	0.05 ± 0.01
QCD background	84700			3800 ± 290

is an approximate improvement of 59% over the preliminary cuts sensitivity.

Separate DTs were originally used for the Technicolor analysis. However, studies of the scale factors needed for these different trees provided large systematic errors for
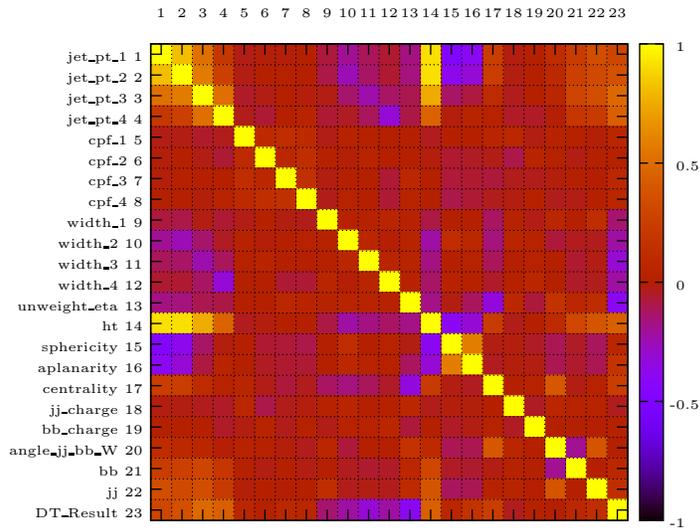
(a) $WH(115)$ signal Monte Carlo sample
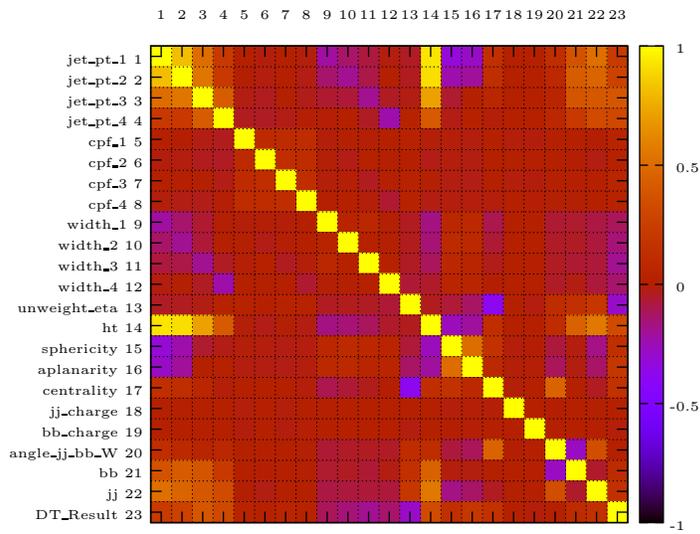
(b) $= 2$ b-tagged data sample

Figure 53: The correlation coefficients between DT variables, the DT output, and the bb and jj invariant masses.
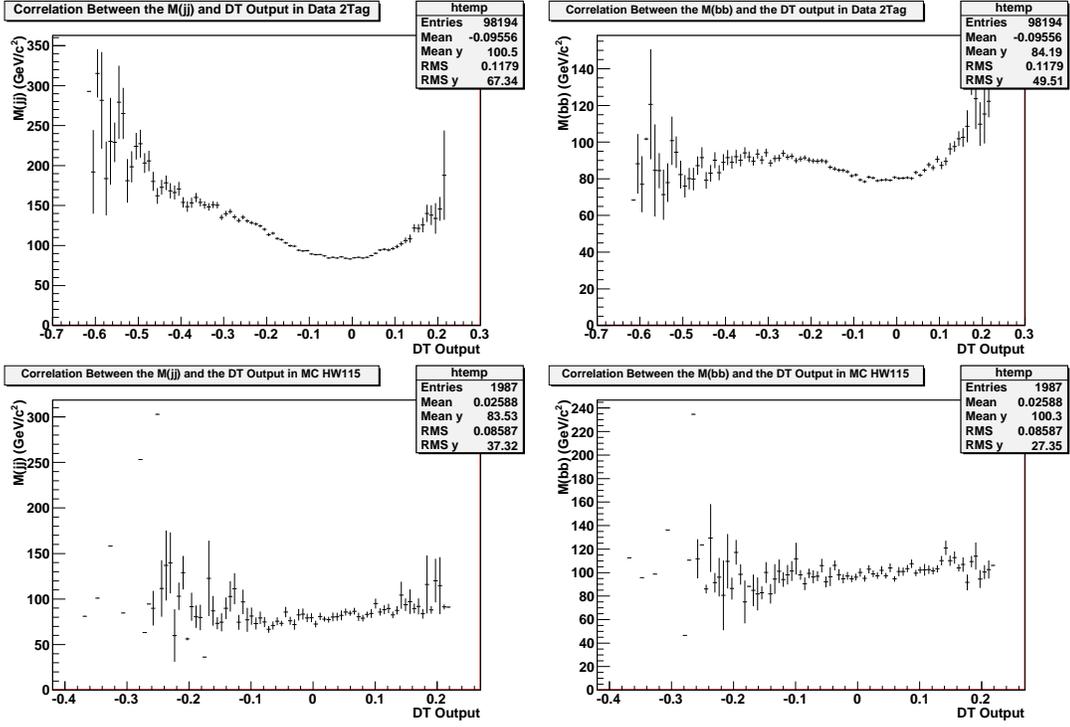
Figure 54: Detailed correlations between the invariant masses and the DT output discriminator in both the data 2Tag sample and the $WH(115)$ MC signal sample. The top two plots show the data =2 tag sample while the bottom two plots show the $WH(115)$ signal Monte Carlo sample. The left column shows the jj dijet mass while the right column shows the bb mass. The y-axis represents the average mass with the error bar indicative of the rms in that particular DT discriminant bin.

the final results. Although the sensitivity as found using the statistical errors only is reduced for the Technicolor analyses with cuts on the WH trained trees, the overall sensitivity is improved. The Technicolor analyses use the VH trained trees for the equivalent mass points with the exception of the TC(140) sample which was processed though the $VH(135)$ decision tree.

The DT discriminant optimized for the Higgs mass of 125 GeV/c^2, and 135 GeV/c^2, along with the WZ and Technicolor samples with the technipion masses of 115 GeV/c^2, 125 GeV/c^2, and 140 GeV/c^2 can be found in the Appendix - Section B.2. The sensitivity scans for the corresponding decision trees can also be found in the same section. Table 11 shows the performance for each of the different DTs trained for each mass point
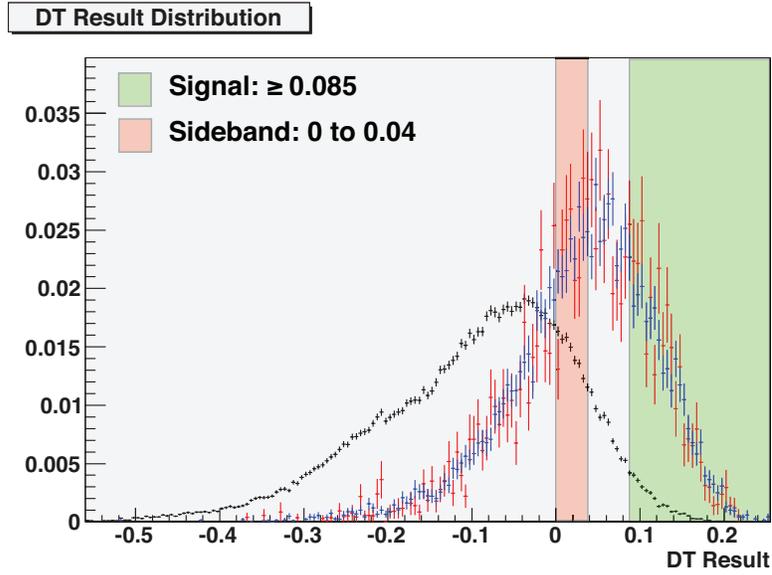
Figure 55: The decision tree discriminant distribution for the WH, ZH, and data set for the $m_H = 115 \text{ GeV/c}^2$ mass point. The red distribution is the WH sample while the blue distribution represents the ZH sample. The data distribution is shown in black. The two shaded regions depict the DT control region (red) and the region of interest (green).
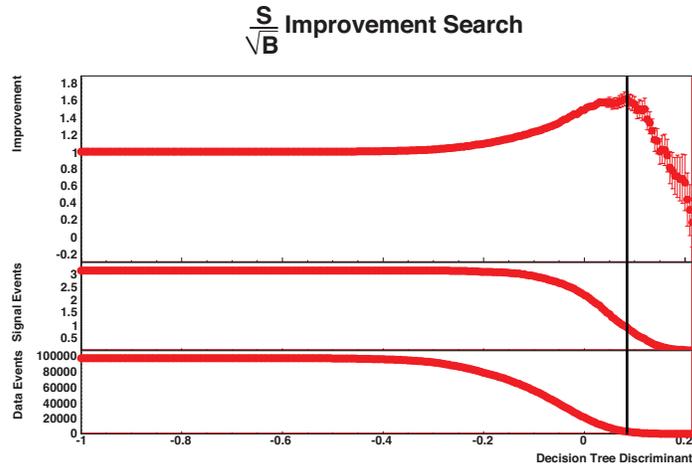
Figure 56: Sensitivity improvement search for the $WH(115)$ decision tree training. The bottom two left axes show the number of signal and background events as a function of a cut on the decision tree discriminant. The top two plots show the Sensitivity $S/\sqrt{B}$ improvement with respect to the preliminary cuts after all cuts. The improvement is the ratio of the sensitivity after the cut on the DT divided by the sensitivity at the preliminary cuts stage. The vertical black line denotes the optimal cut point.

analysis. From this, one can see that for the WZ analysis, we lose sensitivity compared to the preliminary analysis. We choose not to use the DT cut for this analysis.

Mass Point (GeV/c^2)	DT Signal Region	MC Type	Total Efficiency (%)	Number of Predicted Signal Events	Passed Data Events	Sensitivity
115	0.085→1.0	WH	0.97±0.05	0.90±0.05	3064	0.027
		ZH	1.12±0.03	0.62±0.02		
125	0.09→1.0	WH	1.38±0.06	0.80±0.03	4891	0.020
		ZH	1.62±0.04	0.57±0.01		
135	0.07→1.0	WH	1.74±0.06	0.57±0.02	4938	0.014
		ZH	2.01±0.04	0.410±0.009		
N/A	0.13→1.0	WZ	0.69±0.04	2.6±0.1	9651	0.026
115	0.085→1.0	TC	0.62±0.02	3.17±0.09	3064	0.057
125	0.09→1.0	TC	0.71±0.02	5.3±0.1	4891	0.075
140	0.07→1.0	TC	0.74±0.02	4.8±0.1	4938	0.068

Table 11: Performance for the different decision trees trained for each mass point.

7.2   Invariant Mass Cuts

In an effort to increase the sensitivity and correlate the M(jj) and M(bb) distributions, we apply cuts first on the M(jj) distribution to enhance the fraction of events with a W and then on the M(bb)-M(jj) distribution to enhance the particular signal under investigation.

7.2.1 $W \to jj$ distribution in Data

The jj invariant mass distribution from a $t\bar{t}$ enhanced data set can be used to determine the parameters for the dijet invariant mass shape. A set of cuts are applied to a special Electron and Muon skim followed by a fit to the jj invariant mass distribution. The

skim was generated by the Top subgroup by applying both the p17 final jet energy scale and b-tagging to the common samples group EM inclusive and MU inclusive data skims. The following cuts are then applied:

- Transverse missing energy (unbalanced transverse energy) ≥ 20 GeV/c^2

- $|PV_Z| \leq 35$

- ≥ 4 jets with $p_T > 15$ GeV

- at least one jet with $p_T > 20$ GeV

- equal to four jets in event

- a good lepton with at least $p_T > 20$ GeV

- a second lepton with at least $p_T > 15$ GeV

- the second lepton can't be opposite of first

- two NN loose b-tags.

Figure 57 shows the jj invariant mass in this sample. A clear W peak is found which is then fit to a Gaussian for the signal and a threshold function for the background. From the likelihood fit, we find 47 ± 13 events in the Gaussian peak with a mean of (75.0 ± 2.5) GeV/c^2 and sigma of (11.2 ± 2.3) GeV/c^2.

In addition to the study using calorimeter only jets, the invariant mass was also fit for CAL+TRK jets. The CAL+TRK jets combine information from both the calorimeter and the tracking system to improve the overall jet measurement. Here, a similar mean was found, but the sigma was reduced to (9.2 ± 2.1) GeV/c^2. This second study was done for later improvement plans, but the CAL-only jets are used in this analysis.

7.2.2 Invariant Mass Cuts Used in the Analysis

The quantity M(bb)-M(jj) is used since it is not strongly correlated with the decision tree output in the region of interest. This low correlation with the decision tree output,

MINUIT Likelihood Fit to Plot			1&0			MINUIT Likelihood Fit to Plot			2&0	
jj Invariant Mass						jj Invariant Mass (Using CalTrack)				
File: output.root					9-MAR-2008 12:23	File: output.root				9-MAR-2008 12:25
Plot Area Total/Fit	103.00 / 103.00				Fit Status 3	Plot Area Total/Fit	103.00 / 103.00			Fit Status 3
Func Area Total/Fit	104.02 / 104.02				E.D.M. 5.009E-06	Func Area Total/Fit	103.51 / 103.51			E.D.M. 5.589E-06

Likelihood = 29.2   Likelihood = 26.0

χ^2= 29.0 for 28 - 7 d.o.f.,   C.L.= 11.4% χ^2= 25.2 for 28 - 7 d.o.f.,   C.L.= 23.9%

Errors		Parabolic		Minos		Errors		Parabolic		Minos
Function 1: Threshold						Function 1: Threshold				
NORM	0.14133	± 0.5146	- 0.000	+ 0.000		NORM	6.18013E-02	± 0.1995	- 0.000	+ 0.000
∗ OFFSET	15.000	± 0.000	- 0.000	+ 0.000		∗ OFFSET	15.000	± 0.000	- 0.000	+ 0.000
POWER	0.72290	± 1.544	- 0.000	+ 0.000		POWER	1.0642	± 1.344	- 0.000	+ 0.000
COEFF1	-3.54960E-02	± 4.4159E-02	- 0.000	+ 0.000		COEFF1	-4.31738E-02	± 3.8506E-02	- 0.000	+ 0.000
COEFF2	5.98054E-05	± 1.0260E-04	- 0.000	+ 0.000		COEFF2	7.23572E-05	± 9.1169E-05	- 0.000	+ 0.000
Function 2: Gaussian (sigma)						Function 2: Gaussian (sigma)				
AREA	46.569	± 13.44	- 0.000	+ 0.000		AREA	42.297	± 11.88	- 0.000	+ 0.000
MEAN	74.954	± 2.502	- 0.000	+ 0.000		MEAN	74.586	± 2.273	- 0.000	+ 0.000
SIGMA	11.163	± 2.269	- 0.000	+ 0.000		SIGMA	9.1593	± 2.093	- 0.000	+ 0.000

(a) CAL only jets   (b) CAL+TRK jets

Figure 57: Fit of invariant mass of the two non b-tagged jets for $t\bar{t}$ enhanced data set after cuts.

shown in Figure 58, allows a background shape to be constructed not from a mass sideband as before but from a decision tree output sideband. These correlations are also shown for the data for the M(jj) and M(bb) variables in Figure 59 where there is a correlation observed. Figure 60 shows this correlation for the $WH(115)$ signal Monte Carlo for the M(jj), M(bb), and M(bb)-M(jj) quantities. With the signal, you observe no correlations in any of the distributions at higher DT values.

We start by finding a cut on the M(jj) region to emphasize the W signal. A cut window is centered at the reconstructed W mass in the signal Monte Carlo of 79 GeV/c^2. A symmetric window cut is then applied and the sensitivity is then re-evaluated by observing the different efficiencies for the data and the signal after the cut. The optimum window size is then found by determining the highest sensitivity after the cut. The final values for the W mass cut is $61.4 \leq M(jj) < 96.6$ GeV/c^2 which keeps
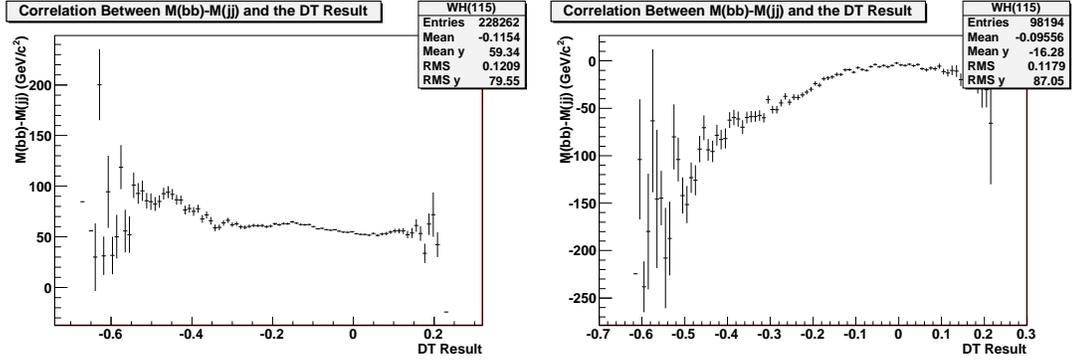
Figure 58: Detailed correlations between the M(bb)-M(jj) variable and the DT output discriminator in both the data 2Tag sample (right plot) and the data 1Tag sample (left plot). The y-axis represents the average of the variable with the error bar indicative of the rms in that particular DT discriminant bin.
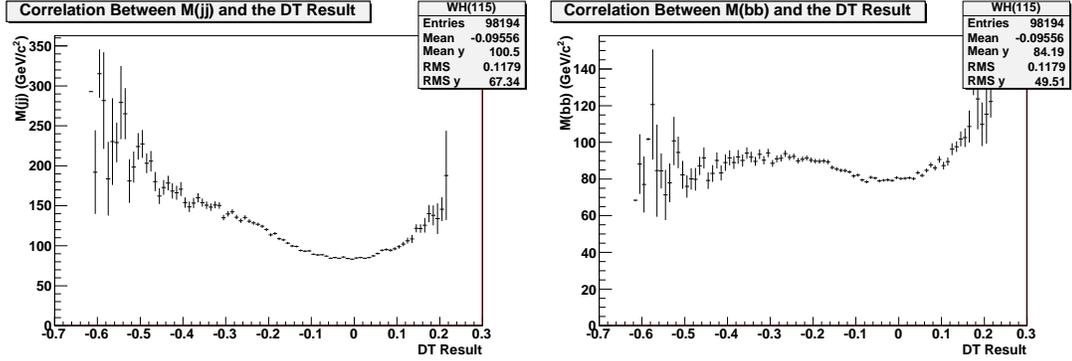
Figure 59: Detailed correlations in the 2Tag data sample between the M(jj) (left plot) variable or M(bb) variable (right plot) and the DT output discriminator. The y-axis represents the average of the variable with the error bar indicative of the rms in that particular DT discriminant bin.

62% of the signal and 29% of the data distribution. This cut does not depend on the Higgs or Z mass and is common among the analyses discussed. We do not re-optimize the window for the ZH analyses but rely on the same cut. This cut retains 52% of the ZH(115) signal.

Figure 61 shows the M(bb)-M(jj) signal MC distributions after the M(jj) cut for each of the analyses. A fit to the WH(115) M(bb)-M(jj) signal MC distribution after the M(jj) cut is shown in Figure 62. The mean of the fitted Gaussian is (24.7 ± 1.1) GeV/c^2 which is as expected the difference in the two means from each of the bb and jj
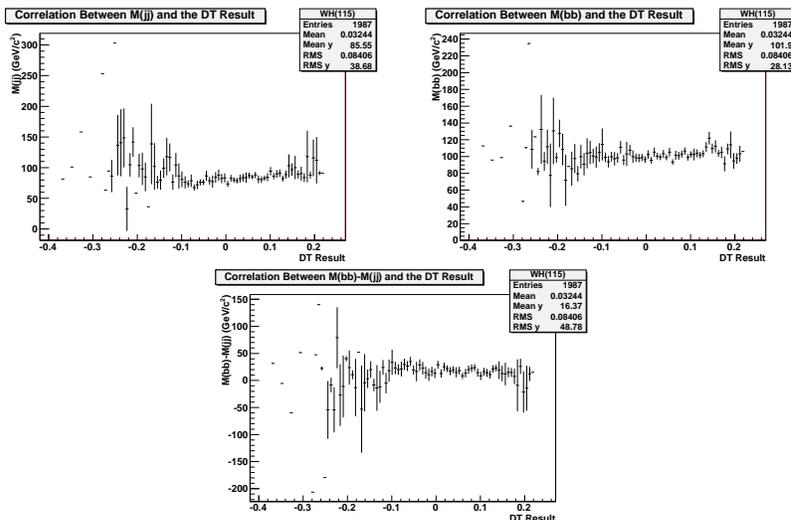
Figure 60: Detailed correlations in the WH(115) Signal MC sample between the M(jj) (left plot) variable, M(bb) variable (center), or M(bb)-M(jj) variable and the DT output discriminator. The y-axis represents the average of the variable with the error bar indicative of the rms in that particular DT discriminant bin.

Analysis	M(jj) (GeV/c^2)	M(bb)-M(jj) (GeV/c^2)
$VH(115)$	61.4→96.6	-6→58
$VH(125)$	61.4→96.6	1→66
$VH(135)$	61.4→96.6	7→78

Table 12: Mass windows used for the different VH mass points.

distributions. The fitted width of (19.3 ± 0.8) GeV/c^2 is approximately the convolution of the two widths. Fits to the different Higgs boson mass signal MC sets show that the widths are very similar, only the mean shifts. We choose a cut on M(bb)-M(jj) which keeps $\simeq$90% of the signal. As stated earlier, the Technicolor channels of this analysis are processed though the VH cut chains due to the complications with the background determination. The TC(115) and TC(125) are processed though the $VH(115)$ and $VH(125)$ cut chains while the TC(140) is processed though the $VH(135)$ cut chain.

The final mass windows used in the analysis are shown in Table 12 while the signal efficiencies, expected signal events, number of events in the signal regions (D_s), and total sensitivities ($S/\sqrt{B}$) are given in Table 13.
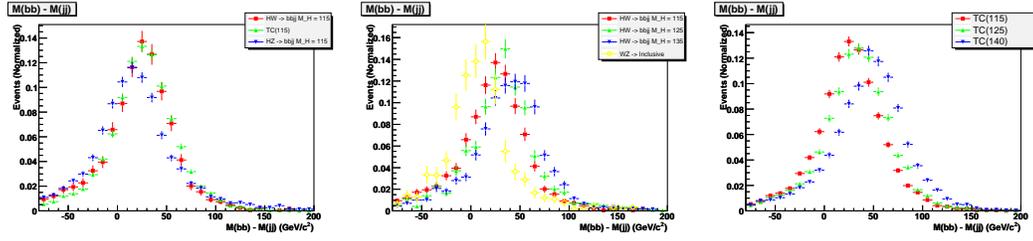
Figure 61: Various combinations of the M(bb)-M(jj) distributions after the preliminary cuts. The collection on the left represents the WH [red squares], ZH [blue triangles], and technicolor [green triangles] samples at a Higgs boson or technipion mass of 115 GeV/c^2. The middle plot shows the $WH(115)$ [red squares], $WH(125)$ [green triangles], $WH(135)$ [blue triangles], and the WZ [yellow circles] samples. The final plot on the right displays the $TC(115)$ [red squares], $TC(125)$ [green triangles], and the $TC(140)$ [blue triangles] samples. Remember that TC stands for the technicolor signal processes.
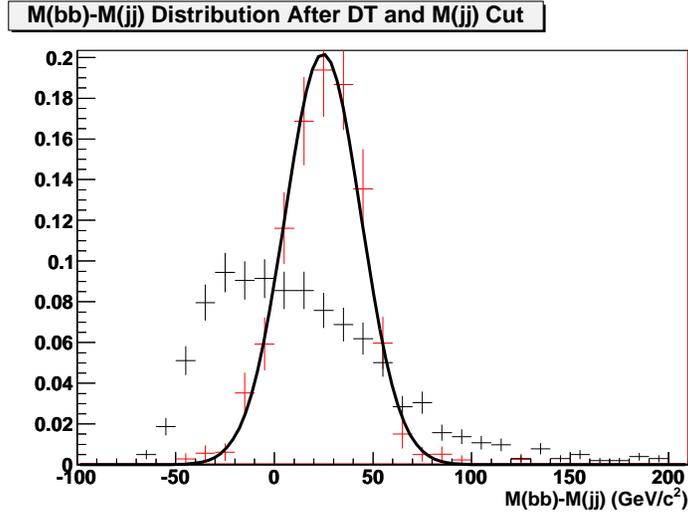
Figure 62: The M(bb)-M(jj) distribution in data after the DT and M(jj) cut is shown in black. The M(bb)-M(jj) distribution from WH(115) MC is shown in red with the smooth black line showing the fit described in the text.

8 Results

This chapter details the method by which the number of signal events was determined and subsequintly converted into a cross section. This cross section is then converted into both expected and observed 95% confidence level upper limits. The number of signal events is determined though a modified background subtraction technique where the background is determined from a data sideband region.

Analysis	Ch	Signal Eff (%)	Expected Signal	D_s	Sensitivity
VH(115)	WH	0.63 ± 0.04	0.59	477	0.039
	ZH	0.49 ± 0.02	0.27		
VH(125)	WH	0.81 ± 0.05	0.46	711	0.027
	ZH	0.73 ± 0.03	0.26		
VH(135)	WH	1.01 ± 0.05	0.33	638	0.020
	ZH	0.88 ± 0.03	0.18		
TC(115)	TC	0.37 ± 0.01	1.93	477	0.088
TC(125)	TC	0.42 ± 0.02	3.13	711	0.117
TC(140)	TC	0.42 ± 0.02	2.79	638	0.109

Table 13: Signal efficiencies, expected events, number of events in the signal region (D_s), and total sensitivities found after the mass window cuts.

8.1 Signal Determination

To determine the number of signal events in the DT signal region after all the cuts have been applied, a modified background subtraction technique is utilized. The background is determined by a scaled DT control region that compensates for the possibility of signal contamination. Since the WZ analysis does not have a DT signal and control region, a separate technique is used and is discussed in Section 9.3. The standard background subtraction is shown in Equation 40,

$$N_s = D_s - B_s \tag{40}$$

where N_s is the number of signal events, D_s is the number of events in the data in the region of interest, and B_s is the number of background events in the same region of interest.

A control region separate from this region of interest is chosen. One can scale the number of background events in this control region, B_b, to estimate the number of

114

events in the region of interest. The new form of the background subtraction is shown in Equation 41, where the SF is the background renormalization scale factor.

$$N_s = D_s - SF \times B_b \tag{41}$$

If the control region did not contain any signal, Equation 41 would be the final form of the signal estimation. However, care must be taken to include the possibility of signal contamination in the control region. The events in the control region can be separated into three classes: background, WH signal, and ZH signal. The relationship between these classes and the total numbers of events in the control region, D_b, are shown in Equation 42.

$$D_b = B_b + S_{b,W} + S_{b,Z} \tag{42}$$

$S_{b,W}$ represents signal contributions from the WH signal and $S_{b,Z}$ the ZH signal. It is easy to estimate the number of signal events from the WH signal in the control region by scaling the number in the region of interest by the ratio of signal efficiencies for the two regions, $R_W = \varepsilon_{b,W}/\varepsilon_{s,W}$ where the b subscript refers to the control region and the s subscript refers to the region of interest. The number of ZH events in the control region can be estimated in the same manner using a ratio, R_Z, built from the efficiencies for the ZH signal in the two regions. N_Z in the region of interest can be estimated from N_W by using the cross-section ratio, σ_{ZH}/σ_{WH}, as a constraint. Placing these substitutions in Equation 42, Equation 43 is found.

$$D_b = B_b + R_W N_{s,W} + R_Z \frac{\sigma_{ZH}}{\sigma_{WH}} N_{s,W} \tag{43}$$

After solving Equation 43 for the number of background events, Equation 44 is found.

$$B_b = D_b - R_W N_{s,W} - R_Z \frac{\sigma_{ZH}}{\sigma_{WH}} N_{s,W} \tag{44}$$

$$= D_b - R_W \gamma N_s \tag{45}$$

The γ in Equation 45 is introduced to collect the various signal efficiency and cross-section terms in order to simplify the final equations. The full form of γ is shown in Equation 46.

$$\gamma = \frac{1 + \frac{R_Z}{R_W} \frac{\sigma_{ZH}}{\sigma_{WH}}}{1 + \frac{\sigma_{ZH}}{\sigma_{WH}}} \tag{46}$$

With this in place and using the concept of scaling the background in the control region to the number in the signal region to estimate the background contribution we get Equation 47.

$$N_s = D_s - SF \left[D_b - R_W \gamma N_s \right] \tag{47}$$

Solving for the number of signal events, Equation 48 is found and this equation is used for the signal determination in all channels with the exception of the WZ channel.

$$N_s = \frac{D_s - (SF)D_b}{1 - (SF)(R_W)(\gamma)} \tag{48}$$

This equation can be checked by looking at the case where there is no signal contribution in the control region. In this case, the denominator is equal to one and the background is simply the scaled version of the number of events in the control region. This is exactly what is shown in Equation 40 where there was no signal contribution taken into account and $B_s = (SF)(D_b)$. The number of signal events found this way will be referred to as $N_s(nc)$.

The error for Equation 48 is found by propagating the errors from the five individual variables; D_s, D_b, SF, R_W, and R_Z. The propagated error on the cross section ratio is

assumed to be small and is therefore ignored. The constituent errors are assumed to be Gaussian due to the large number of events. It is possible to see which errors increase the most with the addition of the signal contamination. The ratio of the number of signal events with and without contamination is present in all of the terms for the relative error contribution. One just needs to look at the additional terms. The relative influence on the final result of the different terms can be seen in Equations 49 to 53 where the relative error has been written as a scaling of the original fractional error in each variable. The individual equations, expressed in this way, help to illustrate how each relative error scales to the full result.

$$\frac{\delta N_{s,D_s}}{N_s} = \left(\frac{D_s}{N_s}\right) \qquad\qquad \times \frac{N_s}{N_s(nc)} \sqrt{\left(\frac{\delta D_s}{D_s}\right)^2} \qquad (49)$$

$$\frac{\delta N_{s,D_b}}{N_s} = \left(\frac{D_s}{N_s} - (1 - (SF)(R))\right) \qquad\qquad \times \frac{N_s}{N_s(nc)} \sqrt{\left(\frac{\delta D_b}{D_b}\right)^2} \qquad (50)$$

$$\frac{\delta N_{s,SF}}{N_s} = \left(\frac{D_s}{N_s} - 1\right) \qquad\qquad \times \frac{N_s}{N_s(nc)} \sqrt{\left(\frac{\delta SF}{SF}\right)^2} \qquad (51)$$

$$\frac{\delta N_{s,R_W}}{N_s} = \left(\left(\frac{\sigma_{WH}}{\sigma_{WH} + \sigma_{ZH}}\right)(SF)(R_W)\right) \qquad\qquad \times \frac{N_s}{N_s(nc)} \sqrt{\left(\frac{\delta R_W}{R_W}\right)^2} \qquad (52)$$

$$\frac{\delta N_{s,R_Z}}{N_s} = \left(\left(\frac{\sigma_{WH}}{\sigma_{WH} + \sigma_{ZH}}\right)(SF)(R_Z)\right) \qquad\qquad \times \frac{N_s}{N_s(nc)} \sqrt{\left(\frac{\delta R_Z}{R_Z}\right)^2} \qquad (53)$$

The order of the relative errors in Equations 49, 50, 51, 52, and 53 are such that the largest scaling is at the top and the smallest is at the bottom. While the largest shift in the relative error is on D_s, the SF contributes the largest relative error in the final result due to the intrinsic error on the scale factor. Assuming that there are approximately 20 signal events in the region of interest and 500 data events in that same region, $D_s/N_s \simeq 25$. If one also assumes that the scale factor is 0.3 and the signal contamination is 0.5, the combination of $(SF)(R) \simeq 0.15$. Combining all of these together into a single error is shown in Equation 54.

$$
\frac{\delta N_s}{N_s} = \frac{N_s}{N_s(nc)} \left[\left(\frac{D_s}{N_s} \right)^2 \left(\frac{\delta D_s}{D_s} \right)^2 \right.
$$
$$
+ \left(\frac{D_s}{N_s} - 1 + (SF)(R) \right)^2 \left(\frac{\delta D_b}{D_b} \right)^2
$$
$$
+ \left(\frac{D_s}{N_s} - 1 \right)^2 \left(\frac{\delta SF}{SF} \right)^2
$$
$$
+ \left(\left(\frac{\sigma_{WH}}{\sigma_{WH} + \sigma_{ZH}} \right) (SF)(R_W) \right)^2 \left(\frac{\delta R_W}{R_W} \right)^2
$$
$$
+ \left. \left(\left(\frac{\sigma_{WH}}{\sigma_{WH} + \sigma_{ZH}} \right) (SF)(R_Z) \right)^2 \left(\frac{\delta R_Z}{R_Z} \right)^2 \right]^{\frac{1}{2}} \tag{54}
$$

While the values of D_s and the D_b are found from the numbers of events passing the cuts on the 2tag data set, R_W, R_Z, and SF are derived quantities. The ratios, R_i are found by dividing the efficiency of the signal Monte Carlo representing the i process in the region of interest by the efficiency of the same signal Monte Carlo in the control region. We then optimize the choice for the DT control region so as to minimize the total error on the number of signal events. We don't want the sideband to be too far from the signal region so we only examine DT regions in the neighborhood of the signal. In addition to being close to the signal region, the correlation between the M(bb)-M(jj) variable needs to be flat which requires us to only investigate DT regions larger than a minimum value. This is determined by eye and for the $WH(115)$ decision tree this minimum value is approximately -0.1. We require there to be at least twice as many events in the control region as the region of interest. Then, we optimize based on the contamination fractions by chosing the region that has the lowest contamination. The sidebands chosen are shown in Figures 55, 71, and 72.

The inputs and the results of the N_s calculation for each of the three VH mass points along with the Technicolor samples is shown in Table 14. For the discussion concerning the derivation in the scale factor please see Section 8.2.

Table 14: Inputs in the calculation of N_s.

Analysis	D_s	D_b	SF	R_W	R_Z	N_s
VH(115)	477 ± 22	1784 ± 42	0.24 ± 0.01	0.24 ± 0.01	0.64 ± 0.05	$54 \pm 25 \pm 26$
TC(115)	"	"	"	0.79 ± 0.04	N/A	$58 \pm 27 \pm 28$
VH(125)	711 ± 27	1940 ± 44	0.34 ± 0.01	0.55 ± 0.06	0.55 ± 0.04	$54 \pm 33 \pm 36$
TC(125)	"	"	"	0.91 ± 0.04	N/A	$63 \pm 39 \pm 42$
VH(135)	638 ± 25	1346 ± 37	0.44 ± 0.02	0.52 ± 0.05	0.56 ± 0.03	$63 \pm 33 \pm 38$
TC(140)	"	"	"	1.00 ± 0.05	N/A	$87 \pm 45 \pm 50$

8.2 Scale Factor Determination

The scale factor, SF, is a multiplier used to determine the number of background events in the region of interest (ROI) from the control region. While there are several different techniques available to determine a scale factor, the technique used in this analysis utilizes an inverted cut in the $M(bb) - M(jj)$ variable in conjunction with an ensemble technique. In order to continue this description, several regions in the $M(bb) - M(jj)$ vs DT Result plane will need to be defined. These regions can be seen in Figure 63.
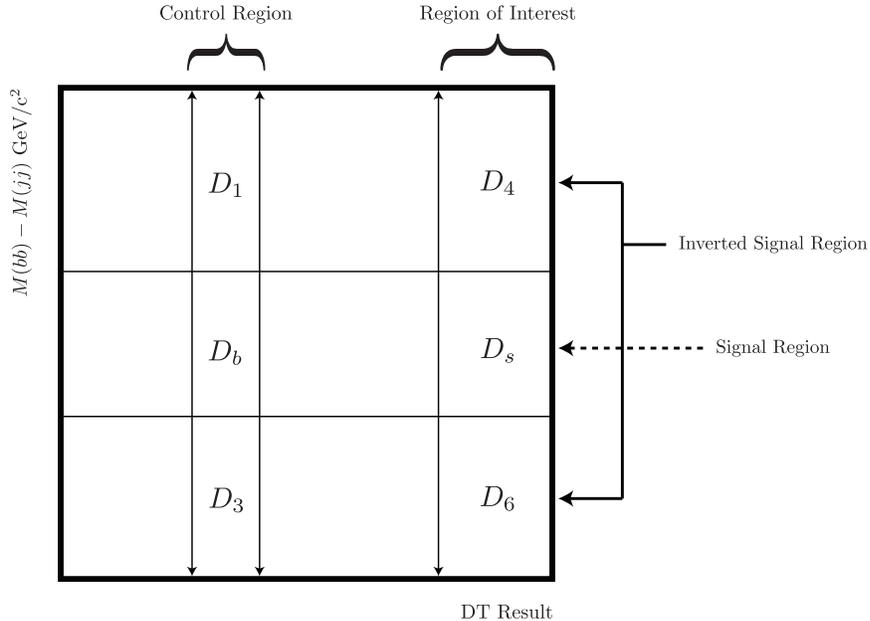
Figure 63: Regions required for the determination of the scale factor (SF) used in the analysis

The regions defined in Figure 63 refer to the event sample after the preliminary

cuts and the $M(jj)$ mass cut. Region 2 is the DT control region while region 5 is the DT region of interest. The scale factor will be used to scale the integrated number of events in region 2 with the expectation that this new scaled number will represent the number of background events in the region of interest, region 5. At this point the scale factor could be defined simply as $SF = D_s/D_b$, where R_i is the number of events in region i, but using the actual signal region in the determination of the SF is not a valid method. We must also account for the fact that there is signal in both regions. To address both of these issues simultaneously, we can derive the scale factor using data from $M(bb) - M(jj)$ sidebands. With this modification, the new scale factor is shown in Equation 55

$$SF = \frac{D_4 + D_6}{D_1 + D_3} \tag{55}$$

In order for this to be successful there should not be large deviations in shape between the $M(bb) - M(jj)$ distributions in the region of interest and the control region in DT. The $M(bb) - M(jj)$ distribution for the control region and the region of interest can be seen in Figure 64 and there is good agreement. In this case, the DT signal region is from 0.085 to 1.0 and the control region is defined from 0 to 0.04 and can be seen in Figure 55.

In order to obtain an uncertainty for this scale factor, which will allow for a systematic error to be placed on the final result, an ensemble technique is performed. Using a Poisson distribution, 10000 random values are chosen for the nominator and the denominator of Equation 55. The parameter for the Poisson distribution is $D_4 + D_6$ for the nominator while $D_1 + D_3$ is used for the denominator. For each iteration of random pulls, a scale factor is determined and added to a collection of scale factors. The final scale factor used in the analysis is the mean value of this distribution of scale factors while the error is set to the standard deviation of the same collection.

This technique can be validated by performing the same procedure in a signal-free
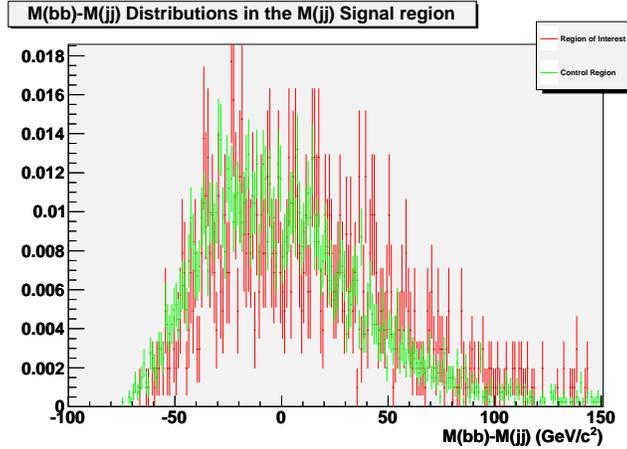
Figure 64: The $M(bb) - M(jj)$ distributions for the control region ($0 \leq$ DT < 0.04) and the region of interest (DT ≥ 0.085) from the $WH(115)$ trained decision tree in the data.

region. In this analysis, this can be accomplished by using the $M(jj)$ sideband regions [$40 \rightarrow 60$ and $100 \rightarrow 120$] instead of the $M(jj)$ signal region [$61.4 \rightarrow 96.6$]. If the difference in the predicted number of events and the actual number of events for D_5 in this independent signal-free region is consistent with zero, then one can say that the technique is valid. As an example, the SF derived for the $VH(115)$ mass point using the $M(jj)$ sideband region instead of the $M(jj)$ signal region is 0.24 ± 0.01. In region 2 for this configuration, there are 1552 events. Multiplying this number of events by the scale factor, the estimated number of events in the signal region (D_s) is 444±24. There are 412±20 events in D_s. The difference in this predicted and actual event count for region 5, called the Δ cross-check, is 26±30 events. This cross check validates the scale factor procedure since 26 ± 30 is consistent with zero. Table 15 lists the scale factors and the validation cross-check for each of the HV mass points, in the Higgs analyses.

Higgs Mass (GeV/c^2)	Scale Factor	Δ Cross-Check
115	0.24±0.01	26±30
125	0.34±0.01	8±35
135	0.44±0.02	5±35

Table 15: Scale factors used for each of the different mass points in the HV analysis along with the validation cross-check discussed in the text.

121

Table 16: Cross section inputs and results. The Technicolor channel efficiencies are for the signal Monte Carlo and the branching ratio for the technipion is set to 1.0 since it was decayed inclusively. The signal efficiency reported in the Technicolor samples is the efficiency for the Technicolor signal Monte Carlo to pass the decision tree and invariant mass cuts even though the column header states it is for the WH or ZH samples.

Analysis	N_s	$\varepsilon_{WH}(\%)$ $\varepsilon_{ZH}(\%)$	$B_H(H \to bb)$	σ (pb)
$VH(115)$	$53.8 \pm 25.1 \pm 26.4$	$0.63 \pm 0.04 \pm 0.13$ $0.49 \pm 0.02 \pm 0.06$	0.73	$9.42 \pm 4.40 \pm 4.83$
$VH(125)$	$53.5 \pm 32.9 \pm 36.0$	$0.63 \pm 0.04 \pm 0.13$ $0.49 \pm 0.02 \pm 0.06$	0.61	$8.23 \pm 5.06 \pm 5.64$
$VH(135)$	$63.4 \pm 33.0 \pm 37.6$	$1.01 \pm 0.05 \pm 0.17$ $0.88 \pm 0.03 \pm 0.11$	0.46	$10.44 \pm 5.43 \pm 6.33$
TC(115)	$58 \pm 27 \pm 28$	$0.37 \pm 0.01 \pm 0.06$	1.0	$22 \pm 11 \pm 12$
TC(125)	$63 \pm 39 \pm 42$	$0.42 \pm 0.03 \pm 0.08$	1.0	$22 \pm 13 \pm 15$
TC(140)	$87 \pm 45 \pm 50$	$0.42 \pm 0.02 \pm 0.07$	1.0	$30 \pm 16 \pm 18$

8.3 Cross Section Results

The main result for this analysis is the cross section for VH inclusive production. A sensitivity for a WV signal is presented in Section 9. The VH cross sections are shown in Table 16. The cross sections are determined from Equation 56.

$$\sigma_{\text{Higgs}} = \frac{N_s}{\mathcal{L}(\varepsilon_{WH}\text{B}_W\text{B}_H + \varepsilon_{ZH}\text{B}_Z\text{B}_H)} \tag{56}$$

For all channels, the luminosity, $\mathcal{L}$, is (1007 ± 61) pb^{-1} and the W and Z branching ratios used are $B_W =$BR$(W \to jj)$=0.68 and $B_Z =$BR$(Z \to jj)$=0.69.[9] In the table, N_s is the number of signal events that was found in Section 8.1. The signal efficiencies, ε_{WH} and ε_{ZH}, include the acceptance only. B_H is BR$(H \to bb)$ as found in the references[15]. The first error in σ is statistical and the second is systematic.

The efficiencies are found by counting the events in the signal MC after all of the cuts for a particular analysis and dividing by the number of generated events. The first error shown on the efficiency includes the statistical error from the MC. The second error is the systematic and is discussed below. The first error on the cross section is determined from the statistical error on N_s. The systematic errors on N_s, $\mathcal{L}$, and ε are added in quadrature and used as the systematic error on the cross section. The errors on N_s are discussed in Section 8.1 and include all background effects. The error on the luminosity is set to the prescribed DØ value for RunIIa as 6.1%.

Systematic effects that we examine on ε come from:

- Jet energy scale

- Jet identification

- Jet resolution

- b jet tagging

- b jet taggability

- Trigger

The prescribed DØ methods are used to find each of these errors. For this, there is a correction made and the Monte Carlo sample is reprocessed using this correction. We then recalculate the efficiency for each signal MC sample after this correction. For each source of uncertainty, we take the largest variation in the efficiency, subtract the nominal efficiency and divide by the nominal efficiency to find the percentage error.

The jet energy scale, jet identification, and jet resolution effects are determined by varying the Monte Carlo signal sample by $\pm 1\sigma$ in each of the three variants according to experiment wide procedures.[68]. The jet identification systematic only has a positive fluctuation and therefore the magnitude of the systematic is based on only this positive fluctuation. The b-tagging systematic is determined by altering the b-tag weight

Table 17: Results of a Gaussian Fit to the M(bb)-M(jj) distribution from the WH(115) signal MC after reprocessing with the prescribed changes for each systematic studied.

Variant	Mean (GeV/c^2)	Sigma (GeV/c^2)
Nominal	25.2 ± 1.7	18.5 ± 1.7
JES Pos	26.9 ± 1.6	17.9 ± 1.5
JES Neg	24.9 ± 1.3	19.0 ± 1.3
RES Pos	26.7 ± 1.9	18.4 ± 1.7
RES Neg	24.9 ± 1.4	17.3 ± 1.3
Jet ID	25.7 ± 1.3	18.8 ± 1.2
BTAG Pos	25.2 ± 2.3	18.5 ± 2.0
BTAG Neg	25.2 ± 1.8	18.5 ± 1.8
Taggability	25.3 ± 1.5	18.7 ± 1.5

applied for the direct b-tagging method by its $\pm 1\sigma$ statistical error and re-evaluating the efficiency as mentioned previously. Taggability is accomplished in much the same way with the exception that it has only a positive variation. Systematic errors for the trigger efficiency were determined from the p17 MSSM multijet analysis upon which the trigger weight code was adopted [62]. The systematic levels found for the trigger in that analysis were approximately 5%. This 5% level has therefore been applied to each of the Higgs mass points.

We check that the signal shapes are unaffected with each of these changes, so that the systematic errors are "flat". Here, we fit the M(bb)-M(jj) signal shapes for the WH(115) MC after each of the systematic shifts to a Gaussian shape. The mean and sigmas are found for each sample and are given in Table 17 and they are shown in Figure 65. The shapes are consistent with each other. This is true for all of the signal samples that are used in this analysis.

The errors from each of the sources are added in quadrature to form a total systematic error on the efficiency. The systematic errors are summarized in Table 18. The WH(115) efficiency has a systematic error of 20.2%.

A 95% C.L. upper limit for each mass point is calculated. A Gaussian distribution is assumed with the mean set to the central value and the width set to the quadrature sum of the statistical and systematic errors on the cross section. The observed limit is

124
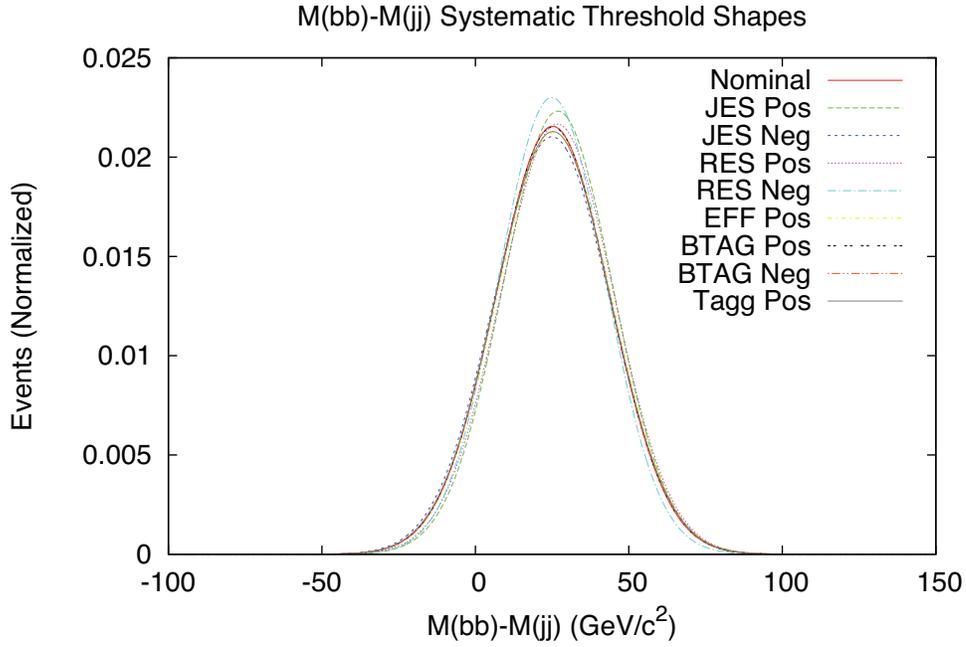
Figure 65: Gaussian fit results for systematic variations to the WH(115) signal MC for the M(bb)-M(jj) invariant mass.

then calculated to be that point which corresponds to 95% of the positive integrated probability region. The expected limit is found in the same mannor by using the same error but setting the mean to zero to represent a background only hypothesis. Table 19 shows the expected and observed limits found for each of the analyses.

The sensitivities presented in Table 19 are smaller than those found after the mass window cuts shown in Table 13. Recal that the sensitivity is defined as the predicted of signal events using the theoretical cross section divided by the estimated error in the signal content. The signal error in this stage of the analysis includes the systematic effects from the background estimation associated with the N_s determination. Table 20 traces the sensitivities found throughout each analysis.

Table 18: Efficiency systematic errors in (%) for each of the analyses.

Channel	JES	Jet ID	Jet Resolution	b-Tag	Taggability	Trigger	Total
$WH(115)$	9.1	8.8	13.8	5.1	2.1	5.0	20.2
$ZH(115)$	1.5	1.6	3.8	5.3	8.8	5.0	12.2
$WH(125)$	9.9	11.3	8.6	5.0	1.8	5.0	18.8
$ZH(125)$	5.0	0.1	4.4	5.3	7.1	5.0	12.2
$WH(135)$	6.3	10.0	10.2	5.2	1.0	5.0	17.2
$ZH(135)$	3.3	2.5	2.7	5.3	8.6	5.0	12.4
WZ	21.4	13.8	11.9	5.6	2.4	5.0	29.2
TC(115)	6.6	1.9	3.2	5.3	13.0	5.0	16.7
TC(125)	7.5	2.5	3.5	5.3	14.7	5.0	18.6
TC(140)	6.5	2.4	5.5	5.1	14.0	5.0	18.0

Table 19: Cross section expected and observed 95% C.L. limits and final sensitivities. The limits are presented in two units. The first is in picobarnes (pb) and the second is in the units of the predicted cross section ($\times$ SM for the VH samples).

			95% C.L. Upper Limit			
			Expected		Observed	
Analysis	σ(pb)	Sensitivity	(pb)	($\times$ Pred.)	(pb)	($\times$ Pred.)
VH(115)	$9.42 \pm 4.40 \pm 4.83$	0.023	12.8	43.7	20.4	69.6
VH(125)	$8.23 \pm 5.06 \pm 5.64$	0.022	9.9	45	16.7	76.1
VH(135)	$10.44 \pm 5.43 \pm 6.33$	0.010	16.3	98.3	24.6	148.0
TC(115)	$22 \pm 11 \pm 12$	0.048	31	41	49	65
TC(125)	$22 \pm 13 \pm 15$	0.053	40	37	57	52
TC(140)	$30 \pm 16 \pm 18$	0.040	47	50	71	74

Table 20: Sensitivities found after each stage of the analysis.

Analysis	Preliminary Cuts	Decision Tree	Mass Window Cuts	Background Estimation
VH(115)	0.017	0.027	0.039	0.023
VH(125)	0.012	0.020	0.027	0.022
VH(135)	0.008	0.014	0.020	0.010
WV	0.026	NA	0.054	0.029
TC(115)	0.045	0.057	0.088	0.048
TC(125)	0.065	0.075	0.117	0.053
TC(140)	0.056	0.068	0.109	0.040

9 WZ Analysis

9.1 Theory

The W and Z bosons of the Standard Model are a required byproduct of the gauge group that defines the electroweak sector, $SU(2) \times U(1)$. The construction of a covariant derivative after the breaking of the electroweak symmetry through a choice of a Higgs field vev, Equation 57, for this gauge group contains two sets of fields: a W_μ field and a B_μ field. The η in Equation 57 is the Higgs field while the v is the chosen vev. Linear combinations of these fields are then combined to produce the standard $W^\pm$ and Z bosons. An additional linear combination accounts for the photon that mediates the electromagnetic interactions. In the Higgs boson discussion presented previously, the Higgs potential was used almost exclusively to derive the results.

$$D_\mu = \partial_\mu \eta + \frac{i}{2}(v + \eta)\left(g'B_\mu + g\tau^i W_\mu^i\right) \tag{57}$$

Using the covariant derivative shown in Equation 57 and the first term in Equation 1, the Lagrangian (without the potential) can be represented by Equation 58 [2].

$$\left(D_\mu \phi'\right)^\dagger \left(D_\mu \phi'\right) + \text{h.c.} = \frac{1}{2}\left(\partial_\mu \eta\right)\left(\partial^\mu \eta\right) + \frac{1}{8}(v+\eta)^2 \chi^+ \left\{(g^2 + g'^2)Z_\mu Z^\mu + 2g^2 W_\mu^+ W^{-\mu}\right\}\chi \tag{58}$$

In Equation 58 several substitutions and simplifications were made. The mass terms for the $W^\pm$ and the Z may be found by looking for terms in the expansion of Equation 58 that are quadratic in the field while the interaction vertices can be found by looking for higher orders in the fields. For a vector field the mass terms have the form of $m^2 F^2$ where F represents the field. In the case of the W, the mass becomes $m_W = \frac{1}{2}vg$. For the Z boson the mass becomes $m_Z = \frac{1}{2}\sqrt{g^2 + g'^2}\nu$.

The production of ZZ, WW, and WZ pairs in the final states comes from two different sources. The first of these is through s-channel processes using the WWW,

WWZ, and ZZZ vertices which are shown in Figure 66. These interactions come from the field strength component of the electroweak Lagrangian [69]. The second of these production modes is the result of an exchange and utilizes the boson to fermion couplings which are also shown in Figure 66.

(a) $W^\pm \to W^\pm Z$ (b) $Z \to W^+ W^-$ (c) $W^\pm \to ff$ (d) $Z \to ff$
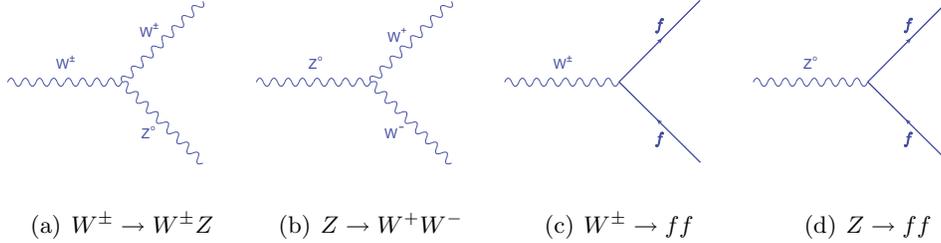
Figure 66: Tree level electroweak triple gauge boson interaction vertices and the tree level electroweak gauge boson fermion vertices.

While the diboson final states can be produced in a number of ways, the production mode used in this analysis is an s-channel process that closely follows that shown in Figure 1. In this process the two fermions from the $p\bar{p}$ merge to form a $W^\pm$ boson propagator. This then couples to the WWZ vertex to produce the final state. The production rates for the dibosons range from 13.5 pb down to 1.6 pb [70]. The largest production rate is for WW and ranges between 12 and 13.5 pb. The WZ process follows the WW process with a production cross section of 3.7±0.3 pb while the ZZ has the lowest production cross section of 1.6±0.1 pb [70].

The branching ratio for the $W^\pm$ boson can be split into two major categories: leptonic at $\simeq 33\%$ and hadronic at $\simeq 67\%$. From world experimental averages, the W boson will decay to an electron plus its neutrino (10.75±0.13)% of the time while the muon and tau variants are (10.57±0.15)% and (11.25±0.20)% respectively [9]. The hadronic (or jet) decay mode accounts for the largest branching fraction with a value of (67.60±0.27)% [9]. A pie chart of these branching fractions is shown in Figure 67. While the hadronic branching fractions for the Z boson is similar to the W boson with a rate of (69.91±0.06)%, the leptonic decay modes are dramatically reduced. In order to maintain charge conservation and not violate lepton number in the decay process, the
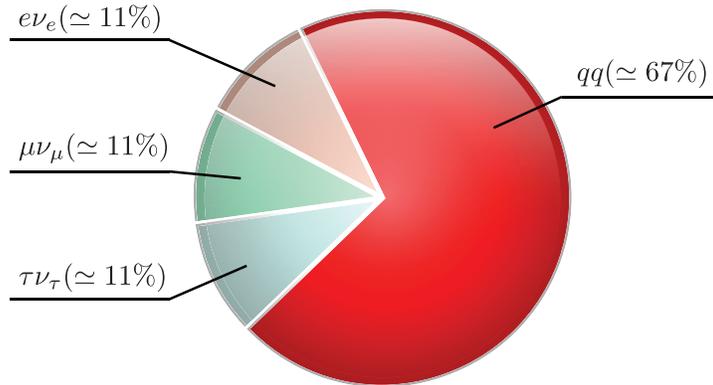
Figure 67: Branching ratios for the various W boson decay modes [9].

Z boson is only allowed to have leptonic decays containing two leptons from the same family with opposite charges. The LEP electroweak working group averages for the $e\bar{e}$ decay accounts for $(3.363\pm0.004)\%$ of the decay modes while the $\mu\bar{\mu}$ and $\tau\bar{\tau}$ modes come at a rate of $(3.366\pm0.007)\%$ and $(3.370\pm0.008)\%$ respectively. These four rates for the Z decay do not add to 100% like one would think. This is due to a large fraction of decays ($\simeq$20%) being considered as invisible. An invisible decay process is a decay with two neutrinos and are therefore undetectable in the detector [9].

9.2   Historical Survey

The final states of the Higgs boson searches are very similar to the final states present in WW, WZ, and ZZ production. To date, all of these processes have been observed at either the CDF or the DØ experiment at the Tevatron. These analyses utilize the fully leptonic decay modes of the two produced bosons and hence rely on the clean leptonic signature. There are several different analyses designed around a semi-leptonic decay, where one of the bosons decays to leptons while the other decays to quarks. These semi-leptonic analyses have been able to produce only upper limits on production and in some cases only had the ability to set limits on the combination $WZ + WW$ due to resolution issues.

The WW diboson signal was observed in 2005 using 252 pb^{-1} of data from the DØ detector [71]. The analysis was designed to use the fully leptonic decay which consisted of two leptons and two neutrinos in the final state. This study was split into the three different sub-analyses based on the lepton flavor; ee, $e\mu$, and $\mu\mu$. The general cuts required in the preliminary stage primarily consisted of two oppositely charged leptons in the final state. The leading lepton in p_T must have a $p_T \geq 20$ GeV/c while the other lepton must have a $p_T \geq 15$ GeV/c. The backgrounds for this analyses come from Z/γ, $t\bar{t}$, WZ, ZZ, and multijet production. In order to suppress the Z contribution, a Z mass window veto cut is applied to the dilepton mass. To suppress the $t\bar{t}$ backgrounds, the scalar sum of the transverse energy in the jets of the event must be below 50 GeV. The multijet background is estimated from data using an orthogonal sample derived by inverting the lepton quality cuts and requiring like-signed leptons. With these cuts in place, the ee channel has 6 data events with an expected 3.42± 0.05 signal events and an expected background of 2.30±0.21 events. The $e\mu$ channel found 15 data events while expecting 11.10±0.10 and 3.81±0.17 events for the signal and background, respectively. The final analysis channel, the $\mu\mu$, is similar to the ee channel with 4 data events and 2.10±0.05 and 1.95±0.41 events for the signal and background, respectively. With a systematic uncertainty of 9 to 12% depending on the analysis channel, the final result of the combined channels using a negative log likelihood technique set the cross section for WW production to be $\sigma(WW) = (13.8^{+4.3}_{-3.8}(\text{stat})^{+1.2}_{-0.9}(\text{syst}) \pm 0.9(\text{lumi}))$ pb. The error on the luminosity is determined directly from the uncertainty in the total integrated luminosity and represents an overall scaling of the data and therefore the cross section. This result has a significance level of 5.2σ which signifies the observation of the process.

While there have been several intermediate analyses performed attempting to determine an improved cross-section result for the WW cross-section, DØ currently has the best measurement with a cross section of $\sigma(WW) = (11.5 \pm 2.1 \pm 0.7)$ pb [72]. This analysis utilized the same decay channels as the previous DØ analysis outlined in the previous paragraph. A major change between the two analyses is the amount of data

utilized. This newer analysis takes into account ~ 1 fb^{-1} while the previous analysis used only 0.25 fb^{-1}. With the addition of more data for the analysis, it was found that the p_T of the WW system in the $t\bar{t}$ backgrounds (and other backgrounds) could be used to increase the sensitivity of the analysis.

The ZZ process was observed at the DØ experiment using the fully leptonic decay where the ZZ decays to $\ell\ell\ell\ell$ [73]. Just as the previous analysis was split into different lepton groups, this analysis is split into three different sub-analyses; $eeee$, $\mu\mu\mu\mu$, and $ee\mu\mu$. The cuts for this analysis are similar for the three different channels. Each channel requires four isolated leptons with a p_T cut of 30 GeV/c for the leading lepton, 25 GeV/c for the second leading lepton and requiring the remaining two to be greater than 15 GeV/c. For the four like flavor lepton channels there must exist a combination of leptons that have an invariant mass greater than 70 GeV/c^2 while the invariant mass of the remaining two leptons is greater than 50 GeV/c^2. In the $ee\mu\mu$ channel this is slightly modified. The invariant mass of the ee pair and the $\mu\mu$ lepton pair are computed without ambiguity. For these two masses, the leading mass must be greater than 70 GeV/c^2 while the other mass must be greater than 50 GeV/c^2. In addition to that, the $ee\mu\mu$ channel imposes an additional angle requirement on the final state leptons to remove $Z \rightarrow \mu\mu$ events in which the μ's emit photons which fake electrons in the calorimeter. The backgrounds for the analysis come from $t\bar{t}$, and W/Z+Jets, and multijet production. The $t\bar{t}$ and other resonant backgrounds are simulated from Pythia while the multijet contribution is determined from the data. This analysis used 1.7 fb^{-1} of data and reports the cross section for ZZ to be $\sigma(ZZ) = (1.75^{+1.27}_{-0.86}(\text{stat}) \pm 0.13(\text{cyst}))$ pb [73]. This result was then combined with the previous version of this analysis and an additional separate analysis to derive a final cross-section of $\sigma(ZZ) = (1.60 \pm 0.63^{+0.16}_{-0.17})$ pb with a significance of 5.7 [73]. To date, this is the best measurement of the ZZ cross-section.

The most recent measurement to come out of the Tevatron complex for $\sigma(ZZ)$ is from the CDF experiment [74]. This analysis uses the largest data set to date of 4.8

131

fb^{-1}. The cuts for the analysis require four isolated final state leptons in the detector with the $p_T \geq 10$ GeV/c. There is an additional requirement that the leading lepton in p_T be greater than or equal to 20 GeV/c. The invariant mass of the different dilepton combinations of the four leptons are then calculated. One of the masses must be in the mass window of 76 to 106 GeV/c^2 while the other dilepton combination must have an invariant mass between 40 to 140 GeV/c^2. Once this requirement has been met, each lepton pair is cross-checked to make sure that each pair has the same lepton flavor and opposite charges. The backgrounds for the analysis are mainly Z+jets and $Z\gamma$+jets while there is a small contribution from $t\bar{t}$ and WZ signal processes. With these cuts in place the data are reduced to 5 candidate events with 4.15±0.02±0.76 expected signal events and 0.042±0.016±0.029 background events. This translates to a cross section of $\sigma(ZZ) = (1.56^{+0.80}_{0.63}(\text{stat}) \pm 0.25(\text{syst}))$ pb with a significance of 5.7σ.

The analyses, and production observations, discussed above have been accomplished using the leptonic decay modes of both the two final state bosons in the event. Due to the small branching fraction of the $W^{\pm}$ and the Z to leptons ($\simeq 10\%$) and the much larger hadronic decay ($\simeq 70\%$), the next logical step is to investigate the semi-leptonic decay channels. In the semi-leptonic decay one of the bosons decays to leptons while the other boson in the final state decays hadronically. This search has been done by both the CDF and the DØ collaboration. In both of these analyses the search was for WV production where the V is either a $W^{\pm}$ or a Z. The CDF analysis reports an upper limit on the cross-section times the branching ratio of 2.88 pb using 1.2 fb^{-1} of data with a significance of 1.7σ [75]. Using 1.07 fb^{-1} of data, the DØ analysis was able to produce a result with a significance of 4.4 which is just shy of observation [76]. The derived cross-section in the DØ analysis for $WW + WZ$ was found to be (20.2±4.5) pb. It should be noted that this represents an inclusive cross-section and can not be directly compared to the CDF analysis without correcting for the corresponding branching ratios. The DØ result is in good agreement with the predicted Standard Model cross section of (16.1±0.9) pb.

9.3 Results

In Section 7.1 we found that the DT did not help increase sensitivity for finding a WZ signal. We examine the potential signal for this channel by placing the dijet invariant mass cuts directly after the preliminary cuts. In Section 6.2 we derived the expected numbers of W's and Z's from several samples of Monte Carlo signals. We are sure we did not include all of the potential sources. However, our bottom up estimates give of order 300 W's and 60 Z's after the preliminary cuts. We also explored using a two dimensional fit in M(jj) versus M(bb). The background shapes couldn't be controlled very well and the number of parameters skyrocketed so that the fits did not converge. Here, we try various methods of reducing backgrounds further to see if a result is possible.

If we make the M(jj) mass window cut discussed previously, we find 25606 events in the W signal region after the preliminary cuts. We define W sidebands to be $45 \leq M(jj) < 60 \, \mathrm{GeV}/c^2$ and $90 \leq M(jj) < 105 \, \mathrm{GeV}/c^2$ which have a total number of events similar to that of our signal region. There are 23495 events in the W sidebands after preliminary cuts. We determine the number of "W" events using a straight sideband subtraction assuming no signal contamination to be, 2111 ± 222. Expectations are that approximately 70% of the true W signal resides in the signal region, but there is significant W contamination in the sidebands. Also, the W and Z signal regions overlap so there are a significant number of Z events in this region, but also Z contamination in the high sideband. Since there are $Z \to jj$ and $W \to bc$ decays in our sample, it is unclear that the two non b-tagged jets should be associated with the W while the two b-tagged jets are associated with a Z. There also could be WW and ZZ in the signal as was found from our bottom up analysis. However, the bulk of the W's in the sample are from top events including single top samples. So our analysis becomes quite murky for the WZ signal. Figure 68 shows results to a fit of the M(bb) mass in the W signal region. We find 602 ± 145 WZ events assuming the signal shape defined by the WZ MC. Clearly, there are events in the Z signal region, but it is completely unclear whether they are Z's, W's, or something else. The confidence level of the fit is also low.

Figure 68: Fit to M(bb) distribution in the data after the preliminary cuts and a cut on the M(jj) mass in the W signal region. The background shape used was from the W sidebands and the signal shape is from the WZ signal MC sample.

We proceed further by making a box cut to enhance the Z region assuming that the two b-jets are from the Z. This signal region is defined as $60 \leq M(bb) < 100$ GeV/c^2. We also define sideband regions from $40 \leq M(bb) < 60$ GeV/c^2 and $100 \leq M(bb) < 120$

134

GeV/c^2. We then can use a two-dimensional technique to obtain the number of signal events which is described in the Appendix (Section C). Using this technique, we find 2843 ± 204 events in the full "9-square" signal region assuming signal contamination in the surrounding boxes and 1378 ± 110 events in the case with no signal contamination. The WZ efficiency found for the total 9-square region is $\simeq 1.6\%$. Using the statistical error of 204 and the expected signal of 6 events, the expected sensitivity is 0.029. Using the result without contamination, one finds 0.054. This assumes that the efficiency can be defined as that for the $WZ \rightarrow$ Inclusive SM only (a dubious assumption). The Monte Carlo generated for the WZ sample was totally inclusive for each of the W and Z decays, therefore, we just determine the efficiency by dividing the number of observed events after cuts by those generated and don't correct for any branching ratios. If one just takes the central box for the signal region, there are 9198 data events. Subtracting the M(jj) sideband regions assuming no contamination gives 1040 ± 132 events with a WZ efficiency of $(0.93 \pm 0.04)\%$ for an expected 3.5 signal events. The expected sensitivity is 0.027, which is found by dividing 3.5 by 132. We find that the two dimensional sideband technique has a better expected sensitivity than with the preliminary cuts alone and is also better than the DT. Using the central box alone provides worse expected sensitivity compared to the two dimensional sideband subtraction, but is equivalent to the DT cut. These sensitivities however are not competitive with analyses from other channels.

10   Conclusion

The *qqbb* four jet final state, while difficult due to multijet backgrounds, presents an interesting final state to study at high energy colliders. This final state is shared by many production processes that are actively being explored and are near either exclusion or discovery. In this analysis, this final state has been exploited to search for a light Higgs boson, WZ diboson production, and evidence for Technicolor. The data acquired for this search were collected between 2002 and 2006 at the DØ detector at the Tevatron accelerator near Chicago, Illinois, and includes a total integrated luminosity of 1007 pb^{-1}. This analysis utilizes a boosted decision tree to reduce background and the final variable used is the difference in the $b\bar{b}$ and the jj invariant masses, M($b\bar{b}$)-M(jj). The Standard Model Higgs boson search performed targets three individual mass points; 115 GeV/c^2, 125 GeV/c^2, and 135 GeV/c^2. Using the procedures outlined in Section 8, the cross sections and the 95% C.L. upper limits have been found for each of these mass points. The 115 GeV/c^2 mass point yields a cross section of (9±4±5) pb producing a 95% C.L. expected upper limit of 12.8 pb and an observed limit of 20.4 pb. The 125 GeV/c^2 and 135 GeV/c^2 mass points correspond to cross sections of (8±5±6) pb and (10±5±6) pb, respectively. These cross sections correspond to 95% C.L. expected (observed) upper limits of 9.9 pb (16.7 pb) for the 125 GeV/c^2 mass point and 16.3 pb (24.6 pb) for the 135 GeV/c^2 mass point.

The most current Tevatron combination analysis reports a combined 95% C.L. expected upper limit of $\simeq$ 2.80 pb with an observed limit of $\simeq$ 4.05 pb for a Higgs boson mass of 115 GeV/c^2 [51]. The 115 GeV/c^2 Higgs boson mass limit of the Tevatron combination has placed a 95% C.L expected (observed) upper limit of 1.78 (2.70) times the Standard Model cross section which can be compared with the upper limit of 43.7 (69.6) times the Standard Model found in this analysis. One of the strongest channels from DØ being used in the combination is the $WH \rightarrow \ell\nu b\bar{b}$ analysis [49]. Using 5 fb^{-1} of data, the $WH \rightarrow \ell\nu b\bar{b}$ analysis has 6.5 expected signal events compared to 707 data

events in the $W + 2$ Jets with 2 b-tags channel which translates to a sensitivity of 0.25 [77]. The final sensitivity of the $qqbb$ analysis presented in this dissertation is 0.023 for the 115 GeV/c^2 VH channel. We can scale our 1 fb^{-1} sensitivity to correspond to the expected sensitivity with 5 fb^{-1} of data to be 0.051. An additional comparison can be performed against an analysis from the CDF collaboration that is similar in design to the $qqbb$ analysis presented in this dissertation [50]. Using 2 fb^{-1}, the CDF analysis has been able to set a 95% C.L. observed (expected) upper limit of 37.9 (37.1) times the Standard Model at the 115 GeV/c^2 mass point. Scaling my analysis from 1 to 2 fb^{-1} of luminosity and assuming that the systematic errors in this analysis do not change, an expected 95% C.L. upper limit of 10.4 pb or 35 times the Standard Model is estimated. Thus, the $qqbb$ analysis technique presented in this dissertation would have a slightly better performance than the CDF analysis.

In parallel with the light Higgs boson search, a search was performed for signatures consistent with Technicolor production. The Technicolor signal samples for the 115 GeV/c^2, 125 GeV/c^2, and 140 GeV/c^2 mass technipions were processed though the 115 GeV/c^2, 125 GeV/c^2, and 135 GeV/c^2 VH analysis chains, respectively. The resulting cross sections for the Technicolor signals were ($22\pm11\pm12$) pb, ($22\pm13\pm15$) pb, and ($30\pm16\pm18$) pb for the 115 GeV/c^2, 125 GeV/c^2, and 140 GeV/c^2 mass points. Converting these cross sections to 95% C.L. upper limits we find an expected upper limit of 31 pb or 41 times the predicted cross section with an observed upper limit of 49 pb or 65 times the predicted cross section for the 115 GeV/c^2 mass point. The 125 GeV/c^2 and 140 GeV/c^2 mass points yield 95% C.L. upper limits of 40 pb (57 pb) and 47 pb (71 pb) for the expected (observed), respectively. Both the DØ and CDF experiments have ongoing Technicolor programs and have produced exclusion zones. The CDF experiment has recently expanded the Tevatron exclusion zone to include the 125 and 140 GeV/c^2 mass points used in this analysis. The 115 GeV/c^2 mass point has not yet been excluded due to the dramatic reduction in cross section when the Technirho mass is less than twice the Technipion mass.

A WZ diboson production search was performed in this analysis. It was found that the decision tree did not improve the sensitivity and therefore was not used in the analysis. Due to the increased predicted cross-section, this channel provided the largest overall sensitivity (0.029) of all the investigated channels in the analysis. However, this sensitivity is not competitive with other WZ analyses at DØ. For more information on this search, see Chapter 9

As with most analyses, the addition of more data will improve the result. If the available data were increased from 1 fb^{-1} to 10 fb^{-1}, the expected upper limit would scale from 12.8 pb to 9.5 pb giving a 26% improvement for the 115 GeV/c^2 VH analysis. While the addition of data to this analysis will improve the results, further enhancements could be implemented such as an improved dijet mass resolution. One could still retain $\simeq 90\%$ of the signal, with tighter dijet mass cuts from an improved dijet mass resolution and therefore, a larger sensitivity could be achieved. Studies have been performed at DØ reporting improvements to the dijet invariant mass resolution on the order of 10% [78]. An improvement to the dijet mass resolution of 10% would increase the post invariant mass cut sensitivity by approximately 10% bringing the 115 GeV/c^2 VH sensitivity from 0.039 to 0.043. In addition to the improved dijet mass resolution, improvements to the critical b-tagging algorithms could be implemented. While the b-tagging algorithm used in this study is based on a neural net, new multivariate b-tagging algorithms are being investigated. The current neural net b-tagging algorithm has an efficiency of approximately 50% with a mistag rate of 0.4%. If the new algorithm improved the efficiency by 10% leaving the fake rate the same, the sensitivity at the preliminary cuts stage would see a 20% improvement. If both the efficiency and the fake rate were to gain a 10% improvement, the resulting sensitivity would increase by approximately 35%. While these analysis improvements depend on an improvement to the event reconstruction tools, it is also possible to add additional cuts in the preliminary stage of the analysis to reduce the overall amount of data but retain a large portion of the signal. A cut falling into this category would be the possibility of an isolated lepton

veto.

The majority of today's Higgs boson searches suffer from relatively low branching ratios in the final states of the decay chain. The analysis presented in this paper provides a method of recovering the large hadronic decay region. With some of the proposed improvements implemented, along with the additional data, this technique holds promise for becoming a competitive channel.

References

[1] ALEPH Collaboration, CDF Collaboration, DØ Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, Tevatron Electroweak Working Group, and SLD electroweak heavy flavour groups. Precision electroweak measurements and constraints on the standard model, 2009.

[2] E. A. Paschos. *Electroweak theory.* Cambridge University Press, Cambridge, UK, 2007.

[3] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13(9):321–323, Aug 1964.

[4] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13(16):508–509, Oct 1964.

[5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13(20):585–587, Nov 1964.

[6] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579 – 588, 1961.

[7] Salam, Abdus, and J. Strathdee. A renormalizable gauge model of lepton interactions. *Il Nuovo Cimento A (1971-1996)*, 11(2):397–435, 09 1972/09/01/.

[8] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19(21):1264–1266, Nov 1967.

[9] C. Amsler et al. PDG. *Phys. Lett. B*, 667(1), 2008.

[10] A. Abulencia et al. Observation of WZ Production. *Phys. Rev. Lett.*, 98(16):161801, Apr 2007.

[11] Michael E. Peskin and Dan V. Schroeder. *An Introduction To Quantum Field Theory (Frontiers in Physics)*. Westview Press, October 1995.

[12] D. Griffiths. *Introduction to Elementary Particles*. John Wiley & Sons, New York, USA, 1987.

[13] D. Perkins. *Introduction to High Energy Physics*. Addison-Wesley, Reading, USA, 2000.

[14] T. Han and S. Willenbrock. QCD corrections to the $pp \to WH$ and ZH total cross sections. *Physics Letters B*, 273(1-2):167 – 172, 1991.

[15] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56–74, 1998, hep-ph/9704448.

[16] S. Dawson. Introduction to electroweak symmetry breaking. 1998, hep-ph/9901280.

[17] J. Ellis et al. The Probable Fate of the Standard Model. *Phys. Lett.*, B679:369–375, 2009, 0906.0954.

[18] Steven Weinberg. Implications of dynamical symmetry breaking. *Phys. Rev. D*, 13(4):974–996, Feb 1976.

[19] Leonard Susskind. Dynamics of spontaneous symmetry breaking in the weinberg-salam theory. *Phys. Rev. D*, 20(10):2619–2625, Nov 1979.

[20] Kenneth Lane. Technihadron production and decay in low-scale technicolor. *Phys. Rev. D*, 60(7):075007, Sep 1999.

[21] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006, hep-ph/0603175.

[22] K. Lane. personal communication, 2010.

[23] Fermilab, Accelerator Division. *Linac Rookie Book*, v2.1 edition, October 1994.

[24] L. W. Alvarez et al. Berkeley Proton Linear Accelerator. 1948. UCRL-236.

[25] Fermilab, Accelerator Division. *Booster Rookie Book*, v4.1 edition, April 2009.

[26] Fermilab, Accelerator Division. *Main Injector Rookie Book*, v1.0 edition, June 2003.

[27] Fermilab, Accelerator Division. *Tevatron Rookie Book*, v2.3 edition, June 2009.

[28] Fermilab Accelerators Division. Tevatron supertable, March 2010. `http://www-bd.fnal.gov/SDAMisc/new_supertableIV.html`.

[29] The upgraded DØ detector. *Nucl. Instrum. Meth. A*, 565(2):463 – 537, 2006.

[30] Cheuk-Yin Wong. *Introduction to High-Energy Heavy-Ion Collisions*. World Scientific, Singapore, 1994. Erratum.

[31] T. Zimmerman et al. The SVX2 readout chip. *IEEE Trans. Nucl. Sci.*, 42:803–807, 1995.

[32] R. Angstadt et al. The Layer 0 Inner Silicon Detector of the DØ Experiment. *ArXiv e-prints*, November 2009, 0911.2522.

[33] V. M. Abazov et al. The muon system of the Run II DØ detector. *Nucl. Instrum. Meth.*, A552:372–398, 2005, physics/0503151.

[34] D. Adams. Finding Tracks. *DØ Note 2958*, 1996.

[35] A. Khanov. HTF: histogramming method for finding tracks. The algorithm description. *DØ Note 3778*, 2000.

[36] Kenneth Bloom and Aaron Dominguez. DØ Tracking Performance at High Luminosity. *DØ Note 4980*, 2006.

[37] E. Busato and A. Bernard. Jet Algorithms in DØ Run II Software: Description and Users Guide. *DØ Note 4457*, 2004.

[38] A. Juste et al. Jet Energy Scale Determination at DØ Run II. *DØ Note 5382*, 2007.

[39] P. Calfayan et al. Muon Identification Certification for p17 data. *DØ Note 5157*, 2006.

[40] Andreas Hocker et al. TMVA: Toolkit for multivariate data analysis. *PoS*, ACAT:040, 2007, physics/0703039.

[41] E. Aguiló et al. Using Boosted Decision Trees to Search for Single Top Quarks in 1 fb^{-1} of Data. *DØ Note 5286*, 2007.

[42] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.

[43] Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. In *In Proceedings of the Thirteenth International Conference on Machine Learning*, pages 148–156, 1996.

[44] K. Hanagaki. *b*-tagging at DØ. In *Hadron Collider Physics 2005*, pages 255–258, 2007.

[45] M. Anastasoaie, S. Robinson, and T. Scanlon. Performance of the NN *b*-tagging Tool on p17 Data. *DØ Note 5213*, 2007.

[46] Ulrich Schwickerath. (Final) Higgs results from LEP. 2002, hep-ph/0205126.

[47] Search for the Standard Model Higgs boson at LEP. *Physics Letters B*, 565:61 – 75, 2003.

[48] P. Bock et al. Lower bound for the standard model Higgs boson mass from combining the results of the four LEP experiments. Technical Report CERN-EP-98-046. CERN-L3-146. CERN-OPAL-TN-503. OPAL-TN-503, CERN, Geneva, Apr 1998. ADLO Note.

[49] DØ Collaboration. Combined Upper Limits on Standard Model Higgs Boson Production from the DØ Experiment in 2.1 – 5.4 fb^1 of Data. *DØ Note 6008-CONF*, 2010.

[50] T. Aaltonen et al. Search for the Associated Production of the Standard-Model Higgs Boson in the All-Hadronic Channel. *Phys. Rev. Lett.*, 103(22):221801, Nov 2009.

[51] The CDF collaboration, The DØ Collaboration, the Tevatron New Physics, and Higgs Working group. Combined CDF and DØ Upper Limits on Standard Model Higgs-Boson Production with 2.1 - 5.4 fb^{-1} of Data, 2009.

[52] Abdallah et al. Search for Technicolor with DELPHI. *The Eur. Phys. J. C*, 22(1):17–29, 09 2001.

[53] V. M. Abazov et al. Search for Techniparticles in e+jets Events at DØ. *Phys. Rev. Lett.*, 98(22):221801, May 2007.

[54] T. Aaltonen et al. Search for Technicolor Particles Produced in Association with a W Boson at CDF. *Phys. Rev. Lett.*, 104(11):111802, Mar 2010.

[55] T. Scanlon. Level 3 b-tagging Triggers for the MSSM Multi-jet Higgs Analysis. *DØ Note 4621*, 2007.

[56] A. Harel. JetID Optimization. *DØ Note 4919*, 2005.

[57] DØ Collaboration. Common analysis format. http://www-d0.fnal.gov/Run2Physics/cs/caf, June 2009.

[58] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003, hep-ph/0206293.

[59] A. Pukhov et al. CompHEP: A package for evaluation of Feynman diagrams and integration over multi-particle phase space. User's manual for version 33. 1999, hep-ph/9908288.

[60] Anders Ryd et al. EvtGen: A Monte Carlo Generator for B-Physics. EVTGEN-V00-11-07, http://hep.ucsb.edu/people/lange/EvtGen/.

[61] R. Brun, R. Hagelberg, M. Hansroul, and J. C. Lassalle. GEANT: Simulation Program for Particle Physics Experiments. User Guide and Reference Manual. 1978. CERN-DD-78-2-REV.

[62] F. Couderc et al. DØ Search for Neutral Higgs bosons at High $\tan\beta$ in Multi-jet Events Using p17 Data. *DØ Note 5341*, 2007.

[63] N. Makovec and J. Grivaz. Shifting, Smearing, and Removing Simulated Jets. *DØ Note 4914*, 2005.

[64] C. Ochando and J. Grivaz. SSR for p17. *DØ Note 5609*, 2008.

[65] F. Couderc, M. Michaut, and B. Tuchming. DØ Search for Neutral Higgs bosons at High $\tan\beta$ in multi-jet events using p17 Data. *DØ Note 5202*, 2006.

[66] Samuel Calvet et al. Search for the Standard Model Higgs boson in the $ZH \rightarrow \nu\nu bb$ channel at $\sqrt{s} = 1.96$ TeV. *DØ Note 5857*, 2009.

[67] M. Begel et al. Measurement of the $t\bar{t}$ cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV in the multi-jet final state. *DØ Note 5057*, 2006.

[68] A. Harel. How to vary jet id efficiencies, energy scale and resolution. https://plone4.fnal.gov/P1/D0Wiki/physics/top/HowToVaryJetCorrections, Feburary 2010.

[69] Cliff P Burgess and Guy D Moore. *The Standard Model: A Primer*. Cambridge Univ. Press, Cambridge, 2007.

145

[70] J. M. Campbell and R. K. Ellis. Update on vector boson pair production at hadron colliders. *Phys. Rev. D*, 60(11):113006, Nov 1999.

[71] V. M. Abazov et al. Production of WZ Events in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV and Limits on Anomalous WWZ Couplings. *Phys. Rev. Lett.*, 95(14):141802, Sep 2005.

[72] V. M. Abazov et al. Measurement of the WW Production Cross Section with Dilepton Final States in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV and Limits on Anomalous Trilinear Gauge Couplings. *Phys. Rev. Lett.*, 103(19):191801, Nov 2009.

[73] V. M. Abazov et al. Observation of ZZ Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. Lett.*, 101(17):171803, Oct 2008.

[74] T. Aaltonen et al. Strong Evidence for ZZ Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. Lett.*, 100(20):201801, May 2008.

[75] T. Aaltonen et al. Search for WW and WZ production in lepton plus jets final state at CDF. *Phys. Rev. D*, 79(11):112011, Jun 2009.

[76] V. M. Abazov et al. Evidence of WW and WZ Production with lepton+jets Final States in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. Lett.*, 102(16):161801, Apr 2009.

[77] G. Bernardi et al. Search for WH associated production using neural networks with 5.0 fb^{-1} of Tevatron data. *DØ Note 5972-CONF*, 2009.

[78] B. Casey. personal communication, 2008.

[79] Peter W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145(4):1156–1163, May 1966.

[80] R. Barate et al. Search for the neutral Higgs bosons of the MSSM in e^+e^- collisions at from 130 to 172 GeV/c^2. *Phys. Lett. B*, 412(1-2):173 – 188, 1997.

[81] R. Barate et al. Observation of an excess in the search for the Standard Model Higgs boson at ALEPH. *Phys. Lett. B*, 495(1-2):1 – 17, 2000.

[82] Marcela Carena and Howard E. Haber. Higgs Boson Theory and Phenomenology, 2002.

[83] A. Bean et al. Search for an Associated Standard Model Light Higgs Boson in the All Hadronic Decay Mode at DØ . *DØ Note 5690*, 2008.

[84] Hai-Jun Yang, Byron P. Roe, and Ji Zhu. Studies of boosted decision trees for miniboone particle identification. *Nucl. Instrum. Meth. A*, 555:370, 2005.

[85] G. Arnison et al. Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider. *Phys. Lett.*, B126:398–410, 1983.

[86] G. Arnison et al. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV. *Physics Letters B*, 122(1):103 – 116, 1983.

[87] Peter W. Higgs. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys. Rev.*, 145(4):1156–1163, May 1966.

[88] T. Aaltonen et al. Search for WW and WZ production in lepton plus jets final state at CDF. *Phys. Rev. D*, 79(11):112011, 2009.

A   DØ Specific Data and Software Processing Information

Inside the DØ environment, there is a lot of specific language and nomenclature used. This section is present to convey, to the members of the DØ collaboration, these specific DØ elements used in this analysis. The DØ experiment has a system, called Sequential Access Metadata or SAM, for collecting data into logical blocks based on meta information. Each block is given an identifier for future referencing and reprocessing. The full data skim used in this analysis was split into three different logical blocks with the identifiers: CSG_CAF_3JET_PASS3_p17.09.03, CSG_CAF_3JET_PASS3_p17.09.06, and CSG_CAF_3JET_PASS3_p17.09.06b.

In order to increase flexibility, the processing of the data (and also the Monte Carlo) was split into two stages. The first of these uses an experiment-wide general analysis framework called CAFe. CAFe provides the experimenter with an environment populated by physics objects and other event specific information such as the instantaneous luminosity. The CAFe stage of this analysis is used to perform jet energy corrections and generate the required inputs for the remainder of the analysis. The second stage uses a custom analysis code which generates the individual distributions and inputs needed for the final stages of the analysis. In addition to the standard modules in the CAFe framework, a custom event processor called `GenSkim` contained in the `hwbbjj` analysis module was used. The combined list of packages and their versions used in this analysis related to the CAFe processing stage are presented below:

- D0RunII: p21.12.00

- cafe_sam: p21-br-05

- cafe: p21-br-27

- tmb_tree: p21-br-51

- jetcorr: p21-br-12

- caf_util: p21-br-97

- caf_util: p21-br-98

- caf_mc_util: p21-br-126

- caf_trigger: p21-br-57

- caf_trigger: p21-br-61

- eff_utils: p21-br-24

- caf_eff_utils: p21-br-10

- caf_mc_util: p21-br-127

- emid_cuts: p21-br-22

- met_util: p21-br-01

- tau_tmb: p21-br-01

- caf_dq: p21-br-03

- dq_util: p21-br-03

- edm_dq: p21-br-03

- dq_defs: v2008-09-11

- emid_eff: v7-preliminary-33

- muid_eff: v04-03-00

- jetid_eff: v03-00-02

- higgs_hb: p17finaljes
  (Adapted code for the trigger
  and 3JET skim simulator)

A.1   Monte Carlo Samples

Table 21: List of Monte Carlo samples used in this analysis.

Process	MC ID	Number of Generated Events
$HW \to bbjj(M_H = 115 \text{ GeV/c}^2)$	78932	103,750
$HW \to bbjj(M_H = 125 \text{ GeV/c}^2)$	78933	101,750
$HW \to bbjj(M_H = 135 \text{ GeV/c}^2)$	78934	100,000
$HZ \to bbjj(M_H = 115 \text{ GeV/c}^2)$	89932	101,500
$HZ \to bbjj(M_H = 125 \text{ GeV/c}^2)$	89933	100,250
$HZ \to bbjj(M_H = 135 \text{ GeV/c}^2)$	89934	107,500
$\pi_{\text{TC}}W \to bbjj(M_\pi = 115 \text{ GeV/c}^2)$	110755	197077
$\pi_{\text{TC}}W \to bbjj(M_\pi = 125 \text{ GeV/c}^2)$	110756	196205
$\pi_{\text{TC}}W \to bbjj(M_\pi = 140 \text{ GeV/c}^2)$	110757	184670
$WZ \to$ SM Inclusive	TOTAL	515,000
	30488, 30489, 33684, 33685, 42212	
$WW \to$ SM Inclusive	TOTAL	510,000
	33681, 33682, 33683	
$ZZ \to$ SM Inclusive	TOTAL	307,500
	30486, 30487, 33686, 33687	
$tb \to Wbb$ (s-channel)	TOTAL	150,000
	30211, 30212, 30213, 30214, 30232	
$tqb \to Wbq$ (t-channel)	TOTAL	149,000
	30539, 30540, 30541	
$t\bar{t} \to$ SM Inclusive	TOTAL	263,302
	31959, 32167	
$Wbb \to jjbb$	80136	209,000
$Wcc \to jjcc$	80139	203,250
$Wjj \to jjjj$	80134	54,750
$\gamma Z \to bb$	TOTAL	105250
	30548	
$Zb \to bbb$	TOTAL	100,000
	51412	
bb, bbj, bbjj	TOTAL	4,728,000
	80975, 80976, 80977, 80978,	
	80980, 80981, 80982, 80983,	
	81452, 81455, 81456, 81457,	
	81459, 81460, 81461, 78050,	
	78052, 81132, 81133, 81134,	
	80979, 80984, 81458, 78051,	
	81572	
cc, ccj, ccjj	TOTAL	611,959
	81152, 81153, 81154	

B   Decision Tree Information

B.1   Optimization

There are two broad techniques used to optimize the training of the decision trees. The first is to alter the input variables, while the second is to alter the parameters of the training algorithm. Both of these approaches were explored in an effort to maximize the performance of the decision tree. The optimizations were carried out for the $m_H = 115$ GeV/c^2 case and it is assumed that the optimization parameters would propagate to the higher masses. The metric used to determine the performance of the decision tree is the integral of the background rejection vs signal efficiency curve, otherwise known as the ROC integral. The efficiencies are determined by processing the testing component of the input samples through the decision tree and evaluating at a particular cut point. The larger the ROC integral is, the better the decision tree performed. A perfectly optimized decision tree would yield a ROC integral of 1.0 indicating a complete separation of the signal and the background decision tree discriminant distributions.

As stated previously, for the first optimization procedure we alter the decision tree variables used for the training. Ten different configurations were attempted that both added and removed different variables. Of these ten configurations, several included the dijet invariant masses, M(jj) and M(bb), in the DT variables. Figure 69 presents the ROC curves used to determine the ROC integrals. These integrals and the descriptions of the variable configurations are shown in Table 22.

Looking through Table 22 and ignoring the configurations that contain M(bb), M(jj), or M($bbjj$), it is seen that altering the variable configuration does not significantly affect the decision tree performance. By removing the bb and the jj charges from the variable list there was a slight improvement. This improvement was not large enough to justify altering the analysis tools to use that new configuration. The results of this cross-check are to use the default, justpt_anghw , configuration of variables listed in Table 9. We choose to leave the dijet masses out of the DT so that we can control their use later and
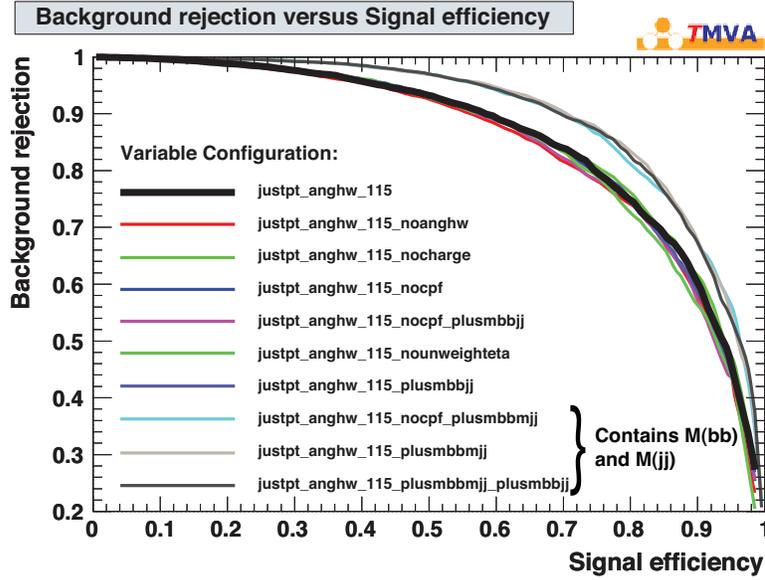
Figure 69: The ROC curves for the various variable configurations in the optimization studies for a Higgs boson mass of $m_H = 115$ GeV/c^2

be able to find control and signal regions for our background subtraction technique.

The second way to optimize the decision tree is to alter the training parameters instead of the input variables. When using the TMVA boosted decision tree package, there are a variety of configuration parameters available. This optimization cross-check studied the effects of five variables (boosting parameter β, boosting algorithm, MaxDepth, nEventsMin, and NTrees) that can strongly affect the performance of the decision tree. The boosting parameter, β, controls the magnitude of the reweighting in the AdaBoost algorithm for each boosted tree [43]. By relaxing this parameter to lower values (less than 1.00), the rate that the tree approaches the maximum performance from the boosting algorithm is reduced. The boosting parameter is scanned from 0.25 to 1.50 in 0.25 unit steps for the purposes of this test. The boosting algorithm itself can be changed from the AdaBoost algorithm to RegBoost, GradBoost, and Bagging. MaxDepth controls how many different levels there can be in each decision tree. While the default value for the MaxDepth is 3, this test used the values of 10, 50, 100, and 1000. At each node there needs to be a minimum number of events in order to allow the splitting

Name	∫ROC	Description
default	85.77	Default configuration, shown in Table 9
noanghw	84.87	Default configuration without the Angle($H,W/Z$) variable
nocharge	85.99	Default configuration without the jj or the bb charge variables
nocpf	85.63	Default configuration without the four charged particle fraction variables
nocpf_plusmbbjj	85.15	Default configuration without the four charged particle fraction variables and adding the four jet invariant mass
nounweighteta	85.00	Default configuration without the unweighted η variable
plusmbbjj	85.68	Default configuration with adding the four jet invariant mass
nocpf_plusmbbmjj	89.69	Default configuration without the four charged particle fraction variables and adding both the M(jj) and M(bb) invariant masses
plusmbbmjj	89.98	Default configuration with adding both the M(jj) and M(bb) invariant masses
plusmbbmjj_plusmbbjj	89.77	Default configuration with adding both the M(jj) and M(bb) invariant masses in addition to the four jet invariant mass

Table 22: The descriptions and the ROC integral values for the different variable test configurations on the $m_H = 115$ GeV/c^2 descision tree

to occur. The variable nEventsMin controls this value and is varied from 20 to 100 in 20 unit steps. The nEventsMin default value used for the analysis was set to 100. The final variable used in the study was the number of boosted trees. In the previous analysis, this value was set to produce 20 boosted trees. The number of trees was varied from 1 to 200 with a non-uniform step. The number of trees used for the analysis by default was 400 which was the default value used in the TMVA package. In order to compare these different tests, the ROC integral is used again to look for large differences in performance. The different ROC curves for the different configurations are shown

in Figure 70. The highest performing configuration is the default with the boosting parameter set to 0.25. It was decided to not use this decision tree due to a strange artifact in the low end of the decision tree discriminant distribution in the data. Since there was no other large gain in performance with the other variable configurations the default values were used. In this default configuration the number of boosted trees is 400, the boosting algorithm is the AdaBoost, the boosting parameter is set to 1.0, and the minimum number of events in a node is 100.
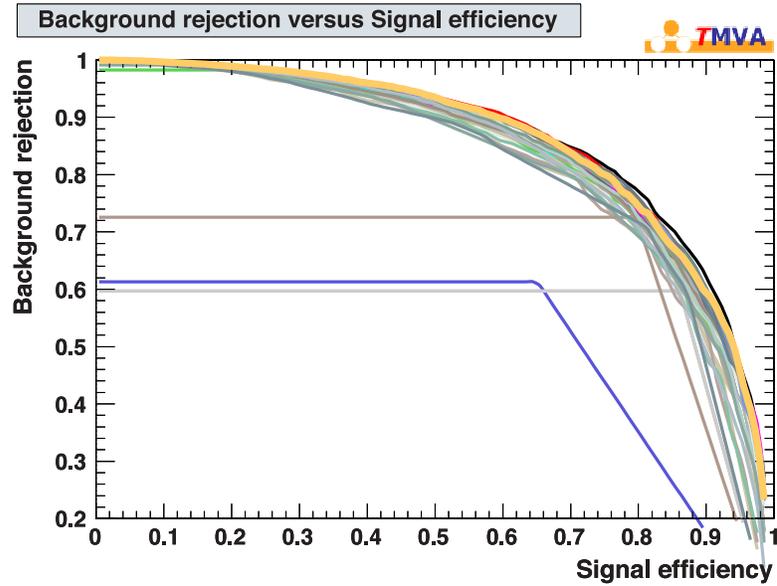
Figure 70: The ROC curves for the various training configurations in the optimization studies for a Higgs boson mass of $m_H = 115$ GeV/c^2. The configuration that was used is shown as a wide orange line.

B.2 Outputs and Sensitivity Scans

This section presents the decision tree discriminant output distributions for the different mass points and signal types used in this analysis. In addition to the DT discriminant distributions, the sensitivity scan is also presented which was used to determine the optimum cut point for the DT analysis stage. Figures 71, 72, and 73 show the DT discriminant optimized for the Higgs mass of 125 GeV/c^2, 135 GeV/c^2, and WZ signals, respectively.
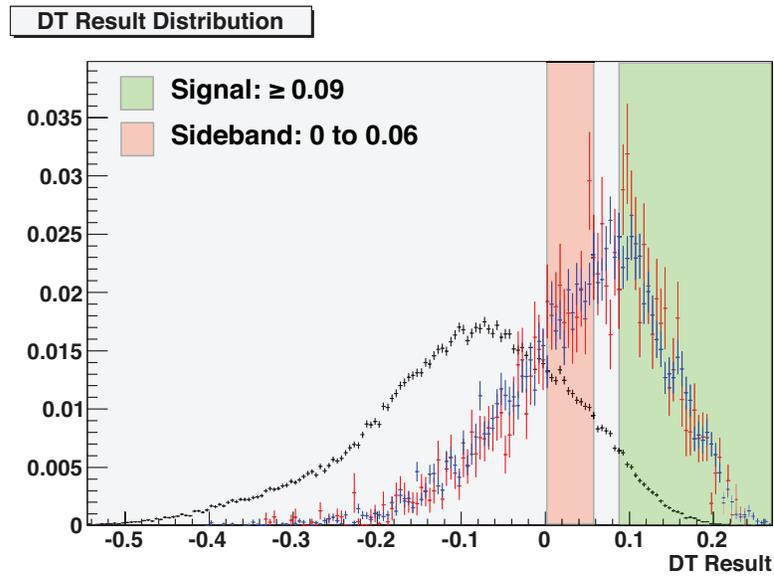
Figure 71: The decision tree discriminant distribution for the WH, ZH, and data set for the $m_H = 125$ GeV/c^2 mass point. The red distribution is the WH sample while the blue distribution represents the ZH sample. The data distribution is shown in black. The two shaded regions depict the DT control region (red) and the region of interest (green).

The sensitivity scans for each of the different Higgs mass points for the VH analyses are found in Figures 74, 75, and 76.
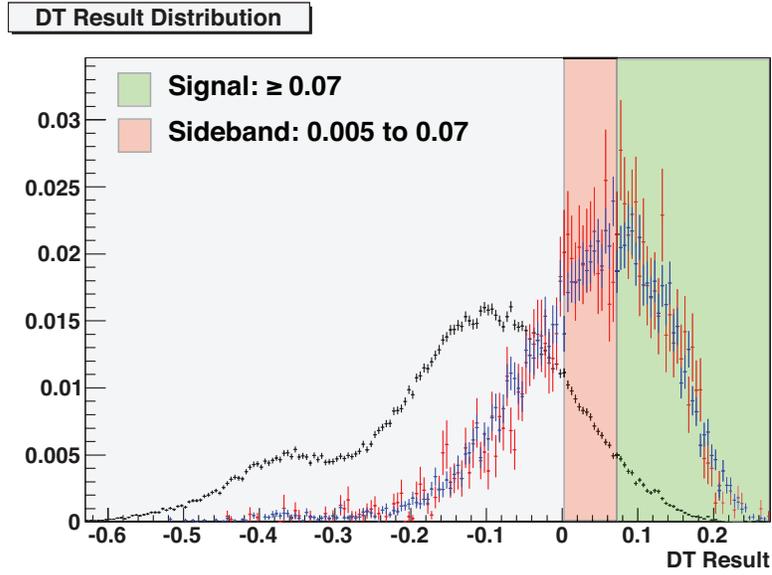
Figure 72: The decision tree discriminant distribution for the WH, ZH, and data set for the $m_H = 135$ GeV/c^2 mass point. The red distribution is the WH sample while the blue distribution represents the ZH sample. The data distribution is shown in black. The two shaded regions depict the DT control region (red) and the region of interest (green).
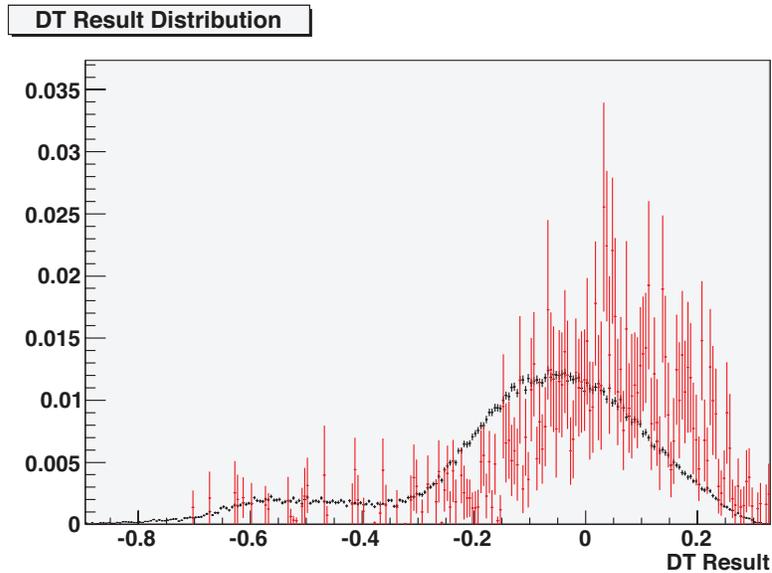
Figure 73: The decision tree discriminant distribution for the WZ analysis. The red distribution is the WZ sample. The data distribution is shown in black.

156
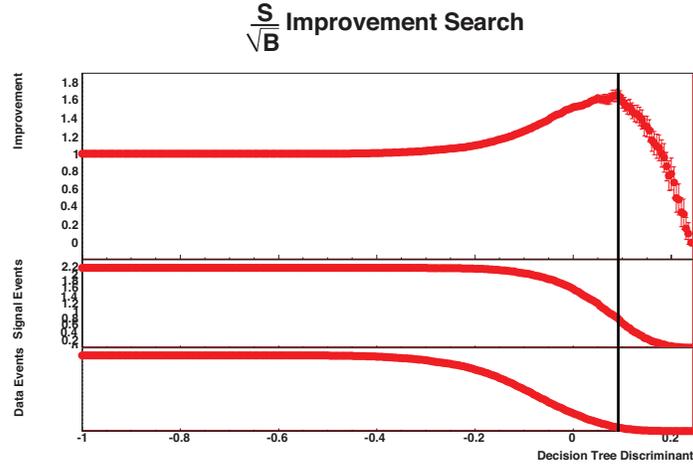
Figure 74: Sensitivity improvement search for the $WH(125)$ decision tree training. The bottom two left axes show the number of signal and background events as a function of the decision tree discriminant. The top two plots show the Sensitivity $S/\sqrt{B}$ Improvement with respect to the preliminary cuts after all cuts. The vertical black line denotes the optimal cut point.
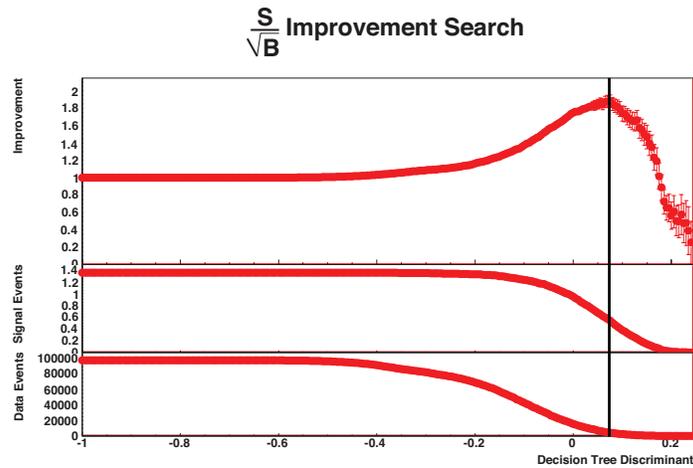
Figure 75: Sensitivity improvement search for the $WH(135)$ decision tree training. The bottom two left axes show the number of signal and background events as a function of the decision tree discriminant. The top two plots show the Sensitivity $S/\sqrt{B}$ Improvement with respect to the preliminary cuts after all cuts. The vertical black line denotes the optimal cut point.
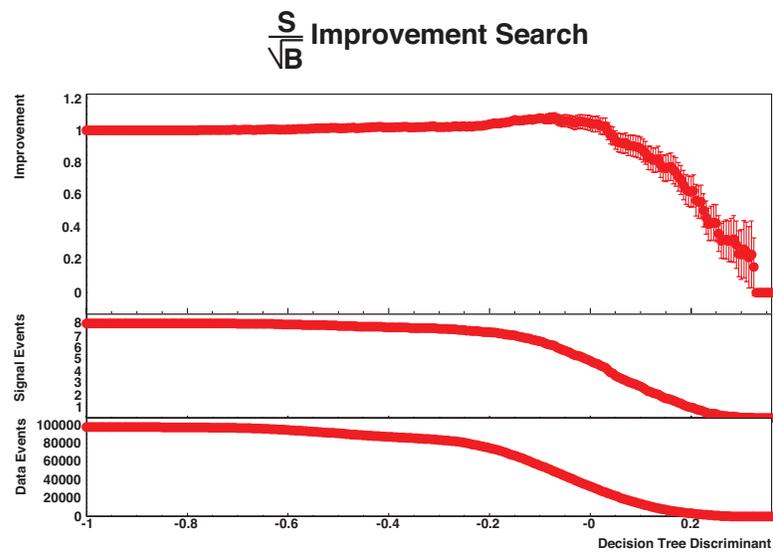
Figure 76: Sensitivity improvement search for the WZ decision tree training. The bottom two left axes show the number of signal and background events as a function of the decision tree discriminant. The top two plots show the Sensitivity $S/\sqrt{B}$ Improvement with respect to the preliminary cuts after all cuts.

B.3 Variable Importance

A figure of merit for determining the performance of the variables in the decision tree is called the variable importance and is determined by the TMVA training algorithm [40]. Table 23 gives the variable importance values for each of the variables in the DT. The best performing variable for the two lower mass Higgs training samples is the unweighted η and the charged particle fraction of the fourth largest jet in p_T rank. At a Higgs mass of 135 GeV/c^2, the best variable becomes scalar sum of the transverse mass, H_T. For the WZ tree, the best variable is again the charged particle fraction of the fourth largest jet in p_T rank.

Table 23: Importance found for each of the decision tree variables used with $WH(115)$, $WH(125)$, $WH(135)$, and WZ signal training. The variables are displayed in order of their effectiveness for the $WH(115)$ training. The rank of each variable for each tree is displayed in the parentheses to the right of each improvement metric.

Variable	Imp. $WH(115)$	Imp. $WH(125)$	Imp. $WH(135)$	Imp. WZ
Unweighted η	0.228 (1)	0.230 (2)	0.217 (3)	0.144 (9)
Fourth Leading CPF	0.182 (2)	0.231 (1)	0.205 (4)	0.231 (1)
Third Leading CPF	0.165 (3)	0.210 (3)	0.250 (2)	0.197 (4)
Second Leading CPF	0.161 (4)	0.185 (6)	0.188 (7)	0.230 (2)
Centrality	0.160 (5)	0.189 (5)	0.198 (5)	0.143 (10)
H_T	0.144 (6)	0.197 (4)	0.251 (1)	0.121 (12)
Leading CPF	0.137 (7)	0.144 (9)	0.141 (11)	0.210 (3)
Third Leading Jet p_T	0.131 (8)	0.169 (7)	0.190 (6)	0.078 (14)
Sphericity	0.127 (9)	0.136 (11)	0.120 (12)	0.076 (15)
Fourth Leading Jet p_T	0.127 (10)	0.143 (10)	0.151 (10)	0.189 (5)
Aplanarity	0.110 (11)	0.109 (13)	0.079 (15)	0.145 (8)
Second Leading Jet p_T	0.107 (12)	0.145 (8)	0.182 (8)	0.057 (20)
Fourth Leading Jet Width	0.088 (13)	0.076 (15)	0.097 (13)	0.185 (6)
Leading Jet p_T	0.084 (14)	0.121 (12)	0.171 (9)	0.011 (26)
Leading Jet η	0.069 (15)	0.078 (14)	0.081 (14)	0.012 (25)
Third Leading Jet Width	0.065 (16)	0.068 (18)	0.075 (17)	0.169 (7)
Angle(H, W)	0.056 (17)	0.075 (16)	0.064 (18)	0.136 (11)
Second Leading Jet Width	0.056 (18)	0.069 (17)	0.077 (16)	0.109 (13)
Second Leading Jet η	0.049 (19)	0.057 (20)	0.059 (20)	0.048 (21)
Leading Jet Width	0.047 (20)	0.060 (19)	0.060 (19)	0.075 (16)
Third Leading Jet η	0.041 (21)	0.048 (21)	0.049 (21)	0.069 (17)
Fourth Leading Jet η	0.028 (22)	0.036 (22)	0.033 (22)	0.036 (22)
bb Charge	0.021 (23)	0.026 (23)	0.022 (23)	0.026 (23)
Fourth Leading Jet ϕ	0.012 (24)	0.008 (28)	0.009 (25)	0.068 (18)
jj Charge	0.011 (25)	0.013 (24)	0.012 (24)	0.009 (27)
Leading Jet ϕ	0.009 (26)	0.012 (25)	0.007 (27)	0.008 (28)
Third Leading Jet ϕ	0.008 (27)	0.011 (26)	0.006 (28)	0.013 (24)
Second Leading Jet ϕ	0.004 (28)	0.009 (27)	0.007 (26)	0.060 (19)

B.4   Before and After DT variable Comparisons

Using the $WH(115)$ training, a comparison of the DT variables before and after the DT cut is made is shown for: the $= 2$ b-tagged data (Figures 77,78,79,80) and the $WH(115)$ signal Monte Carlo samples (Figures 81,82, 83,84). The data distributions show marked differences before and after the DT cut. With the exception of the p_T variables, the signal Monte Carlo DT distributions do not vary before or after the decision tree cut is applied.

(a) Sphericity

(b) Aplanarity

(c) Centrality

(d) H_T
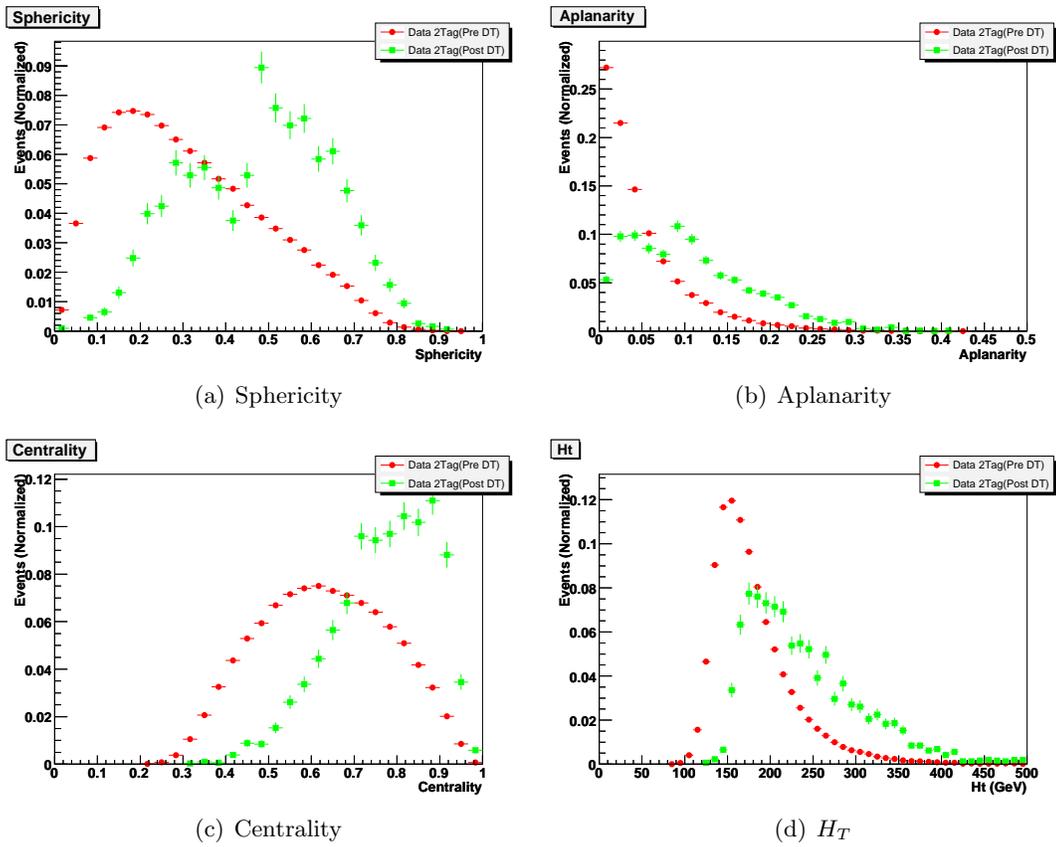
Figure 77: Distributions for events from the $= 2$ b-tagged data after preliminary cuts (red circles) and after the DT output cut (green squares).

According to the predictions from the various background Monte Carlo samples, we find that the QCD and $t\bar{t}$ samples contribute the majority of the events after the decision tree cut is applied. While we do not expect the background QCD Monte Carlo

161
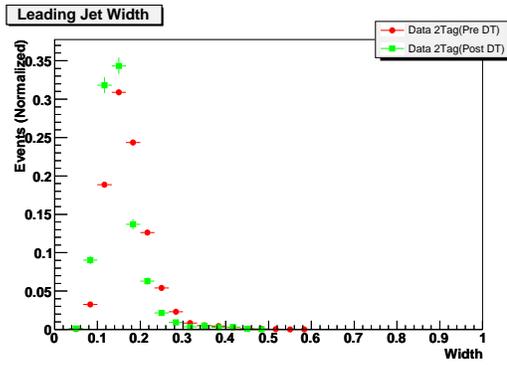
(a) Jet Width for the Leading Jet
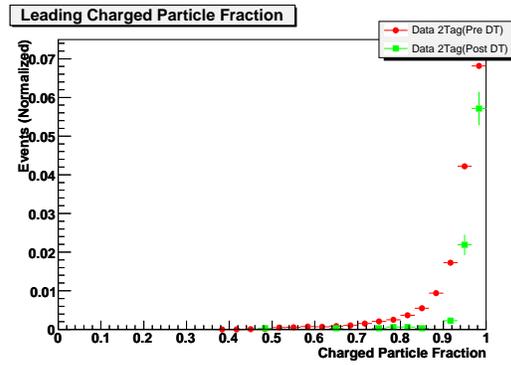
(b) Charged Particle Fraction 1 wrt Primary Vertex
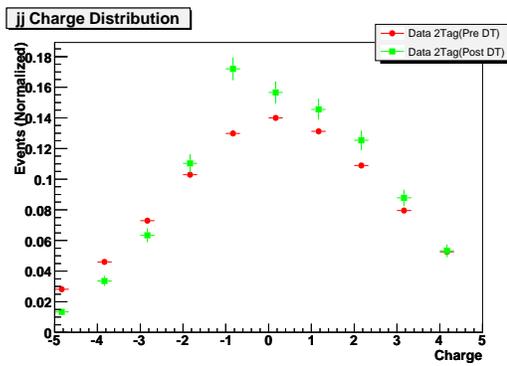
(c) The jj Charge
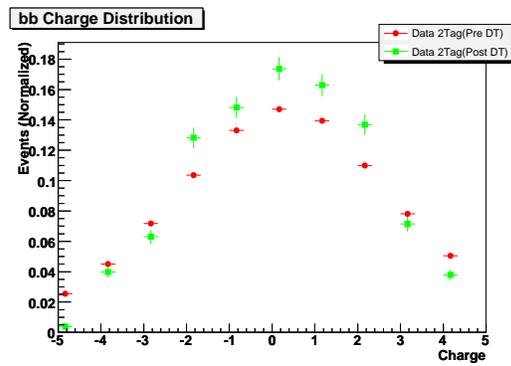
(d) The bb Charge

Figure 78: Distributions for events from the $= 2$ b-tagged data after preliminary cuts (red circles) and after the DT output cut (green squares).
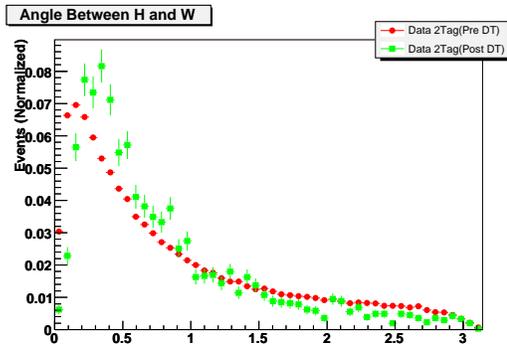
(a) Angle between the jj pair and the $b\bar{b}$ pair in the jj rest frame
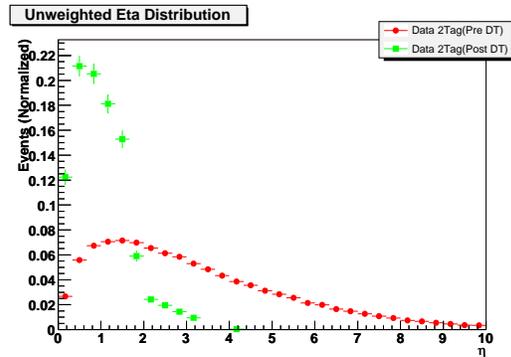
(b) Unweighted η distribution

Figure 79: Distributions for events from the $= 2$ b-tagged data after preliminary cuts (red circles) and after the DT output cut (green squares).

162

(a) Leading jet p_T (b) Second leading jet p_T

(c) Third leading jet p_T (d) Fourth leading jet p_T
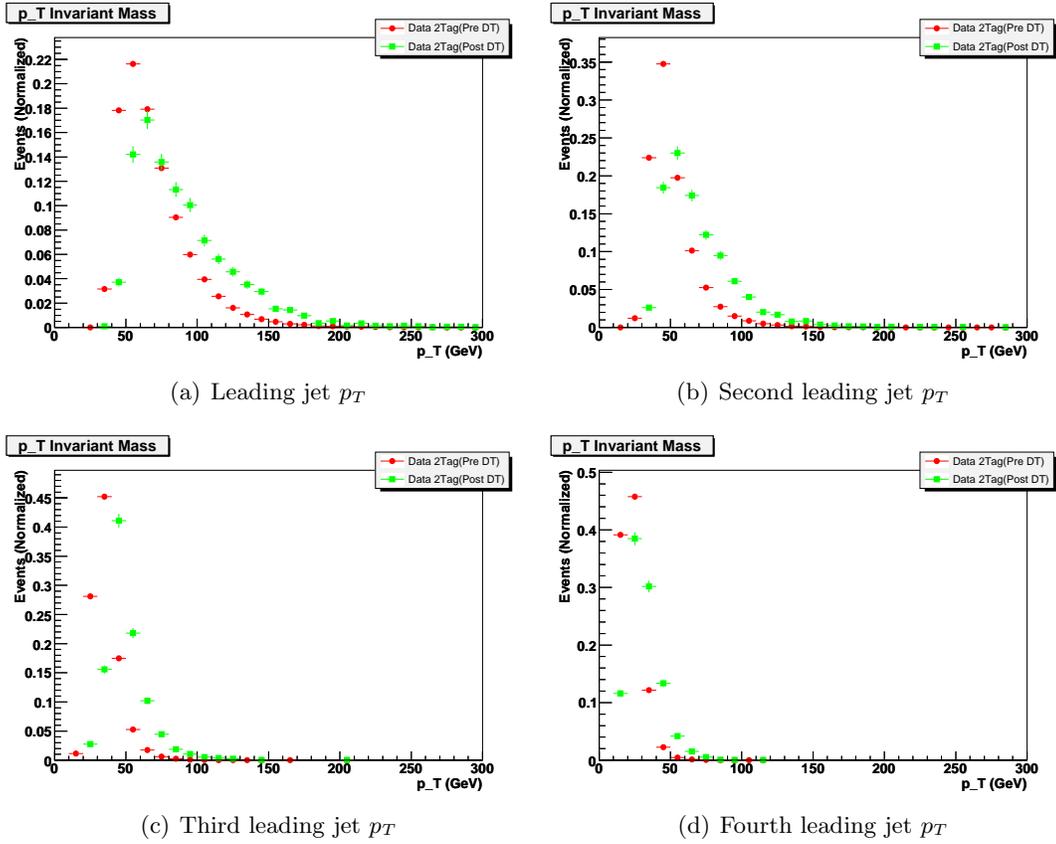
Figure 80: Transverse Momentum distributions for events from the $= 2$ b-tagged data after preliminary cuts (red circles) and after DT output cut (green squares).

sample to match the data, we provide plots to compare the variable distributions for the data with those from the QCD and the $t\bar{t}$ background MC samples. Figures 89,90, 91,and 92 show the distributions in the DT variables after the DT cut, while Figures 93,94, 95, and 96 show these DT variables after the preliminary cuts.

163
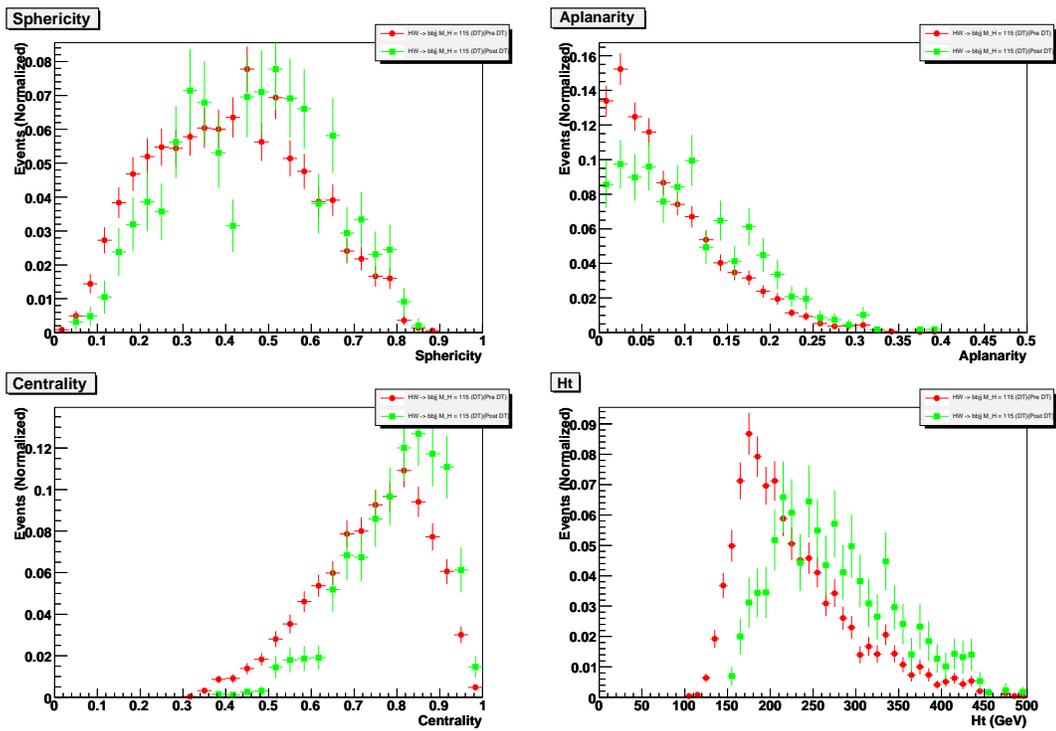
Figure 81: Distributions for events from the $WH(115)$ signal MC after preliminary cuts (red circles) and after the DT output cut (green squares).
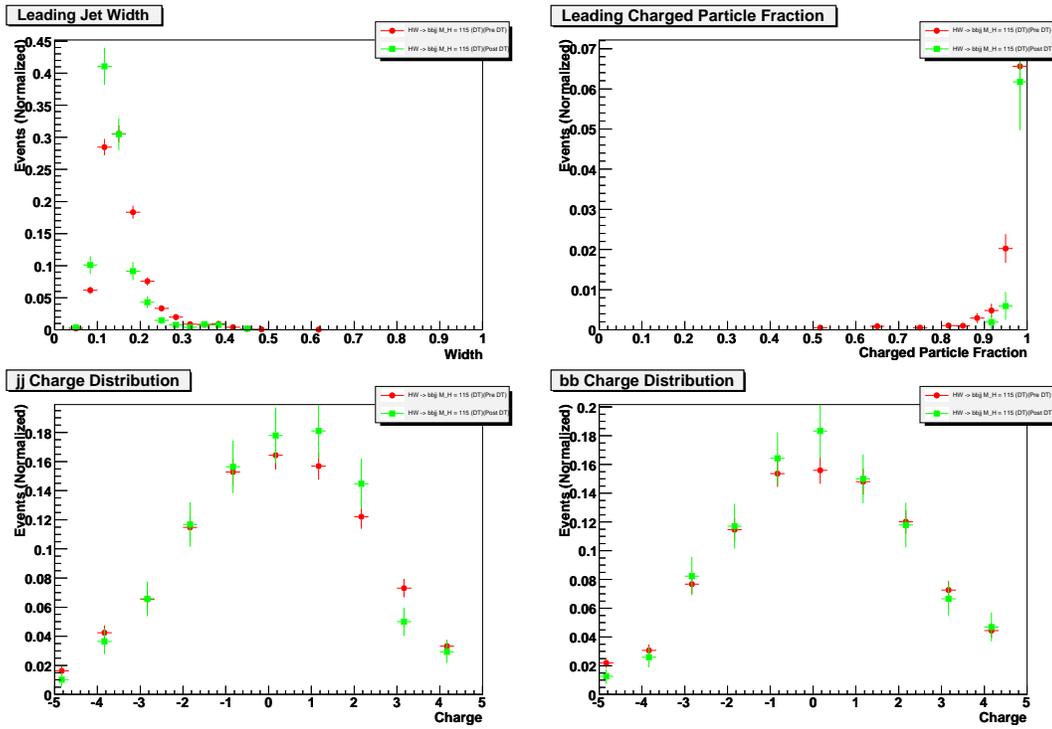
Figure 82: Distributions for events from the $WH(115)$ signal MC after preliminary cuts (red circles) and after the DT output cut (green squares).
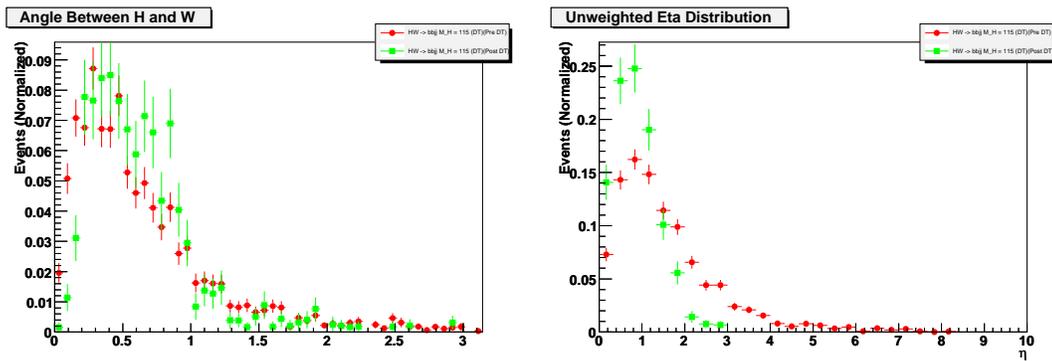
Figure 83: Distributions for events from the $WH(115)$ signal MC after preliminary cuts (red circles) and after the DT output cut (green squares).

165

(a) Leading jet p_T

(b) Second leading jet p_T

(c) Third leading jet p_T

(d) Fourth leading jet p_T
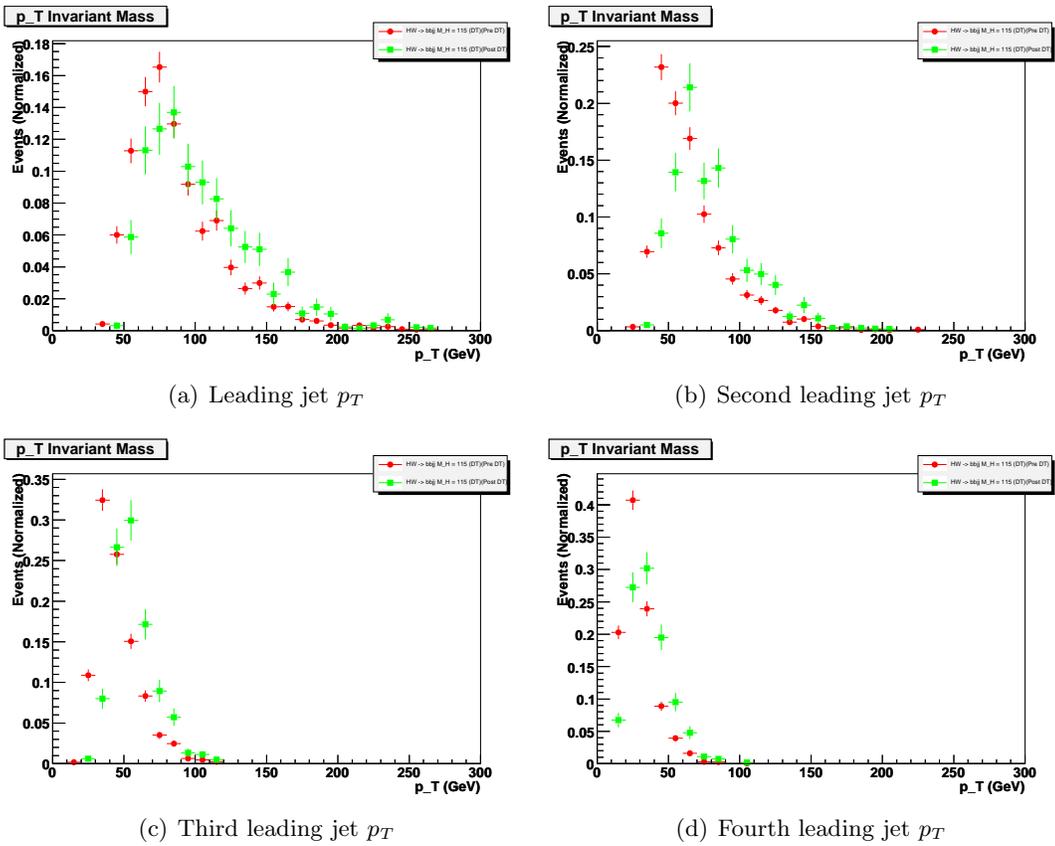
Figure 84: Transverse Momentum distributions for events from $WH(115)$ signal MC after preliminary cuts (red circles) and after DT output cut (green squares).

B.5   Decision Tree Training Input Distributions

The distributions in this section represent the input distribution for the variables used in the decision tree. While the distributions presented here are for the $WH(115)$ decision tree, they are representative of the other channels. The data sample is the 1 b-tagged sample as described in the decision tree section while the signal is represented by the $WH(115)$ Monte Carlo sample. It should be noted that these distributions are the training set taken from the three different subsets defined for the decision tree training and evaluation.
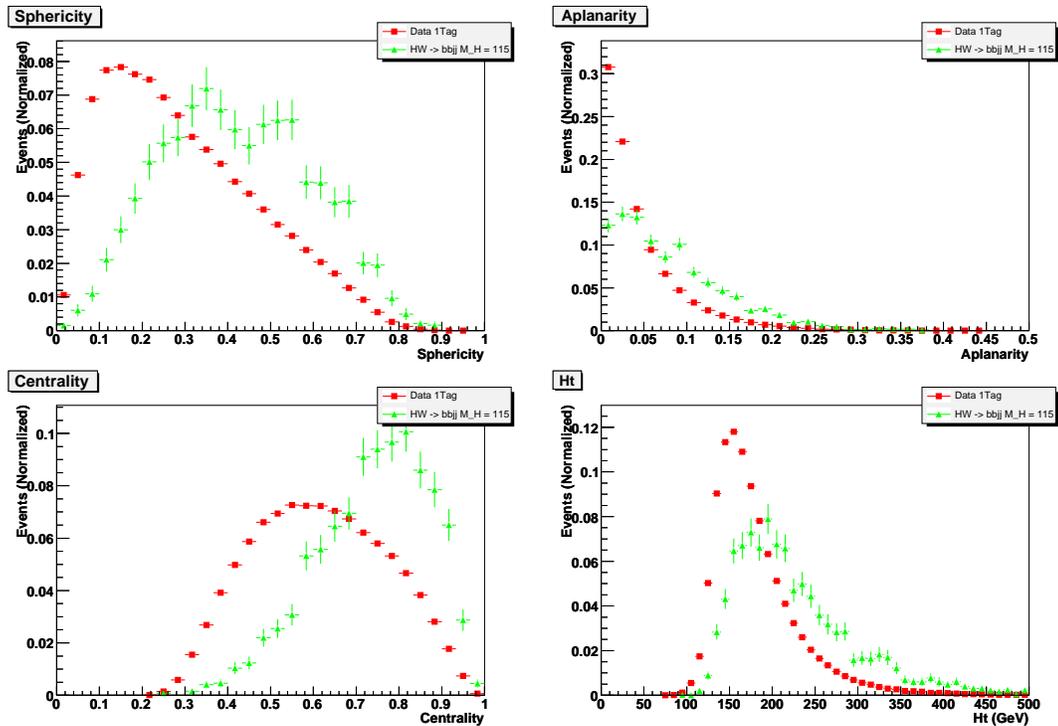
Figure 85: Topological variable distributions for events used in the $WH(115)$ decision tree training phase. The red squares represent the data 1 b-tagged sample while the green triangles are from the $WH(115)$ signal Monte Carlo sample.
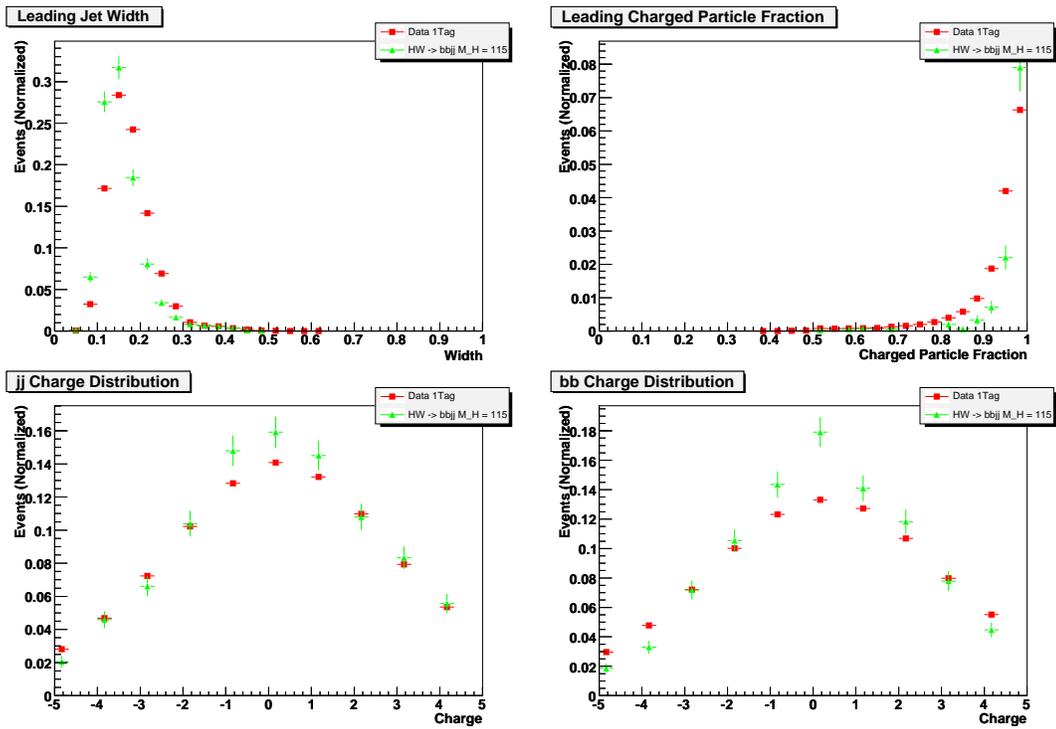
Figure 86: Distributions for events used in the $WH(115)$ decision tree training phase. The red squares represent the data 1 b-tagged sample while the green triangles are from the $WH(115)$ signal Monte Carlo sample.
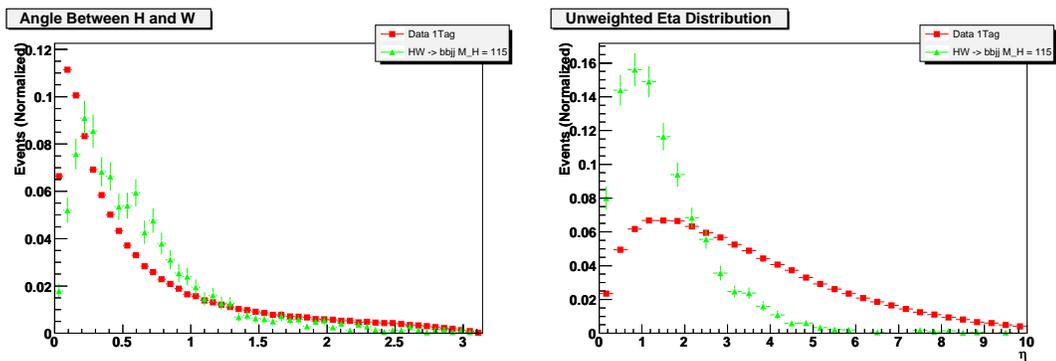
Figure 87: Distributions for events used in the $WH(115)$ decision tree training phase. The red squares represent the data 1 b-tagged sample while the green triangles are from the $WH(115)$ signal Monte Carlo sample.
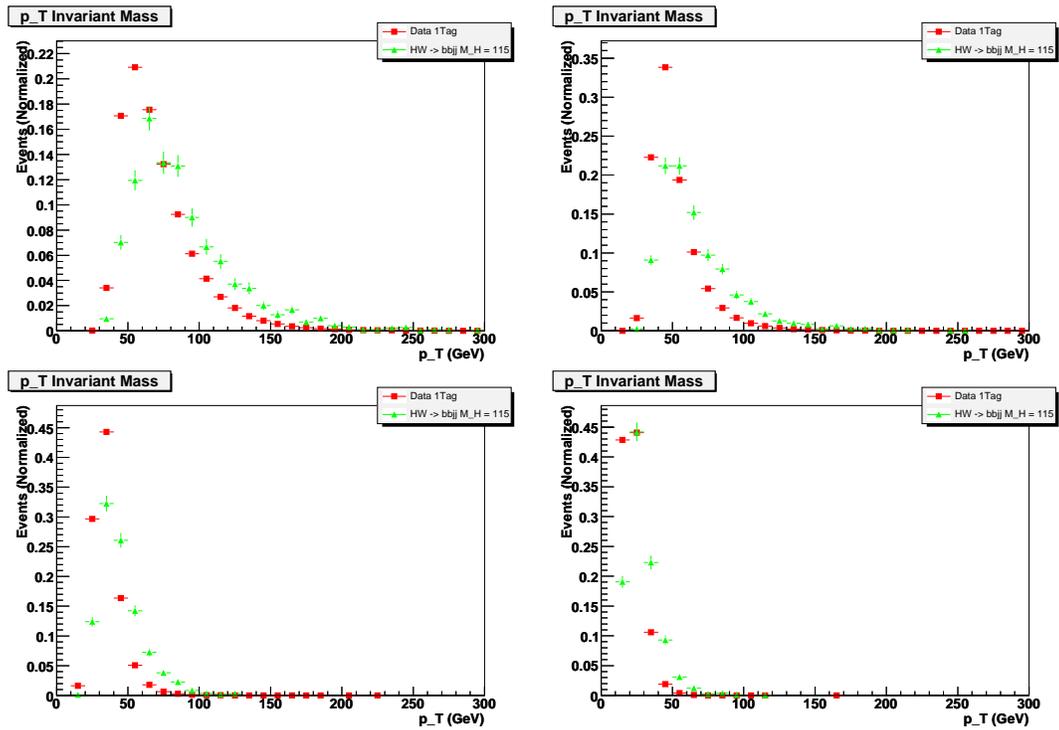
Figure 88: Transverse momentum distributions for events used in the $WH(115)$ decision tree training phase. The red squares represent the data 1 b-tagged sample while the green triangles are from the $WH(115)$ signal Monte Carlo sample.

B.6 Comparison of Data, QCD, and $t\bar{t}$ Monte Carlo After DT Cuts
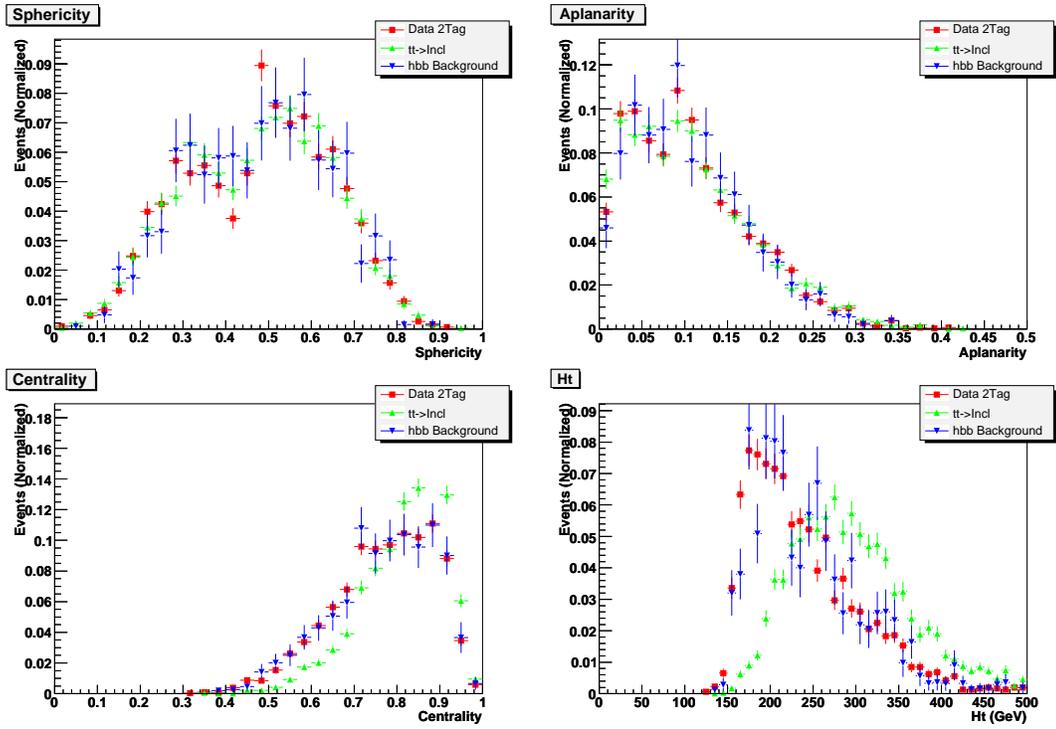
Figure 89: Topological variable distributions for events after the DT cuts from the $= 2$ b-tagged data (red squares), the QCD background (blue down diamond), the $t\bar{t}$ background (green up diamond).
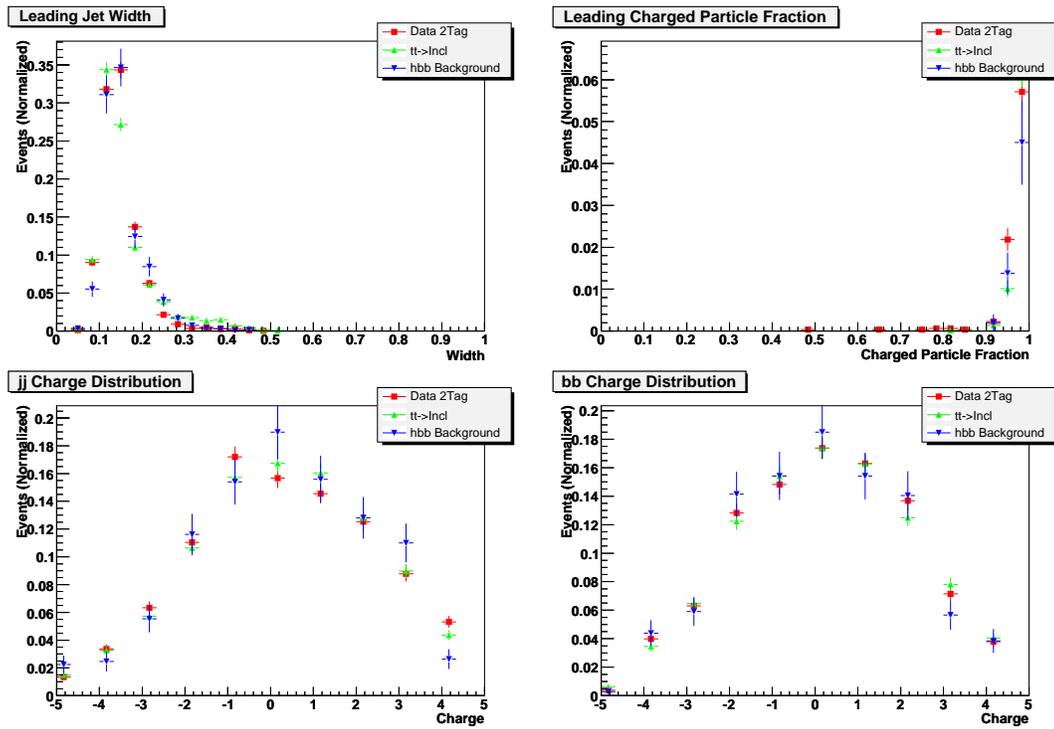
Figure 90: Distributions for events after the DT cuts from the $= 2$ b-tagged data (red squares), the QCD background (blue down diamond), the $t\bar{t}$ background (green up diamond).
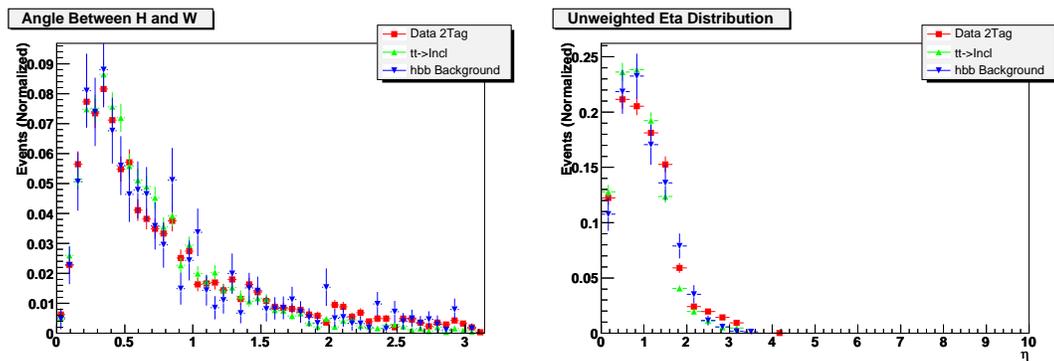
Figure 91: Distributions for events after the DT cuts from the $= 2$ b-tagged data (red squares), the QCD background (blue down diamond), the $t\bar{t}$ background (green up diamond).

171

(a) Leading jet p_T

(b) Second leading jet p_T

(c) Third leading jet p_T

(d) Fourth leading jet p_T
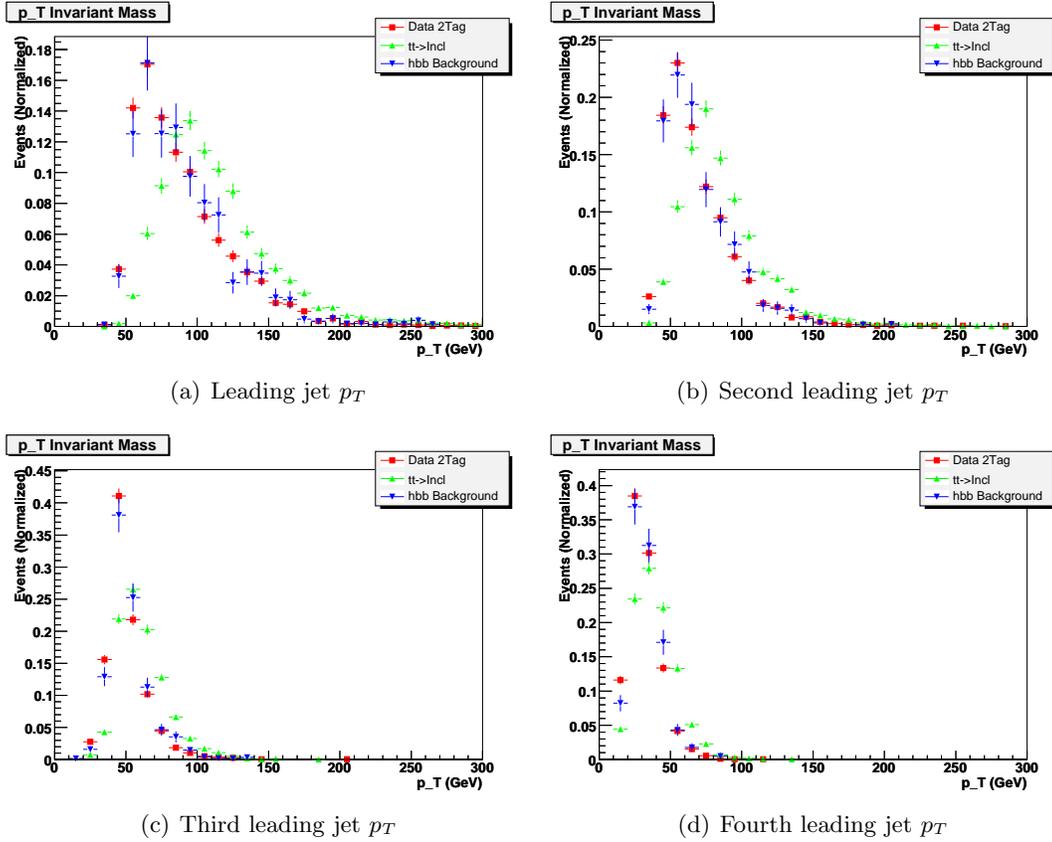
Figure 92: Transverse Momentum distributions for events after the DT cuts from the $= 2$ b-tagged data (red squares), the QCD background (blue down diamond), the $t\bar{t}$ background (green up diamond).

172
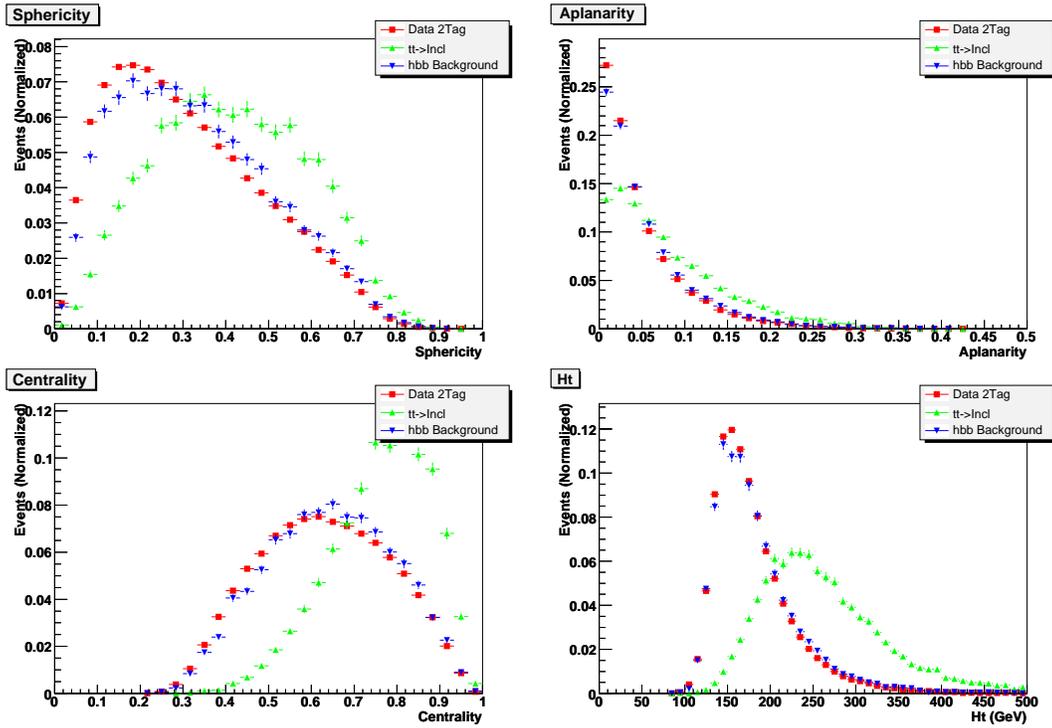
Figure 93: Topological variable distributions for events after the preliminary cuts from the $= 2$ b-tagged data (red squares), the QCD background (blue down diamond), the $t\bar{t}$ background (green up diamond).

173
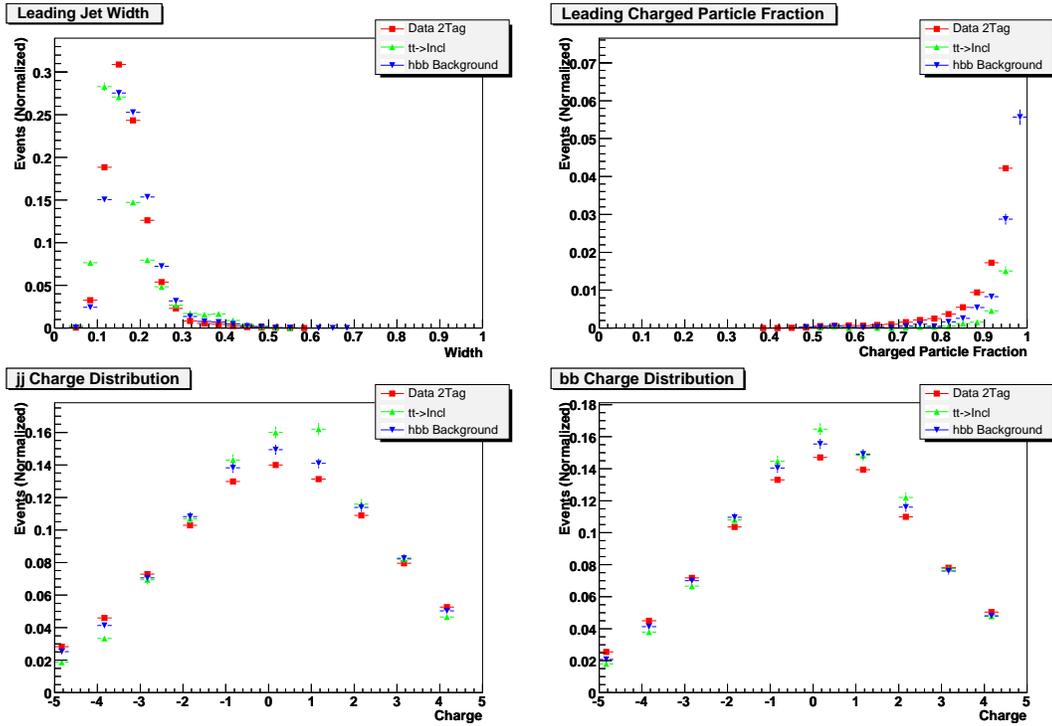
Figure 94: Distributions for events after the preliminary cuts from the $= 2$ b-tagged data (red squares), the QCD background (blue down diamond), the $t\bar{t}$ background (green up diamond).
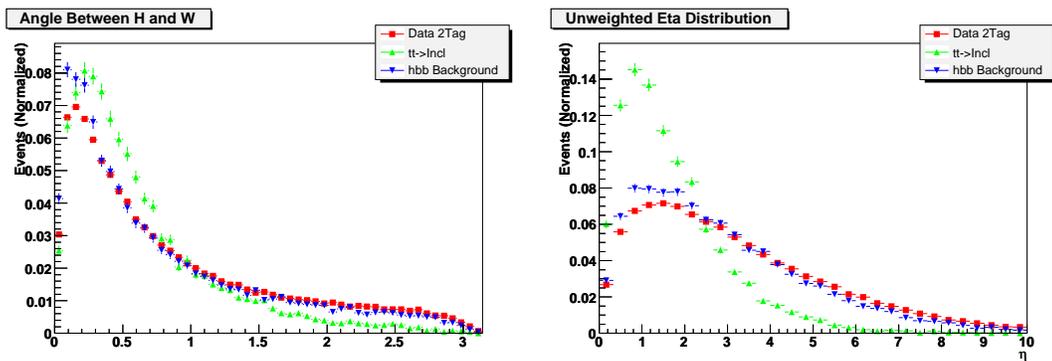
Figure 95: Distributions for events after the preliminary cuts from the $= 2$ b-tagged data (red squares), the QCD background (blue down diamond), the $t\bar{t}$ background (green up diamond).

(a) Leading jet p_T

(b) Second leading jet p_T

(c) Third leading jet p_T

(d) Fourth leading jet p_T
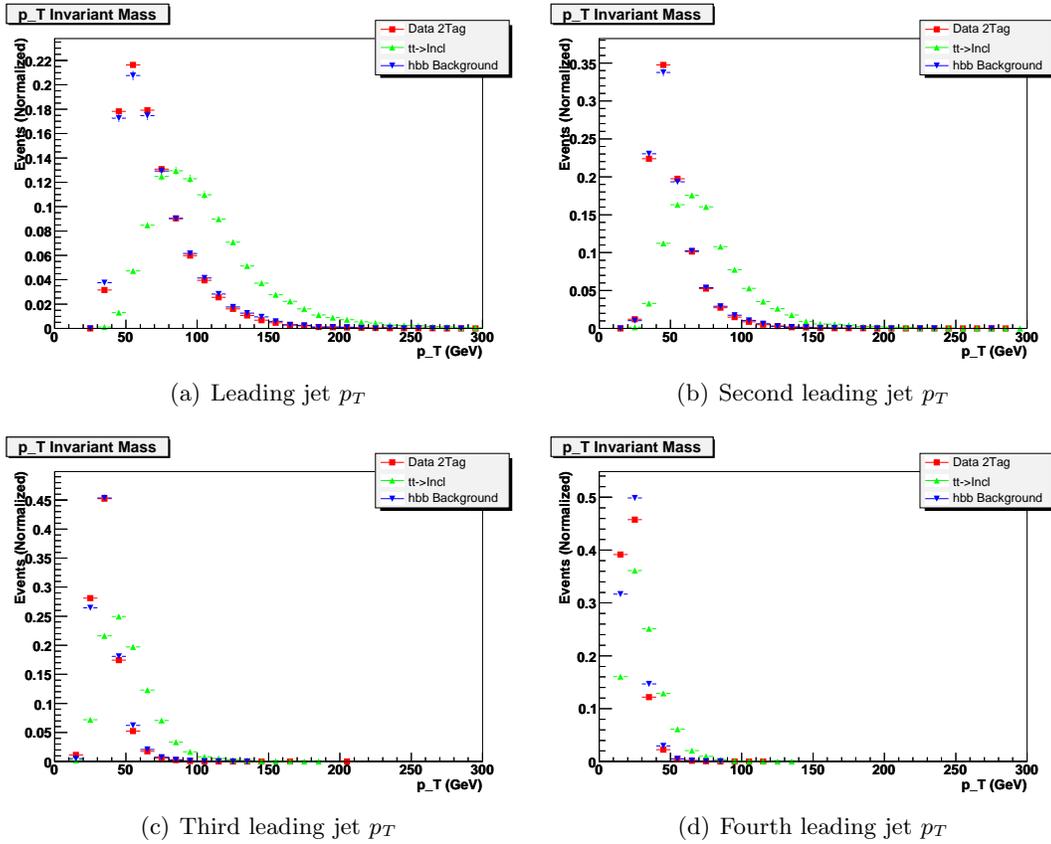
Figure 96: Transverse Momentum distributions for events after the preliminary cuts from the $= 2$ b-tagged data (red squares), the QCD background (blue down diamond), the $t\bar{t}$ background (green up diamond).

175

C   2D sideband subtraction method for the WZ Channel

One technique to determine the amount of signal and background a data sample contains is to use a procedure called sideband subtraction. In this technique the phase space of the data is split into a signal region plus at least one other bin that will act as a sideband. Here we describe a technique using a two-dimensional sideband subtraction method to determine the amount of signal present in a data sample.

For an example take the $b\overline{b}$ invariant mass as the variable and split it into three regions, one from 50 GeV/c^2 to 75 GeV/c^2, the second from 75 GeV/c^2 to 125 GeV/c^2, and the third from 125 GeV/c^2 to 150 GeV/c^2. An initial assumption, to be relaxed later, is that the two sideband bins do not contain any signal events. It is found from Monte Carlo that this is not true due to a background signature from mis-tagged b jets. With this assumption in place the number of events in the sideband is equal to the number of background events in that region. By determining the average background density in the two sidebands an approximate background density can be found for the signal region. To determine the total number of background events in the signal region this average background density needs only to be multiplied by the volume of the signal region. The number of signal events in the data set is then the total number of events in the signal bin minus the number of background events in that bin.

As stated earlier an oversimplification was made by assuming that there were no signal events in the sideband regions. A non-trivial amount of signal can be accounted for by determining ahead of time the efficiency in each bin from a signal Monte Carlo sample. Once this is determined the number of background events in each bin is found from Equation 61 where N_{iB} is the number of background events in the bin, N_s is the total number of signal events in the signal plus sideband regions, and ε_i is the efficiency for the signal Monte Carlo in that bin. The new background event density is found now by taking the sum of this equation over the two sideband bins and dividing by the total area.

In the full version of the background subtraction two variables were used, M($b\bar{b}$) and M(jj). Each of these two variables was split into three different regions, two sidebands and one signal, which gives a total of nine bins. The M($b\bar{b}$) and M(jj) variables are each split into the three regions. The choice of the region boundaries reflects an implicit limitation of the algorithm which is the assumption of a linear background model. While the distributions in data are certainly not linear across the entire region a smaller region is approximately linear. The derivation of the equation for N_s and its associated error are presented below.

C.1   General Case

Definitions

A_i	Area of the cell
ε_i	Signal efficiency for the cell i
N_{MC}	Number of MC events that make it into the plot
N_s	Total number of signal events
N	Number of data events
N_i	Number of data events in cell i
N_{iB}	Number of background events in cell i
N_{iS}	Number of signal events in cell i
z	Index of the signal cell (the middle cell)
q	Number of cells

Assumptions

$$\sum N_i = N \tag{59}$$

$$\sum A_i = A \tag{60}$$

$$N_{iB} = N_i - N_s \varepsilon_i \tag{61}$$

$$N_{zB} = \frac{\displaystyle\sum_{i=1,i\neq z}^{q} N_{iB}}{\displaystyle\sum_{i=1,i\neq z} A_i} A_z \tag{62}$$

By combining Equations 61 and 62 we can obtain an expression for N_s, Equation **??**.

$$N_z - N_s \varepsilon_z = A_z \frac{1}{q-1} \sum_{i=0,i\neq z}^{q} N_{iB} \tag{63}$$

$$= A_z \frac{1}{q-1} \sum_{i=0,i\neq z}^{q} \frac{N_i - N_s \varepsilon_i}{A_i}$$

$$= A_z \frac{1}{q-1} \left[\sum_{i=0,i\neq z}^{q} \frac{N_i}{A_i} - N_s \sum_{i=0,i\neq z}^{q} \frac{\varepsilon_i}{A_i} \right]$$

$$= A_z \frac{1}{q-1} \sum_{i=0,i\neq z}^{q} \frac{N_i}{A_i} - A_z N_s \frac{1}{q-1} \sum_{i=0,i\neq z}^{q} \frac{\varepsilon_i}{A_i}$$

$$N_z - A_z \sum_{i=0,i\neq z}^{q} \frac{N_i}{A_i} \frac{1}{q-1} = N_s \left[\varepsilon_z - A_z \frac{1}{q-1} \sum_{i=0,i\neq z}^{q} \frac{\varepsilon_i}{A_i} \right]$$

$$\therefore \ N_s = \frac{N_z - A_z \frac{1}{q-1} \displaystyle\sum_{i=0,i\neq z}^{q} \frac{N_i}{A_i}}{\varepsilon_z - A_z \frac{1}{q-1} \displaystyle\sum_{i=0,i\neq z}^{q} \frac{\varepsilon_i}{A_i}} \tag{64}$$

C.2 Uncertainty Derivation

The uncertainty in N_s is shown in Equation 65.

$$\delta N_s^2 = \overbrace{\left(\frac{\partial N_s}{\partial N_z}\delta N_z\right)^2}^{A} + \overbrace{\left(\frac{\partial N_s}{\partial \varepsilon_z}\delta \varepsilon_z\right)^2}^{B} + \overbrace{\frac{\partial N_s}{\partial \varepsilon_i}\frac{\partial N_s}{\partial \varepsilon_j}\delta\varepsilon_i\delta\varepsilon_j CC_{ij}^{\varepsilon}}^{C} + \overbrace{\frac{\partial N_s}{\partial N_i}\frac{\partial N_s}{\partial N_j}\delta N_i\delta N_j CC_{ij}^{N}}^{D}$$
$$+ \underbrace{\frac{\partial N_s}{\partial \varepsilon_z}\frac{\partial N_s}{\partial \varepsilon_i}\delta\varepsilon_z\delta\varepsilon_i CC_{zi}^{\varepsilon}}_{E} + \underbrace{\frac{\partial N_s}{\partial N_z}\frac{\partial N_s}{\partial N_i}\delta N_z\delta N_i CC_{zi}^{N}}_{F} \qquad (65)$$

In the above error equation, the matrix CC_{ij}^{ε} represents the correlation coefficient matrix for the efficiencies while the matrix CC_{ij}^{N} represents the same for the number of data events in the individual bins. The correlation coefficients for N_i are assumed to be zero in the off-diagonal regions and simply one along the diagonal. The correlations for the efficiencies are not that simple. Since the total efficiency sums to unity then the efficiencies are non-trivially correlated. In the event that there are only two bins then the correlation between the two efficiencies would be negative one exactly. When adding a third bin information is lost as to in which bin the compensation is done, which reduces the correlation. By performing this bin addition argument ad infinitum one could state that the correlation asymptotically approaches zero. To estimate the correlation coefficient for the error calculation, the negative of a decaying exponential function is used since it adheres to these properties. It is fixed so that the correlation coefficient is equal to negative one in the case of two bins. In the error equation the sums over i and j are assumed. The errors in N_i are assumed to be Gaussian due to their large numbers and errors in the efficiencies are assumed to follow that of Equation 66.

$$\delta\varepsilon_i = \sqrt{\varepsilon_i\frac{1-\varepsilon_i}{N_{\text{MC}}}} \qquad (66)$$

The individual derivatives can be written in the following manner.

$$\frac{\partial N_s}{\partial N_z} = \frac{1}{\varepsilon_z - A_z \frac{1}{q-1} \sum\limits_{i=0, i \neq z}^{q} \frac{\varepsilon_i}{A_i}} \tag{67}$$

$$\frac{\partial N_s}{\partial N_{i,i \neq z}} = -\frac{1}{\varepsilon_z (q-1) - A_z \sum\limits_{j=0, j \neq z}^{q} \frac{\varepsilon_j}{A_j}} \tag{68}$$

$$\frac{\partial N_s}{\partial \varepsilon_z} = -\frac{N_z - A_z \frac{1}{q-1} \sum\limits_{i=0, i \neq z}^{q} \frac{N_i}{A_i}}{\left(\varepsilon_z - A_z \frac{1}{q-1} \sum\limits_{i=0, i \neq z}^{q} \frac{\varepsilon_i}{A_i} \right)^2} \tag{69}$$

$$\frac{\partial N_s}{\partial \varepsilon_{i,i \neq z}} = \frac{N_z - A_z \frac{1}{q-1} \sum\limits_{j=0, j \neq z}^{q} \frac{N_j}{A_j}}{\left(\varepsilon_z - A_z \frac{1}{q-1} \sum\limits_{j=0, j \neq z}^{q} \frac{\varepsilon_j}{A_j} \right)^2} \left(\frac{A_z}{A_i} \right) \frac{1}{q-1} \tag{70}$$

By combining these derivatives the error terms, A-F, are found to be:

$$A = \frac{N_z}{\left(\varepsilon_z - A_z \frac{1}{q-1} \sum\limits_{I=0,I\neq z}^{q} \frac{\varepsilon_I}{A_I}\right)^2} \tag{71}$$

$$B = \left[\frac{N_z - A_z \frac{1}{q-1} \sum\limits_{i=0,i\neq z}^{q} \frac{N_i}{A_i}}{\left(\varepsilon_z - A_z \frac{1}{q-1} \sum\limits_{i=0,i\neq z}^{q} \frac{\varepsilon_i}{A_i}\right)^2}\right]^2 \left(\varepsilon_z \frac{1-\varepsilon_z}{N_{\mathrm{MC}}}\right) \tag{72}$$

$$C = \left[\frac{N_z - A_z \frac{1}{q-1} \sum\limits_{i=0,i\neq z}^{q} \frac{N_i}{A_i}}{\left(\varepsilon_z - A_z \frac{1}{q-1} \sum\limits_{i=0,i\neq z}^{q} \frac{\varepsilon_i}{A_i}\right)^2}\right]^2 \frac{1}{q-1} \frac{A_z^2}{A_i A_j} \sqrt{\left(\varepsilon_i \frac{1-\varepsilon_i}{N_{\mathrm{MC}}}\right)\left(\varepsilon_j \frac{1-\varepsilon_j}{N_{\mathrm{MC}}}\right)} CC_{ij}^{\varepsilon} \qquad i \text{ or } j \neq z$$

$$\tag{73}$$

$$D = \left[\frac{A_z \frac{1}{q-1}}{\varepsilon_z - A_z \frac{1}{q-1} \sum\limits_{m=0,m\neq z}^{q} \frac{\varepsilon_m}{A_m}}\right]^2 \frac{1}{A_i A_j} \sqrt{N_i N_j} CC_{ij}^{N} \qquad i \text{ or } j \neq z \tag{74}$$

$$E = -\left[\frac{N_z - A_z \frac{1}{q-1} \sum\limits_{m=0,m\neq z}^{q} \frac{N_m}{A_m}}{\left(\varepsilon_z - A_z \frac{1}{q-1} \sum\limits_{m=0,m\neq z}^{q} \frac{\varepsilon_m}{A_m}\right)^2}\right]^2 \frac{1}{q-1} \frac{A_z}{A_i} \sqrt{\left(\varepsilon_z \frac{1-\varepsilon_z}{N_{\mathrm{MC}}}\right)\left(\varepsilon_i \frac{1-\varepsilon_i}{N_{\mathrm{MC}}}\right)} CC_{zi}^{\varepsilon} \qquad i \neq z$$

$$\tag{75}$$

$$F = -\frac{\frac{A_z}{A_i} \frac{1}{q-1}}{\left(\varepsilon_z - A_z \frac{1}{q-1} \sum\limits_{m=0,m\neq z}^{q} \frac{\varepsilon_m}{A_m}\right)^2} \sqrt{N_z N_i} CC_{zi}^{N} \qquad i \neq z \tag{76}$$

C.3 Special Case (1D, 3 bins)

In an effort to test the validity of the above derivation a simple case is studied. This simple case parallels that of the simple case given in the textual introduction of this section. This case is one dimensional in the $b\bar{b}$ invariant mass with the bin configuration of $50 \to 75$, $75 \to 125$, $125 \to 150$. Along with using only a single variable, the Monte

Carlo efficiencies for the different bins will be such that there is not a signal in the sidebands, this corresponds to $\varepsilon_1 = \varepsilon_3 = 0$ and $\varepsilon_2 = 1$, where $z = 2$ for this case. The equation for N_s reduces to the following:

$$N_s = \frac{N_z - A_z \frac{1}{q-1} \sum\limits_{i=0,i\neq z}^{q} \frac{N_i}{A_i}}{\varepsilon_z - A_z \frac{1}{q-1} \sum\limits_{i=0,i\neq z}^{q} \frac{\varepsilon_i}{A_i}} \longrightarrow N_2 - A_2 \frac{1}{2} \left[\frac{N_1}{A_1} + \frac{N_3}{A_3} \right] \tag{77}$$

The error in this case reduces to the following form.

$$\delta N_s = N_z + A_z^2 \frac{1}{4} \left[\frac{N_1}{A_1^2} + \frac{N_3}{A_3^2} \right] \tag{78}$$

If a further simplification is introduced such that the sidebands have the same area, and that area is half of the signal area. (which is the situation in this simple example) the equations further reduce to their simplest forms.

$$N_s = N_2 - (N_1 + N_3) \tag{79}$$

$$\delta N_s^2 = N_2 + (N_1 + N_3) = N \tag{80}$$

D   Cross Checks

D.1   Background Shape

The multijet background Monte Carlo generated from a combination of bb, bbj, $bbjj$, bbc, cc, ccj, and $ccjj$ samples can be used to compare the data distributions after the DT cut. As discussed previously in Section 6.1, we used the combination ratios derived from the p17 MSSM hbb multijet analysis and find that the dominant component is $bbjj$. The multijet background is 78% $bbjj$ after the preliminary cuts and this increases to 82% after the DT cut for the WH(115) DT. Figure 97 shows the M(bb)-M(jj) distributions for the 2-tag data and this Monte Carlo QCD background sample after the M(jj) window cut in the DT control region. Based on the statistical errors in the figure, we see decent agreement.
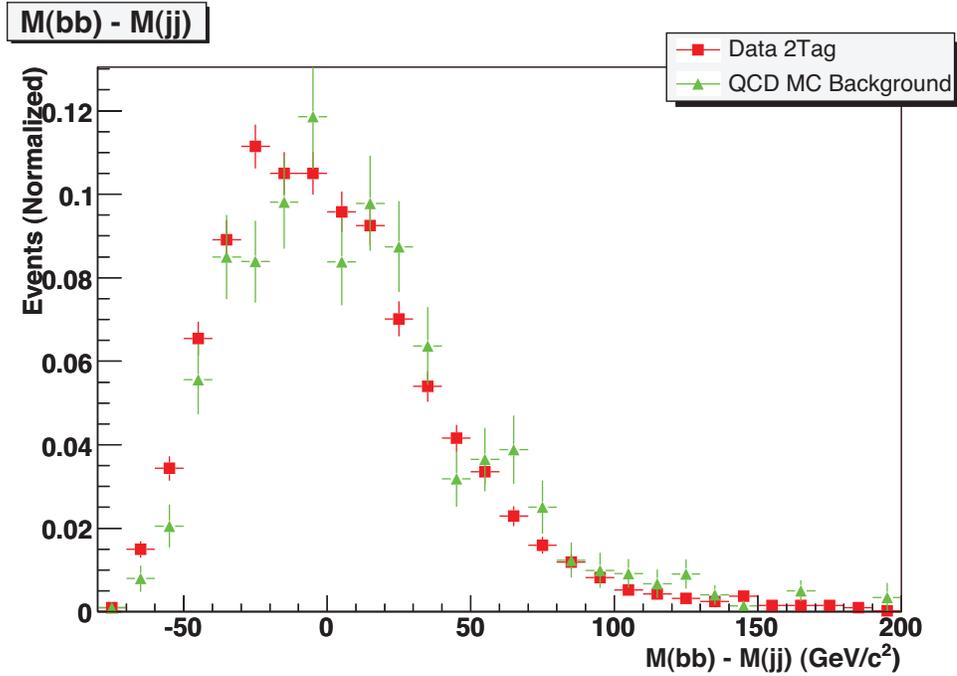
Figure 97: The M(bb)-M(jj) distribution after the M(jj) window cut in the DT control region. The red squares show the data 2-tag and the green triangles show the hbb QCD MC background sample for the WH(115) trained DT.

D.2 Cross Section Results

The signal cross section can be derived in two ways, the first including the effects of the signal contamination in the control region and the second way would be without including this effect. As a cross check, the cross sections are calculated to look for large deviations. Here, $N_s(nc) = D_s - (SF)(D_b)$. Table 24 compares the results using this assumption with those from Table **??**. It is seen that the values for the same mass are consistent between the two effects.

Analysis	N_s	$N_s(nc)$	σ(pb)	$\sigma(nc)$ (pb)
$VH(115)$	$54 \pm 25 \pm 26$	$47 \pm 22 \pm 23$	$9.4 \pm 4.4 \pm 4.8$	$8.2 \pm 3.8 \pm 4.2$
$VH(125)$	$54 \pm 33 \pm 36$	$43 \pm 27 \pm 29$	$8.2 \pm 5.1 \pm 5.6$	$6.7 \pm 4.1 \pm 4.6$
$VH(135)$	$63 \pm 33 \pm 38$	$49 \pm 25 \pm 29$	$10.4 \pm 5.4 \pm 6.3$	$8.0 \pm 4.2 \pm 4.9$
$TC(115)$	$58 \pm 27 \pm 28$	$47 \pm 22 \pm 23$	$23 \pm 11 \pm 12$	$18 \pm 8.5 \pm 10$
$TC(125)$	$63 \pm 39 \pm 42$	$43 \pm 27 \pm 29$	$22 \pm 13 \pm 15$	$15 \pm 9 \pm 11$
$TC(140)$	$87 \pm 45 \pm 50$	$49 \pm 25 \pm 29$	$30 \pm 16 \pm 18$	$17 \pm 9 \pm 11$

Table 24: The numbers of signal events after background subtraction assuming signal contamination in the DT control region (N_s), assuming no contamination ($N_s(nc)$), and the measured cross sections.